Advances in Spatial Science

Randall Jackson
Peter Schaeffer
Editors

Regional Research Frontiers - Vol. 1

Innovations, Regional Growth and Migration

Springer

Advances in Spatial Science

The Regional Science Series

More information about this series at http://www.springer.com/series/3302

Randall Jackson • Peter Schaeffer
Editors

Regional Research Frontiers - Vol. 1

Innovations, Regional Growth and Migration

Springer

Editors
Randall Jackson
Regional Research Institute
West Virginia University
Morgantown
West Virginia, USA

Peter Schaeffer
Division of Resource Economics
and Management
Faculty Research Associate
Regional Research Institute
West Virginia University
Morgantown, WV, USA

ISSN 1430-9602 ISSN 2197-9375 (electronic)
Advances in Spatial Science
ISBN 978-3-319-50546-6 ISBN 978-3-319-50547-3 (eBook)
DOI 10.1007/978-3-319-50547-3

Library of Congress Control Number: 2017936673

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The idea for this book emerged as we prepared the celebration of the 50th anniversary of the Regional Research Institute (RRI) at West Virginia University in 2016. The Institute was founded in 1965, and the personalities who helped shape it include founding director William Miernyk, Andrew Isserman, Luc Anselin, Scott Loveridge, and Randall Jackson. The Institute reflected the research focus and personalities of each of these directors, flavored by the diversity of personalities and scholarship of others with RRI ties. Yet throughout its history, the primary mission remained: engaging in and promoting regional economic development research, with a special emphasis on lagging and distressed regions. RRI scholars have come from economics, geography, agricultural and resource economics, urban and regional planning, history, law, engineering, recreation and tourism studies, extension, etc. Over the half century of RRI's existence, regional research has grown and developed dramatically, with members of the Institute contributing to scholarship and leadership in the profession. Reflecting on the history of the RRI made us wonder about the next 50 years of regional research, so we decided to ask colleagues in our field to share their thoughts about issues, theories, and methods that would shape and define future regional research directions. Many responded to our call for contributions, and in the end we accepted 37 chapters, covering many aspects of regional research. Although the chapters are diverse, several share common ideas and interests, so we have grouped them into seven parts. As with most groupings, of course, there are chapters whose content would have been appropriate in more than one part.

The large number of contributions resulted in a much greater number of pages than planned, but their quality made us reluctant to cut some or to significantly shorten them. We are, therefore, grateful to Johannes Glaeser, Associate Editor for Economics and Political Science at Springer, and to the Advances of Spatial Sciences series editors, for suggesting that we prepare two volumes instead of only one, as initially proposed. We also thank Johannes Glaeser for his advice and support throughout the process of preparing the two volumes. Volume 1 carries the subtitle "Innovations, Regional Growth and Migration" and contains 20 chapters in its four parts. In addition to the topics named in the subtitle, Volume 1 also contains three chapters on disasters, resilience, and sustainability, topics that are of growing interest to scholars, policy makers, and agency and program administrators alike. The subtitle of Volume 2 is "Methodological Advances, Regional Systems Modeling and Open Sciences." Its 17 chapters are organized into the three parts named in the volume's subtitle. The two volumes are roughly equal in length.

The chapters reflect many of the reasons why research methods and questions change over time. A major reason for recent developments in regional research is the digital revolution, which made vastly increased computational capacities widely available. This made possible methodological advances, such as spatial econometrics or geographic information systems (GIS), but perhaps more importantly, it changed fundamentally the way empirical modeling is conducted. Furthermore, it has become possible to integrate different tools, such as spatial econometrics and GIS, and generate graphical displays of complex relationships that enrich our analyses and deepen our understanding of the processes that underlie empirical patterns. Overall, the impact of technological changes on regional research has been pervasive and, judging by the contributions to this volume, will likely continue to be so, and this can be seen in most book parts. In *Modeling Regional Systems*, the chapters' authors rely on recently developed methodological tools and approaches to explore what future research directions could be. In the part *Disasters and Resilience*, Yasuhide Okuyama proposes a future modeling system that would be unthinkable without modern computational tools. All contributions in the part *Spatial Analysis* depend heavily on computational spatial analytical tools, including visualization (e.g., Trevor Harris' contribution on exploratory spatial data analysis). Particularly interesting in this context is the part *Open Source and Open Science*, because it is dealing with aspects of the computational revolution and the Internet that are only now starting to become a major force in our fields, and the collective development and integration of software proposed by Jackson, Rey, and Járosi is still in its infancy.

The evolution of technologies not only drives much of societal change but also has changed how we look at economic growth. While early models of economic growth focused on the capital-labor ratio and treated technology as an exogenous variable, current research in economic growth includes technology as an endogenous variable and stresses entrepreneurship. It is, therefore, not surprising to see an entire part focused on technology, innovation, and entrepreneurship. This part confronts gender issues explicitly in the chapter by Weiler and Conroy, further reflecting changing social attitudes. Gender issues are also addressed in the *Regional Growth*, *Regional Forecasts*, *and Policy* part. As Chalmers and Schwarm note, gender is still a relatively neglected topic in regional research, but social trends and forces will likely increase the attention it receives in the future.

The digital revolution that made mobile phones ubiquitous has also had another important effect, namely the emergence relatively recently of "big data" (e.g., the chapters by Newbold and Brown, and Harris). Even more importantly, vastly improved communication technologies and faster means of transportation are changing the nature of agglomeration. Timothy Wojan reminds us that Alfred Marshall anticipated some of these changes more than a century ago, a remarkable feat of foresight. Because of improved communication technologies, the gap between geographic and social distance is likely to widen in the future, particularly among the highly skilled. Those of us working in research settings at universities or institutes are already experiencing this phenomenon, as it has become common to collaborate with distant colleagues, a sharp contrast to the case until the late twentieth century. It seems certain that the impact of digital technologies on traditional views of geographical space as separation and differentiation will raise new regional research questions. Woodward provides a complement to Wojan's chapter when he speculates about the effects of the interplay of agglomeration and automatization, which is yet another example of the pervasive influence of technology on the future of spatial organization of our societies.

Wojan is not the only one looking to the past to glance into the future. David Bieri studies neglected contributions in regional monetary economics of such foundational scholars of regional research as Lösch and Isard. His chapter presents a genealogy of regional monetary thinking and uses it to make a strong case for renewed attention over the next 50 years to this neglected branch of our intellectual family tree.

While most regional scholars are well aware of the impacts of the digital revolution, there is less awareness of the impacts of an ongoing demographic revolution. This may be because the revolution is far advanced in the economically most successful countries, mostly the members of the Organisation for Economic Co-operation and Development (OECD). But while England became the first country to be more urban than nonurban in the mid-nineteenth century, the world as a whole has reached this threshold less than 10 years ago. Indeed, urbanization in the southern hemisphere is proceeding at a very rapid pace that poses significant policy challenges in the affected nations. As part of industrialization and urbanization, the world is also experiencing a dramatic decline in effective fertility, with the number of births per female of child-bearing age declining. Since longevity is increasing, this is resulting in demographic structures unlike any in the past. This phenomenon is most advanced and dramatic in places such as Germany, Japan, and most recently China—where government policies contributed mightily to demographic restructuring—and challenges the future of public social safety programs, particularly provisions for the financial security of the elderly and their healthcare. In such cases, immigration may be seen as a way to slow the transition from a predominantly young in the past to a much older population. Franklin and Plane address issues related to this unprecedented demographic shift.

Migration, domestic and international, is also of growing importance because of the disruptions caused by industrialization in many countries. The "land flight" that once worried today's industrial powers is now occurring in the southern hemisphere. Migration is also fueled by political change in the aftermath of the end of colonialization. The new nations that emerged were often formed without regard for historic societies and traditions, and tensions that had been held in check have sometimes broken out in war between neighboring countries or civil war. As a result, the world as a whole has seen an increase in internally displaced persons as well as refugees who had to leave their home countries. In an overview of directions in migration research, Schaeffer, therefore, argues for more work on migrations that are rarely completely voluntary because traditional models have been developed primarily for voluntary migrations.

Demographic shifts are also driving reformulations and advances in *Regional Systems Models*, as evidenced by new directions in household modeling within the chapter on household heterogeneity by Hewings, Kratena, and Temurshoev, who touch on these and enumerate a comprehensive research agenda in the context of dynamic interindustry modeling, and Allen and his group identify pressing challenges and high potential areas for development within computable general equilibrium models. Varga's chapter contributes to this part's topic and to technological change, as his Geographic Macro and Regional Impact Modeling (GMR) provides explicit mechanisms for capturing the impacts of innovation and technology.

The chapters in these volumes reflect the changing world that we live in. While some new directions in regional research are coming about because new technologies allow us to ask questions, particularly empirical questions that once were beyond the reach of our capabilities, others are thrust upon us by political, economic, social, demographic, and environmental events. Sometimes several of these events combine to effect change. A primary task of a policy science is to provide guidelines for the design of measures to address problems related to change. So far, regional researchers seem to have been most successful in making progress toward completing this task in dealing with environmental disasters, addressed in the *Disasters and Resilience* part. Rose leverages decades of research in regional economic resilience to lay the foundation for this part.

These chapters will certainly fall short of anticipating all future developments in regional research, and readers far enough into the future will undoubtedly be able to identify oversights and mistaken judgements. After all, Kulkarni and Stough's chapter finds "sleeping beauties" in regional research that were not immediately recognized, but sometimes required long gestation periods before becoming recognized parts of the core knowledge in our field, and Wojan and Bieri also point to and build upon contributions that have long been neglected. If it is possible to overlook existing research, then it is even more likely that we are failing to anticipate, or to correctly anticipate, future developments. Nonetheless, it is our hope that a volume such as this will serve the profession by informing the always ongoing discussion about the important questions that should be addressed by members of our research community, by identifying regional research frontiers, and by helping to shape the research agenda for young scholars whose work will define the next 50 years of regional research.

Morgantown, WV

Randall Jackson
Peter Schaeffer

Contents

Part I Technology, Innovation, Gender, and Entrepreneurship

1 Opportunities and Challenges of Spatially Distributed Innovation Imaginariums 3
Timothy R. Wojan

2 Exploring Innovation Gaps in the American Space Economy 21
Gordon F. Mulligan, Neil Reid, John I. Carruthers, and Matthew R. Lehnert

3 Future Shock: Telecommunications Technology and Infrastructure in Regional Research 51
Tony H. Grubesic

4 Mobility and Technology Research: From the Industrial Revolution to Flying Vehicles in 2050 71
Roger R. Stough

5 Entrepreneurship, Growth, and Gender 85
Tessa Conroy and Stephan Weiler

Part II Regional Growth, Regional Forecasts, and Policy

6 Agglomeration and Automation in the Twenty-First Century: Prospects for Regional Research 97
Doug Woodward

7 Designing Policies to Spur Economic Growth: How Regional Scientists Can Contribute to Future Policy Development and Evaluation 119
Carlianne Patrick, Amanda Ross, and Heather Stephens

8 Regional Science Research and the Practice of Regional Economic Forecasting: Less Is Not More ... 135
Dan S. Rickman

9 Energy for Regional Development ... 151
Paulo Henrique de Mello Santana

10 Regional Perspectives on Public Health ... 161
Sara McLafferty and Alan T. Murray

11 New Approaches to Gender in Regional Science ... 175
Katherine Chalmers and Walter Schwarm

12 Identifying Sleeping Beauties in the Lore of Regional Science ... 183
Rajendra Kulkarni and Roger R. Stough

13 Regional Policy and Fiscal Competition ... 199
Santiago M. Pinto

14 Back to the Future: Lösch, Isard, and the Role of Money and Credit in the Space-Economy ... 217
David Bieri

Part III Diasters and Resilience

15 Economic Resilience in Regional Science: Research Needs and Future Applications ... 245
Adam Rose

16 Disaster and Regional Research ... 265
Yasuhide Okuyama

17 Regional Sustainability and Resilience: Recent Progress and Future Directions ... 277
Elena G. Irwin, Tim Jaquet, and Alessandra Faggian

Part IV Migration, Demography, and Human Capital

18 Directions in Migration Research ... 299
Peter V. Schaeffer

19 Human Capital Research in an Era of Big Data: Linking People with Firms, Cities and Regions ... 317
K. Bruce Newbold and W. Mark Brown

20 The View from Over the Hill: Regional Research in a Post-Demographic Transition World ... 329
Rachel S. Franklin and David A. Plane

Editors and Contributors

About the Editors

Randall Jackson is professor, Department of Geology and Geography, West Virginia University (WVU), and Director of the Regional Research Institute. His primary research interests are regional industrial systems modeling; energy, environmental, and economic systems interactions; and regional economic development. He is an adjunct professor in WVU's Department of Economics and Division of Resource Management, and in Geography at the Ohio State University (OSU). Previous faculty positions were at OSU and Northern Illinois University. Dr. Jackson earned his PhD in geography and regional science from the University of Illinois at Urbana-Champaign in 1983.

Peter Schaeffer is professor, Division of Resource Economics and Management, West Virginia University (WVU). His primary research interests are regional economic policy, international labor migration, job mobility, natural resource management, and historic preservation. He is a faculty research associate in WVU's Regional Research Institute and adjunct professor in the Department of Economics. Previous faculty positions were at the Universities of Colorado–Denver, Illinois at Urbana–Champaign, and one year as visiting professor at the Swiss Federal Institute of Technology–Zurich. Dr. Schaeffer earned the Ph.D. in economics from the University of Southern California in 1981.

Contributors

W. Mark Brown Economic Analysis Division Statistics Canada, Ottawa, ON, Canada

David Bieri School of Public & International Affairs, Virginia Tech, Blacksburg, VA, USA

John I. Carruthers Sustainable Urban Planning Program, The George Washington University, Washington, DC, USA

Katherine Chalmers California State University, Sacramento, CA, USA

Tessa Conroy Department of Agricultural and Applied Economics, Center for Community Economic Development, University of Wisconsin-Madison/Extension, Madison, WI, USA

Paulo Henrique de Mello Santana Engineering, Modelling and Applied Social Sciences Center, ABC Federal University, Sao Paulo, Brazil

Alessandra Faggian Department of Agricultural, Environmental and Development Economics, The Ohio State University, Columbus, OH, USA

Rachel S. Franklin Brown University, Providence, RI, USA

Tony H. Grubesic Center for Spatial Reasoning and Policy Analysis, Arizona State University, Tempe, AZ, USA

Elena G. Irwin Department of Agricultural, Environmental and Development Economics, The Ohio State University, Columbus, OH, USA

Tim Jaquet Department of Agricultural, Environmental and Development Economics, The Ohio State University, Columbus, OH, USA

Rajendra Kulkarni George Mason University, School of Public Policy, Fairfax, VA, USA

Matthew R. Lehnert Ph.D. Program in Spatially Integrated Social Sciences, University of Toledo, Toledo, OH, USA

Sara McLafferty Department of Geography & GIScience, University of Illinois at Urbana-Champaign, Champaign, IL, USA

Gordon F. Mulligan School of Geography and Development, University of Arizona, Tucson, AZ, USA

Alan T. Murray Department of Geography, University of California at Santa Barbara, Santa Barbara, CA, USA

K. Bruce Newbold School of Geography & Earth Sciences, McMaster University, Hamilton, ON, Canada

Yasuhide Okuyama University of Kitakyushu, Kitakyushu, Japan

Carlianne Patrick Department of Economics, Georgia State University, Atlanta, GA, USA

Santiago M. Pinto Federal Reserve Bank of Richmond, Richmond, VA, USA

David A. Plane University of Arizona, Tucson, AZ, USA

Neil Reid Department of Geography and Planning, University of Toledo, Toledo, OH, USA

Dan S. Rickman Oklahoma State University, Stillwater, OK, USA

Adam Rose Price School of Public Policy, and Faulty Affiliate, Center for Risk and Economic Analysis of Terrorism Events (CREATE), University of Southern California (USC), Los Angeles, CA, USA

Amanda Ross Department of Economics, Finance, and Legal Studies, University of Alabama, Tuscaloosa, AL, USA

Peter V. Schaeffer Division of Resource Economics and Management, Faculty Research Associate, Regional Research Institute, West Virginia University, Morgantown, WV, USA

Walter Schwarm Demographic Research Unit, California Department of Finance, Sacramento, CA, USA

Heather Stephens Resource Economics and Management, West Virginia University, Morgantown, WV, USA

Roger R. Stough Schar School of Policy and Government, George Mason University, Fairfax, VA, USA

Stephan Weiler Department of Economics, Colorado State University, Fort Collins, CO, USA

Timothy R. Wojan Economic Research Service, U.S. Department of Agriculture, Washington, DC, USA

Doug Woodward Department of Economics, Moore School of Business, University of South Carolina, Columbia, SC, USA

Part I
Technology, Innovation, Gender, and Entrepreneurship

Chapter 1
Opportunities and Challenges of Spatially Distributed Innovation Imaginariums

Timothy R. Wojan

1.1 Introduction

Envisioning what research questions will emerge in the spatial analysis of innovation over the next 50 years is fraught with uncertainty. Perhaps the safest bet is to identify those constructs and hypotheses that currently command a wide degree of agreement and predict that these consensuses will have disappeared by 2066. To the extent that "scientific truths" tend to endure for as long as their progenitors (Azoulay et al. 2015; Planck 1949), new frontiers might be most clearly delineated by the shadows cast by the current crop of star regional scientists. Unfortunately, ideas regarding the geography of innovation do not usually fall into clear white/black or sun/shade dichotomies. Just postulating a seeming opposite might not get one very far.

The contrarian views investigated in this chapter rely on two different strategies for helping to illuminate the as yet lightly investigated counter-arguments. We begin with Alfred Marshall's counterfactual musings on the declining importance of proximity as the costs of communicating new ideas fall. Now that the huge reductions in communications costs have been realized, empirical analysis has been able to put Marshall's conjecture to the test. The notion that proximity has become

The views expressed are those of the author and are not attributable to the Economic Research Service or the U.S. Department of Agriculture.

T.R. Wojan
Economic Research Service, U.S. Department of Agriculture, Washington, DC, USA
e-mail: twojan@ers.usda.gov

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_1

much less important to the transmission of new ideas runs counter to conventional wisdom that regularly invokes Marshall's 1890 argument. Ideas in the literature at odds with the conventional wisdom but that reinforce the Marshallian counterfactual are also discussed.

The second strategy contrasts the ideas of two seminal thought leaders: Joseph Schumpeter and John Dewey. The ideas of Schumpeter will be easily recognized in a discussion of an innovation economy and the governance requirements of an entrepreneurial state. The contribution of Dewey to ideas about governance as it relates to innovation will be much less familiar to readers but presents a fecund alternative path that is much more compatible with emerging models of distributed innovation. The alignment of the persuasive Schumpeterian linear model of innovation, data collection focused on the required inputs and expected outputs of this model, and an implicit social contract amongst the stakeholders of that model (Jasanoff 2009; see Bush 1945) reinforced the belief that innovation is a function of entrepreneurial leadership. The counterargument that innovation is a function of the collective ingenuity of a community of actors confronting a problem has yet to be mainstreamed.

Combining the counter-arguments from the two strategies provides a potentially rich but thinly investigated pathway for the spatial analysis of innovation. User innovators who now have access to both deep reservoirs of information through the Internet and to like-minded problem-solvers through social media may spur radical innovation without any need for entrepreneurial leadership as conventionally understood. The objective of this chapter is to provide an outline of this alternative pathway for innovation that can be contrasted with the dominant linear model of innovation. There are three reasons why pursuit of this alternative track may be advantageous for the regional study of innovation. First, in a rapidly urbanizing world, large agglomerations that are the preferred locus of innovation in the linear model will become increasingly commonplace. Identifying factors that amplify the rare sparks of genius wherever they occur encompasses all innovation instead of limiting analysis on the basis of density—an increasingly ad hoc criterion. Second, the struggle within regional science between the relatively amorphous construct of "community" and more concrete constructs defined by purely spatial relationships is reinvigorated. Finally, the reintroduction of community opens up consideration of the types of governance structures that promote socially valued innovation, departing from the implicit assumption that innovation is universally good.

1.2 Conventional Wisdom

If the conventional wisdom regarding the geography of innovation was limited to academic discussion, then any critique would likewise be academic. However, there is clear evidence that conventional wisdom imbues thinking on policy related to the best way to promote innovation and economic growth. Indeed, the notion that innovation is best explained by scientism—the idea that scientific and engineering truths that are "out there" waiting to be discovered is the true source of innovation—has not only guided economists and regional scientists but forms the foundation for Federal innovation policy (Jasanoff 2009; Phelps 2013). The economic study of innovation has been dominated by a linear model of hard inputs such as science and engineering personnel, and R&D expenditures, motivated by the rational pursuit of monopoly profit where the output is most reliably represented by patents. Despite the critique that patents of new inventions do not adequately capture the concept of innovation, the wide availability of patent data has made them convenient proxies.

But while economists and regional scientists may disagree over validity of patents as a proxy for innovation and the need for collecting additional measures that can more fully capture the concept of innovation, there is widespread agreement over the salient characteristics of the geography of innovation. Feldman and Kogler (2010, p. 381) distill these characteristics into eight stylized facts:

- Innovation is spatially concentrated
- Geography provides a platform to organize innovative activity
- Places are not equal: Urbanization, localization, and diversity
- Knowledge spillovers are geographically localized
- Knowledge spillovers are nuanced, subtle, pervasive, and not easily amenable to measurement
- Local universities are necessary but not sufficient for innovation
- Innovation benefits from local buzz and global pipelines
- Places are defined over time by an evolutionary process

With the exception of references to universities and global pipelines, these stylized facts provide a direct connection between Marshall's (1890) seminal discussion of industrial districts and the current state of the art regarding the geography of innovation. The main twentieth century embellishment to the Marshallian story of localized knowledge spillovers is the addition of agglomeration and urbanization economies that derive benefits from the cross-pollination of ideas from related or seemingly unrelated sectors (Glaeser et al. 1992).

The presumed stickiness of information—the added cost of acquiring, transferring or using information in a new location—and the heavy reliance on local information and knowledge essential to inventors and entrepreneurs explains the focus on particular cities to fully understand innovation in an industry (Feldman and Kogler 2010). The same factor that Marshall identified in 1890—the costliness of transferring information from place to place—appears as prominent now as it was then. There are a number of explanations for why the cost of transferring information related to economic innovation has remained high despite drastic reduction in cost and phenomenal expansion of capabilities in communications technology since 1890. If some interactions required for the substantive transfer of knowledge and information are not reliably conveyed by communications technologies, then huge reductions in cost for other types of information transfer may matter little.

1.2.1 Empirical Challenges to Conventional Wisdom

Packalen and Bhattacharya (2015) point to a brilliant insight by Alfred Marshall regarding the conditions that convey significant innovation advantages to agglomeration. In the late nineteenth century, physical proximity was the dominant vector for conveying new ideas. And thus, the larger agglomerations would naturally provide much more fertile ground for the dissemination and recombination of novelty. But Marshall also recognized that this advantage might be altered by the "cheapening of the means of communication" leading ultimately to knowledge production being dependent "on the aggregate volume of production in the whole world." It is indisputable that the means of communication today are drastically cheaper than in 1890 and, in some cases, were wholly unimagined.[1] And yet this critical caveat from Marshall—that the benefits of agglomeration might be dependent on the costs of communication—has largely been explained away in the literature (Glaeser 2010; Florida et al. 2008).

Packalen and Bhattacharya test Marshall's intuition by applying content analysis to the universe of U.S. patent applications to identify word sequences that represent idea inputs to invention. By identifying the first occurrence of each idea input, they are able to calculate the Age of Idea Inputs for every patent. Because the benefits of agglomeration are believed to be strongest for the transmission, discussion, and evaluation of new ideas, they construct a dummy variable to identify those patents that are in the Top 5% by Age of Newest Idea Input which they regress against population density of the Primary Metropolitan Statistical Area (PMSA) where the inventors lived. The odds ratio of the coefficient estimate represents how much more likely inventions in large cities are built on new ideas relative to average sized cities. For the period relevant to when Marshall was writing (1880s–1910s), residents in large cities were 20% more likely to use new ideas in patents. The coefficient declines in the 1920s–1960s (15%) and 1970s–1980s (8%) periods. In the most recent period of the 2000s, the odds ratio is still statistically significant representing a 4% increase but is no longer robust to alternative specifications using Total Population rather than Population Density as the independent variable or an instrumental variable specification.

The empirical tests of Marshall's intuition—that the advantages of agglomeration might decline as communication technology improved—open up the possibility that innovation using new knowledge may take place in areas other than where the new knowledge is produced. However, the Packalen and Bhattacharya paper does not directly address that issue. Given the focus on patents, the most that can be concluded is that physical proximity no longer appears to be an essential vector for the transmission of new knowledge to the creation of the newest knowledge.

[1] Teleporting does not yet exist but the ability to virtually transport a 3-dimensional object over the Internet only requires a 3D scanner to send digital information to a 3D printer located anywhere in the world.

A recent paper by Capello and Lenzi (2014) makes the necessary distinction between the creation of knowledge—in the form of R&D expenditures—and innovation—in the form of shares of firms introducing new products or services. The measure of innovation is constructed from responses to the fourth round of the Community Innovation Survey. Their findings extend the Marshallian intuition by demonstrating that some regions with very limited knowledge creation capabilities may still support a relatively high share of innovative firms. Most importantly, a relatively high share of innovative firms have a much larger impact on GDP growth in those regions that have limited knowledge-creation capacity. The dual findings that innovation appears to be more highly dispersed than knowledge creation activities, and that the impact of innovation on GDP growth is also larger than the impact from R&D expenditures suggests that the linear model of innovation with its strong orientation to scientism may actually provide a very limited explanation of how innovation is related to economic growth.

1.2.2 Conceptual Challenges to Conventional Wisdom

The linear model of innovation does appear to be aptly suited to an implicit confirmation bias in regional science that begins from the premise that place matters. Because the institutions, inputs, and outputs in the linear model are spatially concentrated, the geography of (linear model) innovation is easily identified and easily tested. A much more difficult model of innovation for regional scientists is the one developed by Edmund Phelps (2013) as a counter to the scientism of the linear model. Rather than place innovation at the intersection between scientific/engineering discoveries and entrepreneurial leadership, Phelps envisions a much more inclusive imagining of new products, processes and uses by consumers, craftsmen, technicians, along with professional scientists and engineers. The challenge to solve vexing problems and the creative spark emerge as motivations for innovation that extend beyond the pursuit of monopoly profits. These problem contexts for innovation—or “imaginariums”—might also describe a Marshallian industrial district along with many other types of economic dynamism. It is the inability of the linear model to explain the emergence of innovation and economic dynamism in unexpected places that raises the biggest questions.

The historical record that is wholly inconsistent with the linear model of innovation is the rapid productivity growth in early nineteenth century America. Scientism would predict that the far superior scientific and engineering knowledge in Europe at the time, along with much deeper markets for exploiting profitable new innovations would have given the nod to England. And yet productivity growth in the sparsely populated U.S., with a very small contingent of tinkerers, eclipsed England before mid-century. The frontier may have had something to do with the success of these New World imaginariums as new challenges were frequent with few constraints from convention or established ways of doing things. A more recent example of dynamism in unexpected places is the surprising rise of Finland as the world-leading pioneer of cellular technology. Clearly the scientific and engineering expertise in electronics in Japan, Western Europe, or North America eclipsed that of a little known, diversified Finnish manufacturing conglomerate. And the number of consumers in those three large markets exceeded the Finnish population of four million by more than a couple orders of magnitude. But the problem context of being able to communicate without a landline while at a remote summer cottage appears to have been the spark that would eventually change the world (Steinbock 2001).

1.3 A More Comprehensive View of Innovation

It is important to reiterate that the goal of developing alternatives is not to displace the geography of (linear model) innovation. Rather, the goal is to illuminate other types of innovation that are occurring, some of which may be much more relevant to understanding innovation processes that emerge in unexpected places.

Recognition that the linear model of innovation only tells part of the story is not new. As early as 1991 economic historian and science policy scholar Nathan Rosenberg made the seemingly unequivocal claim that "[e]veryone knows that the linear model of innovation is dead." He did not add "Long live the linear model of innovation" to that pronouncement though he very well could have. The resilience of the linear model of innovation, despite recognized shortcomings, is explained by the seamless way it integrates the entrepreneurial theory of Schumpeter with conventional neoclassical economic theory (Phelps 2013).

The strongest argument for expanding the economic study of innovation beyond the linear model comes from Nobel laureate William Baumol. Baumol's (2010) discussion of radical and incremental innovation focuses on the microfoundations behind the division of innovative labor. Incremental innovation is aptly suited to the linear model where the systemic investigation of options to improve performance or lower cost—often through the application of new scientific discoveries—is more likely to result in an acceptable rate of return on R&D investment. As profit maximizers, incumbent firms will invest R&D in areas that have the highest profit potential, conditional on relatively high chances of success. And since their competitors are engaged in similar efforts, failure to pursue incremental innovation risks the survival of the firm. In contrast, radical innovation has the potential for very high payoffs accompanied by very high risk of failure. Radical innovation is aptly suited to the entrepreneurial firm where an individual or small team conceive of a genuinely novel product or process. Baumol's theoretical argument is compelling and gets heavy empirical reinforcement from the long list of disruptive technology innovations by small firms that have profoundly changed the world we live in.[2]

[2]Baumol cites a U.S. Small Business Administration report that lists roughly 100 highly disruptive technologies developed by small firms including the airplane, air conditioning, microprocessor,

The main takeaway is that innovative entrepreneurship is a necessary but largely discounted dimension of economic theory and applied economic analysis.

The regional science implications of the Baumol call to study the complementary forms of incremental and radical innovation might be muted, if noted at all. While the availability of data consistent with the linear model of innovation has resulted in a geography of innovation dominated by large, global cities, the presumption has been that these are the same places where innovative entrepreneurship should also be most dynamic and most prevalent. A local symbiosis is implicit in Baumol's mainly aspatial discussion of the need for large incumbents to outsource radical innovation activities. In this case, even if the linear model of innovation does not provide a comprehensive story of innovation, it may still provide a comprehensive identification of where incremental and radical innovation take place.

However, there are two aspects of innovative entrepreneurship that counter an implicit local symbiosis story of the co-location of radical and incremental innovation. Two dominant sources of innovative entrepreneurship come from: (1) users who either extend capabilities or develop entirely new uses of products; or (2) through serendipity when a different problem environment sparks a completely new approach. Since users of products are likely to be much more spatially dispersed than the sites where the products were made, user innovation is also likely to be a fairly dispersed process. For serendipity, being physically and cognitively separate from those places of recognized expertise may accelerate the generation of true novelty. While both paths to innovation have been recognized in the literature, these paths are consistently discounted as being too sparse and too random to be worthy of study.

Eric von Hippel (2005) has devoted his career to the study of user innovation and has built up an impressive body of evidence. His distinction between users and manufacturers is quite simple: users are firms or individuals that expect to benefit from using a product or service while manufacturers expect to benefit from selling a product or service. Beginning with case studies of particular industries with lead users of products such as scientific instruments, the research provided empirical evidence that the linear model was missing a lot. For example, the Science and Engineering Indicators published annually by the National Science Foundation consistently show that patent productivity as a function of R&D expenditures is supposedly very high in the scientific instruments industry. However, since upwards of 80% of innovations in scientific instruments are developed by users in academia and other industries, simply attributing patents in the industry to industry R&D expenditures seriously misconstrues that industry's inventive process (von Hippel 1976). Most industries will not be user dominated as is the case with scientific instruments, but user innovation is present in all industries. The development of the Internet, social media, 3D printing, and simulation software for rapid prototyping are all reinforcing the trend toward democratizing innovation (von Hippel 2005).

personal computer, supercomputer, high resolution x-ray and CAT scanner, vacuum tube, and integrated circuit.

As the technology for searching the codified global knowledge base improves, and as the tools for translating ideas into prototypes get cheaper, the barriers to innovation will also decline. A survey of consumer innovators in the U.K., U.S., and Japan estimated the total expenditures on "grassroots R&D" in the billions of dollars, exceeding corporate R&D expenditures in the U.K. and comprising a third of corporate R&D expenditures in the U.S. (von Hippel et al. 2011).

Recent collections of data on user innovators that are representative of the population suggest not only that it comprises an important component of innovation, but that it also has the potential to transform users into producers. The 2014 Economic Research Service (ERS) Rural Establishment Innovation Survey includes questions to determine if businesses were founded to market goods or services that were originally produced for their own use. The population of interest is all establishments with five or more employees in nonfarm tradable sectors. User entrepreneurs make up 5.59% of the establishments in this population, accounting for 6.12% of employment for a total of roughly 2.9 million jobs (USDA-ERS 2015). These percentages are likely conservative estimates as respondents in large establishments are less likely to be familiar with circumstances surrounding the founding of the business. However, even with this potential for bias in large establishments, these data confirm that at least 3.23% of user entrepreneur establishments have grown to employ more than 100 workers, compared to 6% for all other establishments. Preliminary geographical analysis of user entrepreneur establishment confirms that they are found throughout the settlement hierarchy. Additional analysis will examine the extent to which these user entrepreneurs are more likely to emerge in industrial clusters or other locations of specialized knowledge relative to more greenfield locations.

Industrial, and now occupational (Florida 2002), specialization is presumed to be a critical component of the geography of innovation as the focus provided by a large number of people in a place thinking about similar problems should accelerate finding solutions to those problems. The counter-argument—that the freshness of thinking about these problems is enhanced in new environments using tools that may be foreign to more specialized locations—might be intriguing but would be seemingly impossible to test. This makes research on the characteristics of winners of science innovation tournaments all the more compelling. Jeppesen and Lakhani (2010) provide strong empirical evidence that both "technical marginality" and "social marginality" are consistently associated with winning broadcast search tournaments. Broadcast search tournaments consist of well-defined but complex algorithmic or computational challenges with high uncertainty of finding an optimal solution. Examining 166 such contests that received submissions from 12,000 scientists, winners were more likely to come from fields of expertise somewhat distant from the problem field (technical marginality) as well as being more likely to be women, which is interpreted as a proxy for social marginality given the strong gender bias in the sciences (Jeppesen and Lakhani 2010).

The most interesting result of this line of research for regional science comes from the NASA Tournament Lab that held a contest to devise an algorithm that would select the optimal medical kit for space travel based on simulated medical event data provided by NASA (Boudreau et al. 2011). The problem required a software solution trading off mass and volume against sufficient resources to minimize the risk of medical evacuation. After significant effort, the kit optimization algorithm developed by NASA took 3 hours. NASA researchers were "blown away" by the winning solution that performed the kit optimization in 30 seconds.

But the winner did not come from a space agency center like Washington DC, Paris, or Moscow, but from a former Eastern Bloc satellite country (Lahani 2015). Anecdotally, the winning extreme value outcomes in other contests came from similarly unexpected places suggesting that spatial marginality might accompany technical and social marginality as predictors of tournament success. But even without statistical evidence of this, the empirical confirmation that "optimal marginality" appears to be an important factor in radical innovation (McLaughlin 2001; Jeppesen and Lakhani 2010) points to a large blind spot for the linear model of innovation: concentrated specialization that aggressively explores the core of a field may inhibit lateral exploration of radical solutions.[3]

The fact that R&D labs are the main clients of the broadcast search tournaments studied by Jeppesen and Lakhani (2010) suggests that conventional practice is not waiting for the linear model of innovation to catch up. A recent special issue of the *Journal of Economic Geography* on "Knowledge creation—local building, global accessing" identifies a number of traditional and emerging constructs such as international trade fairs, crowd-sourcing, and listening posts that are accelerating realization of the Marshallian conjecture that knowledge production will ultimately be dependent "on the aggregate volume of production in the whole world" (Maskell 2014; Bathelt and Cohendet 2014; Bathelt and Gibson 2015). And empirical evidence is now available suggesting that collaboration with distant interlocutors can substitute for proximate knowledge spillovers (Grillitsch and Nilsson 2015). Whether this rapidly emerging literature on non-local sourcing of knowledge and information is the crest of a wave or a short-lived fad remains to be seen. What is most notable is the vintage of the conventional wisdom counterargument regarding "the stickiness of information"—for example, the most recent reference to this construct in the review article by Feldman and Kogler (2010) is 22 years old (cf. von Hippel 1994), which corresponds with the founding of the World Wide Web Consortium.

1.4 From Locale to Community

The emergence and growing importance of communities that are not defined by physical proximity casts localization of the archetypal Marshallian industrial district

[3]Two papers investigate the possibility that the serendipitous interaction that is thought to be the key advantage of urban agglomerations and clusters may in fact promote lock-in to conventional ways of thinking about problems (Boschma 2005; Fitjar and Rodriguez-Pose, forthcoming).

in a different light. Given Marshall's focus on economics, it is understandable that he would emphasize economic concepts like tacit knowledge rather than sociological constructs of reciprocity and trust that underlie community. Quite simply, the competitive advantages that clusters are thought to confer may be overdetermined. Framed in the vernacular of Marshall's evocative explanation, "[t]he 'secret', thus, of local clusters may reside much more in the relational aspects of community (i.e. as one spatial form of knowing through communities) than on the balance between tacit and codified knowledge." (Amin and Cohendet 2005). The local cluster may be better understood as conferring benefits from tacit knowledge *and* tacit relations.

Empirically, communities that are defined by spatial proximity may be thought of as a dummy construct where the effects of spatial proximity are measurable but no explicit explanation of the components of those effects is provided. Advances in behavioral and experimental economics help illuminate the types of relations we might expect to emerge organically as a byproduct of proximate interaction every day (Beinhocker 2006). For example, the empirical regularity of strong reciprocity—the common observance of conditional cooperation and altruistic punishment in experimental prisoner dilemma games—suggests that relations that are not assumed *a priori* for economic agents do emerge spontaneously as tacit relations. In contrast, spatially distributed communities that form in the interest of innovating to solve particular problems will not necessarily develop the norms, expectations, and reciprocal relations simply as a byproduct of "coming together" at a distance. As a result, the construct of community has been at the center of how open source, user innovators, communities of practice, and other virtual communities are constituted and operate (West and Lakhani 2008).

The spatially distributed community that most readers should be familiar with is the academic community with the four basic elements of community defined by universalism, organized skepticism, disinterestedness, and communalism (Merton 1942). The last two elements require elaboration as they provide the starkest contrast with the proprietary model that is used to understand most economic behavior. Disinterestedness suggests that member are rewarded for actions that appear to be selfless and communalism requires common ownership of scientific discoveries in which ownership of intellectual property is waived in exchange for recognition and esteem. Within the proprietary model, exclusive property rights allow an organization to capture value from discoveries that are manifest in commercialized products. The elements that typically define spatially distributed innovation communities are a hybrid of the academic and proprietary models (Shah 2006). Motivation of members is often from a self-interested need to solve a specific problem that is more in line with the proprietary model but violates the universalism and disinterestedness elements of the academic model. However, the self-interestedness of community members is ameliorated by the sharing of discoveries consistent with the communalism element from the academic model. The value that community members place on making important discoveries independent of any material gain is also consistent with the communalism element. Organized skepticism is also a common element in some innovation communities that can give them an advantage over proprietary models where the best short-term profit maximizing solution may not provide the best technical solution, evidenced by the closed source dictum to "sell first, fix later."

The simplest way to summarize the hybridization is to define the objectives of the proprietary model as instrumental (i.e., activities that are instrumental to maximizing profit) and the academic model as constitutive (i.e., activities that are constitutive of being an academic).

Innovation communities are an amalgam of instrumental and constitutive objectives of solving problems for personal gain but where the activity is also constitutive of the identity of members. In the words of one such community member, "in an open source community, not one answer is forced on anyone. Everything is up for discussion and change—all the time . . . it's empowering and it leaves room for new people to come in and make improvements and changes . . . That dynamic just doesn't exist in communities around tightly licensed corporate code . . . " (as quoted in Shah 2006, p. 1000). The same constitutive aspect of productive communities emerges in Michael Piore's resolution of the seeming conflict between competition and cooperation in Marshallian districts in the Third Italy:

> The openness of the production process, and of the innovations in the instruments of production, becomes almost a prerequisite for their existence. If production is to serve as an arena of discourse and a stage for action [the activity through which men reveal themselves to other men, and through which they achieve meaning as individuals], the interlocutors-and the audience obviously have to be allowed to enter the theatre (Piore 1990, p. 66).

Spatially distributed innovation communities help make explicit the motivations and governance that may also characterize many localized innovation communities but whose contribution might more easily be attributed to generic information spillovers. But by bringing out the essential role of community in various types of innovation, spatially distributed innovation challenges the adequacy of the linear model of innovation as a proxy or catchall of innovation that matters.

The social contract that has undergirded the linear model of innovation since World War II is similarly premised on a combination of the academic model and proprietary model, but in a way that treats these components as separable (Jasanoff 2009). Basic science under the academic model would be largely autonomous to pursue curiosity-driven research agendas meeting the rigors of peer review. Applied science and engineering would fall within the proprietary model, directed by entrepreneurial leadership to commercialize discoveries that contributed most to social welfare and corporate profits. Within this social contract, the process of innovation is not dependent on community but on two undemocratic searches: the search for truth and the search for profits. The governance of innovation within the linear model is driven by the market, where valued innovations survive and redundant or obsolete innovations perish, which may be accompanied by regulation to address information asymmetries. However, the imaginative aspect of innovation—determining what new technologies should do—is limited to entrepreneurial leadership that attempts to foretell the desires of consumers.

The Schumpeterian model that is implicit in the linear model likely does explain a good portion of innovation, but that portion would appear to be declining over the post-war period as problems requiring innovative solutions are increasingly contentious (e.g., cloning, climate change; Jasanoff 2009) and as innovations are increasingly dependent on the imaginative resources of communities to keep pace in directing them (von Hippel 2005). Addressing the first aspect—the increasing ethical challenges of innovation—is beyond the scope of this chapter. However, to the extent that the second aspect is reliant on the deliberative capabilities of a community, alternatives to the linear model of innovation may be more likely to support more democratic modes of technology assessment. The focus here is on how more inclusive, distributed forms of innovation have performance advantages over focused, expert-centered linear innovation and the governance structures that might be most conducive to promoting this type of innovation.

Both academic and popular research make the compelling case that protection of new ideas that pre-empts connection to a panoply of other new—often half-baked—ideas can limit innovation (Moser 2013; Johnson 2010). The metaphor that Johnson uses to motivate his discussion of the increasing importance of distributed innovation is the liquid network. The same network construct that defines how a new idea is recognized by the brain as a novel firing of synapses has its fractal parallel in the liquid network where multiple hunches from disparate brains collide.

The "places" of these liquid networks may range from coffee houses to World Fairs, and from conference rooms in corporate R&D labs to open source chat rooms. Innovation in this case is driven by the combined imaginations of many rather than by a singular commercial creative genius that defines entrepreneurial leadership.

A counter to Schumpeter's political economy argument that entrepreneurial leadership is the essential motive force comes from his contemporary, John Dewey. Dewey's view of the human imagination as the key to developments in science, technology and all other human endeavors had a profound effect on education. These same ideas in the context of a mass production economy in the first half of the twentieth century, driven by large increases in labor productivity and material wealth, had little purchase. In the current context of an innovation economy the insights from Dewey deserve reappraisal. The most powerful argument for this reappraisal comes from the title of Eric von Hippel's book *Democratizing Innovation*. The book powerfully demonstrates how innovation is becoming more distributed and potentially more inclusive—a descriptive use of democratizing—but does not address the potential functional advantages of democracy to innovation.

Dewey's notion of democracy as a way of life is perhaps easiest to understand as a counter to Schumpeter. In contrast to the classical notion that the truth of the universe is "out there" for elite scientists to discover, the pragmatic notion is that all discovery calls on "a new audacity of imagination." By focusing on imagination—a universal capability of all people—the emphasis is on the contribution of every individual. "Creative Democracy"—an essay written during the rise of Fascism and totalitarian states—makes the strong case that democracy is not merely a form of government that makes periodic demands on the civic duty to vote but instead permeates all social relations.

> Democracy is a way of life controlled by a working faith in the possibilities of human nature. Belief in the Common Man is a familiar article in the democratic creed . . . Democracy as compared with other ways of life is the sole way of living which believes wholeheartedly in the process of experience as an end and as mean; as that which is generating the science which is the sole dependable authority for the direction of future experiences and which releases emotions, needs and desires so as to call into being the things that have not existed in the past. For every way of life that fails in its democracy limits the contacts, the exchanges, the communications, the interactions by which experience is steadied while it is also enlarged and enriched (Dewey 1940, p. 2).

In its simplest terms, Dewey is expanding the dictum of "government of the people, by the people and for the people." The emphasis on the processes of origination and bringing new things into existence could also be expressed as "innovation of the people, by the people and for the people." The democratization of innovation that Dewey envisions is a continual work in progress. What is provided is a grand scheme of how this might work but is remiss in providing the organizational or institutional designs required to implement it (Sabel 2012). The governance of many open source communities and Marshallian industrial districts already point to ways in which democracy as a way of life is an essential component of community dynamism (Shah 2006; Piore 1990). However, developments in technology over the next 50 years are likely to make the democratization of innovation both more possible and more essential.

1.5 Opportunities and Challenges of Spatially Distributed Imaginariums

The opportunities of spatially distributed imaginariums are brought in bold relief by how the "world of production" and the "world of innovation" are stylized by conventional wisdom. The world of production is thought to be flat as a result of the rapid diffusion of information and production technologies that has enabled an unparalleled modernization in low-wage countries. In contrast, the world of innovation is thought to be spiky resulting from the agglomeration of highly specialized knowledge in global cities that are widely regarded as the essential engines of economic growth (World Bank 2009). What has yet to be widely recognized is the potential that the rapid diffusion of information technologies and design and prototyping technologies has for substantially raising the plateau of innovation outside of global cities. The fact that this potential was first introduced by Marshall more than 100 years ago makes conventional wisdom's current blind spot all the more blatant.

What is undeniable is that the resources for atomized invention and innovation outside of large agglomerations are expanding at a rapid rate (von Hippel 2005; Kennedy 2016):

> The exciting news for consumer-innovators is that it is getting steadily easier to commercialize an innovation oneself; you need not give up an attractive job or career you already have. Companies can be hired to produce your design in volume, to accept and process customers' orders and payments and to ship the completed product to the customers for you as well. It is a far cry from the all-consuming entrepreneurial effort that was required to perform these tasks in earlier days. In effect, the way has now been opened for the innovating consumer to be a "casual entrepreneur" (von Hippel 2005).

Recognition of the largely untapped innovation resource of lead users by business has resulted in the development of innovation platforms that allow users to demonstrate new uses of current products or modifications that expand a product's capability (von Hippel 2005). Unorganized versions of this have emerged in the form of hacks—clever new uses of products, a mode of exaptation and invention particularly suited to the modular products from IKEA. The Maker movement—represented by *Make* magazine, Maker Faire and makerspaces popping up in libraries and schools—point to a renewed interest in tinkering and inventing that harkens back to the heyday of *Popular Mechanics,* only this time with integrated circuits, lasers, and 3D printers.

These innovation tools, that are increasingly accessible to consumer innovators, will similarly expand the capabilities of small and medium sized businesses. However, starting as manufacturers (or producers) their focus of concern will be their ability to market innovations in products they sell. Here too capabilities are expanding. The tremendous growth in digital globalization provides new avenues for small firms to not only target niche markets but also to develop global networks (Manyika and Lund 2016). As familiarity with digital interaction increases, and as norms and protocols evolve to facilitate mutually beneficial interaction, the conventional wisdom on the advantages of face-to-face contact will need to be reassessed (Storper and Venables 2004).

What is sure to lag the development of new communications modes and technologies is data on just how they are being used. The huge benefit of localized innovation imaginariums noted above is that interaction can be assumed as an outcome of proximity. The data needs of studying spatially distributed imaginariums is just beginning to be addressed (Grillitsch and Nilsson 2015; USDA-ERS 2015). As the evidence base is in its infancy, it is helpful to posit some stylized conjectures suggested by findings in the literature reviewed in this chapter as a parallel to the stylized facts provided by Feldman and Kogler (2010):

- The creative spark required for disruptive innovation is randomly distributed
- Specific geographical contexts may facilitate the transformation of great ideas into implemented solutions
- Active search protocols that combine seemingly unrelated ideas increasingly displaces proximity engendered serendipity as the primary source of exaptations
- Some types of knowledge spillovers remain geographically localized

- Knowledge spillovers that can be spatially distributed increasingly exploit the genius of crowds
- Local universities that pursue curiosity-driven, normal science become irrelevant for innovation while universities that facilitate transdisciplinary approaches to wicked problems spur local innovation
- Socially desirable innovation benefits from democracy as a way of life.

The challenges or threats to spatially distributed innovation imaginariums all begin from the undemocratic tendencies of concentrated power that limit contacts, exchanges, communications and interactions. Barriers to entry are the most visible manifestation of this. Evidence from OECD countries suggest that the advantages to incumbency have never been higher as the share of firms under 5 years of age is at an all-time low (Criscuolo et al. 2014). In the urban U.S., the share of firms under 5 years declined from 46% of all firms in 1990 to less than 35% in 2012. The trend for rural firms is even more dire, declining from 42 to 27.2% over the same time period (US Census 2013). Despite the steady decline, young firms are still the sole source of net employment growth when aggregates of all OECD countries are examined, and the predominant source of net employment growth in the U.S. (Criscuolo et al. 2014). This suggests that the economy is not rewarding age, longevity, or experience in some way that breaks with historical patterns. Rather, the data suggest that the toe-hold for start-ups is becoming ever more tenuous.

One possible explanation for these trends is the displacement of small employer firms by freelance contract workers. Since non-employer firms are not included in the datasets examining business dynamics, the statistics may be an artifact of the rapid rise of the "gig economy." The possible implication of this are mixed: some types of spatially distributed innovation may be conducive to such freelancing, such as innovation tournaments that double as freelance rating agencies (Jeppesen and Lakhani 2010). On the other hand, increased importance of the gig economy to innovation would shift the residual claims to innovation from the innovative entrepreneur to incumbent firms, which retain the intellectual property rights to discovery made under contract.

The more explicit case of concentrated power limiting exchanges and interaction to the detriment of innovation is the crafting of intellectual property law and gaming intellectual property law to thwart entry by the construction of patent thickets (Foray 2004). Regulation, licensing and standard setting can also work at cross-purposes by favoring incumbents with poorer technical solutions but with more resources to lobby legislators or third-party auditors.

The scenario that provides the most compelling case for the necessity of democratizing innovation comes directly from the evolution of technology itself. The imminent, though never quite realizable, threat of robotics and artificial intelligence making huge swaths of the workforce redundant is likely to come to fruition sometime over the 50 years after publication of this book (Smith and Anderson 2014). The most conservative view suggests that the transformation will happen much later than most think simply because the mindset that will allow massive automation will evolve much more slowly than the technology that would allow it to happen. Distributed innovation as discussed here has the potential to order the innovative, cognitive, and manual divisions of labor in a way that resolves the potential conflicts between productivity enhancing robots and the common man. The mindset that may do the trick is the same one that John Dewey elaborated nearly 100 years ago. If imagination is the essential input to the process of innovation, and if all persons are capable of unleashing that imagination, then the human mind will remain the primary source for its continued social evolution.

In conclusion, it is important to place the democratization of innovation within the framework of regional science. Consider the pronouncements of the 2009 World Development Report (World Bank 2009). Although the book does not reflect the views of many regional scientists and economic geographers, it does provide a very thoughtful and cogent summary of new economic geography. That summary also comports with the prediction of the linear model of innovation. The key insight is that innovation is an activity limited to large urban agglomerations while lower order areas in the "portfolio of places" provide more routine production, logistics and service functions. If, in fact, innovation and imagination are only of consequence in global cities, then the prospect of fulfilling, meaningful economic lives everywhere else by 2066 is very dim indeed.

References

Amin A, Cohendet P (2005) Geographies of knowledge formation in firms. Ind Innov 12(4):465–486. doi:10.1080/13662710500381658

Azoulay P, Fons-Rosen C, Graff Zivin JS (2015) Does science advance one funeral at a time? NBER working paper no 21788, December. http://www.nber.org/papers/w21788. Accessed 16 May 2016

Bathelt H, Cohendet C (2014) The creation of knowledge: local building, global accessing and economic development—toward an agenda. J Econ Geogr 14:869–882

Bathelt H, Gibson R (2015) Learning in 'organized anarchies': the nature of technological search processes at trade fairs. Reg Stud 49(6):985–1002

Baumol W (2010) The microtheory of innovative entrepreneurship. Princeton University Press, Princeton, NJ

Beinhocker ED (2006) The origin of wealth: evolution, complexity and the radical remaking of economics. Harvard Business Press, Cambridge, MA

Boschma R (2005) Proximity and innovation: a critical assessment. Reg Stud 39(1):61–74

Boudreau KJ, Lacetera N, Lakhani KR (2011) Incentive and problem uncertainty in innovation contests: an empirical analysis. Manag Sci 57(5):1–21

Bush V (1945) Science—the endless frontier, a report to the president on a program for postwar scientific research. United States Government Printing Office, Washington, DC

Capello R, Lenzi C (2014) Spatial heterogeneity in knowledge, innovation, and economic growth nexus: conceptual reflections and empirical evidence. J Reg Sci 54(2):186–214

Criscuolo C, Gal PN, Menon C (2014) The dynamics of employment growth: new evidence from 18 countries. OECD science, technology and industry policy papers, no 14. OECD Publishing. http://dx.doi.org/10.1787/5jz417hj6hg6-en. Accessed 14 Sept 2015

Dewey J (1940) Creative democracy: the task before us. In: John Dewey and the promise of America, Progressive education booklet, no 14. American Education Press. http://pages.uoregon.edu/koopman/courses_readings/dewey/dewey_creative_democracy.pdf. Accessed 1 Feb 2016

Feldman M, Kogler D (2010) Stylized facts in the geography of innovation. In: Hall BH, Rosenberg N (eds) Handbook of the economics of innovation, vol 1. Elsevier, Amsterdam, pp 381–410
Fitjar RD, Rodiguez-Pose A (forthcoming) Nothing is in the air. Growth Chang. doi:10.1111/grow.12161
Florida R (2002) The rise of the creative class: And how it's transforming work, leisure, community, and everyday life. Basic Books, New York
Florida R, Gulden T, Mellander C (2008) The rise of the mega-region. Camb J Reg Econ Soc 1(3):459–476. doi:10.1093/cjres/rsn018
Foray D (2004) Economics of knowledge. MIT Press, Cambridge, MA
Glaeser EL (ed) (2010) Agglomeration economics. University of Chicago Press, Chicago
Glaeser E, Kallal HD, Scheinkman J, Shleifer A (1992) Growth in cities. J Polit Econ 100(6):1126–1152
Grillitsch M, Nilsson M (2015) Innovation in peripheral regions: Do collaborations compensate for a lack of local knowledge spillovers? Ann Reg Sci 54:299–321. doi:10.1007/s00168-014-0655-8
Jasanoff S (2009) Governing innovation. Paper presented at knowledge in question symposium. www.india-seminar.com/2009/597/597_sheila_jasanoff.htm. Accessed 16 May 2016
Jeppesen LB, Lakhani KR (2010) Marginality and problem solving effectiveness in broadcast search. Organ Sci 20:1016–1033
Johnson S (2010) Where good ideas come from: the natural history of innovation. Riverhead Books, New York
Kennedy P (2016) Inventology: how we dream up things that change the world. Houghton Mifflin Harcourt, New York
Lahani KR (2015) Innovation field experiments: empirical insights and policy implications from organizing innovation contests. NBER innovation and policy conference, 14 Apr, Washington, DC
Manyika J, Lund S (2016) Globalization for the little guy. McKinsey Global Institute. http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/globalization-for-the-little-guy. Accessed 31 Jan 2016
Maskell P (2014) Accessing remote knowledge—the roles of trade fairs, pipelines, crowdsourcing and listening posts. J Econ Geogr 14:883–902
McLaughlin N (2001) Optimal marginality: innovation and orthodoxy in Fromm's revision of psychoanalysis. Sociol Q 42(2):271–288
Merton RK (1942) Science and technology in a democratic order. J Leg Pol Sociol 1:115–126
Moser P (2013) Patents and innovation: evidence from economic history. J Econ Perspect 27(1):23–44
Packalen M, Bhattacharya J (2015) Cities and ideas. NBER working paper no 20921, January. http://www.nber.org/papers/w20921. Accessed 16 May 2016
Phelps E (2013) Mass flourishing: how grassroots innovation created jobs, challenge, and change. Princeton University Press, Princeton, NJ
Piore M (1990) Work, labour and action: work experience in a system of flexible production. In: Pyke F, Becattini G, Sengenberger W (eds) Industrial districts and inter-firm cooperation in Italy. International Institute for Labour Studies, Geneva, pp 52–74
Planck MK (1949) Scientific autobiography and other papers. Philosophical Library, New York
Rosenberg N (1991) Critical issues in science policy research. Sci Public Policy 18(6):335–346
Sabel C (2012) Dewey, democracy, and democratic experimentalism. Contemp Pragmat 9(2): 35–55
Shah SK (2006) Motivation, governance, and the viability of hybrid forms in open source software development. Manag Sci 52(7):1000–1014
Smith A, Anderson J (2014) AI, robotics, and the future of jobs. Pew Research Center. http://www.fusbp.com/wp-content/uploads/2010/07/AI-and-Robotics-Impact-on-Future-Pew-Survey.pdf. Accessed 1 Feb 2016
Steinbock D (2001) The Nokia revolution. Amacon Books, New York

Storper M, Venables AJ (2004) Buzz: face-to-face contact and the urban economy. J Econ Geogr 4:351–370

von Hippel E (1976) The dominant role of users in the scientific instrument innovation process. Res Policy 3:212–239

von Hippel E (1994) “Sticky information” and the locus of problem solving: Implications for innovation. Manag Sci 40:429–439

von Hippel E (2005) Democratizing innovation. MIT Press, Cambridge, MA

von Hippel E, Ogawa S, Jong JPJ (2011) The age of the consumer-innovator. MIT Sloan Manag Rev 53(1):27–35

US Census (2013) Business dynamics statistics http://www.census.gov/ces/dataproducts/bds/data.html. Accessed 16 May 2016

USDA-ERS (2015) 2014 Rural establishment innovation survey, Unpublished data

West J, Lakhani K (2008) Getting clear about communities in open innovation. Ind Innov 15(2):223–231. doi:10.1080/13662710802033734

World Bank (2009) World development report 2009: reshaping economic geography. World Bank, Washington, DC

Timothy R. Wojan is senior economist, Rural Economy Branch, Economic Research Service, U.S. Department of Agriculture. His primary research interests are innovation in rural areas, the attraction and impact of the creative class, contributions of the arts and cultural capital to development, statistical inference with small or limited samples, and the role of multi-level governance in development. Previous position were at the Governance Directorate of the OECD and at TVA Rural Studies, University of Kentucky. Dr. Wojan earned the Ph.D. in agricultural and applied economics from the University of Wisconsin-Madison in 1996.

Chapter 2
Exploring Innovation Gaps in the American Space Economy

Gordon F. Mulligan, Neil Reid, John I. Carruthers, and Matthew R. Lehnert

2.1 Introduction

The field of regional science emerged out of a need to understand the geographic disposition of economic activity within and among metropolitan areas—what Isard (1956) called the "spatial physiognomy" of development. Evolving on its own, with little or no intervention on the part of planners and other urban policy makers, the space economy has produced some of the most organized systems on the planet. For example, the northeast corridor—nothing less than an American "Prometheus," as described in Gottmann's classic (1961) text *Megalopolis*, an exploration of the Boston, MA—Washington, DC conurbation—is configured along an almost perfectly linear NE $\times$ SW axis $\sim$750 kilometers in length. A stylized map of the region, consisting of the population weighted $\{x, y\}$ coordinates of census block

G.F. Mulligan (✉)
School of Geography and Development, University of Arizona, Tucson, AZ, USA
e-mail: mulligan@u.arizona.edu

N. Reid
Department of Geography and Planning, University of Toledo, Toledo, OH, USA
e-mail: neil.reid@utoledo.edu

J.I. Carruthers
Sustainable Urban Planning Program, The George Washington University, Washington, DC, USA
e-mail: jic@gwu.edu

M.R. Lehnert
Ph.D. Program in Spatially Integrated Social Sciences, University of Toledo, Toledo, OH, USA
e-mail: matthew.lehnert@rockets.utoledo.edu

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_2

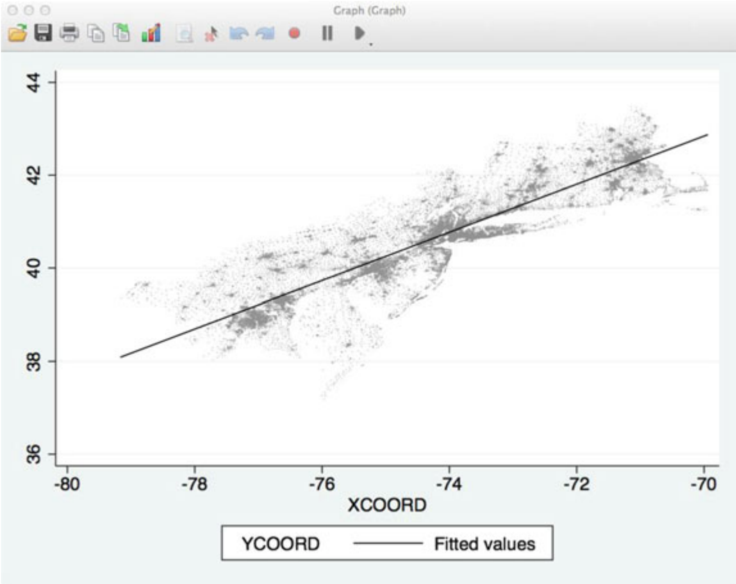


Fig. 2.1 Prometheus unbound

groups calculated by Carruthers et al. (2012), is shown in Fig. 2.1.[1] There are about 36,000 northeastern census block groups shown in the graph, and the correlation coefficient between the *x*-coordinates (longitude) and *y*-coordinates (latitude) is 0.91; moreover, a linear regression of latitude on longitude, plus a constant, yields an r^2 value of 0.82. Not only is the northeast region an economic titan, it is one of the most structured objects on the planet—so structured that it might be considered as natural as anything else created by life on earth. Though this organization is enforced, in part, by physical geography, the overall form is one of spatial agglomeration: economic forces acting to create a massive pattern of urbanization that is fundamental to the prosperity of the United States (Fujita 1989; Krugman 1991a, b, 1997; Fujita et al. 1999; Zenou 2009; Glaeser 2011; Fujita and François 2013).

At the national scale, a vast and enduring central place hierarchy has emerged out of the same (Gabaix 1999; Ioannides and Overman 2003, 2004; Gabaix and Ionniades 2004) agglomeration processes that gave rise to the *Megalopolis*. The overall pattern is shown in Fig. 2.2. While clearly the pattern is ordered around geographic advantage—for example, the transportation routes of the Great Lakes—and disadvantage—for example, the remoteness of the Great Plains (Krugman 1991a, b, 1997; Fujita et al. 1999)—the system is clearly organized (Christaller 1966; Lösch 1954; Beckmann 1968; Pred 1977; Berry and Parr 1988) and rank-

[1]For further details, see Renner et al. (2009).

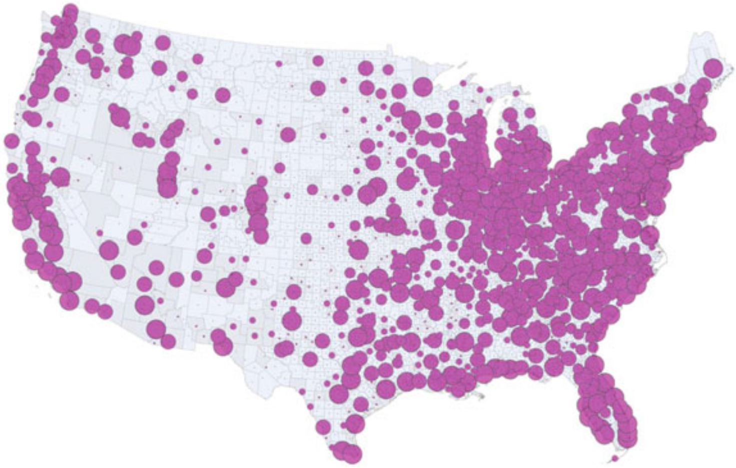


Fig. 2.2 The American city system

ordered (Zipf 1949) in the manner long predicted by location theorists (Mulligan et al. 2012). Nevertheless, the world has evolved since early theoretical frameworks were advanced—and the space economy along with it. Hyperbole from the heady days of the Internet explosion aside (see, for example, Cairncross 2001), economic growth in advanced economies like that of the United States has become increasingly uneven, demanding new frameworks for explaining and predicting the processes and outcomes of economic geography and urbanization.[2] The world itself is urbanized (Angel 2012) with spatial structures looking a lot like those depicted in Fig. 2.1 developing across the globe (Peirce et al. 2008; Nelson and Land 2011; Angel et al. 2012; Batty 2013). Increasingly, these frameworks are grounded in the concept of agglomeration economies, which give rise to specialization and patterns of development that are often unpredictable—new economic geography models typically predict "catastrophic" outcomes (see Head and Mayer 2004)—and highly specialized (Behrens and Robert Nicoud 2015; Duranton and Puga 2015; Combes and Gobillon 2015; Carlino and Kerr 2015).

With this history in mind, the present chapter presents an analysis examining the fact that, over the past few decades, the United States space-economy has separated into "have" and "have not" urban regions. Demographers, beginning with Frey (2002), have also noticed this break and have gone so far as to suggest that it may be irreversible. In the realm of economic geography, Florida (2014) and Moretti (2004, 2012), among others, have suggested that a permanent break has

[2]In the natural sciences, a good analogy is the relationship between Newton's laws of motion, which, to this day work well for many practical applications, but—their enduring power and utility—were shown to be incomplete and wholly supplanted by Einstein's theory of general relativity.

arisen in metropolitan America between growing areas that learn and innovate and declining areas that do not. This notion has been reinforced in a recent case study, by Storper et al. (2015), where the divergent paths of development taken by San Francisco and Los Angeles are documented in great detail. Still others have noted that a similar break has formed in the nation's non-metropolitan areas between the more urbanized micropolitan centers and the more isolated rural areas. In 50 years' time, these disparities should be even starker as growth and change typically follow dependency paths that have been established at earlier times.

Today, a number of private and public agencies monitor various aspects of economic performance in the nation's more than 350 metropolitan areas. Several websites exist where one city can be directly compared to another using different attributes of their labor markets: unemployment rates, average wages and salaries, recent employment growth, and the like. Increasingly, though, attention is being given to the notion of innovation—a term capturing the rise of high-tech industries and the importance of those factors that sustain high-tech industries: knowledge, creativity, advanced skills, entrepreneurship, research and development, patent production, technology transfer, and communications infrastructure. In fact, a growing number of observers now suggest that metropolitan areas should establish and nurture innovation ecosystems that include key actors (corporations, universities, etc.), service providers, venture capitalists, networks among the actors, and good local or regional governance. However, outside of various case studies, not much is known about how the complex features of innovation ecosystems might vary from one metropolitan center to the next.

This chapter argues that key input, output, and contextual attributes are combined locally to create each city's innovation ecosystem. These attributes not only differ from one place to another, but the manner in which they are combined varies by location. As a start on the project 20 variables were selected in order to discern how these innovation ecosystems currently vary across the nation's very largest metropolitan areas. Using multivariate techniques, these production ecosystems were deconstructed in order to reveal the underlying dimensions of innovation that are common to all of the nation's 350-plus metropolitan areas. Once the primary dimension of general innovation has been identified for all metros, the other $n - 1$ less important dimensions will reveal how specific cities differentially combine their knowledgeable and educated workforces to produce patents, engage in entrepreneurship, and create high value-added outputs. As a consequence, the overall innovative index of any metropolitan economy can be estimated by first generating its score on each of the latent dimensions and then, second, adding up those performance scores across all of the dimensions.

Today—despite advances in technology and the reduced *tyranny of distance* in advanced economies—geography matters more than ever (de Blij 2012). Cities and regions that are at the "have not" end of the spectrum described above may be able to improve their performance by pursuing policy strategies that enhance their connectivity around the globe. Economists increasingly emphasize competitive cities (see Glaeser and Joshi-Ghani 2015) as a means of achieving progress and the most recent *State of the World* report, published by the Worldwatch Institute, highlights the need for urban policy to not only be environmentally sound—but also keyed towards connectivity and inclusivity. Looking toward the future of regional science, it seems that competitiveness, inclusivity, and connectivity on the global stage will be fundamental to regional success, even as local innovation remains a main driver. Even as the analysis contained in this chapter illustrates how innovation works to establish regional productivity, the findings are used to look to a future wherein the world is smaller and evermore interconnected.

2.2 Background

There is, not surprisingly, a large literature dedicated to trying to understand why some cities and regions are more successful than others in building economies that are both vibrant and resilient and which provide a high quality of life for their residents. While no one would disagree with the premise that innovation plays a vital role in urban well-being, there are disagreements as to how to foster innovation within metropolitan economies (see Brueckner 2011 and Cheshire et al. 2014 for wide-ranging and recent overviews of the state of economic thinking on cities).

One school of thought, championed primarily by Richard Florida, emphasizes the importance of creativity, diversity and tolerance as key drivers of innovative economies. The basic argument is that such values drive innovation and urban growth by establishing a creative environment which attracts and retains the brightest and smartest—and most innovative—people (Florida 2014). Not surprisingly perhaps, Florida and his colleagues have found evidence in support of these ideas. For example, Lee et al. (2010) examined the role of human capital, creativity and diversity, and industry mix in explaining variations in innovation across 284 metropolitan statistical areas (MSAs). Their findings suggest that variations in innovation across MSAs are related to differences in levels of human capital, creativity and diversity while, at the same time, are not related to differences in industrial mix.

The impacts of not being able to attract workers with the right skill sets are discussed by Lautman (2011) who suggests that a shortage of qualified workers is creating a zero-sum labor market that results in communities "stealing" talent from each other. This scenario is the result of two failures. First, low fertility rates among the baby boom generation resulted in a labor shortage and second, the education system failed to equip the available labor force with the requisite skills demanded by the modern economy. Lautman argues that communities unable to attract qualified workers have bleak economic futures. Florida et al. (2011) found evidence that analytical and social intelligence skills are associated with higher wages and that individuals with such skills tend to be located in larger metropolitan areas. In sharp contrast, physical skills are associated with lower wages where such skills are being concentrated in smaller metropolitan areas. Florida et al. argue that skills are more important than either education or human capital in explaining geographic differences in wages (Florida et al. 2011). Members of the creative class (and by extension the communities in which they live) are more resilient than other workers to economic downturns (Gabe et al. 2012). In the case of the Great Recession, for example, this greater resiliency has been attributed to the fact that the creative classes held occupations that were less likely to be negatively impacted by the recession, and indeed, were bolstered by post-recession structural change. Those with creative class occupations were less likely to be unemployed during the study period, 2006–2011, than those holding non-creative working or service class occupations. Much of the work on the economic resiliency of the creative class suggests that it is the possession of skill sets, rather than formal education that is important (Gabe et al. 2012).

The role of institutions of higher education, particularly universities, has also been examined extensively in the literature. University contributions to economic development have been categorized into generative and developmental roles, with the former emphasizing knowledge-driven economic development and the latter focusing on capacity building (Gunasekara 2006). The generative role is derived from the Triple Helix concept in which the role of universities is expanded beyond the traditional ones of teaching and research to include economic development (Etzkowitz and Leydesdorff 2000). Through knowledge transfer and commercialization activities, in particular, universities can make important contributions to both local and regional economic development. These activities often manifest themselves in licensing agreements and spin-off companies. Not surprisingly, some universities and regions are more successful than others in creating spin-off companies (Calzonetti and Reid 2013). In regions with a track record of generating spin-off companies, there can be considerable benefit to the local economies. For example, the University of Utah created 188 spin-off companies between 1970 and 2010, 61% of which still had operations in Utah. These spin-offs have had a positive impact on the state economy through the creation of a significant number of jobs that pay well above the state average (Crispin 2010).

Knowledge spillovers have also been credited with contributing to entrepreneurship and innovation within urban areas. The Knowledge Spillover Theory of Entrepreneurship (KTSE) suggests that entrepreneurship is not exogenous but rather the result of the presence of knowledge spillovers. In other words "entrepreneurial behavior is a response to profitable opportunities from knowledge spillovers" (Acs et al. 2013: 759). Opportunities for entrepreneurial activity in the form of start-up companies present themselves when incumbent firms create knowledge that they do not commercialize (Audretsch and Keilbach 2007). This knowledge is appropriated by entrepreneurs who utilize it to establish start-up firms. Capitalizing on these knowledge stocks, however, requires a particular set of "skills, aptitudes, insights, and circumstances" that are neither ubiquitous nor uniformly distributed through the population (Acs et al. 2009). Carlino et al. (2007) explored the relationship between patent intensity (patents per capita) and employment density across American metropolitan areas.[3] Their findings suggest that the rate of innovation is enhanced

[3]It is worth noting that patent counts can be problematic for cross-country comparisons for a number of reasons. For example Japan applied a higher standard for judging innovation with the

where employment density is higher, the local economy is more competitive (many small rather than a few large firms), manufacturing jobs account for a larger share of total employment, and a larger share of the adult population have a college degree. Shapiro (2006) examined the role of college graduates on MSA employment growth. He found that a 10% increase in a MSAs concentration of college-educated residents was associated with a 0.8% increase in subsequent employment growth. College graduates impacted employment growth by enhancing productivity growth and the quality of life in an MSA.

While national scale studies can be highly informative and often bring valuable insights to the processes that underpin variations in urban economic performance, it is also important to remember that place, particularly the uniqueness of place, matters. All urban areas are influenced by their history and, as a result, are susceptible to lock-in and path dependency (David 1985; Arthur 1989). Each city has its own story to tell. At the local level it is important to understand this uniqueness and the historical backdrop of how a particular city reached the particular position that it now occupies. For example, Reese et al. (2014) document the negative influence of race relations and a “regimeless” governance culture in Detroit’s economic fortunes. In an equally compelling piece on Boston, Glaeser (2005) recounts that city’s fascinating economic history and highlights the critical role played by human capital, institutions of higher education, and labor force skill-sets in the city’s post-1980 renaissance.

2.3 Empirical Analysis

The main intent of the empirical analysis is to reveal the substantially different ways that large American cities now produce their various goods and services. A mixture of input, output, and contextual variables are used in the multivariate analysis of 352 metropolitan areas at one point in time. These groups of variables measure: (1) the quality of the workforce, (2) the incidence of entrepreneurship, (3) the intensity of patent production, (4) overall innovation, (5) overall productivity, and (6) the metropolitan context for innovation.

2.3.1 Workforce

Six different variables are chosen here (1) to reflect general creativity, (2) to capture the general benefits that flow from higher education, and (3) to highlight the special advantages of attracting and maintaining youthful, educated workers. First,

result that it is more stringent in awarding patents than a number of other countries, including the United States. See de Rassenfosse et al. (2016).

Richard Florida's widely-known twin measures of creative workers are adopted (see martinprosperity.org/tag/creativity-index). The share of metropolitan workers in the creative class, CCLASS, is used to capture the importance of having city workers engaged in the fields of science and technology, design and architecture, and various other professional areas. Alternatively, the creativity index, CREATE, is used to measure the incidence of the 3Ts—technology, talent, and tolerance—in the local environment. Although highly correlated (Pearson's $r = 0.80$), these variables are not identical. The third variable in this cluster is the standard measure of human capital, DEGREE, which reflects the percentage of the MSA adult population 25 years of age and older having a bachelor's degree. This variable correlates significantly (in the 0.55–0.60 range) with Florida's variables, but the incidence of college degrees is relatively high in certain places, like retirement areas, that are not very innovative.

Three other variables are used to address the special importance of attracting university-educated workers in the 18–44 age cohort. The annual Leading Locations studies provide ordinal data regarding this youthful "prime workforce" and both the current level and recent change in this demographic group are addressed (see www.areadevelopment.com). So data are adopted for: (1) the percentage share of the prime workforce, PRIMWF, in the total workforce; (2) the 3-year percentage change, CHPRIM, in the prime workforce; and (3) the 1-year percentage rate of in-migration, INPRIM, for this prime workforce. To facilitate the interpretation, all the ranks are reversed so that higher scores are better. PRIMWF and INPRIM correlate strongly with the variables in the first cluster; however, CHPRIM is a lot more volatile and does not correlate with any other variable in the entire group.

2.3.2 Entrepreneurship

While entrepreneurship is widely known to be a key in regional growth, there is some disagreement about how best to measure that activity. Three approaches appear to dominate the current literature: (1) self-employment rates, (2) new product formation rates, and (3) new business startup rates. Much research has argued that startups, (usually) taken as those businesses opened in the previous year, are an especially important part of Schumpterian "creative destruction" dynamics. Recent studies have focused on how activities (STEM) of a more technical nature are especially important in the creation of new jobs. Moreover, several observers have argued that high-tech multipliers are much higher than other multipliers. The analysis here makes use of data provided in the report prepared by Hathaway (2013) for the Kauffman Foundation. This report is especially valuable because it differentiates general startups GENTRE from high-tech startups TENTRE, although the levels do correlate highly ($r = 0.95$ in 2010). Here startup density figures, or location quotients using a national base, are given for each measure of entrepreneurship in 1990 and 2010. The 2000–2010 changes in those concentration measures, CHGENT and CHTENT, are estimated by subtracting the 1990 densities from the 2010 densities and dividing that change in half. The degree of correlation for the density changes (r = 0.78) is somewhat lower than for the densities.

2.3.3 Patents

Patent production has been studied for years, especially in the more advanced economies. The analysis here uses the figures disclosed for utility patent grants in all classes, as provided annually by the United States Patent and Trademark Office (2015). Here the metropolitan location is attributed to the residence of the first-named inventor. The website www.uspto.gov provides figures for the patent volumes, which correlate highly with city size, and these figures must be standardized to address local specialization. The raw patent figures PATENT for 2010 and 2000 were transformed into patent densities LQPATE for each year by computing location quotients using the local and national population figures. Next, the change in patent density CHLQPA was computed as the 2010 location figure minus the 2000 location quotient. Patent densities correlate significantly with entrepreneurship densities (r is approximately 0.50 in both instances), but there is virtually no correlation between the alternative measures of change. Finally, a composite variable ENTPAT was computed as GENTRE*LQPATE to capture the possible synergy effects between entrepreneurial activity and patent production in 2010.

2.3.4 Overall Innovation

A comprehensive view of metropolitan innovation is given by the composite variable, INNOVA, which is calculated by Statsamerica (see www.statsamerica.org). This index is comprised of four distinct sub-indices—human capital, economic dynamics, productivity and employment, and economic well-being—where each sub-index is in turn comprised of several performance indicators. The human capital sub-index comprises 30% of the total score and, as considered above, this reflects both the age and the education of the workforce. The economic dynamics sub-index, which comprises a further 30%, is different from all of the other variables used here because it addresses venture capital and broadband connection. The other two sub-indices are to a large degree captured in the variables that are considered below. The basic problem with this composite variable is that many of the input data are known to be highly correlated (redundant), but it remains unclear where this double-counting actually occurs.

2.3.5 Overall Productivity

The nation's metropolitan areas vary tremendously in terms of their per capita productivity. In fact, gaps in urban productivity have been growing steadily for decades. To capture this aspect of the nation's metropolitan landscape, the analysis makes use of productivity figures provided each year by the Bureau of Economic Analysis (see www.bea.gov). Highly innovative cities generally have high levels of gross domestic product per capita, GRDPPC, but not all highly productive cities are in fact highly innovative. This occurs, in part, because some cities remain productive in high value-added industries that are in the declining part of their product cycle. In our case, the degree of correlation between GRDPPC and INNOVA is $r = 0.64$.

2.3.6 Metropolitan Context

Four other variables are included in the multivariate analysis in order to provide suitable context for the different variables that have been discussed above. The first of these is metropolitan population, POPULA, where it is known that the largest cities, like New York and Chicago, have very different production ecosystems—with highly beneficial internal and external scale economies—than the smallest cities, like Yuma and Dothan. A second variable, WAGSAL, addresses average wages and salaries across the cities, and is adopted to recognize that qualitative differences exist in the metropolitan workforces. Moreover, new patent production and entrepreneurship tend to be promoted in those areas that already enjoy higher wages and salaries. Two other variables are chosen to control for the highly variable composition of both metropolitan income and employment in the United States. One variable, EARRAT, measures the ratio of earned income to total income and thereby distinguishes retirement places and depressed areas, with high amounts of non-earnings income, from the nation's most vibrant and productive cities. A second variable, EMPRAT, measures the ratio of total employment to population and distinguishes those productive places with high (gross) labor force participation rates from those unproductive places with low participation rates. These variables are all taken from or manufactured from BEA data for the year 2010.

2.3.7 Empirical Results

Table 2.1 shows the factor loadings for each of the 20 variables across 10 latent dimensions (factors) after applying a Varimax rotation. These numbers basically represent the degree of correlation of the variables with the different orthogonal

Table 2.1 Loadings for the ten orthogonal factors

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
CCLASS	0.65	0.32	0.07	0.13	0.10	0.42	0.02	0.00	0.01	0.35
CREATE	0.75	0.20	−0.04	0.21	0.01	0.32	0.03	0.04	0.15	0.34
DEGREE	0.81	0.15	−0.01	0.20	0.20	−0.08	0.05	0.06	0.29	−0.13
PRIMWF	0.66	0.19	0.08	0.18	0.26	0.56	−0.06	−0.10	−0.01	−0.06
CHPRIM	−0.02	−0.02	0.05	0.02	0.12	0.09	0.02	−0.99	0.01	0.01
INPRIM	0.86	0.09	0.15	0.11	0.18	−0.07	0.05	−0.01	−0.03	0.05
GENTRE	0.45	0.44	0.38	0.16	0.18	0.17	−0.02	−0.02	0.58	0.16
CHGENT	0.04	0.10	0.93	−0.13	0.05	−0.09	−0.09	−0.05	0.12	−0.07
TENTRE	0.45	0.41	0.39	0.12	0.17	0.14	−0.01	−0.02	0.60	0.12
CHTENT	0.06	−0.09	0.92	−0.05	0.05	0.09	−0.02	−0.05	0.12	−0.07
PATENT	0.14	0.55	−0.13	0.75	0.12	0.05	0.09	−0.05	0.05	0.13
LQPATE	0.30	0.88	−0.05	0.05	0.10	0.14	0.14	0.04	−0.04	0.11
CHLQPA	0.03	0.17	−0.04	0.03	0.00	−0.02	0.98	−0.01	0.00	0.02
ENTPAT	0.15	0.90	0.07	0.08	0.11	0.02	0.10	−0.01	0.28	0.04
INNOVA	0.60	0.46	−0.03	0.21	0.21	0.23	−0.10	0.03	0.22	0.32
GRDPPC	0.24	0.21	0.01	0.26	0.76	0.21	0.05	−0.04	0.13	0.39
POPULA	0.17	0.04	−0.09	0.95	0.03	0.11	−0.02	−0.06	0.07	0.06
WAGSAL	0.27	0.35	−0.16	0.44	0.35	0.13	0.06	−0.01	0.15	0.60
EARRAT	0.08	−0.10	−0.02	0.13	0.42	0.80	−0.02	−0.13	0.11	0.12
EMPRAT	0.30	0.09	0.16	−0.03	0.85	0.26	−0.02	0.01	0.05	−0.06

dimensions. The ten rotated factors, identified later, are arranged in order of their declining importance in accounting for the variance in the 20 × 352 data matrix; together the ten factors account for more than 92% of that variance. Many variables load heavily and few lightly on Factor 1: hence, it is recognized as the primary underlying dimension of the factor analysis. In fact, this single factor accounts for nearly 20% of the variance in the data set. The frequent large loadings reveal that this factor represents youthful, educated, and talented workers who are very active in entrepreneurship, patent production, and overall innovation.

The other latent dimensions of the analysis are identified in Table 2.2, where the amount of the total variance accounted for by each factor is also shown. Here the second factor, F2, which accounts for 14.5% of the variance, captures those creative workers that are very involved in patent production despite the fact that these workers have lower levels of university education than in the first factor. Factor scores, which are standardized, are computed in order to see how the various metropolitan areas perform on each factor. Table 2.3 shows that Ithaca (3.02) and Ann Arbor (2.57), both university-dominated cities, are ranked highest on F1, while San Jose (13.49) and Boulder (6.10), both well-known high-tech centers, are ranked highest on F2. The various loadings in Table 2.1 indicate that while Ithaca and Ann Arbor have somewhat younger workforces, San Jose and Boulder are much more productive in patents. Looking down the top-15 lists for F1 and F2, two places—Ann Arbor and Corvallis—actually appear twice. So these two cities are identified as being especially innovative in that entrepreneurship, patent production, and GDP per capita are all rated very highly.

The revealed factors or dimensions basically deconstruct the various metropolitan economies in ten separate ways. In a sense, then, the city's score on each dimension reveals something unique or important about the innovation ecosystem of that metropolitan area. Some metros are productive but not so entrepreneurial, others are entrepreneurial but not so active in patent production, while still

Table 2.2 Interpreting the ten orthogonal factors

Factor	%Variance	Interpretation
F1	19.7	Young, educated, and highly talented innovators
F2	14.5	High patent production with high entrepreneurship
F3	10.7	Ascending entrepreneurial
F4	9.8	Large cities with diverse patents and entrepreneurs
F5	9.6	Young workers in very high valued-added industries
F6	7.6	Young, talented workers in high value-added industries
F7	5.2	Ascending in patent production
F8	5.1	Significant rise in young workers
F9	5.0	High entrepreneurship but low patent production
F10	4.8	Talented workers with moderate entrepreneurship and patents

F8 is flipped (reversed) for interpretation reasons

Table 2.3 Top-15 metro areas on each factor

Factor 1	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*
Ithaca	San Jose	Boulder	New York	Midland TX
Ann Arbor	Boulder	Ft Collins	Los Angeles	Casper
Charlottesville	Boise City	Cheyenne	Chicago	Bridgeport
Corvallis	Corvallis	Missoula	San Francisco	Sioux Falls
Olympia	Burlington VT	Corvallis	Philadelphia	Ocean City
Gainesville FL	San Francisco	Champaign	Boston	Columbus IN
College Station	Santa Cruz	Bend	Dallas	Charleston WV
Naples	Rochester MN	Dover	Houston	Dubuque
Durham	Austin	Sioux Falls	Miami	Napa
Tucson	Seattle	Seattle	Seattle	San Jose
Lawrence	Bremerton	Colorado Sp	San Jose	Fargo
Iowa City	Rochester NY	Grand Junction	Atlanta	Lake Charles
Prescott	Fort Collins	Boise City	Detroit	Atlantic City
Athens	Ann Arbor	Kansas City	Washington	Sioux Falls
Washington	Raleigh	Crestview	Phoenix	Madison
Factor 6	*Factor 7*	*Factor 8*	*Factor 9*	*Factor 10*
Manchester	Bremerton	Coeur d'Alene	Boulder	Bridgeport
Provo	Corvallis	Harrisonburg	Manchester	Durham
Worcester	San Jose	Warner Robins	Provo	Trenton
Burlington VT	Seattle	Florence	Denver	Washington
Ames	Burlington VT	Lawrence	Salt Lake City	Midland TX
Rochester MN	Rochester MN	Hot Springs	Washington	Hartford
Des Moines	Bloomington IL	Casper	Dallas	Anchorage
Columbus OH	Reno	Kingsport	Fort Collins	Kokomo
Raleigh	Santa Fe	Bay City	Colorado Sp	San Francisco
Morgantown	Champaign	Spartanburg	Phoenix	Hinesville
Springfield IL	Spartanburg	Ogden	Las Vegas	Kennewick
Atlanta	Iowa City	Michigan City	Orlando	Palm Bay
Akron	Portland ME	Nashville	Ann Arbor	Boston
Santa Cruz	Tucson	Grand Forks	Huntsville	Seattle
Greenville NC	Provo	San Angelo	Cheyenne	Huntsville

Factor F8 has been flipped for interpretation

others are simply large and perform about average in both patent production and entrepreneurship.

After examining all of the top-15 lists in Table 2.3, it becomes apparent that some metropolitan areas appear much more frequently than others. For instance, San Francisco appears on the top-15 lists for F2, F4, and F10 and Seattle appears on the top-15 lists for F2, F3, F4, and F10. So these two large metros demonstrate a strong similarity in their overall innovative ecosystems. In fact, a vector of the ten ranked scores (where lower is better) provides a simple signature for the innovation ecosystem of each and every metropolitan area. In the case of San Francisco, this vector is (181, 6, 89, 4, 75, 229, 15, 137, 87, 9), and in the case of Seattle this vector is (185, 10, 10, 10, 117, 104, 4, 91, 25, 14). Here the only significant rank difference is seen on F6 where Seattle has substantially younger workers. The pattern of ranks in both of these vectors is very different from that of a moderately innovative place like Tucson, AZ (10, 135, 316, 121, 324, 267, 13, 83, 270, 131) or a weakly innovative place like Dalton, GA (351, 131, 113, 193, 169, 241, 278, 102, 45, 84).

Earlier in the chapter it was noted that an innovation index can be constructed by assessing performance on each dimension and then aggregating this performance across the different orthogonal dimensions. Factor scores, FS, which are standardized across the cities, can be used to gauge performance on each dimension and then these factor scores can be summed. However, there is some disagreement about whether or not these scores should be weighted by the importance of each factor before the summation occurs. So in the first instance the unweighted index INNOV1 is calculated as:

$$\text{INNOV1} = \text{FS}_1 + \text{FS}_2 + \cdots + \text{FS}_{10}$$

while in the second instance the weighted index INNOV2 is calculated as:

$$\text{INNOV2} = \text{w}_1\text{FS}_1 + \text{w}_2\text{FS}_2 + \cdots + \text{w}_{10}\text{FS}_{10}$$

where the weights are calculated by taking either the square root of each dimension's (rotated) eigenvalue or the square root of the percentage of the variance that is accounted for by that dimension. So the first index treats all factors equally, while the second index places greater emphasis on those factors that are found to be more important in the analysis. However, this added importance might only reflect the fact that the redundancy of having more highly correlated variables is being extracted by those factors with larger eigenvalues.

An examination of the results indicates that our methodology penalizes those metros that have experienced downturns in either entrepreneurship or patent production in recent times. If a metropolitan area has a large negative score on Factor 5 or Factor 7, then its overall performance can be severely compromised. Considerable drops in entrepreneurship depress the overall innovation scores for Riverside, CA and Provo, UT while considerable drops in patent production depress the overall innovation scores for Austin, TX and Boise City, ID. So if the analyst preferred an index that did not address change, then both INNOV1 and INNOV2 could be recalculated with the dimensions F5 and F7 completely removed from the formula.

Table 2.4 Top 10% innovators: unweighted and weighted scores

Rank	City	INNOV1	City	INNOV2
1	Boulder	17.68	Boulder	53.21
2	San Jose	14.94	San Jose	50.43
3	Seattle	14.11	Seattle	37.34
4	San Francisco	11.74	San Francisco	35.74
5	Washington	11.71	Washington	32.97
6	Denver	9.30	Corvallis	30.23
7	Corvallis	9.11	New York	29.19
8	Durham	8.20	Denver	23.29
9	New York	8.00	Durham	22.77
10	Portland OR	7.11	Boston	21.49
11	Boston	6.28	Fort Collins	20.01
12	Kansas City	6.27	Burlington VT	19.90
13	Salt Lake City	6.26	Raleigh	19.39
14	Des Moines	6.11	Trenton	17.34
15	Cheyenne	5.78	Portland OR	17.30
16	Raleigh	5.61	Des Moines	17.07
17	Chicago	5.52	Kansas City	16.81
18	Burlington VT	5.48	Cheyenne	16.70
19	Cedar Rapids	4.91	Los Angeles	15.45
20	Los Angeles	4.87	Atlanta	15.10
21	Sioux Falls	4.82	Huntsville	15.04
22	Atlanta	4.79	Chicago	14.77
23	Huntsville	4.43	Colorado Sp	14.72
24	New Orleans	4.42	Champaign	14.67
25	Charlotte	4.39	Bremerton	14.61
26	Philadelphia	4.16	Philadelphia	14.34
27	San Diego	4.09	Salt Lake City	13.92
28	Baltimore	4.04	San Diego	13.46
29	Jacksonville	4.02	Richmond	13.45
30	Provo	3.96	Rochester MN	13.18
31	Casper	3.95	Missoula	12.87
32	Bridgeport	3.94	Bridgeport	12.52
33	Richmond	3.86	Minneapolis	12.48
34	Minneapolis	3.83	Provo	12.35
35	Portland ME	3.81	Santa Fe	12.33

INNOV2 is calculated using the square root of the percentage of the variance extracted by each factor

In any case, Table 2.4 shows these innovation indices for the top 35 metropolitan areas. Figures 2.3 and 2.4,which accompany the table, show the cities that make it on the list twice and once, respectively. Here the scores on Factor 8 have been inverted or flipped as was noted earlier. The lists are much the same, but there are

Fig. 2.3 Metropolitan areas listed twice in Table 2.4

Fig. 2.4 Metropolitan areas listed once in Table 2.4

seven metros (20% of the total) on one list that are not on the other; in fact, the degree of correlation between the two indices across all of the metropolitan areas is $r = 0.92$. So 28 cities qualify as a top 10% innovator on both indices, while a further 14 cities qualify as a top 10% innovator on one of the two indices.

When using INNOV1 Boulder, CO stands out alone at #1 (its score is 17.68), San Jose and Seattle then group closely together at #2 and #3 with scores between 14.0 and 15.0, and then San Francisco and Washington group together at #4 and #5. After that the scores slowly descend until reaching the two least innovative cities: Naples, FL (−4.95) and Punta Gorda, FL (−5.79). Alternatively, when using INNOV2, Boulder (53.21) and San Jose (50.43) distance themselves from the remainder of the pack; five other places have scores ranging between 29.0 and 38.0, and only then does a pattern of steady decline appear in the weighted index. Now the two least innovative cities in the nation are identified as being Sumter, SC (−14.11) and Sheboygan, WI (−14.20).

Although this can only be a rough indicator of geographic differences in metropolitan innovation, the average index values for the 117 Sunbelt cities are calculated to be INNOV1 = −1.14 and INNOV2 = −0.52, while the average index values for the 235 Snowbelt cities are INNOV1 = 0.53 and INNOV2 = 0.26. So the Sunbelt appears to be marginally less innovative than the Snowbelt. Moreover, separate OLS regressions show that this spatial distinction is significant for INNOV1 at the 0.05 level and for INNOV2 at the 0.15 level. However, once these OLS regressions include metropolitan population as a second independent variable, the Snowbelt-Sunbelt distinction becomes significant at the 0.05 level in both instances; moreover, the new results indicate that a strong positive relationship exists between innovation and metropolitan population in both cases. The nation's most innovative metropolitan areas are also mapped out in order to discern whether or not these metros follow some sort of geographic pattern.

One of the major shortcomings of indices is that they often remain descriptive tools and are infrequently used to make testable inferences—in this case about other attributes of metropolitan labor markets. Regional science should become interested in how the two constructed innovation indices correspond to the most widely-used measures of health in labor markets: those being the current unemployment rate, recent employment growth, recent wage growth, and recent per capita productivity growth. Moreover, the field should be interested in how the performances of our two indices compare to the performances of the three input variables CCLASS, CREATE, and INNOVA that were used in their construction.

For assessing (and not predicting) recent growth, the time period 2007–2013 was selected, in order to include several years both before and after the year 2010. Table 2.5 shows some Pearson correlation coefficients between the two innovation indices calculated above and these four important labor market attributes. The table also shows correlation coefficients for the three input variables just mentioned. Evidently, a strong relationship holds between all the measures and the 2010 unemployment rate across the nation's metropolitan areas, where metros with more talent or higher innovation rates enjoyed significantly lower unemployment rates. Very clearly, too, the five measures are strongly correlated with recent employment

Table 2.5 Correlations between innovation measures and labor market attributes

Index	Un rate 2010	Job growth	Wage growth	GDP growth
INNOV1	−0.29**	0.29**	0.13*	0.11*
INNOV2	−0.36**	0.32**	0.12*	0.08
CCLASS	−0.33**	0.20**	0.01	0.06
CREATE	−0.18*	0.20**	−0.11*	−0.05
INNOVA	−0.29**	0.21**	−0.01	0.02

Growth is during 2007–2013; ** significant at 0.01 level; * significant at 0.05 level.; n = 352

growth, although the weighted index INNOV2 outperforms the others by a fair amount. However, the two new indices are clearly superior to the three input variables in the areas of wage growth and per capita GDP; in fact, CREATE actually has a significant negative relationship with growth. Evidently, innovation is now a determining factor in various aspects of metropolitan growth and change, but the full extent of its effects will have to be established by more sophisticated econometric models. But, for present purposes, it is worth stating that OLS regressions, using logarithms, indicate that a 1% increase in either innovation index decreases the unemployment rate by approximately 0.15% and increases employment growth by approximately 0.13%.

2.4 Challenges to Regional Science

The multivariate analysis undertaken above clearly indicates that U.S. metropolitan areas currently exhibit a great amount of heterogeneity in the composition of their high-tech economies. This means, among other things, that the prospects for future metropolitan growth are highly polarized where, very clearly, there will be winners and losers. These findings suggest that regional science is now confronted with several theoretical problems and empirical issues that must be addressed before policymakers can discern the degree to which the metropolitan space-economy will be divided into two broad welfare camps. These topics can be roughly categorized as follows: (1) understanding how urban creativity leads to innovations; (2) understanding how these innovations create new industries; (3) understanding how these new industries are transformed into old industries; and (4) understanding how global connectivity can capitalize on local agglomeration forces. At the same time, it will be important for regional science to understand how human capital, startups, and patent production, among other factors, interact in order to drive city growth during all three of these stylized phases or time periods.

2.4.1 Creativity

Human creativity is uncharted territory for regional science because, outside of a few contributions by people like Ake Andersson (2009), little has been written about why new ideas often appear only at certain places and only at certain times. Here, the discipline of psychology has much to offer and regional science should begin thinking—much like the field of behavioral economics has—about how to extend standard models in psychology to incorporate features of the economic landscape like population density and the age-gender composition of the population.

A good place to begin is with the research of Dean Simonton (1999) on Darwinian approaches to creativity. In one paper Simonton (1997) outlines a two-step model that addresses time-specific rates for both ideation (new ideas) and elaboration (their application). This leads to the construction of a productivity curve $P(t)$ for each person where:

$$P(t) = c\,[\exp(-at) - \exp(-bt)]$$

Here, a is the ideation rate and b is the elaboration rate and both of these rates have negative signs because the fixed stocks (representing all combinatorial possibilities) of new ideas and their elaborations are both being continuously exhausted over time. The constant $c = abm/(b - a)$ where m represents the maximum number of ideational combinations the person can conceive in a lifetime. So the genius or long-lived person can conceive of many more possibilities than can the average or short-lived person and m can be set much higher for these cases. So the stylized productivity curve rises quickly at first, then peaks a number of years later (usually at 20–25 career years), and eventually falls, but at an ever-decreasing rate. The rates for ideation and elaboration ($0 < a,b < 1$) will both vary from one activity (industry) to another so the two-step model actually exhibits a fair bit of flexibility.

It would be a useful—but not entirely a straightforward—exercise to expand this individual model to include several matters of interest to regional scientists. One important step would involve identifying where so-called geniuses prefer to live. Our current thinking is that superstars prefer living in highly urbanized environments in order to enjoy interaction with other highly endowed people and to maximize their opportunities to find a compatible partner. However, there are indications that many of these highly talented people actually prefer some mixture of urban and non-urban lifestyles. At the very least, there is a need to better understand how superstars selectively demand the key attributes provided by these two very different residential settings. Simonton's model could be modified to portray a hierarchy of creativity by using the shift parameters in his model while recognizing that large cities have more industries than small cities. Another consideration would be some appreciation of demographic heterogeneity. Metropolitan areas have different numbers of persons in various age-gender cohorts and Simonton's model could be adapted to address this fact. So the city-wide productivity curve for new ideas and elaborations would necessarily aggregate across all the various individual productivity curves for each resident of the city. One more important consideration would be population density. A lot of the thinking in the NEG framework is based on the notion of externalities arising from agents that are located close together in space. Presumably Simonton's model could be modified so that both of the rates *a, b* would rise with higher density and this in turn would raise the city's aggregate productivity curve to a higher level.

2.4.2 New Industries

Economists now regularly differentiate between those innovations that are more important for subsequent growth and those that are not. Following Schumpeter, the former—and much more influential group of commercial applications—are usually called disruptive or radical innovations. Uber, the new taxi service, is an example of a deliberately disruptive business model that seeks to transform an industry as old as vehicles themselves through innovation and networking (Cramer and Krueger 2016). Such innovators, though often unwelcome, are usually not only rewarded but often transform existing practices and rules of business organization and industrial/service production. The emerging literature on this topic reminds one of Thomas Kuhn's vision of scientific progress, where normal science and revolutionary science become separated by entirely different paradigms

It is increasingly apparent that some firms, cultures, nations, and cities are simply more open and receptive to new ideas and practices than others (Hofstede 1997). Furthermore, there are variations in the ability of a firm to take new ideas and utilize them—in other words absorptive capacities vary across individuals, firms, industries, and cultures (Cohen and Levinthal 1990; Spithovena et al. 2011). The field needs a better understanding of the determinants of absorptive capacity at the scales of both the individual and the firm (Schmidt 2005). Of course, this position has been established in the metropolitan setting by Richard Florida who has long maintained that the tolerance exhibited by the creative class is very important for sustained urban growth. But mainstream economists now recognize the importance of such factors as openness and tolerance and have begun to develop models where agents actually choose between incremental and radical innovations depending upon their innate attributes, including their age, level of human capital, and the attitudes they hold toward risk (Acemoglu et al. 2015).

Radical innovations, which can be stymied by both pecuniary and non-pecuniary means, are absolutely critical for continued economic success in our cities, and these are the innovations that will increasingly differentiate the fortunes of our metropolitan areas well into the future. Along these lines, Storper et al. (2015) have argued that one of the reasons Greater San Francisco has outpaced Greater Los Angeles in recent decades is because the private and public officials of the Bay area were generally open to new ideas and practices, and were also able to cooperate with one another in a constructive way. In other words, innovations are fostered when openness occurs in both the firms and the wider institutions (networks) of society. While Social Network Analysis (SNA) has been utilized by social scientists (mainly sociologists) since the 1940s (Freeman 2011), it is only more recently that geographers and regional scientists have started to use SNA to enhance our understanding of innovation and regional economic development (Fritsch and Kauffeld-Monz 2010; Ter Wal and Boschma 2009). SNA holds considerable potential to unlock significant insights into the role of networks in the innovation process.

An obvious challenge to regional science is to develop models that shed light on how the location-specific balance between normal and disruptive innovations is reached. As in the handful of firm-level innovation models that already exist, visible attributes like age and education of the workforce will clearly be important drivers of both the generation and acceptance of radical innovations. These agent-based models should also allow key personnel to move upwards in the same firm or, alternatively, move sideways to a different firm in order to accomplish their disruptive innovations. As a result, externalities will likely once again play an important role in these sorts of urban innovation models. The overall risk-avoidance attitude of the urban population will also be an important factor in this process. Moreover, adopting a learning-by-doing approach, those cities that already enjoy successful high-tech industries will be more open to radical innovations than others, and this factor alone should introduce an element of circular and cumulative change to the future geography of the most important innovations.

2.4.3 Transformation of Industries

One of the greatest challenges to regional science over the next 50 years will be to understand the ever-changing nature of industries as they first appear, then mature, and eventually become declining industries The traditional capitalist model, based on the notion of the product cycle (Vernon 1966), is that new industries generally arise only in those places with favorable incubation properties—high human capital, support services and more—and then diffuse outward to other locations with less costly labor as they become sunset industries and the production process is rationalized. By the 1990s this model, originally envisaged to serve domestic purposes, had been nicely adapted to explain the global dispersal of industry from the core manufacturing regions of Anglo-America and Western Europe to the world's more peripheral developing regions. However, this stylized model has now proved to be somewhat inadequate for several reasons.

Mainly, manufacturing has simply become less important—relatively speaking—in advanced economies where more jobs are now available in trade and the services. In fact, most top-20 lists of future occupations will mention a litany of service jobs, particularly those that are health-related, across the various private and public service industries (see for example United States Department of Labor 2015). It seems, too, that industry numbers might rise or fall at certain locations but occupation numbers remain a lot more stable.

More theoretically, as captured in the thinking of Dixit-Stiglitz-type models, consumers have shown a great preference for more and more variety in the different goods and services that national economies now provide. In fact, this demand for variety will likely drive the next industrial revolution in goods production as identity-minded consumers strive to differentiate their personal products as much as possible. There is evidence that many consumers are primarily interested in goods and services that they consider *authentic*—with authenticity being defined in terms of products that are made by small-scale independent specialist producers, often using high quality inputs (Kovács et al. 2014; Lewis and Bridger 2000). The explosive growth of the craft beer market in the United States and elsewhere exemplifies this trend (Moore et al. 2016). Twenty-one percent of beer sales in the United States in 2015 were accounted for by the craft segment (Brewer's Association 2016). The rising popularity of craft beer and other craft products creates a market opportunity that has also been explained in terms of resource partitioning theory (Carroll et al. 2002) and has led to what Jovanovic (2001, 108) refers to as "variety proliferation". A complex interplay of factors seems to be driving the demand for these products including the demand for customized goods and services, anti-mass production sentiments, and the status conferred by being seen to be the owners/consumers of certain products (Carroll et al. 2002). These changing market dynamics may mean challenging times ahead for some large-scale producers. In the case of the brewing industry, large-scale brewers have already been jolted into action: AB InBev's $104 billion takeover of SABMiller is partly a response to declining market share in the U.S. and other western markets. Purchasing SABMiller gives AB InBev access to the former's production and distribution networks in future growth markets, such as Africa (Shadbolt 2015; Forbes, 2015). Changing consumer pattern is partly driven by growing importance of the millennial demographic. Described as "confident, self-expressive, liberal, upbeat and open to change" (Pew Research Center 2010: 1), Millennials number 75.4 million and are the largest age cohort in the United States (Fry 2016). Understanding the millennial consumer will be critical to future corporate success—as consumers, millennials are highly dependent upon social media for information, feel that there is too much power vested in large corporations, prefer to purchase from companies that support solutions to specific social issues, and often seek the opinions of peers (often via social media) before making a purchasing decision (Barton et al. 2012; Fromm et al. 2011; Cone Communications 2013; Winograd and Hais 2014).

Finally, nearly all goods production has become highly roundabout where vast international supply chains—networks—are normally involved in the production of final consumer goods. But, when communication and transportation allow just-in-sequence assembly, the pervasive logic of the product cycle model will invariably assert itself. However, the dispersal of new industries to the periphery might be delayed for some time whenever skill-intensive, batch production in the center still provides cost advantages over unskilled, Fordist production in the periphery. This fact is evident in the fashion industry where the production of women's clothing is only now departing from Los Angeles which, in contrast with New York, chose to specialize in low-quality, fast fashion clothing. In fact, New York's alternative choice on the 'fashion divide' has served to encourage strong jobber-contractor relationships that in turn led to problem-solving cooperation among the different agents in that industry. So, at the very least, regional science must find ways of recognizing that different industries will mature or evolve in somewhat different ways. Perhaps a typology for industry maturation is needed, where it is recognized that location is critical to production but only at certain junctures in each industry's complex product cycle. Moreover, regional science must improve our understanding of how successful, healthy regions are able to reinvent their export bases along the lines already noted by Douglass North and Hans Blumenfeld some 60 years ago.

2.4.4 Modeling Interconnectivity Agglomeration

This chapter opened by presenting what is perhaps the most productive urban agglomeration in the history of humanity—namely the American Megalopolis—and the system of cities in which it is situated. Looking toward the future, regional science must find ways of modeling technological agglomeration, or the manner in which such dense urban agglomerations are bound together across the globe. This is especially true in an economy that is dominated by creative industries in metropolitan areas. The cross-sectional analysis developed earlier in the chapter is very useful for providing snapshot insights into the nation's metropolitan space economy but more sophisticated longitudinal models are needed to uncover the long term trajectories of both economic and population growth across those hundreds of metropolitan areas— and how they are woven together nationally and internationally.

Perhaps the simplest way to begin is to draw an analogy with the research on regional production functions and hypothesize that productivity growth across cities depends upon both patent production and business startups, along with controls for various initial conditions. This, of course, raises various issues related to measurement and specification (see below). In any case, a straightforward linear regression model has been estimated whose reduced form is as follows;

$$\mathrm{PROD}_t = f\,(\mathrm{PROD}_{t\text{-}1}, \mathrm{PATN}_{t\text{-}1}, \mathrm{PROP}_{t\text{-}1}; \mathrm{M}_{t\text{-}1})$$

where PROD is productivity (real GDP per capita), PATN is the volume of patents, PROP is the number of proprietors operating single-person businesses, and M is a vector of initial conditions, including metropolitan population, climate attributes, and a Sunbelt versus Snowbelt distinction. The three key variables were all standardized to make them scale-free: city productivity was changed into a ratio using the national average for productivity (in the appropriate years) while both patents and proprietors were converted into densities (by dividing each by its city population) before being changed into ratios using the appropriate national density figures.

Over the 12-year period, 2001–2013, proprietorship (0.885) has a coefficient that is nearly twice that for patents (0.415), so there is a prima facie case for startups being a much more important contributor to GDP growth than trademarks, copyrights, and the like. However, when a system of three equations is solved for a Carlino-and-Mills adjustment model, then over the very long run—as convergence in the coefficients takes place—patents are seen to be about four times as important as proprietorship. Seven of the nine coefficients in this 3 by 3 adjustment model are significant and positive; however, with the lag of 12 years, proprietors exhibit a negative impact on patent growth and patents exhibit a negative impact on proprietorship growth. So, as the direct and indirect effects unfold over the very long run, it seems that the importance of patents eventually overshadows that of startups in the metropolitan productivity model.

The estimation was repeated with shorter lags in order to assess the pattern of coefficients before, during, and after the events of the Great Recession. During 2001–2005, patents and proprietors both had a positive effect on one another but, during both 2005–2010 and 2010–2013, the two negative effects identified above were once again prominent. So is it possible that productivity growth since the Recession has taken on a different form where many metropolitan areas, largely in the Snowbelt, are specializing in patents while others, mostly in the Sunbelt, are specializing in business startups? Only more extensive research will answer that question. In addition, the two direct effects on productivity growth that were mentioned earlier were apparent in the first and third time periods where, if anything, the effects of startups were even stronger in the later as opposed to the earlier years (the ratio of coefficients climbed from approximately 2:1 to approximately 5:1). However, during the second time period, proprietorship was negatively related to productivity growth, probably because the recessionary events of 2007–2009 were so much more pronounced in the cities of the Sunbelt states.

This preliminary analysis sheds light on several issues that will challenge regional science over the upcoming years. To begin, analysts will have to decide if the specification of this simple model is appropriate or not; in other studies it has been found that entrepreneurs seem to occupy an intermediary position between patent activity and the creation of jobs by incumbent firms (Tsvetkova 2015). In the same way, some patent activity might translate directly into higher productivity through incumbent firms, but other patent activity might require new proprietors in order to see the light of day. And, of course, there is the standard observation that all patents are not equal. Besides, the use of proprietors to represent startups in the estimation is by itself problematic; other data sources will have to be mined to get a more accurate picture of annual business startups in these metropolitan areas. Finally, econometric experimentation should clarify what the appropriate time lags are for these or other sorts of metropolitan productivity growth models.

The dimension of global connectivity can be modeled in the spirit of contagion in financial markets (see, for example, Kolb 2011). Just as financial contagion can be broken down within and across regions—whether regions defined within a nation or among them—so, too, can innovation. Regional scientists (see Kelejian et al. 2006; Hondroyiannis et al. 2009) have already started making good use of spatial theory and methods to study contagion and similar frameworks would extend to innovation in a natural way. Ultimately, both research on and, indeed, the practice of innovation must recognize that regions—whether established titans like the megalopolis or emerging ones like those in China—can leverage advantages derived from spatial agglomeration globally. This is perhaps the next large question horizon for regional science in terms of the matter of this chapter: how can localized spatial agglomeration be transferred across networked global space? Some thinkers (Khanna 2016) seem to believe that the answers to this question have important implications for the future of the global economy. If so, regional science is a field uniquely positioned to help explore, understand, and predict what is yet to come.

2.5 Summary and Conclusion

This chapter set out several specific goals, in the context of the history of city systems and advanced urban economies. First, it argued that key input, output, and contextual attributes are combined locally to create each city's innovation ecosystem. These attributes not only differ from one place to another, but the manner in which they are combined varies by location. The project then settled on 20 variables in order to discern how these innovation ecosystems currently vary across the nation's very largest cities. Next, using multivariate techniques, production ecosystems were deconstructed in order to reveal the underlying dimensions of innovation that are common to all of the nation's 350-plus metropolitan areas. Once the primary dimension of general innovation was identified for all metropolitan areas, the other $n - 1$ less important dimensions were used to reveal how specific metros differentially combine their knowledgeable and educated workforces to produce patents, engage in entrepreneurship, and create high value- added outputs. As a consequence, the overall innovative index of any metropolitan economy could then be estimated by first generating its score on each of the latent dimensions and then, second, adding up those performance scores across all of the dimensions. The results produced a clear set of innovation centers, spread evenly across the United States. Not only do these findings square with contemporary theory on agglomeration economies, as explained by Mulligan et al. (2012), they line up nicely with older, less behaviorally motivated theories of central place hierarchies.

To conclude, it is worth pointing out several areas for future research. First, while satisfactory for present purposes—as is often the case—more and better data are key to refining this project; ideally, the work could be replicated using micro data on firms. Microdata would facilitate the development of behavioral models of firm agglomeration, which could then be productively used to explore gaps in innovation and performance both across and within metropolitan areas. Looking within metropolitan areas would be especially interesting to better understand the kind of mega-regions (see Nelson and Land 2011) identified in the introduction and that are increasingly relevant within the policy frame. Identifying the wage gradients that motivate firms to decentralize (Brueckner 2011) would be especially useful for understanding the processes that give rise to the “fine structure” of contemporary land-use patterns. Clearly, such evolution needs to take advantage of time series data, and look at two or more points in time in order to chart the trajectory of growth and change within regions as their fortunes rise and fall. Further, future research would do well to trace the contours of the actual business world by looking specifically at *Fortune-500* companies, and their respective headquarter-cities—plus by examining occupational (as opposed to employment) related data. Finally, such work might also look at particular industrial outcomes, including patents and/or other products of innovation.

Acknowledgements The authors thank Randy Jackson and Peter Schaeffer for their helpful comments on a previous draft of this chapter.

References

Acemoglu D, Akcigit U, Celik MA (2015) Young, restless and creative: openness to disruption and creative innovations. SSRN Electron J. doi:10.2139/ssrn.2392109

Acs ZJ, Braunerhjelm P, Audretsch DB, Carlsson B (2009) The knowledge spillover theory of entrepreneurship. Small Bus Econ 32:15–30

Acs ZJ, Audretsch DB, Lehmann EE (2013) The knowledge spillover theory of entrepreneurship. Small Bus Econ 41:757–774

Andersson A (2009) Economics of creativity. In: Karlson C, Andersson A, Cheshire P, Stough R (eds) New directions in regional economic development. Springer, New York

Angel S (2012) Planet of cities. Lincoln Institute of Land Policy, Cambridge, MA

Angel S, Parent J, Civco DL, Blei AM (2012) Atlas of urban expansion. Lincoln Institute of Land Policy, Cambridge, MA

Arthur WB (1989) Competing technologies, increasing returns, and lock-in by historical events. Econ J 99:116–131

Audretsch DB, Keilbach M (2007) The theory of knowledge spillover entrepreneurship. J Manag Stud 44(7):1242–1254

Barton C, Fromm J, Egan C (2012) The millennial consumer: debunking stereotypes. The Boston Consulting Group, Boston, MA

Batty M (2013) The new science of cities. MIT Press, Cambridge, MA

Beckmann MJ (1968) Location theory. Random House, New York

Behrens K, Robert-Nicoud F (2015) Agglomeration theory with heterogeneous agents. In: Duranton G, Henderson JV, Strange W (eds) Handbook in regional and urban economics. North-Holland, Amsterdam

Berry BJL, Parr JB (1988) Market centers and retail location. Prentice-Hall, Englewood Cliffs, NJ

Brewer’s Association (2016) Small and independent brewers continue to grow double digits—Brewers Association. In: Brewers Association. https://www.brewersassociation.org/press-releases/small-independent-brewers-continue-grow-double-digits/. Accessed 1 Aug 2016

Brueckner JK (2011) Lectures on urban economics. MIT Press, Cambridge, MA

Cairncross F (2001) The death of distance: how the communications revolution is changing our lives. Harvard Business Press, Brighton

Calzonetti FJ, Reid N (2013) The role of universities in supporting new technology industries through commercialization spin-off activities. Stud Reg Sci 43:7–23

Carlino GA, Chatterjee S, Hunt RM (2007) Urban density and the rate of invention. J Urban Econ 61:389–419

Carlino G, Kerr WR (2015) Agglomeration and innovation. In: Duranton G, Henderson JV, Strange W (eds) Handbook of regional and urban economics. North-Holland, Amsterdam
Carroll GR, Dobrev SD, Swaminathan A (2002) Organizational processes of resource partitioning. Res Organ Behav 24:1–40
Carruthers JI, Hepp S, Knaap G-J, Renner RN (2012) The American way of land use: a spatial hazard analysis of changes through time. Int Reg Sci Rev 35:267–302
Cheshire PC, Nathan M, Overman HG (2014) Urban economics and public policy: challenging conventional wisdom. Edgar Elgar, Northhampton, MA
Christaller W (1966) Central places in southern Germany. Prentice-Hall, Englewood Cliffs, NJ
Cohen WM, Levinthal DA (1990) Absorptive capacity: a new perspective on learning and innovation. Adm Sci Q 35:128
Combes PP, Gobillon L (2015) The empirics of agglomeration. In: Duranton G, Henderson JV, Strange W (eds) Handbook of regional and urban economics. North-Holland, Amsterdam
Cone Communications (2013) 2013 cone communications social impact study: the next cause evolution. Cone Communications, Boston, MA
Cramer J, Krueger AB (2016) Disruptive change in the taxi business: the case of Uber. Am Econ Rev 106:177–182. doi:10.1257/aer.p20161002
Crispin JE (2010) The economic impact of startup companies and invention licensees originating from research at the University of Utah. Utah Bus Econ Rev 70:1–7
David PA (1985) Clio and the economics of QWERTY. Am Econ Rev 75:332–337
de Blij HJ (2012) Why geography matters: more than ever. Oxford University Press, Oxford
de Rassenfosse G, Jaffe A, Webster E (2016) Low-quality patents in the eye of the beholder: evidence from multiple examiners. NBER Working Paper. doi:10.3386/w22244
Duranton G, Puga D (2015) Urban land use. In: Duranton G, Henderson JV, Strange W (eds) Handbook of regional and urban economics. North-Holland, Amsterdam
Etzkowitz H, Leydesdorff L (2000) The dynamics of innovation: from National Systems and “Mode 2” to a Triple Helix of university–industry–government relations. Res Policy 29:109–123
Florida R (2014) The creative class and economic development. Econ Dev Q 28:196–205
Florida R, Mellander C, Stolarick K, Ross A (2011) Cities, skills and wages. J Econ Geogr 12:355–377
Forbes (2015) Forbes. In: Forbes. http://www.forbes.com/sites/greatspeculations/2015/12/15/how-the-potential-ab-inbev-sabmiller-deal-focuses-on-africa/#1bebe5713a6e. Accessed 1 Aug 2016
Freeman L (2011) The development of social network analysis—with an emphasis on recent events. In: Scott J, Carrington P (eds) The Sage handbook of social network analysis. Sage, Los Angeles, CA
Frey WH (2002) Three Americas: the rising significance of regions. J Am Plan Assoc 68:349–355
Fritsch M, Kauffeld-Monz M (2010) The impact of network structure on knowledge transfer: an application of social network analysis in the context of regional innovation networks. Ann Reg Sci 44:21–38
Fromm J, Lindell C, Decker L (2011) American millennials: deciphering the enigma generation. Barkley, Kansas City, MO
Fry R (2016) Millennials overtake Baby Boomers as America’s largest generation. In: Pew Research Center RSS. http://www.pewresearch.org/fact-tank/2016/04/25/millennials-overtake-baby-boomers/. Accessed 1 Aug 2016
Fujita M (1989) Urban economic theory: land use and city size. Cambridge University Press, Cambridge
Fujita M, Krugman P, Venables A (1999) The spatial economy: cities, regions and international trade. MIT Press, Cambridge, MA
Fujita M, François TJ (2013) Economics of agglomeration: cities, industrial location, and globalization. Cambridge University Press, Cambridge
Gabaix X (1999) Zipf’s law for cities: an explanation. Q J Econ 114:739–767
Gabaix X, Ionniades YM (2004) The evolution of city size distributions. In: Henderson JV, Thiesse JF (eds) Handbook of regional and urban economics. North-Holland, Amsterdam

Gabe T, Florida R, Mellander C (2012) The creative class and the crisis. Camb J Reg Econ Soc 6:37–53

Glaeser E, Joshi-Ghani A (2015) The urban imperative toward competitive cities. Oxford University Press, Oxford

Glaeser EL (2005) Reinventing Boston: 1630–2003. J Econ Geogr 5:119–153

Glaeser EL (2011) Triumph of the city: how our greatest invention makes us richer, smarter, greener, healthier, and happier. Penguin Press, New York

Gunasekara C (2006) Reframing the role of universities in the development of regional innovation systems. J Technol Transfer 31:101–113

Hathaway I (2013) Tech starts: high-technology business formation and job creation in the United States. Kauffman Foundation Research Series: Firm Foundation and Economic Growth. www.kauffman.org. Accessed 1 Aug 2016

Head K, Mayer T (2004) The empirics of agglomeration and trade. In: Henderson JV, Thiesse JF (eds) Handbook of regional and urban economics. North-Holland, Amsterdam

Hofstede GH (1997) Cultures and organizations: software of the mind. McGraw-Hill, New York

Hondroyiannis G, Kelejian HH, Tavlas GS (2009) Spatial aspects of contagion among emerging economies. Spat Econ Anal 4:191–211. doi:10.1080/17421770802625965

Ioannides YM, Overman HG (2003) Zipf's law for cities: an empirical examination. Reg Sci Urban Econ 33:127–137

Ioannides YM, Overman HG (2004) Spatial evolution of the US urban system. J Econ Geogr 4:131–156

Jovanovic B (2001) Fitness and age: review of Carroll and Hannan's demography of corporations and industries. J Econ Lit 39:105–119

Kelejian HH, Tavlas GS, Hondroyiannis G (2006) A spatial modelling approach to contagion among emerging economies. Open Econ Rev 17(4):423–441

Khanna P (2016) Connectography: mapping the future of global civilization. Random House, New York

Kolb RW (ed) (2011) Financial contagion: the viral threat to the wealth of nations. Wiley, New York

Kovács B, Carroll GR, Lehman DW (2014) Authenticity and consumer value ratings: empirical tests from the restaurant domain. Organ Sci 25:458–478

Krugman P (1991a) Increasing returns and economic geography. J Polit Econ 99:483–499

Krugman PR (1991b) Geography and trade. MIT Press, Cambridge, MA

Krugman P (1997) Development, geography, and economic theory. MIT Press, Cambridge, MA

Lautman M (2011) When the boomers fail. Logan Square Press, Albuquerque, NM

Lee SY, Florida R, Gates G (2010) Innovation, human capital, and creativity. Int Rev Public Admin 14:13–24

Lewis D, Bridger D (2000) The soul of the new consumer: authenticity—what we buy and why in the new economy. N. Brealey Pub, London

Lösch A (1954) The economics of location. Yale University Press, New Haven

Moore MS, Reid N, McLaughlin R (2016) The location determinants of microbreweries and brewpubs in the United States. In: Cabras I, Higgins D, Preece D (eds) Brewing, beer and pubs: a global perspective. Palgrave McMillan, London, pp 182–204

Moretti E (2004) Human capital externalities. In: Henderson JV, Thiesse JF (eds) Handbook of regional and urban economics. North-Holland, Amsterdam

Moretti E (2012) The new geography of jobs. Houghton Mifflin Harcourt, New York

Mulligan GF, Partridge MD, Carruthers JI (2012) Central place theory and its reemergence in regional science. Ann Reg Sci 48:405–431

Nelson AC, Land RE (2011) Megapolitan America: a new vision for understanding America's metropolitan geography. APA Press, Chicago, IL

Peirce NR, Johnson CW, Peters F (2008) Century of the city: no time to lose. Rockefeller Foundation, New York

Pew Research Center (2010) Millennials: a potrait of generation next. Pew Research Center, Washington, DC

Pred A (1977) City systems in advanced economies. Halsted Press, New York
Reese LA, Sands G, Skidmore M (2014) Memo from motown: is austerity here to stay? Camb J Reg Econ Soc 7:99–118
Renner RN, Carruthers J, Knaap GJ (2009) A note on data preparation procedures for a nationwide analysis of urban form and settlement patterns. Cityscape J Policy Dev Res 11:127–136
Schmidt T (2005) What determines absorptive capacity. In: DRUID tenth anniversary summer conference 2005—dynamics of industry and innovation, Denmark
Shadbolt P (2015) Brewery battle: AB InBev and the craft beer challenge. BBC News, 13 Oct. http://www.bbc.com/news/business-34383721
Shapiro JM (2006) Smart cities: quality of life, productivity, and the growth effects of human capital. Rev Econ Stat 88:324–335
Simonton D (1997) Creative productivity: a predictive and explanatory model of career trajectories and landmarks. Psychol Rev 104:66–89
Simonton D (1999) Origins of genius. Oxford University Press, Oxford
Spithoven A, Clarysse B, Knockaert M (2011) Building absorptive capacity to organise inbound open innovation in traditional industries. Technovation 31(1):10–21
Storper M, Kemeny T, Makarem NP, Osman T (2015) The rise and fall of urban economies. Stanford University Press, Stanford, CA
Ter Wal AL, Boschma RA (2009) Applying social network analysis in economic geography: framing some key analytic issues. Ann Reg Sci 43(3):739–756
Tsvetkova A (2015) Innovation, entrepreneurship, and metropolitan economic performance: empirical test of recent theoretical propositions. Econ Dev Q 29:299–316
United States Department of Labor (2015) Fastest growing occupations. Occupational outlook handbook, 17 Dec. https://www.bls.gov/ooh/fastest-growing.htm
United States Patent and Trade Office (2015) Calendar year patent statistics: listing of all United States metropolitan and micropolitan areas, total utility counts, 2000–2013. www.uspto.gov. Accessed 1 Aug 2016
Vernon R (1966) International investment and international trade in the product cycle. Quar J Econ 80:190–2017
Winograd M, Hais M (2014) How millennials could upend wall street and corporate America. Governance studies at brookings, Feb. https://www.brookings.edu/wp-content/uploads/2016/06/Brookings_Winogradfinal.pdf
Zenou Y (2009) Urban labor economics. Cambridge University Press, New York
Zipf GK (1949) Human behavior and the principle of least effort. Addison-Wesley, Cambridge

Gordon F. Mulligan is professor emeritus, School of Geography and Development, University of Arizona. His primary research interests are in economic geography, with special interests in urbanization, housing, location theory, regional development, and spatial behavior. He is past book review editor of the Annals of Regional Science and past co-editor of the Journal of Regional Science, and has been an RSAI Fellow since 2011. Previous faculty positions were at the University of Washington, Queen's University, and Flinders University. Dr. Mulligan earned the Ph.D. in geography from the University of British Columbia in 1976.

Neil Reid is Professor of Geography and Planning and Director of the Jack Ford Urban Affairs Center at the University of Toledo. His primary research interests are economic geography with special interests in industrial location and local economic development. He serves as the Executive Director of the North American Regional Science Council, Editor for the Americas for Regional Science Policy and Practice, and Book Review Editor for Economic Development Quarterly. He has held the position of Visiting Professor in the European Faculty of Engineering at Częstochowa University of Technology in Poland where he taught courses in industrial location and regional development. He holds a Ph.D. in Geography from Arizona State University.

John I. Carruthers is the Director of the Sustainable Urban Planning Program at the George Washington University. His areas of research include urban and regional economics; sustainable urban development; environmental quality of life; economic geography; geospatial and econometric analysis. He holds an adjunct faculty appointment in the Department of Urban Planning and Engineering at Yonsei University, in Korea. Dr. Carruthers has served as an elected member of the RSAI Council (2009–2011) and the North American Regional Science Council (2011–2014) and is in the Board of the Western Regional Science Association. He earned is MS and doctorate, respectively, at the University of Arizona and the University of Washington. Regional Science is his home field.

Matthew R. Lehnert is a PhD student in Spatially Integrated Social Science (SISS) at the University of Toledo. His primary research interests are in international development, with special interests in the spatial dimensions of human development, poverty, and education in Morocco. He has served in adjunct positions at Owens Community College and Al Akhawayn University. Currently he holds a research assistantship with the North American Regional Science Council.

Chapter 3
Future Shock: Telecommunications Technology and Infrastructure in Regional Research

Tony H. Grubesic

3.1 Introduction

By definition, *telecommunication* refers to the transmission of words, sounds or images over great distances, typically in the form of electromagnetic signals (RH 2015). Historically, telecommunication has taken many forms, ranging from drums, to smoke signals, to the telegraph, radio and Internet (Huurdeman 2003). As detailed by Gillespie and Williams (1988), telecommunication is often lumped together with other "distance shrinking" technologies such as air transport—technologies that serve to reduce the friction of distance and generate a functional convergence of space and time. Specifically, Gillespie and Williams (1988, p. 1317) note that "when the time and effort required to communicate over 10,000 miles is indistinguishable from that required to communicate over 1 mile, the convergence of time and space has taken place at a fairly profound scale."

The impacts of telecommunication have challenged many of the fundamental notions of regional science and neoclassical economics. In explaining how cities, metropolitan areas and larger regional systems grow, traditional approaches have relied upon central place theory (Christaller 1993) and/or location theory (Weber 1929), both of which assume that there is an unavoidable need to physically transport both people and goods between places. Here, the friction of distance across the Euclidean plane is associated with significant costs in both time and money—dictating urban form and structure, as well as the economic behavior of people, firms and industry (Graham and Marvin 1996). In short, as physical distance between places increases, their overall levels of interaction decrease (i.e., distance decay). Although this is often true, the strict limitations of these theories were exposed,

T.H. Grubesic
Center for Spatial Reasoning and Policy Analytics, Arizona State University, Phoenix, AZ, USA
e-mail: grubesic@asu.edu

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_3

particularly with the advent of advanced telecommunications technologies such as the Internet. As detailed by Graham and Marvin (1996, p. 56), instead of assuming that “the only way goods and services could be consumed, or communications could take place was through physical transportation,” telecommunications could be viewed as “a means of moving information, services, capital and labor power.” It is unfortunate, then, that for many years the role of telecommunications was largely ignored in much of regional science and neoclassical economics, even though the assumption of perfect information for producers and consumers was rooted within. For example, how was knowledge about the economic system acquired? This is a particularly important question considering that transaction costs are assumed to be zero within these types of systems (Nicol 1985). Zero costs assume that the collection of information about a system and making transactions within the system can be done without any effort or time expended. This is somewhat ironic, because if there is any single technology that would come close to facilitating perfect knowledge within a market, drastically lower transaction costs and facilitate the movement of information, services, capital and labor power without heavy physical impediments, such as distance, it is telecommunications.

With the pioneering work of Gillespie and Williams (1988), Ota and Fujita (1993), Warf (1995), Graham and Marvin (1996), Capello and Nijkamp (1996), Gaspar and Glaeser (1998), Rietveld and Vickerman (2004), as well as the more recent work by Caragliu et al. (2011), Mack and Rey (2014), Grubesic and Mack (2015), and many others, telecommunications is now widely recognized as a critical component to the development of regions. Many of the classic physical limitations of distance and location remain, but technologies such as broadband, wireless communications and satellite systems continue to both enhance and facilitate the exchange of information, knowledge and communication—the valuable goods of the information economy (Moss and Townsend 2000). The purpose of this chapter is to explore key challenges in the continued development of the telecommunications ecosystem, highlighting how future technologies might impact regions. Challenges that will be discussed include: (1) infrastructure deployment, (2) shifting needs and uses of telecommunications systems, and, (3) future technologies. Implications for regional research and policy will also be addressed.

3.2 Infrastructure Deployment

The process associated with deploying telecommunications infrastructure is complex. Take, for example, the Telecommunications Act of 1996, which deregulated the telecommunications industry in the United States. It was hoped that this legislation would provide consumers and businesses with the benefits of more provider choice, higher quality of service and lower prices. Over time, the price for voice communications certainly fell, but costs, choices and overall quality of service associated with platforms such as broadband have failed to keep the U.S. apace with much of the developed world. For example, the most recent Organization for

Economic Co-Operation and Development (OECD) statistics (2014) suggest that the U.S. ranks 16th in the world for wireline broadband penetration, behind countries such as Greece, the Czech Republic, Canada and many others. Further, although it is difficult to compare pricing across countries, the U.S. is clearly not a price leader (OECD 2014). Broadband is more expensive and slower in the United States than many other countries.

The fundamental problem with telecommunications infrastructure deployment in the U.S. is that it has been left to private companies, such as Verizon, AT&T and many of the other regional wireline juggernauts. This is not to say that investment has been lacking. For example, recent estimates suggest that wireline providers invested $78 billion in broadband during 2014, with the industry having invested $1.4 trillion since 1996 (US Telecom 2015). But, there are disparities with how these monies are invested, particularly from a regional perspective. As detailed by Grubesic (2015), although broadband is widely deployed in the United States, there are significant variations in both the speed and quality of available service, particularly when comparing urban and rural/remote areas. Specifically, in large, densely populated areas, such as San Francisco, Philadelphia and Chicago, network providers are eager to rollout the latest (and greatest) network platform enhancements and associated technologies. Smaller communities, particularly those located in exurban or rural regions, are much slower to receive network upgrades and often find themselves at least one generation behind more urbanized areas. In remote regions of the U.S., communities are often two or more generations behind the urban leaders, if broadband is available at all. In fact, many of the most isolated locales in the U.S. rely upon satellite broadband connections, because reliable wireline and/or wireless (i.e., cellular or WiMax) is not available (Grubesic and Mack 2015). These stubborn gaps in access, quality of service, pricing and speed can be attributed to the dominance of private telecommunication infrastructure provision in the U.S. Companies like Verizon, Comcast and Cox want a rapid return on their infrastructure investments, so they cater to thick, urban markets, rather than thin, rural markets (Grubesic 2006; Grubesic et al. 2011). It is unlikely that these geographic disparities in telecommunications infrastructure will be ameliorated in the near future.

3.2.1 The Future of Infrastructure Deployment

One of the more troubling facets of existing broadband infrastructure is known as the first-last-mile bottleneck (Kazovsky et al. 2011). In a nutshell, as the demands for bandwidth continue to increase, a reliance upon older copper infrastructure for the delivery of that bandwidth is becoming a weak link in the system. Although copper provides basic functionality, the future will require and rely upon fiber optic networks. Consider, for example, the most high-bandwidth flavor of digital subscriber lines, VDSL2. It delivers downstream speeds of approximately 100 Mbps with upstream speeds approaching 30 Mbps (Kazovsky et al. 2011). These speeds pale in comparison to the most basic broadband passive optical network (BPON), which offers speeds of 622 Mbps downstream and 155 Mbps upstream. Even with the advent of repeater technologies that help mitigate the geographic service restrictions of xDSL and the emerging bonding technologies that connect several telephone lines with multiple xDSL modems to enhance bandwidth (Gorshe et al. 2014), there is simply no way that any last mile technology reliant upon copper can compete with an all fiber system.

There are several reasons why it is important to acknowledge the first-last-mile bottleneck. First, the world is shifting away from an Internet of computers toward an Internet of things (IoT) (Mattern and Floerkemeier 2010). As more devices become connected to the Internet, including thermostats, refrigerators, door locks and lighting systems, the demand for bandwidth and its use will increase exponentially. Second, there are millions of supervisory control and data acquisition systems (SCADA) used throughout the world to remotely operate and monitor infrastructure systems (e.g., water and gas distribution, road conditions, electrical grid, etc.) that are reliant upon last-mile connectivity (Murray and Grubesic 2013). As more of these systems arrive online and the associated demands for data from these systems increases, copper first-last-mile infrastructure will no longer be adequate and it will be replaced with fiber. Third, as detailed by Grubesic and Mack (2015), the costs to produce and install optical fiber are decreasing steadily, gradually pushing the industry toward all-fiber solutions. Consider, for example, the rather archaic installation process currently required for fiber. Small channels are cut into the asphalt along streets (compromising its overall integrity), earth is trenched and excavated, conduit is installed, the fiber is threaded through the conduit, the earth is replaced and the street is repaved. This is an extremely expensive and time consuming process that creates significant disruptions to travel, commerce and life, especially in urban areas (Specht 2015; Canon 2014). However, emerging techniques, such as micro-trenching are quickly gaining favor in many locations, as are installations within existing infrastructure such as water pipelines and sewer systems using drones and associated robotic vehicles. In time, this will greatly expand the optical fiber footprint, bringing exceptionally high bandwidth connectivity to the masses and easing local/regional disruptions.

Finally, it is important to address the issues associated with wireless telecommunications systems and their associated infrastructure. Some believe that the death of wireline broadband has already begun (Worstall 2013a). Specifically, Worstall (2013b) argues that with the increasing bandwidth available from cellular systems, it makes no sense to install expensive optical fiber networks that may, at some point in the future, lock regional economies into an antiquated platform. This is, after all, exactly what happened to landline telephone systems (Worstall 2013b). Or is it? It is easy to forget that plain-old-telephone-service (POTS) has existed since 1876 and that the first long distance line was installed just 1 year later in 1877. What is the historical value of POTS and its network? Frankly, the economic, cultural, social and political value would be so large, it might be impossible to estimate for a region, let alone an entire country. Today, with less than 20% of U.S. households maintaining a landline, and the Federal Communications Commission making plans to phase-out the existing analog network, it is easy to suggest that optical fiber will suffer the same slow death 140 years from now. *It might*. However, wireless infrastructure will not be the slayer of wireline systems anytime soon. There are several reasons for this. First, wireless systems are completely reliant upon optical fiber networks for data transit and routing. Wireless networks are not autonomous systems that operate independently of wires—they need fiber. Second, wireless systems are heavily constrained by the availability of spectrum, and spectrum is in short supply (Vantage Point 2015). This means that wireless providers have very few options for accommodating a high density of customers or their bandwidth demands. The capacity simply does not exist. Third, wireless systems are highly susceptible to interference during transmission. Obstacles such as mountains, vegetation and weather all function as impediments to high-quality wireless connections. Thus, while it is provocative to suggest that the end is near for wireline networks, and that the U.S. and private companies should not waste their resources on wireline systems, the future is (and will remain) fiber for a long time to come.

3.3 The Shifting Needs and Uses of Telecommunications Systems

Most readers are likely familiar with Moore's Law. In 1965, Gordon Moore, the director of research at Fairchild Semiconductor and future co-founder of Intel, wrote a paper (Moore 1965) in which he observed that the doubling of components per integrated circuit would continue to occur at regular intervals, approximately every 18 months. As detailed by Schaller (1997), this observation has been self-fulfilling to some degree, with the entire industry using this benchmark as the roadmap for growth and development. Further, although projections suggest that physics (i.e., quantum theory and thermodynamics) will ultimately collapse Moore's law within the next 20–40 years (Lloyd 2000; Powell 2008), there may be room to grow toward the Bekenstein bound, which is the upper limit on the amount of information that can be contained and/or processed within a finite region of space that has a finite amount of energy (Bremermann 1962, 1965; Bekenstein 1981).

A lesser known, but equally intriguing observation is Nielsen's law of Internet bandwidth (Nielsen 1998), which suggests that the average user's bandwidth will grow 50% per year. This, of course, is much slower that Moore's law, but there are several good reasons for the lag. First, as detailed above, the telecommunication companies responsible for installing the physical plant (e.g., fiber) are relatively conservative. Thus, although they are not afraid to invest, installations take time and there is a significant lag between investment and service rollout for many places. Second, Nielsen (1998) notes that average users are relatively reluctant to spend money on bandwidth. For most users, a modestly fast connection will suffice. As a result, the return on investment for bandwidth improvements and associated infrastructure is slow. Research has also shown that the demand for broadband services is elastic (Goolsbee 2002; Galperin and Ruzzier 2013), contingent upon household income, installation fees and monthly service rates. Finally, as the broadband user base broadens, the demand for low-end services will grow faster than premium services. Because most existing physical networks can handle these modest bandwidth demands, even with a growing user base, the incentives to rapidly improve the network are lacking.

One underlying problem with Nielsen's law is that it is user-focused, assuming that bandwidth demand is rooted in the individual (or collective) needs of human users. Although this has been true during the first 20 years of the commercial Internet, the landscape is changing quickly. As the IoT begins to take root, bandwidth demand is going to increase exponentially and the way in which telecommunications systems are used is going to shift dramatically. In fact, the internet of things is going to quickly become a misnomer, being replaced by the internet of everything (IoE).[1]

Consider, for example, the current landscape of demand for the IoT. Many readers will be familiar with the Nest thermostat, a WiFi enabled device that interactively learns the heating and cooling patterns of households and can be operated remotely from a smartphone. Aside from an occasional software update, which may require 50 MB per month upload and 10 MB per month download (Nest 2015), the daily operations of the thermostat require only a few hundred bytes of data transfer. In fact, this is true for most appliances currently comprising the IoT—they are small, fixed, power efficient devices with low bandwidth demands that occasionally require a routine software update. Current estimates suggest that only 1% of all bandwidth is allocated to the IoT (Lawson 2014) and that there are approximately 10 billion devices connected (globally) to the Internet (Bradley et al. 2013). Again, this is going to shift dramatically within the next decade, as barriers to connectivity start to decrease. In fact, with the shift from Internet Protocol version 4 (IPv4) to IPv6, more than 340 undecillion addresses[2] become available for devices, people and things.[3] Further, much like a social network, there is value in connectivity. Bradley et al. (2013) estimate that the IoE will be worth nearly $14.4 trillion in value by 2022. The question is, how does one connect all of these devices? In addition, what are the implications for how telecommunications systems are designed and used with the IoE?

3.3.1 The Future of IoE and Telecommunications Systems

As noted previously, the current IoT is primarily comprised of fixed devices that "sip" bandwidth supplied by household broadband connections. This includes devices such as thermostats, refrigerators, lighting systems, door locks and other

[1] IoE refers to connected people, processes, devices and things (Bradley et al. 2013).

[2] 340,282,366,920,938,463,463,374,607,431,768,211,456.

[3] IPv6 is the next iteration of Internet Protocol and will replace IPv4. IPv6 provides a unique, 128 bit numerical IP address for any device that is connected to the Internet.

thing within the household. Of course, other computing and/or mobile devices can also reside in a household that do more than sip bandwidth, examples include devices that stream Internet content (e.g., Roku or Apple TV). The future of the IoE, however, will exhibit a radical shift away from fixed devices in the ecosystem, to mobile "things" that require ubiquitous access to the Internet. This will include automobiles, bicycles, clothing, sunglasses, recreational equipment (e.g., golf clubs, soccer balls), and cameras of every size and for every application, unmanned aerial systems, personal food allergy testing equipment, etc. Currently, the only networks that could come close to handling this need for ubiquitous connectivity are the existing 3G and 4G wireless systems. The problem with this scenario is that few users can (or will be able to) afford the bandwidth required for their personal IoT ecosystem, especially as the IoE shifts from devices that sip bandwidth to data hungry devices such as driverless cars. Wireless data remains expensive, especially in the U.S. Recent statistics indicate that it costs about $10 for 2–5 GB of mobile data in Copenhagen, but $45 for the same amount of data in the U.S. (OTI 2014). Moreover, there is no wireless network, at least as they are currently constructed in the U.S., that will be able to handle bandwidth demand generated from an estimated 50+ billion devices to be connected to the network by 2020.

In anticipation of increasing demand for the IoE, some are speculating that regional cellular networks, specifically designed for IoE, will solve spectrum shortages by making sure the growing ecosystem does not overload existing networks (Chamberlain 2015). While this seems like an appealing option for alleviating the strain on existing wireless systems, the complications associated with new network protocols might limit the ability of IoE things to communicate with each other (Chamberlain 2015). Thus, critics of the IoE-specific networks suggest that improvements to local WiFi, cellular wide area networks and Bluetooth personal networks should suffice (Jawanda 2014). More importantly, the value of the IoE ecosystem is largely contingent upon users capitalizing on the bring-your-own-device (BYOD) model that their current lives are centered upon (Jawanda 2014). BYOD values interoperability and reflects the prevailing preferences of customers, in that they want devices that work with every other device and/or application they care about.

5G, the next generation of mobile technology, will likely have a profound impact on regions and will be the key to realizing the growth and interoperability of the IoE. For now, however, nobody has a firm grasp on exactly what 5G is going to be. The standards and associated technologies to be embedded within 5G are still being worked out (Boccardi et al. 2014). That said, we a rough idea of what providers and the massive consumer market would like to be included in the fifth generation wireless platform:

1. 5G will be fast—perhaps 10 GB per second fast. This is roughly 66 times faster than existing 4G technologies (150 MBps). Fast enough, in fact, for autonomous cars to safely and seamlessly navigate the streets and for the 5G network to support billions of IoE devices, sensors and things.

2. 5G will support augmented reality, where ambient and/or contextual data about a place will be projected onto car windshield, or one's personal eyewear. This requires nearly instantaneous data delivery and 5G will lower latencies to approximately 1 ms, greatly departing from the 25 ms delays in current 4G systems.
3. The infrastructure support systems will change dramatically. Instead of solely relying upon large, macro-cell antennae, which typically exhibit transmission coverages of approximately 35 km in simple terrain, 5G networks will use heterogeneous networks that include an assortment of cells, from macrocells (35 km) to microcells (2 km), to picocells (200 m) and femtocells (10 m) to accommodate the IoE (Boccardi et al. 2014). More importantly, the size reduction of cells helps spectral efficiency, allowing systems higher rates of frequency reuse (Chin et al. 2014).
4. Providers are also moving toward a system architecture that splits the uplink and downlink communication channels from a single base station to multiple, heterogeneous cells. The larger cell will maintain connectivity and mobility for a user and associated devices (uplink), while the smaller cell will provide the high-bandwidth data connection (downlink). This will allow for higher spectrum efficiency, throughput, and energy savings (Chin et al. 2014). More importantly, when no data is required, the small cell can be switched off.
5. Device-to-device (D2D) communications will be supported more formally. Although this can already occur in unlicensed spectrum, using technology such as Bluetooth, 5G will allow for cellular base stations to help manage D2D and machine-to-machine (M2 M) communications, helping to avoid intra-cell interference (Chin et al. 2014).

There are many other potential innovations being considered for 5G, including the use of Massive Multiple-Input/Multiple-Output (MIMO) antenna technologies that can accommodate three dimensional beam forming. This would allow for capacity increases and the improvement of radiated energy efficiency (Larsson et al. 2014). Millimeter wave technology (Rappaport et al. 2013) may also help alleviate existing spectrum limitations. Although mmWave spectrum is limited and it is highly sensitive to blockages, optimal arrangements antennae and the use of emerging antenna technology will help minimize blockages, especially for micro and picocells arranged in indoor areas.

Most analysts believe that 5G will debut sometime between 2020 and 2022. In the meantime, 4G networks still have plenty of life. Estimates suggest that wireless operators will spend upwards of $1.7 trillion on their 4G networks between 2014 and 2020, with the anticipation of reaping a return on those investments (Best 2014). Further, during the transitional phase in the early 2020s, when 4G systems begin competing with 5G systems (and eventually being replaced by 5G), there will be problems with consumer pricing models (Bouras et al. 2014). As a result, providers are not necessarily in a rush to bring 5G to market. As detailed earlier, there are many technological and market-based hurdles to overcome.

3.4 Future Technologies

Of all the potential technological innovations that we are likely to see in the next several decades, there are two developments that will represent the pinnacle of modern telecommunications advances: holography and brain-to-brain communications. Although these types of communication are fairly standard in popular science fiction movies (see Star Wars, for example), the technological hurdles associated with making these forms of telecommunication a reality are very much real, but progress is being made.

3.4.1 Holograms

Consider, for example, holography. In the 1977 film Star Wars, Princess Leia records a holographic message which is displayed by R2-D2 for Ben Kenobi. In this instance, the hologram was displayed by R2-D2 without any additional equipment (e.g., a screen) and it was viewed without any special eyewear or lighting. In other words, the message was a true hologram. This, not surprisingly, is extremely difficult to accomplish. Holography is a process by which wavefronts of light are recorded and then reconstructed in both amplitude and phase information to generate three-dimensional (3D) images (Collier 2013). In most instances, holograms are some type of photographic recording of a light field and not an image formed through more traditional lenticular or autostereoscopic 3D display technologies (Huff and Fusek 1980). Holograms require fixed wavelengths of light to work properly, which is why lasers are used as the light source for holograms (Collier 2013). Only recently have advances in holography provided a way to generate these types of displays, in 3D, that can be viewed without any additional equipment (e.g., glasses). The recent work of Li et al. (2015) uses graphene[4] for generating 3D holographic images. Specifically, graphene oxide is pulsed with a laser to facilitate the process of photoreduction. As the graphene oxide is heated/pulsed by the laser, a nanoscale pixel is created that is capable of bending the light to create a holographic image. These small tweaks to the refractive index[5] can create many different optical effects (Gu 2015), including 3D holograms that are visible to the naked eye. Although the images created by Li et al. (2015) remain small, approximately 1 cm in size, they anticipate that the process for generating these holograms will scale, eventually accommodating much larger images (Gu 2015), but this will take time.

Although we are years away from realizing a 3D display like the one seen in Star Wars, there are many interesting commercial and military applications for

[4]Graphene is a two dimensional, hexagonal lattice of bonded carbon atoms that is 100 times stronger than steel, can carry 1000 times for electricity than copper and has a pronounced field effect (Colapinto 2014).

[5]The refractive index is a measure of how much light passes through a medium or material.

holography that are changing the way in which strategic planning is approached, as well as telecommunication for large public events and performances. One example, although not directly related to telecommunications, is the use of holographic prints for strategic planning. Using a combination of digital printing technologies, as well as commercial planning platforms, such as computer aided drawing (CAD), geographic information systems (GIS) and remotely sensed imagery, holographic prints can be generated to create a realistic 3D view of an urban landscape, the interior or an office building or an architectural plan (Saenz 2010). This is done by embedding several hundred thousand individual 3D views onto a single sheet, where the parallax allows for the image to look completely three dimensional. Although these images are not directly immersive, such visualizations can be used both to enhance and extract valuable regional planning information, facilitating interaction and collaboration for strategic planning (Groenendyk 2013).

One of the more stunning and useful applications of holography for telecommunications is holographic telepresence (HT). Although the technology used for generating HT is continually evolving, HT displays rely upon a transparent screen for display. Some readers may recall the 2012 Coachella Valley Music and Arts Festival, where Tupac Shakur (who died in 1996), appeared on stage and performed as a hologram (Ngak 2012). Michael Jackson was next, performing at the Billboard Music Awards in 2014 (Respers 2014). In both of these cases, the artists' performances were pre-rendered and recorded for the performance. Today, however, holographic telepresence is increasingly relying upon broadband networks for real-time interactions, performances and video conferencing. For example, fans of late-night television may recall that Jimmy Kimmel used holographic telepresence for hosting his show live from Los Angeles and Nashville (hologram) simultaneously. Although the results were not particularly convincing or realistic (see http://tinyurl.com/omsuq8r), the technology will improve. A more impressive application was recently staged by Musion, which generated a holographic telepresence in London using 4G wireless networks. The devil, of course, is in the details. Musion used eight SIM cards and a codec that takes signals from the four largest wireless providers in London to generate the hologram (see http://tinyurl.com/pqbw5nb). It worked well, but this type of arrangement is both expensive and difficult to organize for every day applications.

Current limitations aside, holographic telepresence has massive potential for changing the way people interact and communicate. In time, HT will be immersive. For now, Microsoft and Digital Video Enterprises (DVE) have already partnered together to create a holographic presentation room that allows participants (both real and holographic) to make "real" eye contact with each other and interact with 3D objects suspended in mid-air from Microsoft Windows applications using their Kinect technology. This type of interaction is vastly superior to existing video conference systems, providing a more personal and engaging platform for long-distance meetings (Summit 2015). Further, from the perspective of regional science, the possibilities embedded within holographic telepresence may, finally, have notable impacts on travel behavior (e.g., eliminating some demand for long-distance travel), facilitate telemedicine in real-time using 3D holographic models for instruction or emergency intervention, or using holograms for aiding geographically remote, robotic control for navigation or disaster relief applications. In short, there are endless possibilities and their potential impacts to regional development, commerce and transportation may be remarkable.

3.4.2 Brain-to-Brain Communication

A second technological advance for telecommunication worth detailing is non-invasive brain-to-brain (B2B) communication. Of course, the invasive version of any brain science is scary (but intriguing) stuff. For example, invasive cortical stimulation can be used to promote recovery of function after a stroke (Plow et al. 2009). That said, there is some disagreement between the efficacy of invasive versus non-invasive approaches (Hummel et al. 2008). Here, we are focusing on non-invasive techniques to facilitate communication between humans. Many readers might be curious to know what the rationale is for B2B connectivity. First and foremost, most existing forms of communication are rather inefficient. Thoughts, emotions, instructions and/or processes must be communicated between humans using words, symbols, visual and/or auditory cues. This creates a temporal lag, where thoughts are converted to discernable information, then conveyed between humans. Once conveyed, the information must be processed and then considered and/or acted upon by the humans involved. All phases of this process take time. Temporal lags are further exacerbated when communications take place across great geographic distances. Agarwal et al. (2010) notes that even for hospitals, communication inefficiencies can drastically inflate operating costs. It is estimated that for a typical 500 bed hospital, over $4 million in costs must be absorbed due to inefficiencies associated with conveying information within the hospital. This ranges from reading/writing on medical charts, allocating the appropriate care at the correct time, and simply tracking down nurses, physicians or other specialists for assistance (Agarwal et al. 2010).

Second, existing forms of communication are always subject to misinterpretation. We have all suffered from this, saying or writing something that did not accurately convey a sentiment or thought. Further, as detailed by Rao et al. (2014, p. 1), “a great deal of the information that is available to our brain is not introspectively available to our consciousness, and thus cannot be voluntarily put into linguistic form.” For example, a pro-basketball player cannot simply tell a child how to exactly position a ball on their hand and use their legs, arms and torso to shoot a 3-pointer. Even when an idea or concept can be expressed linguistically or through symbols, there may be problems with translation, language barriers and/or a propensity for misinterpretation because of cultural biases or associated limitations. B2B technology seeks to bypass the use of these archaic forms of communication, instead opting for neural code, mediated by computers and transmitted across the Internet.

Recent research has demonstrated that there are a variety of ways for decoding (i.e., reading) and writing (i.e., encoding) digital information from neural activity. In fact, most brain-to-brain interfaces (BBI) leverage recording techniques, such as electrocorticography (Blakely et al. 2009) and magnetoencephalography (Hämäläinen et al. 1993; Mellinger et al. 2007), to capture neural activity. There are also a number of techniques used to stimulate brain activity, including transcranial magnetic stimulation (Hallett 2000) and cochlear implants (Starr and Brackmann 1979). The problem with all of these technologies is that some type of invasive implant is required. However, the work of Grau et al. (2014) and Rao et al. (2014) uses non-invasive technologies to accomplish brain-to-brain connectivity. For example, Grau et al. (2014, p. 1) establish "internet-mediated B2B communication by combining a BCI (brain-computer-interface) based on voluntary motor imagery-controlled electroencephalographic (EEG) changes with a CBI (computer-brain-interface) inducing the conscious perception of phosphenes (light flashes) through neuronnavigated, robotized transcranial magnetic stimulation." Put more simply, the researchers transmitted the words "hola" and "ciao" between the brain of a human located in India and the brain of another person located in France using the Internet and a host of other non-invasive decoding and encoding technologies. Rao et al. (2014) used a combination of electroencephalography and transcranial magnetic stimulation for brain-to-brain communication between two humans. The two test subjects were able to cooperatively perform a task (embedded within a computer game) using *only* the direct brain-to-brain interface as a channel of communication.

Thus far, the B2B experiments and associated results are relatively small in scale and scope, limited to one or two individuals. However, the non-invasive technologies hold much promise for making neural communication a reality. For example, although existing experiments have been limited to *one-to-one* scenarios, *one-to-many* or *many-to-one* connections for B2B interaction will soon be a reality (Rao et al. 2014). There are, or course, ethical implications associated with emerging B2B interfaces, including significant privacy concerns. As detailed by Trimper et al. (2014), privacy issues are highly relevant for any technology that can extract meaningful information from a person. For example, privacy for neural information should be no different than one's personal genetic information, or data gathered from tools like functional magnetic resonance imaging (fMRI) that measures brain activity and associated changes with blood flow given a particular stimulus (Nishimoto et al. 2011). More importantly, because neural activity can be transmitted over the Internet, when B2B communications become more pervasive, there are dangers associated with third parties hacking neural transmissions or the neural devices themselves. Amazingly, medical devices such as pacemakers have already been hacked using software-based radio attacks (Halperin et al. 2008). Issues associated with neural coercion, cognitive enhancement or the transmission of misinformation and/or false memories are also a concern (Ramirez et al. 2013). Indeed, the premise of the movie Inception (2010), where corporate espionage is conducted by implanting ideas into a target's subconscious mind may soon be a reality.

The implications of brain-to-brain communications for regions and their development are significant, particularly for transportation. As detailed above, the power of holographic telepresence is that it gives the audience the impression that real interaction is taking place. The nuances of body language and facial expressions provide a significantly more personal and engaging platform from which to conduct business and interact. As holography improves, it will likely replace a small percentage of business travel. Now, imagine the implications of direct, B2B neural communications, once the technology has improved beyond its existing, relatively simple framework. Hyper-detailed thoughts, emotions and intellect will be transferred over telecommunications networks, potentially providing more intimacy and meaningful exchange of information than would be possible in the standard face-to-face meetings of today. In fact, it is difficult to speculate on exactly how rich these neural exchanges will be, but it is likely that B2B communications will reduce the need for travel. More importantly, B2B communications are the most likely future technology to move transaction costs toward zero within a classic economic system (Nicol 1985).

3.5 Discussion and Conclusion

Telecommunications technologies and their associated infrastructure systems are continually improving. Just over 150 years ago, pundits were debating whether or not information transmitted over the telegraph was too fast to be reliable. Consider the following, gleaned from a New York Times article published in 1858 (LaFrance 2014, p. 1):

> Superficial, sudden, unsifted, too fast for the truth, must be all telegraphic intelligence. Does it not render the popular mind too fast for the truth? Ten days brings us the mails from Europe. What need is there for scraps of news in ten minutes? How trivial and paltry is the telegraphic column?

This text was written in response to the first transatlantic telegraph line installed between the Valentia Island in western Ireland to Heart's Content in eastern Newfoundland. Today, 20 years past the emergence of the commercial Internet, academic engineers and those working in the private sector are trying to develop techniques to reduce millisecond latencies on wireless telecommunication networks so that self-driving cars will not have any technical difficulties in navigating congested urban environments. Clearly, the importance of telecommunication, its associated infrastructure and application will continue to exert an influence on the economic, cultural, political, social and environmental fabric of regions. Future technological innovations will be no less controversial or be less prone to skepticism than previous technologies, including the telegraph, telephone or the Internet. If there is one major difference between the past and current technologies and those that are currently emerging and/or under development, it is the speed at which these technologies are being developed, tested and brought to market. For example, over

30 years elapsed between the first telegraph message and Alexander Graham Bell patenting the electromagnetic telephone (1876). It was another 90 years before the basics of the Internet (e.g., packet switching) were conceived of and applied to telecommunications systems in the U.S. (Abbate 2000). Today, advances are more numerous and more immediate. For example, standards for second-generation wireless technology were introduced in 1991, yet third (3G) and fourth (4G) generation technologies emerged in 2001 and 2006, respectively (Qualcomm 2014). 4G Long Term Evolution (LTE) standards were ratified in 2008 and 4G LTE networks were rolled-out by 2010 in the United States. As detailed previously, 5G represents the most significant overhaul to wireless standards, maybe ever, but it too is nearing rollout, likely in 2020.

3.5.1 A Regional Perspective

It would be easy to assume that upon the wide release of these technologies and their platforms, few barriers would exist (or emerge) regarding their adoption and use. For example, telecommunication providers will push strongly for consumers to upgrade their handsets to 5G capable devices. 5G data rates will be expensive at first, but as more providers rollout the technology and consumers migrate to 5G, economies of scale and competition will lower costs. The same can be said for holograms and brain-to-brain communication devices. Initial rollouts of this technology will be expensive, used primarily by corporate customers, but over time, the technology will diffuse and be adopted at the household or individual level.

The problem with these simplistic assumptions is that the influence of regions and place are completely overlooked. There is a long history of regions recasting, modifying and/or impacting technologies and their provision in wildly unexpected ways. This can happen through formal, policy-driven channels, or through informal entrepreneurial and/or grassroots movements.

Formal channels include policy mechanisms that can either promote, or hinder the widespread rollout, use or adoption of emerging technologies (Stoneman and Diederen 1994). For example, 21 U.S. states have regulations that either heavily restrict or prohibit municipalities from offering video, telecommunications or broadband services. In addition, these laws prevent municipalities from purchasing, leasing, maintaining, operating or building the infrastructure or facilities for providing video, telecommunications or broadband services (Gillette 2005; CBN 2015; Grubesic and Mack 2015). This is clearly a regional problem. For example, in 2011, the state of North Carolina passed the "Level Playing Field Act", which mandated that cities in the state cannot expand communications services beyond their corporate limits, that municipal telecommunications companies would have to pay the same fees that the government charges private companies and that cities cannot subsidize the communication service with other revenue sources (Billman 2015). In short, this act ensures that virtually no city can get its own network up and running in North Carolina.

In a recent plot twist, however, the FCC preempted this law in North Carolina and a similar one in Tennessee, ruling that “the need for broadband is everywhere, even if the business case is not” and that “it is clear that the combination of requirements effectively raises the cost of market entry so high as to effectively block entry and protect the private providers that advocated for such legislation from competition.” (Billman 2015). The state of North Carolina recently filed a lawsuit in federal court asking that the preemptive efforts of the FCC be reviewed. The case is ongoing.

Regardless of the outcome, the case in North Carolina exemplifies the somewhat strange and twisted regional machinations that occur when private companies attempt to provide a public good or service, including broadband (Grubesic 2006). Specifically, corporate entities often look to build, expand, and profit from local markets, but end up facing a constellation of counterforces (and policies) that attempt to safeguard local interests, ensure equity and protect communities. In such situations, it is often more realistic to establish public–private cooperation to ensure the needs to both groups are met (Girth 2014). This is especially true in North Carolina, where the local populace has already seen exactly how successful a municipal effort can be. In fact, the Greenlight broadband system in Wilson, NC, which was established prior to the 2011 Act, is widely considered a best practice example of municipal broadband for the United States (O’Boyle and Mitchell 2012).

From a grassroots/entrepreneurial perspective, many communities are forced to think creatively about technology to meet the needs of their local populace. Consider the plight of rural, Appalachian Ohio (Washington, County) during the early 2000s. There were no fiber-optic backbones (or plans for them) to support broadband telecommunication to its residents or businesses (Grubesic 2001). At the time, average costs to install fiber exceeded $20,000 per mile in areas where the terrain was not challenging, but given Washington County’s location and the complex topography of the Ohio River Valley, fiber installation costs had the potential to increase by a factor of 10 (Grubesic 2001). As a result, large telecommunication providers were not interested in expanding to Washington County, fearing that return on investment would be too low (Grubesic 2003). In their place, Sequelle Communications Alliance (SCA) was formed and plans were hatched to rollout wireless broadband services to the region (Sequelle 2004).

Unfortunately, accumulating enough spectrum to support broadband connections was difficult for SCA. Given the rapid growth of cellular communication systems in the U.S., especially during the mid-2000s, licensed spectrum was nearly impossible to obtain. In a creative twist, SCA made arrangements to utilize licensed spectrum from Mountain State College in Parkersburg, West Virginia. The spectrum was originally allocated for television applications, but the Federal Communication Commission (FCC) granted permission to switch the spectrum from analog to digital (Grubesic 2001) to support broadband services. In the end, the project was an abysmal failure. Although SCA received $3.295 million from the U.S. Department of Agriculture, a $600,000 grant from West Virginia’s Development Office and $400,000 in loans from the Mid-Ohio Valley Regional Council, the broadband service never manifested for the region (Dampier 2010). Worse, many lawsuits were filed, fraud charges were levied and the president of Mountain State College admitted to obstruction of justice and concealing facts in the case (Dampier 2010).

Although the SCA broadband particular project failed operationally, at its core, it was a conceptual triumph. Since the early 2000s, many other regions have taken innovative and entrepreneurial approaches for recasting and appropriating broadband technologies to fit their specific needs. For example, in the state of Minnesota, the most populous regions (e.g., Minneapolis, St. Paul, their suburbs, etc.) have a long history of excellent broadband service. The more rural areas of the states, however, had far slower service, if it existed at all (Grubesic 2008). Instead of tolerating this disparity and hoping that large broadband providers would eventually expand to the region, several telecommunication cooperatives[6] took matters into their own hands, including Paul Bunyan Communications (PBC) in Bemidji, Minnesota (Dampier 2012). In 2009, PBC began installing their own fiber and providing 25 Mbps to the community. In 2014, PBC embarked on a project to bring 1 Gbps speeds to the community (PBC 2015). PBC is not alone, many other coops and smaller telecommunications providers throughout the United States have followed suit, using a combination of wireline and wireless platforms to bring broadband service to their communities. In fact, since 2009, the USDA (2015) has provided more than $77 million to help such efforts.

Regions and telecommunication are inextricably linked and will continue to influence each other in both expected and unexpected ways. This chapter explored three major facets of telecommunication technologies and infrastructure. The first dealt with infrastructure provision and deployment. Even with all of the technological advances occurring at the ends of the cable plant (e.g., devices, routing equipment, lasers, etc.), optical fiber is, and will continue to be, the workhorse of telecommunications systems (both wired and wireless) for the conceivable future. Fiber is extremely fast, reliable, high capacity and relatively easy to install. In short, fiber provides significant room to grow (both in its capabilities and geographic footprint) as a transport technology for telecommunication systems. Second, the shifting needs and uses of telecommunications systems were documented, with particular attention given to the coming wave of demand associated with the Internet of Things and the IoE. Ubiquity is the key concept here, because the IoE requires a nearly continuous connection to the Internet to facilitate operations and exchange of information from massive sensor networks, adaptive energy allocation systems, home appliances and other connected devices. As a result, a more robust and ubiquitous wireless communication infrastructure will be required. Current development trajectories suggest it will be hierarchical in nature, using macro-to-femto level systems that enable the control and coordination of high bandwidth demands in the field. This also includes developments that facilitate device-to-device communications. Future telecommunications technologies, including holograms and brain-to-brain connections were detailed. Although there is much

[6]Cooperatives are businesses that are owned and run jointly by its members, who share the profits or benefits.

work to be done before these technologies are commercially mainstream, both offer the potential to reduce physical travel and facilitate more meaningful and intimate interactions between humans. Only time will tell if these technologies manifest accordingly. Finally, this chapter highlighted several key examples of how regions impact telecommunication systems and technology, emphasizing both formal policy mechanisms and entrepreneurial approaches to overcoming local challenges. Regardless of how future technologies manifest, leveraging a combination of efficient and meaningful policy with creative approaches for addressing regional telecommunication needs will be of paramount importance.

References

Abbate J (2000) Inventing the internet. MIT Press, Cambridge, MA

Agarwal R, Sands DZ, Schneider JD, Smaltz DH (2010) Quantifying the economic impact of communication inefficiencies in US hospitals. J Healthc Manag 55(4):265–281

Bekenstein JD (1981) Universal upper bound on the entropy-to-energy ratio for bounded systems. Phys Rev D 23(2):287

Best J (2014) The race to 5G: inside the fight for the future of mobile as we know it. TechRepublic. http://www.techrepublic.com/article/does-the-world-really-need-5g/

Billman JC (2015) North Carolina sues FCC over municipal broadband. Indyweek. http://tinyurl.com/hrcpwdf

Blakely T, Miller KJ, Zanos SP, Rao RP, Ojemann JG (2009) Robust, long-term control of an electrocorticographic brain-computer interface with fixed parameters. Neurosurg Focus 27(1):E13

Boccardi F, Heath RW, Lozano A, Marzetta TL, Popovski P (2014) Five disruptive technology directions for 5G. IEEE Commun Mag 52(2):74–80

Bouras C, Kokkinos V, Papazois A (2014) Financing and pricing small cells in next-generation mobile networks. In: Wired/wireless internet communications. Springer, Cham, pp 41–54

Bradley J, Barbier J, Handler D (2013) Embracing the internet of everything to capture your share of $14.4 trillion. http://tinyurl.com/cklzuh9

Bremermann HJ (1962) Optimization through evolution and recombination. In: Self-organizing systems. Sparton Books, Washington, DC, pp 93–106

Bremermann HJ (1965) Quantum noise and information. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol 4. University of California Press, Berkeley, CA, pp 15–22

Canon S (2014) Google fiber construction disrupts as it modernizes Kansas city. http://tinyurl.com/za3eaug

Capello R, Nijkamp P (1996) Telecommunications technologies and regional development: theoretical considerations and empirical evidence. Ann Reg Sci 30(1):7–30

Caragliu A, Del Bo C, Nijkamp P (2011) Smart cities in Europe. J Urban Technol 18(2):65–82

CBN (2015) Community broadband network. Community network map. http://www.muninetworks.org/communitymap

Chamberlain R (2015) IoT could demand a whole new, dedicated network—here’s why. VentureBeat. http://tinyurl.com/npu26bx

Chin WH, Fan Z, Haines R (2014) Emerging technologies and research challenges for 5G wireless networks. IEEE Wirel Commun 21(2):106–112

Christaller W (1933) Central places in Southern Germany (trans: Baskin CW). Prentice-Hall, London[1966]

Collier R (2013) Optical holography. Elsevier, Waltham, MA

Colapinto J (2014) Material question. The New Yorker. http://tinyurl.com/lbqxzm4
Dampier P (2010) Another guilty plea in rural west Virginia wireless broadband caper—$2.4 million ripoff. http://tinyurl.com/jaw3h7l
Dampier P (2012) Want rural 21st century broadband? Form a co-op or wait indefinitely for someone else to provide it. http://tinyurl.com/9ezr89g
Galperin H, Ruzzier CA (2013) Price elasticity of demand for broadband: evidence from Latin America and the Caribbean. Telecommun Policy 37(6):429–438
Gaspar J, Glaeser EL (1998) Information technology and the future of cities. J Urban Econ 43(1):136–156
Gillespie A, Williams H (1988) Telecommunications and the reconstruction of regional comparative advantage. Environ Plan A 20(10):1311–1321
Gillette SE (2005) Municipal wireless broadband: hype or harbinger. South Calif Law Rev 79:561
Girth AM (2014) What drives the partnership decision? Examining structural factors influencing public-private partnerships for municipal wireless broadband. Int Public Manag J 17(3):344–364
Goolsbee A (2002) Subsidies, the value of broadband, and the importance of fixed costs. In: Crandall R, Alleman JH (eds) Broadband: should we regulate high-speed internet access? Brooking Institution Press, Washington, DC, pp 278–294
Gorshe S, Raghavan A, Starr T, Galli S (2014) Broadband access: wireline and wireless-alternatives for internet services. Wiley, New York
Graham S, Marvin S (1996) Telecommunications and the city: electronic spaces, urban places. Routledge, London
Grau C, Ginhoux R, Riera A, Nguyen TL, Chauvat H, Berg M et al (2014) Conscious brain-to-brain communication in humans using non-invasive technologies. PLoS One, 9(8):e105225
Groenendyk M (2013) Emerging data visualization technologies for map and geography libraries: 3-D printing, holographic imaging, 3-D city models, and 3-D model-based animations. J Map Geogr Libr 9(3):220–238
Grubesic, T. H. (2001). A spatial analysis of internet accessibility. Doctoral dissertation, The Ohio State University
Grubesic TH (2003) Inequities in the broadband revolution. Ann Reg Sci 37(2):263–289
Grubesic TH (2006) A spatial taxonomy of broadband regions in the United States. Inf Econ Policy 18(4):423–448
Grubesic TH (2008) The spatial distribution of broadband providers in the United States: 1999–2004. Telecommun Policy 32(3):212–233
Grubesic TH (2015) The broadband provision tensor. Growth Chang 46(1):58–80
Grubesic TH, Mack EA (2015) Broadband telecommunications and regional development. Routledge, New York
Grubesic TH, Matisziw TC, Murray A (2011) Market coverage and service quality in digital subscriber lines infrastructure planning. Int Reg Sci Rev 34(3):368–390
Gu M (2015) Graphene champions the next generation 3D display technology. http://tinyurl.com/jy484yx
Hallett M (2000) Transcranial magnetic stimulation and the human brain. Nature 406(6792):147–150
Halperin D, Heydt-Benjamin TS, Ransford B, Clark SS, Defend B, Morgan W et al (2008) Pacemakers and implantable cardiac defibrillators: software radio attacks and zero-power defenses. In: IEEE symposium on security and privacy, SP 2008, May 2008. IEEE, Oakland, CA, pp 129–142
Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV (1993) Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev Mod Phys 65(2):413
Huff L, Fusek RL (1980) Color holographic stereograms. Opt Eng 19(5):195691
Hummel FC, Celnik P, Pascual-Leone A, Fregni F, Byblow WD, Buetefisch CM et al (2008) Controversy: noninvasive and invasive cortical stimulation show efficacy in treating stroke patients. Brain Stimul 1(4):370–382

Huurdeman AA (2003) The worldwide history of telecommunications. Wiley, New York
Inception (2010) http://www.imdb.com/title/tt1375666/
Jawanda S (2014) We can handles devices, but can we handle data? Raconteur. http://tinyurl.com/p6vcpnb
Kazovsky LG, Cheng N, Shaw WT, Gutierrez D, Wong SW (2011) Broadband optical access networks. Wiley, Hoboken
LaFrance A (2014) In 1858, people said the telegraph was 'too fast for the truth'. The Atlantic. http://tinyurl.com/p372ww5
Larsson E, Edfors O, Tufvesson F, Marzetta T (2014) Massive MIMO for next generation wireless systems. IEEE Commun Mag 52(2):186–195
Lawson S (2014) Bandwidth-sipping IoT steers clear of the net neutrality debate—for now. PC World. http://tinyurl.com/nztaw7e
Li X, Ren H, Chen X, Liu J, Li Q, Li C et al (2015) Athermally photoreduced graphene oxides for three-dimensional holographic images. Nat Commun 6:6984
Lloyd S (2000) Ultimate physical limits to computation. Nature 406(6799):1047–1054
Mack EA, Rey SJ (2014) An econometric approach for evaluating the linkages between broadband and knowledge intensive firms. Telecommun Policy 38(1):105–118
Mattern F, Floerkemeier C (2010) From the internet of computers to the internet of things. In: From active data management to event-based systems and more. Springer, Berlin, pp 242–259
Mellinger J, Schalk G, Braun C, Preissl H, Rosenstiel W, Birbaumer N, Kübler A (2007) An MEG-based brain–computer interface (BCI). NeuroImage 36(3):581–593
Moore GE (1965) Cramming more components onto integrated circuits. Electron Mag 38(8):114–117
Moss ML, Townsend A (2000) The internet backbone and the American metropolis. Inf Soc J 16(1):35–47
Murray AT, Grubesic TH (2013) Fortifying large scale, geospatial networks: implications for supervisory control. In: Crisis management: concepts, methodologies, tools and applications, 3rd edn. Information Science Reference, Hershey, pp 224–246
Nest (2015) Technical specifications for the 2nd generation nest thermostat. http://tinyurl.com/h4h54ac
Nicol L (1985) Communications technology: economic and spatial impacts. High Technol Space Soc 28:191–209
Nielsen J (1998) Nielsen's law of internet bandwidth. Nielsen Norman Group. https://www.nngroup.com/articles/law-of-bandwidth/
Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL (2011) Reconstructing visual experiences from brain activity evoked by natural movies. Curr Biol 21(19):1641–1646
Ngak C (2012) Tupac Coachella hologram: behind the technology. CBS. http://tinyurl.com/hauqub9
O'Boyle T, Mitchell C (2012) Wilson's greenlight leads North Carolina in fast internet. Institute for Local Self-Resilience. http://tinyurl.com/j929qyb
Open Technology Institute. (2014) The cost of connectivity 2014. http://tinyurl.com/p6ob8p2
Organization for Economic and Co-Operative Development [OECD] (2014) OECD broadband portal. http://tinyurl.com/pm6gqf8
Ota M, Fujita M (1993) Communication technologies and spatial organization of multi-unit firms in metropolitan areas. Reg Sci Urban Econ 23(6):695–729
Paul Bunyan Communications (2015) History. http://paulbunyan.net/cooperative/history.html
Plow EB, Carey JR, Nudo RJ, Pascual-Leone A (2009) Invasive cortical stimulation to promote recovery of function after stroke a critical appraisal. Stroke 40(5):1926–1931
Powell JR (2008) The quantum limit to Moore's law. Proc IEEE 96(8):1247–1248
Qualcomm (2014) The evolution of mobile technologies. http://tinyurl.com/nn655of
Ramirez S, Liu X, Lin PA, Suh J, Pignatelli M, Redondo RL et al (2013) Creating a false memory in the hippocampus. Science 341(6144):387–391
Rao RP, Stocco A, Bryan M, Sarma D, Youngquist TM, Wu J, Prat CS (2014) A direct brain-to-brain interface in humans. PLoS One 9(11):e111332

Rappaport TS, Sun S, Mayzus R, Zhao H, Azar Y, Wang K et al (2013) Millimeter wave mobile communications for 5G cellular: it will work! IEEE Access 1:335–349
Respers L (2014) Michael Jackson's hologram: creepy or cool? CNN. http://tinyurl.com/lybh4k8
RH (2015) Telecommunications. Dictionary.com Unabridged. Random House, New York. http://www.dictionary.com/browse/telecommunications. Accessed 24 Jan 2017
Rietveld P, Vickerman R (2004) Transport in regional science: the "death of distance" is premature. In: Fifty years of regional science. Springer, Berlin, pp 229–248
Saenz A (2010) 3D hologram prints invade military, commercial design. Singlularity hub. http://tinyurl.com/3xeajcy
Schaller RR (1997) Moore's law: past, present and future. IEEE Spect 34(6):52–59
Sequelle (2004) Sequelle communications alliance deploys aperto networks in a 2.5 GHz broadband wireless rus network across West Virginia and Ohio. http://tinyurl.com/jpkfh85
Specht PA (2015) Google, AT&T fiber installations draw complaints in Raleigh, Cary. News Observer. http://tinyurl.com/oqw5mox
Starr A, Brackmann DE (1979) Brain stem potentials evoked by electrical stimulation of the cochlea in human subjects. Ann Otol Rhinol Laryngol 88(4):550–556
Stoneman P, Diederen P (1994) Technology diffusion and public policy. Econ J 104(425):918–930
Summit Integration Systems [Summit] (2015) Introducing holographic telepresence: is it the future? http://tinyurl.com/zjf8y62
Trimper JB, Wolpe PR, Rommelfanger KS (2014) When "I" becomes "We": ethical implications of emerging brain-to-brain interfacing technologies. Front Neuroeng 7:4
United States Department of Agriculture [USDA] (2015) USDA announces funding for rural broadband projects. http://tinyurl.com/gq32ywt
US Telecom (2015) Broadband investment. http://tinyurl.com/z4jo74p
Vantage Point (2015) Wireless broadband is not a viable substitute for wireline broadband. http://tinyurl.com.kr8g9hl
Warf B (1995) Telecommunications and the changing geographies of knowledge transmission in the late 20th century. Urban Stud 32(2):361–378
Weber A (1929) Alfred Weber's theory of the location of industries. University of Chicago Press, Chicago
Worstall T (2013a) The beginning of the death of wired broadband. Forbes. http://tinyurl.com/k9dr6ya
Worstall T (2013b) Why high speed fiber is becoming irrelevant. Forbes. http://tinyurl.com/bm2r28v

Tony H. Grubesic is a Professor in the College of Public Service & Community Solutions, Arizona State University and Director of the Center for Spatial Reasoning & Policy Analytics. His research interests include GIScience, spatial analysis, urban and regional development, as well as environmetrics and telecommunications policy. Grubesic has held faculty positions at the University of Cincinnati, Indiana University, Drexel University and Oregon State University. Grubesic earned his Ph.D. in geography and regional science from the Ohio State University in 2001.

Chapter 4
Mobility and Technology Research: From the Industrial Revolution to Flying Vehicles in 2050

Roger R. Stough

4.1 Introduction

This chapter summarizes the ways in which mobility has been and will be affected by technological change over a relatively long period from the late 1800s to 2050 and the implications of this for research in regional science. The review covers the period from the Industrial Revolution to the mid-2000s. Out of necessity the scope will be broad, encyclopedic and somewhat general. The extrapolation of how mobility will be affected by technological change in the first half of the twenty-first century will be the primary and concluding focus of the analysis.

A major conclusion of the interpretive analysis of this chapter is that mobility of people and things (e.g., cargo) has been heavily influenced by technology and that this impact has grown stronger and exerted increasingly more impact over time. Given these general conclusions, regional science mobility research will continue to require adjustment if not radical transformation. In short, it will continue to be dynamic. This chapter seeks to draw conclusions about the nature and need for regional science mobility research over the next 30 years.

The chapter is organized chronologically around eras of technological change. So, in the last half of the nineteenth century such technologies as the steam engine and the emergence of wireline information transmission were the focus. In the twentieth century, technologies that led to changes in transportation and communication options, such as automobiles (internal combustion engines), airplanes (jet engines), space travel (rocketry), high speed rail, and information technology and telecommunications, were important. Issues of micro level mobility, unlike earlier developments of a more macro level, began to take on importance as well as

R.R. Stough
Schar School of Policy and Government, George Mason University, Fairfax, VA, USA
e-mail: rstough@gmu.edu

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_4

increased attention being given to the support needs for various dependency groups, such as the aged, children, wounded and the handicapped, commuters, etc., as the issue of equity assumed greater importance. The twenty-first century analysis extrapolates the fusion of new and enhanced information technology (IT), energy, transportation and telecommunications technologies leading to enhanced vehicles for moving about, i.e., automobiles, airplanes, space travel, prosthetics, mobile telecommunications and the increasing substitution of communication for travel.

It is useful before embarking on the analysis to define some terms. Mobile things are those that are "capable of moving or being moved readily". The phrase "capable of moving" implies underlying intentionality or motivation and thus movement of animated things, such as people or animals; the term "being moved" may be viewed as movement of things (but in some cases, humans as well) and, the term "readily" implies moving with little constraint (or moving less "readily"), depending on the level of barriers to movement. With mobility there are a variety of physical and institutional barriers, thus, there is always a concern with efficiency of moving (or being moved). And while technology aims to reduce barriers to mobility, there may also be unintended consequences that in fact increase constraints despite enhancing movement. In this context, mobility's sister concept, accessibility, becomes important. Accessibility may be defined as "approachability" or "being able to be reached or approached" (Vocabularity.com); and in the context of mobility, it implies either a constraining influence or a complement to moving about.

4.2 Mobility and the Industrial Revolution up to 1900

4.2.1 Technology

Prior to the twentieth century, mobility of people and things was supported mostly by horse and horse and buggy, boats on inland and coastal waterways, walking, wind or human powered ships and railroads. So mobility, for the most part, was limited due to the slow nature of these supporting modes of movement and limited communication about scheduling and arrival times. The invention of the steam engine reduced this constraint somewhat in that it enabled much faster rail and sea travel. This reduced the time to specific places but had a marginal impact on enhanced local mobility as travel by rail was confined to the nodes or city regions that it connected. Other mobility support services, such as horses and carriages, were required once one reached points closer to the destination. In short, even in the early twentieth century most mobility and travel was by walking, horse and/or horse and buggy, and some railroads and steam ships for ocean and river travel. Despite the more macro level context that railroad and steamship travel represented, they improved mobility over long distances measurably.

4.2.2 Research and Modeling

Research in the 1800s regarding mobility was mostly in map making and improving tools to increase mobility of humans and cargo with the new steam powered vehicles. But logistic support for human and cargo travel was elementary as telegraphing, which replaced the "Pony Express" in the U.S., was of use for city to city communication and along railroads, but many areas and locales were not linked in and telephoning was just beginning to expand beyond limited availability in the largest cities and was mostly for commercial use at the end of this period.

Johann Heinrich von Thunen's publication "The Isolated State" (1826) is often recognized as the first formal theory of location of economic activities. It is important for this analysis in that it formulated a theory of efficient location of agricultural land uses based on transport cost, i.e., mobility efficiency. von Thunen's work, as we will see, was later used as a platform for creating a theory of urban land use (Alonso 1964). In short, mobility related research leading up to the twentieth century was limited with the only major formal theoretical foundation for research being the work of von Thunen.

4.3 Mobility in the Twentieth Century

4.3.1 Technological Change

The twentieth century witnessed huge technological change with corresponding huge expansionary impacts on mobility and the modes created to support moving about at all scales, including international, national, regional and local. For example, rail based local transportation made possible the development of suburbs at the local level. Further, automobiles invented around the turn of the century had become highly sophisticated vehicles not only from their design and motive technology but also from the addition of telecommunications and advanced IT by the dawn of the twenty-first century. Airplanes evolved after the early 1900s to a capability of flying passengers and cargoes half way around the world, thereby connecting literally all major urban places in the world by direct air travel by the end of this century. Moreover, air travel was by then largely guided by autopilots, including both take-off and landing supported by advanced IT and telecommunications technology. It was not uncommon to fly from Chicago to China or such destinations as Japan, Australia, India, Indonesia, Turkey, etc. in less than 1 day, conduct business and return to Chicago the next day. Although possible, most travelers prefer to have more time at the destination to recover from jet lag.

Railroads that had become common by 1900 evolved into supporting huge and complex mobility networks as well as superfast trains (Shinkassen (Japan), TGV and Ice trains in Europe, and Acela in the U.S.) enabling travel from origin to destinations in extended urban corridors (Gottman 1961) or to multiple nodes in national and multinational networks, often in times no longer than required to make trips by plane up to distances of as much as 500 miles. Yet places not on these networks required, at least in part, other forms of travel, including motorized vehicles to reach the destination.

Those born early in the twentieth century, living near its end and on into the first few years of the twenty-first century, could look back to a time when most travel or mobility was still by horse and wagon, railroads, and walking and bicycling from an era of fast automobiles, fast trains, jet airplanes including SSTs and the increased likelihood of space travel, or at least tourism, to beyond the Earth's atmosphere. Building technology also contributed resulting in, for example, tunnels over 30 miles in length. Interstate highways, pioneered in Germany, markedly reduced travel time even without technological progress in automobile design and operation. Further, IT and telecommunications technology advanced and evolved to support what some call smart cars, planes and trains for mobile humans and cargoes. Such smart vehicles have continuous contact between different sub-systems via electronic sensors that provide corrections to malfunctions, internal and external safety threats and information about constraints and barriers to mobility. This technology also enables such mobility supporting vehicles with continuous connectivity to other places, agents and information centers, thus enabling mobile actors to execute tasks remotely including monitoring the transport of cargoes. The latter part of the twentieth century and the early part of the twenty-first century witnessed the gradual merging of mobility of people and things with mobile communication systems, which enabled complementarity and substitution of each for the other (Moktarian 1990).

4.3.2 Research and Modeling

Mobility related research and theory, like mobility itself, evolved and expanded enormously during the twentieth century. Theory moved beyond the agricultural land use work of von Thunen to urban land use models, including the work of Alonso (1964), which extended the earlier agricultural model to the urban context. But other theoretical land use models appeared such as the Park-Burgess Concentric Zone Model (Park 1915; Park et al. 1925) and the Hoyt sectoral land use model (Hoyt 1939), and later the multi-nuclear model (Harris and Ullman 1945), which gradually replaced the assumption of a core dominated city as the primary model. The Christaller (1966), Losch (1940) and Weber (1929) models also appeared in the first half of the twentieth century, which in different ways assumed minimizing transport costs or achieving efficient transport; i.e., mobility was important and played a critical role in defining where things (manufacturing plants, cities, towns and the structure of landscapes) would or should be located to create efficient mobility of humans and things to access these places. At the base, the objective of all of these models was achieving mobility efficiency and that accessibility *ceteris paribus* would be optimal under models that assumed a maximizing function for mobility.

In addition to these mostly land use models, network optimization and spatial interaction models were designed to guide the development and management of transportation systems as well as to locate facilities to enhance accessibility. Later after the mid-century mark, network optimization (Boyce et al. 1983; Goldner 1971; Ben-Akiva et al. 1996; Lowry 1964), transportation simulation (Barrett et al. 1999; Clark and Holm 1987; Landis 1992; Southworth 1995; Mackett 1990; Nagel et al. 1999; Wegener 1982) and spatial interaction models (Haynes and Fotheringham 1984) had evolved that enabled modeling and simulating the optimization of mobility as well as accessibility. Parallel to this work, equity emerged as an important variable in the design and management of urban space leading to research like that of David Harvey (1973). So, models attempting to maximize mobility and access were increasingly required to include factors that could or might constrain equal access to all groups (other abled, children and youths, aged, ethnic minorities, women and other disadvantaged groups). Further, the last half of the century witnessed the rise of concern for environmental integrity as the provision of clean air and water became major factors that needed to be considered in attempts to maximize mobility and related efficiency of movement (EPA 2000; Wegener 1996). Hence, network models included variables in addition to transport cost that impacted mobility and with models seeking to optimize mobility efficiency subject to equity and environment factors as well as cost of fuel. Further, an increased interest in institutional factors in mobility was observed (Rietveld and Stough 2005).

4.3.3 Integration and Fusion of Technology

Parallel to the overall concern with ensuring mobility and accessibility were the rapid advancements in IT and telecommunications technologies, such as the Internet and mobile communication, and their merging with computer technology. This development was initially called Intelligent Vehicle Highway Systems—IVHA and later was renamed as simply Intelligent Transportation Systems (ITS) (Stough 2001). In the 1990s and on into the twenty-first century, communication support systems evolved to the stage where mobility and accessibility information was now available in near real time at most places and for most agents. In turn, this capability had begun to serve as a major complement and substitute for trip taking (Button et al. 2006; Stough et al. 2003; Moktarian 1990). In addition to personal mobile technology, video technologies were emerging that enabled virtual conferencing as a substitute for personal transport and mobility. One highly significant outcome of this was an increasing ability to telework and teleconference, thereby substituting remote communication for trip taking to the office and other meetings with communication. This began to occur increasingly not only for local metropolitan level travel but also over increasingly long distances, e.g., transcontinental or international. But this is what was occurring in the first decade of the twenty-first century. In the next part of the chapter we turn the clock forward 30–40 years to the age of flying vehicles and beam energy.

4.4 Mobility Going Forward into the Twenty-First Century

4.4.1 Technological Change

For a portal into the future, we now extrapolate technological change as a means of identifying the likely impact of new transportation technologies on regional science and its models as they evolve in the continuing effort to enhance mobility efficiency and equity. The world has witnessed an ever increasing rate of technological change since the Industrial Revolution and similarly a huge impact on human mobility. There is little reason to believe that technology will not continue to evolve at an increasing rate other than the possibility of huge global natural and or man-made disasters, which we assume will not occur. The basic assumption upon which this part of the chapter rests is that technological change will continue to occur at an increasing rate and that it will in turn impact mobility and accessibility accordingly.

What will a scenario of life 25–35 years from now look like? The major technologies we have seen that have impacted mobility over the past 150 years or so have been in the area of energy and related propulsion, transportation, information technology and telecommunications. We focus on these going forward.

Energy sourcing, production and use is in the midst of a major transformation with emphasis on moving away from more traditional combustion fossil fuels to alternative sources, such as solar, wind, bio-fuels, natural gas, nuclear and fuel cells. It is likely that these will come to dominate energy markets over the next several decades. But in the background there are other alternatives, such as highly improved and safer nuclear power and even nuclear fusion possibly being available by mid-century (Mahaffey 2009; Nuttal 2005). Further, beam energy (Landis 2004; Marx 1966) or power produced in solar and or nuclear plants located at the edge of the Earth's atmosphere may be available by 2050 or sooner, whereby energy collected and produced in outer parts of the earth's atmosphere or beyond will be beamed to earth to support a variety of energy demands and mobility enhancements.

This transformation in energy supply and use, when coupled with other technologies as described below, opens enormous opportunities for enhanced mobility that will be needed to support movement in cities of the future that may far exceed the size of the world's largest mega cities of today (International Federation of Surveyors 2010). It is not unreasonable to expect to witness cities with populations of as much as 50–60 million by 2050 (Lewis 2007) given that Tokyo, the world's largest megacity, has about 37 million residents today. To support efficient mobility in urbanized complexes of this size will require use of underground subways or tube transport at much higher levels than today and likely a more intense use of air space over cities. We already see the leading edge of this possibility with the development and use of drones (Beard et al. 2005) for not only defense but also to support emergency response, disaster recovery, real estate sales, police and fire services, retail delivery logistics, urban and transport planning and management, and so on. Below we discuss the possibility of flying personal vehicles as a possible solution to using air space more intensively for local travel. In the meantime, we take a digression to discuss the evolving fusion of energy, telecommunications, IT and computer technologies and how this development might impact surface mobility over the coming decades. This discussion is important because the same technology that is evolving to support land based two-dimensional travel by auto or high speed and mag lev trains will be important for the age of flying vehicles.

4.4.2 Integration and Fusion of Technologies

Perhaps the most important transport technological change ongoing in this early part of the twenty-first century is the integration or fusion of computer, information and telecommunications technologies with evolving motive technology, such as hybrid and all electric vehicles, as they move beyond the novelty phase. Personal and commercial autonomous vehicles are already being tested in seven states (National Conference of State Legislatures 2015). The application of this technology is in a nascent, if not embryonic, stage of development with perhaps the most important immediate or near term actual application being to create a backup-system for driver vehicles that intervenes when a safety, security of collision threat appears eminent. For example, some personal and commercial vehicles already have very sophisticated onboard collision avoidance technology. This fusion of technologies enables global positioning, way finding and guidance especially in remote, unchartered or unknown areas, but soon these way finding systems will be fully erected on geo-location data bases for the entire world. Further, even today's vehicles enable continuous connectivity to communication via on-board telephoning and messaging, meaning that drivers and riders can be in nearly complete communication with family, friends, colleagues and clients in real time and at almost all times.

Looking ahead 25 or so years gives every reason to assume that personal and commercial travel, and thus, mobility will be monumentally transformed. Vehicles will be autonomous and rarely driven by humans and, thus, driving, or should we say riding, behavior will be transformed beyond our current experience. For example, a recent report (Delaware Institute 2016) envisions sports stadia reaching a capacity of 250,000 seats with considerably reduced footprints as participants will arrive by autonomous vehicles that drop them off before and pick them up after the event. Thus, the need for vast parking areas at the stadium will no longer be required. Smaller land requirements mean that a stadium can be more inexpensively located in downtown areas thereby opening new avenues for optimizing aggregate mobility.

Driving to distant destinations may become more realistic given the possibility of being picked up in the evening and bedding down in an autonomous vehicle that arrives at the destination the next morning, thus enabling the rider to sleep all during the trip. Further, autonomous vehicles by definition drive themselves once a target destination(s) is provided, thus enabling those with health or other barriers to "drive" and travel easily by car, such as dependent populations (e.g., children and the aged). Also, travel and traffic related accidents are expected to be hugely reduced, if not totally eliminated, due to collision avoidance technology, thus helping the US Federal Highway Administration (FHWA) reach its ultimate goal of no traffic deaths on U.S. roadways (FHWA 2015).

On board communication and visualization technology will enable passengers to be entertained and to entertain, do business, reach friends and relatives and work on team projects anywhere in the world while in route to a destination and/or even substitute trip taking with enhanced quality communication systems. Mobility should, thus, be enhanced at the same time that productivity and recreation time is increased. Commercial transport and mobility will be different as well. It is possible within the next 10 years that commercial transport will be largely accomplished by autonomous vehicles and by 2050 solely by driverless vehicles with on-board communication technology that connects the sender with the recipient client as needed if problems arise with the protection of the cargo, deliveries, etc. As well, passenger and freight trains will operate autonomously and provide the same quality of communication services that personal and commercial vehicles will have.

4.4.3 Flying Vehicles and Beam Technology

There are energy and transport technologies on the drawing board that by 2050 will enable the creation of lightweight vehicles that fly over many of the world's urban locations. Such vehicles are expected to weigh as little as 200 pounds (DaVinci Institute 2016) because they will not have an onboard source of propulsion. These machines will be propelled by beam energy supplied from outer atmosphere power plants fueled by collection and transmission of solar or nuclear produced energy or a hybrid of these sources (DaVinci Institute 2016; Landis 2004; Marx 1966). How likely is this scenario? Much of the technology for such flying vehicles is already available, and the air traffic control system technology is essentially known for managing a large fleet of simultaneously flying vehicles as might be required for commuting to work or for shopping and entertainment in metropolitan airspace. That said, beam energy production and applications that make light flying vehicles possible are in a nascent stage of development. So, the critical technical barrier for personal flying vehicles is an available external source of energy.

The air traffic control system required for managing and controlling the flight of a large fleet of personal vehicles would draw upon existing two-dimensional land based traffic control models and the technology that is envisioned to safely manage "free flight" of current air transport (DaVinci Institute 2016). Such free flight management systems have been envisioned for some time (GAO 2001), and the U.S. Federal Aviation Administration (FAA) has long sought to develop the technology to manage such a flight system in an ideal world with onboard collision avoidance technology. In the European Union, a hybrid system of free flight already exists with some airway control as planes near landing or at takeoff. Such a hybrid system would be required to manage flying vehicles over urban locations in the future.

One system that has been envisioned is that flight would be channeled into circular patterns with different elevations used for flights to different directional locations in urban areas, a system DaVinci calls “directional layering of airspace” (DaVinci Institute 2016, p. 2). At takeoff these personal planes would rise under centrally controlled guidance technology to an elevation for the intended destination and would then be maintained in that flight pattern until nearing the destination when an autonomous central controller would guide the landing. A type of free flight would be possible at the different elevations and control would be central for takeoff and landing. These light vehicles would have on board parachutes to ensure a safe landing if technical difficulty or an accident occurred during the flight. Collisions with such advanced on-board avoidance and traffic control systems for takeoff and landing will likely have advanced to the point where collisions and breakdowns would be rare events in 2050.

4.4.4 Research and Modeling

What does all of this visioning about what might be possible mean for regional science research? One would expect that land based two-dimensional traffic control systems coupled with on-board vehicle communication, navigation and collision avoidance technology would serve to manage the fleet of largely autonomous personal and commercial vehicles by 2050. Driver autos will be rare events. Travel and mobility will be safer, more efficient and pleasurable, and contribute to higher levels of human productivity. The traffic control systems of the future will be at base hybrid in nature, including both central control systems that manage the fleet to maximize or optimize system-wide mobility efficiency as well as manage equitable use and access while at the same time ensuring safety and security. Safety and security will be managed, in part, by on-board navigation and collision avoidance technologies. At the same time, a central control system that is in communication with every operating vehicle at all times, i.e., has perfect information at all times, can intervene and redirect or channel vehicles into alternative routes as hot spots arise in the system, from the rare collisions or technical breakdowns that may occur. The traffic control systems of the future will operate in faster than real time and be erected on dynamic algorithms as proposed in the late twentieth century (Friesz et al. 1996) and by Kulkarni et al. (1996) using an approach borrowed from particle physics called spin glass modeling (Kulkarni et al. 1996) that enables not only real time monitoring of traffic system performance but anticipates the beginnings of “hot spots” or bottlenecks that lead to congestion. Such a modeling approach revolves around two basic parameters: (1) identification of the beginnings of hot spots and, (2) estimation of how fast and large the bottleneck will become. Such information can be used to dynamically manage a large urban surface transport system optimally by intervening and re-routing traffic flow to support optimally efficient mobility and accessibility.

The Friesz et al. (1996) and Kulkarni et al. (1996) approaches offer insight into the faster than real time systems that will be required for the management of travel in the mid-twenty-first century. Travel, and thus mobility, will rely on methods that are erected on perfect information about all the vehicles operating in a geographic area and thus can anticipate bottlenecks and reroute vehicles to ensure optimal traffic flows despite hot spots that may develop (of course, subject to a variety of constraints including ensuring equity goals and environmental integrity). Further, onboard navigation and collision avoidance systems will prevent most collisions and accidents, and because all vehicles and their systems are known by the central controller, adjustments needed to ensure that optimal system functioning can be achieved. The current systems for land and air based traffic control systems do not have the luxury of perfect information about all the vehicles operating in an arena. Consequently, they cannot be operated in real or faster than real time to optimize travel and subject to, for example, safety and environmental integrity constraints. The systems of the future will be ones that ensure that travel system operation is optimized in terms of an objective function that not only optimizes efficiency, equity and safety because such systems anticipate problems and intervene to avoid them. Traffic management and control models of the future will operate in faster than real time and will be totally dynamic unlike the systems of today and the past that do not have the luxury of perfect information.

One might ask with such a system's operation how equity in mobility and access would be achieved. An example helps illustrate this. Today urban road pricing applications have become fairly common. While some argue that this results in little other than to create "Lexus lanes", i.e., lanes that enable those who wish to pay an extra demand related fee to move more continuously and efficiently through urban space, recent evidence finds that both those who can and will pay and those who cannot pay gain efficiency from such investments (Malone 2014). Those electing to use tolled options gain efficiency because pricing is used to ensure that travel is unimpeded; those who do not use the priced option also gain efficiency in that reduced use of the non-tolled infrastructure occurs because of those who choose to use priced lane(s) opt out, thereby increasing efficiency of both user groups and, thus, a more equitable outcome for those not using, or financially unable to use, tolled lanes. In short, providing priced lanes seems to create a socially optimal outcome regarding mobility efficiency. Traffic control systems by 2050, in the age of autonomous vehicles, can manage a surface travel system and infrastructure that includes priced lane options. All that is needed is a decision to use or not-use the priced lanes. The system will manage such distributed programming of thousands of vehicles and their routing.

The mobility supporting systems of the future will be dynamic and operate in faster than real time. Research in the future will increasingly be aimed at enhancing efficiency, equity and safety just as it has been historically since almost the time of von Thunen in the early nineteenth century. What has changed is that unlike in von Thunen's era, it will be possible to assume that perfect or near perfect information will exist regarding all the elements in a transportation system. Once that assumption is satisfied, it is a matter of programming the system to achieve optimum system outcomes. Such programming will likely take on a superstructure along the lines of optimization modeling whereby an objective function, e.g., efficiency, maximization is subject to equity and safety constraints or multiple objectives are moved in and out of the objective function. At the same time, operational programming will be required to identify and rectify bottlenecks or hot spots in the system with sophisticated extensions of the Friesz et al. (1996) dynamic systems model or the Kulkarni et al. (1996) spin glass modeling approach.

4.5 Discussion and Conclusions

4.5.1 Discussion

To this point in the essay the focus has been upon future mobility supported by individual modes of transportation. The reader may legitimately ask "what sorts of public or public subsidized transportation in urban, extra urban or inter-city contexts will exist in 2050 and how they will impact mobility?"

Regarding the extra urban and inter-city contexts, it is likely that high speed rail will probably have been upgraded to magnetic levitation for greater speed that will provide travel for distances up to perhaps 1000 miles or more with commercial air serving longer trips. Even magnetic levitation trains will have been transformed significantly from the nascent technology used today for example, a solar powered magnetic levitation fast train has been proposed that would link Phoenix and Tucson, Arizona (Inhabitat 2010) and an even grander concept, proposed by Elon Musk, is a 700 mile per hour solar hyperlink magnetic levitation train that would span the U.S. from Atlantic–Pacific coastal locations at low cost and which would require only a few hours for the trip (Inhabitat 2013). Places with high levels of sunlight may have adopted such trains by 2050. The effects of such developments would be to reduce the cost and time of longer distance travel and thus increase mobility and accessibility. It is an important aside here to note that by 2050 parabolic flight will likely support intercontinental flights thus reducing trip times to an hour or two at most from say New York to Sydney Australia.

At the intra-urban level a question still remains as to what form local public transportation will take by 2050. It would be a stretch to envision that large capacity flying vehicles would be affordable or efficient for moving people and cargo about in metropolitan space. What makes the vision of individual flying vehicles attractive is that it would be convenient and speed up travel for individuals or quite small families or groups. Although, it is probable that flying drones would handle a lot of logistics and reverse logistics support.

Most likely some adaptation of magnetic levitation tube or subway transit systems would provide public support for travel in city regions. The future in 2050 may also be that advanced individual vehicles may provide linkage for "last mile" destinations when these distances are too far to conveniently walk to, given such subway systems cannot directly link all possible destinations. Such individual vehicles may be highly advanced powered bicycles or other individualized vehicles like, for example, "Segways" that can be accessed at major subway destinations?

4.5.2 Conclusions

We have seen the evolution of mobility related research advance from the elementary efficiency model of von Thunen to a vision of models that control and manage faster than real time models that optimize a mix of mobility objectives such as efficiency, equity and safety. It is of note that the basic objectives of mobility research are not likely to change much even by 2050, rather, what has changed is the technology that creates not only highly efficient vehicles and related models which in turn enable vast improvements in mobility. Technology has evolved from steam engines supporting improvements in rail and sea transport and related mobility, to a vision that provokes believability when we look out a generation or so into the future.

Future mobility will likely be supported by robotic and central control systems whereby travelers are, at best, observers. Given this, it will be important to ask whether this represents a future where freedom is highly constrained. Some would argue that such a future creates more freedom as the traveler still chooses the route and mode but does not have to bother with the details of mobile behavior. Yet there also are potential privacy issues that are just beginning to surface in the "Big Data" era, because systems that are erected on perfect information threaten privacy of the individual traveler, and in turn, freedom. Perhaps mobility research in the age of autonomous vehicles and perfect information guided central control systems will focus more on issues like protection of privacy and freedom than on achieving improved mobility as the envisioned systems suggest that mobility optimization will be so well managed that it will be a secondary concern. Emerging issues like freedom and privacy may only be forerunners that hint at a future where many engineering and technological problems are relatively easily addressed while society tradeoffs in terms of actualizing personal intentional behavior actually dominate!

References

Alonso W (1964) Location and land use: toward a general theory of land rent. Harvard University Press, Cambridge, MA

Barrett CL, Beckman RL, Berkbigler KP, Bisset KR et al (1999) Transportation analysis simulation system (TRANSIMS). Version TRANSIMS-LANL-1.0. Volume 0: Overview. LA-UR 99-1658. Los Alamos National Laboratory, Los Alamos, NM

Beard RD, Kingston M, Quigley M et al (2005) Autonomous vehicles technology for small fixed-wing UAVs. J Aerosp Comput Inf Commun 2(1):92–108

Ben-Akiva ME, Bowman JL, Gopinath D (1996) Travel demand model system for the information era. Transportation 23:241–266
Boyce DE, Chen KS, Lee YJ, Lin KT (1983) Implementation and computational issues for combined models of location, destination, mode, and route choice. Environ Plan A 15:1219–1230
Button K, Stough R (with Bragg M, Taylor S) (eds) (2006) Telecommunications, transportation, and location. Edward Elgar, Cheltenham
Christaller W (1966) Central places in Southern Germany. Prentice-Hall, Englewood Cliffs
Clark M, Holm E (1987) Microsimulation methods in spatial analysis and planning. Geogr Ann 69B:145–164
DaVinci Institute (2016) 2050 and the future of transportation frictionless vehicles and binary power will define transportation in the world ahead. Davinci Institute, Westminster, CO
Delaware Institute (2016) http://www.futureof.org/wp-content/uploads/The-Future-of-Sport-2015-Report.pdf
EPA (US Environmental Protection Agency) (2000) Projecting land-use change: a summary of models for assessing the effects of community growth and change on land use patterns. EPA/600/R-00/098. US Environmental Protection Agency, Office of research and Development, Cincinnati, OH. http://www.epa.gov/ecocommunity/tools/reportfinal3.pdf
FHWA (2015) http://www.safety.fhwa.dot.gov/tzd/cx_21cities_ml_0611megacities.html. Accessed 4 July 2016
Friesz T, Bernstein D, Stough R (1996) Dynamic systems, variational inequalities and control theoretic models for predicting time-varying urban network flows. Transp Sci 30(1):14–31
GAO (2001) National airspace system: free flight tools show promise, but implementation challenges remain. GAO-01-932 report, Washington, DC. http://www.gao.gov/assets/240;231996.pdf
Goldner W (1971) The Lowry model heritage. J Am Inst Plann 25:73–76
Gottman J (1961) Megalopolis: the urbanized northeast seaboard of the United States. Twentieth Century Fund, New York
Harris CD, Ullman E (1945) The nature of cities. Ann Am Acad Pol Soc Sci 242:7–17
Harvey D (1973) Social justice and the city. Johns Hopkins Press, Baltimore
Haynes KE, Fotheringham AS (1984) Gravity and spatial interaction models. Sage, Thousand Oaks, CA
Hoyt H (1939) The structure and growth of neighborhoods in American cities. Federal Housing Administration, Washington, DC
Inhabitat (2010) http://wws.inhabitat.com/transportation-tueseday-arizona-startup-unveils-solar-bullett-train
Inhabitat (2013) http://www.nhabitat.com/solar-powered-maglev-concept-train-for-bombardier/
International Federation of Surveyors (FIG) (2010) Rapid urbanization and mega cities: need for spatial information management (Research study by FIG Commission 3). International Federation of Surveyors, Copenhagen
Kulkarni R, Stough R, Haynes K (1996) Spin glass and the interactions of congestion and emissions: an exploratory step. Transp Res C 4(6):407–424
Landis JD (1992) BASS II: a new generation of metropolitan simulation models. Working paper 573. Institute of Urban and regional Development, University of California, Berkeley, CA
Landis G (2004) Interstellar flight by particle beam. Acta Astronaut 55(11):931–934
Lewis M (2007) Megacities of the future. http://www.forbes.com/2007/06/11/megacities-population-urbanization-biz. Accessed 4 July 2016
Losch A (1940) Die raumliche ordung der wirtschaft. English edition: Woglom WH, Stolper WF (trans) (1954) The economics of location. Yale University Press, New Haven
Lowry I (1964) A model of metropolis. RM-4035-RC. Rand Corporation, Santa Monica, CA
Mackett RL (1990) Comparative analysis of modelling land-use transport interaction at the micro and macro levels. Environ Plan A 22:459–475
Mahaffey J (2009) Atomic awakening: a new look at the history and future of nuclear power. Pegasus Books, San Jose, CA

Malone K (2014) Are "Lexus Lanes" really Lexus lanes? http://www.wlrn.org/post/are-lexus-lanes-really-lexus-lanes
Marx G (1966) Interstellar vehicle propelled by laser beam. Nature 211:22–23
Moktarian P (1990) A typology of relationships between telecommunications and transport. Transp Res 24A(3):231–242
Nagel K. Beckman RJ, Barrett CL (1999) TRANSIMS for transportation planning. LA-UR 98-4389. Los Alamos National Laboratory, Los Alamos, NM. http://transims.tsasa.Lanl.gov/PDF_Files/LAUR98-4389.pdf
National Conference of State Legislatures (2015) http://www.ncsl.org/rsearch/transportation/autonomousvehicles-legislation,aspx
Nuttal WJ (2005) Nuclear renaissance. Taylor & Francis Group, New York
Park RE (1915) The city: suggestions for the investigation of behavior in the city environment. Am J Sociol 20:579–583
Park RE, Burgess E, McKenzie R (1925) The city. University of Chicago Press, Chicago
Rietveld P, Stough R (eds) (2005) Barriers to sustainable transport: institutions, regulations and sustainability. Spon Press, London
Southworth F (1995) A technical review of urban land use-transportation models as tools for evaluating vehicle travel reduction strategies. ORNL-6881. Oak Ridge National Laboratory, Oak Ridge, TN. http://www.tmip.fhwa.dot.gov/clearinghouse/docs/landuse/compendium/ornl_toc.stm
Stough R (ed) (2001) Intelligent transport systems: cases and policies. Edward Elgar, Northhampton, CN
Stough R, Higano Y, Button K, Nijkamp P (eds) (2003) Transport and information systems. Edward Elgar, Cheltenham
von Thunen KH (1826) Der isolierte staat in beziehung auf landwirschaft und national-okonomie. Perthes, Hamburg. English edition: Wartemberg CV (1966) Von Thunen's isolated state. Pergamon, Oxford
Weber A (1929) Alfred Weber's theory of the location of industries. The University of Chicago Press, Chicago
Wegener M (1982) A multilevel economic-demographic model for the Dortmund region. Syst Urbani 3:371–401
Wegener M (1996) Reduction of CO_2 emissions of transport by reorganization of urban activities. In: Hayashi Y, Roy J (eds) Transport, land use and the environment. Springer, Heidelberg, pp 103–124
www.thefolexdictiony.com/mobility
www.vocabularity.com/dictionary/accessibility

Roger R. Stough is university professor, the Schar School of Policy and Government, Vice President for Research and Economic Development (2008–2013) and currently Vice President, of the Regional Science Academy, an affiliate of the Regional Science International Association. His primary research interests are regional economic development and policy, transport policy, and entrepreneurship and innovation policy. He holds an honorary doctorate of philosophy, University of Jonkoping, Sweden.

Chapter 5
Entrepreneurship, Growth, and Gender

Tessa Conroy and Stephan Weiler

5.1 Introduction

All regions have fundamental economic development goals. Perhaps the most foundational aspiration is that of quality job growth, with continuing streams of new employment in industries and niches that are innovative with high labor value-added. Yet in direct contrast to many of the employment pillars of the nineteenth and most of the twentieth centuries, the most dynamic and intriguing of such jobs now come from small firms. Innovation no longer is the sole purview of major industrial research laboratories, but has spread out to the grass-root garages and workbenches across both rural and urban landscapes. Entrepreneurship has become the primary channel through which innovation tests the market frontier, with survival and growth being the clearest form of success.

Macroeconomic research has made clear the surprisingly dominant role of startups in particular in the job creation process. Yet these individual entrepreneurial projects are difficult to discern in the haystack of aggregate economic activity. Regional science has shed light on the key entrepreneurship-growth nexus by offering a more discerning and natural spatial lens through which to examine that key relationship. Yet answering fundamental questions about entrepreneurship, growth, and innovation has simply led to a more penetrating series of questions. In this essay, we sketch an agenda for regional research in this field, first by exploring the next steps in understanding this nexus, then by outlining the exciting prospects of

T. Conroy
Department of Agricultural and Applied Economics, Center for Community Economic Development, University of Wisconsin-Madison/Extension, Madison, WI, USA
e-mail: tessa.conroy@wisc.edu

S. Weiler (✉)
Department of Economics, Colorado State University, Fort Collins, CO, USA
e-mail: Stephan.Weiler@colostate.edu

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_5

exploring the rapidly emerging role of women as innovative entrepreneurs opening whole new networks and market niches.

5.2 Entrepreneurship and Growth

The most recent and compelling vision of entrepreneurship is as the conduit for innovations to shape themselves for the market place. Innovation itself is a process, product, or service that potentially adds new value to the marketplace. But raw innovations, often the product of labs or garages, are rarely immediately ready for the rigors and demands of the market. Entrepreneurs are essential to identify, mold, and maximize a raw innovation's market value (Weiler et al. 2012). In that sense, Steve Jobs was less an innovator than an innovative entrepreneur, with an unusually keen vision of technology's potential niche in the broader marketplace. Thus, identifying and assessing entrepreneurs is critical.

However, entrepreneurship, by its nature, is difficult to measure. Quite simply, startups are small, and arguably even smaller than most statistics indicate. In most data, startups are "born" when the first employee is hired. However, the firm may already have existed as a non-employer proprietorship for some time, as the owners develop their business. Yet the latter are rarely counted in formal statistics, precisely because they are especially hard to count, even though they may well be the original signal of an innovation meeting the market (Weiler and Conroy 2014). In that sense, the "birth" of an employer firm is in fact a major growth step for a set of owners, as it involves adding the employee, her payroll, and supervision/management pieces to the emerging business.

Astonishingly, that employment step is in fact the foundation of the largest element of job creation in the entire economy. Recent research on job creation is very clear. Startups, despite their almost universal small size, are themselves responsible for an extraordinarily large portion—roughly one third—of gross job growth. Although firm survival is recognized to be difficult, with barely half of newborn firms reaching their fifth birthday, younger firms nevertheless universally have higher job growth rates than their more mature counterparts (Decker et al. 2014). So not only are startups creating a significant portion of jobs right as they open, surviving young firms make up much of the rest of the job creation process as well. Meanwhile, job growth in older firms becomes progressively slower.

Recent trends in job growth underscore troubling prospects for these young firms. Job creation rates have been falling over the last three decades, particularly sharply during the Great Recession. Given that the latter was clearly rooted in financial markets, small firms critically dependent on early-stage capital suffered particular hardship (Haltiwanger et al. 2013). High-growth young firms in particular are becoming less numerous since 2000 (Decker et al. 2015). Those difficulties are reflected in the especially sharp drop in job creation both during the recession itself as well as early phases of the recovery—which are not being recouped as the recovery continues. These signals matter, as the flow of marketable innovations depend vitally on the vitality of the entrepreneurial sector. And this sector is also the critical driver of job growth.

Interestingly, while this intuitive link between entrepreneurship and growth is clear, it is challenging to discern empirically. The basic problem is one of endogeneity. Entrepreneurship clearly should be a driver of job growth, but job growth itself creates new market opportunities for entrepreneurs. Regional science was instrumental in addressing and overcoming this vexing statistical problem. Macroeconomic views jumble many markets. Regions offer a better empirical lens with which to explore and determine patterns of interaction between growth and entrepreneurship. Recent availability of publicly available regional data on firm births and deaths now provide regional scientists adequate time-series to definitively attack the endogeneity question (Statistics of U.S. Businesses Employment Change Data, United States Census Bureau). While many, including some major journal editors, believed that disentangling the growth-entrepreneurship endogeneity was not possible, innovative use of instrumental variables is finally allowing the illumination to better understand this key relationship (e.g. Bunten et al. 2015; Conroy and Weiler 2016). Entrepreneurship really does drive job growth.

As we embark on the next 50 years of regional research, the entrepreneurship-growth research agenda is an increasingly exciting one. The just-cited research used a geographic informational structure as its theoretical motivation, where establishment births and deaths reveal valuable increments of market viability. The fact that deaths positively predict future entrepreneurial and subsequent job growth is an especially powerful result, with a series of game-theoretic, Bayesian-revealed preference, and a unifying endogenous growth model (Weiler 2000a; Weiler et al. 2006; Bunten et al. 2015) providing the framework for a theory of geographic informational asymmetries. The attractiveness of identical projects will differ simply based on the amount of market-revealing entrepreneurial activity in the form of births and deaths. Thus, marginalized rural and inner-city regions may get caught in a cycle of neglect and stagnation due to the informational imbalance due precisely to their lack of business activity in the first place.

Yet the flipside of these challenges also offers the unusual opportunity to foster market efficiency and equity simultaneously, two goals that are otherwise almost always in opposition. Addressing the informational market failure not only improves the functioning of the market, allowing the clearer and more certain determination of the prospects for entrepreneurial projects, but also fosters equity by leveling the informational playing field between thinner and thicker markets. Universities are unusually well-placed to provide informational public goods, which may also provide a new vision for land-grant universities in particular (Weiler 2000b). In that sense, the research agenda offers the exciting opening for an especially-applied research agenda.

But there is far more terrain to forge in the empirical field as well. Given the recent evidence that information has decreasing returns to evolving networks, as indicated by market scale and gender results in similar growth inquiries, the question of which sectors are most impacted by these network informational effects becomes important. Births and deaths may reveal the viability of a number of different aspects of business success, from local labor pool to supplier networks, but the success of an on-site service such as dining or retail is obviously much more closely linked to the local geographic market itself. In that vein, the information effect should have a stronger manifestation in a retail establishment than an export-manufacturing establishment, for example. This hypothesis could be assessed empirically.

Furthermore, given the obvious constraining role that the recent financial crisis played in handicapping small firms from playing their job creation role, the relative roles of different forms of financing need to be evaluated. In particular, the changing landscape of banking has reduced the relationship-lending role that small community banks used to play in small business development (Conroy et al. 2016). This consolidation trend as well as the flow of small business loans can be explicitly tested against trends in entrepreneurship. Innovative "fintech"—such as crowdfunding—is anecdotally having a significant impact on entrepreneurial loan options; this proposition could also be tested empirically. More in the business-school anecdotal tradition, entrepreneurs may be moving (back to) Friends, Family, and Fools—as well as credit cards—to finance their operating expenses.

The increasing flow of regional-level data on entrepreneurship should help fuel this agenda, as should the incorporation of more methodologies. In particular, Geographical Information Systems (GIS) techniques could be leveraged without endangering privacy through localized aggregations of firms and labor pools. County-level data continue to provide an imperfect but useful standard for market areas, but GIS mapping technology could offer a wider range of aggregations while also being more precise about key networks. Computable general equilibrium models could also be brought into this research arena, with further efforts to explore an entrepreneurial sector within existing industrial and labor market segments. By providing sectoral- and household-level lenses that regional data alone currently cannot offer, the integration of CGE's could generate a more detailed sense of critical job flows and regional growth.

5.3 Women Entrepreneurs as the New Pioneers of Innovative Entrepreneurialism

Women entrepreneurs are proving to be an underutilized economic resource with the innovative potential to generate significant regional employment effects. Even based on straightforward statistics, the job creation power of women-owned businesses is evident. In the years preceding The Great Recession, privately held women-owned firms added nearly half of a million jobs to the economy while other privately held firms lost jobs (U.S. Department of Commerce 2010). The number of women-owned businesses has also been growing rapidly, outpacing the growth of male-owned businesses in recent decades. This distinction is particularly noteworthy, given the just-cited recent deceleration of firm and job creation. The accelerated performance of women-owned businesses during a time when other businesses slowed or stagnated testifies to the importance of women entrepreneurs to the economy. Their businesses are a promising source of future job creation, though they are still underrepresented in innovation and entrepreneurship.

In light of the evidence that women entrepreneurs are key to job creation, understanding what drives them and what barriers may impede their innovative potential is critical. The distributions of knowledge, resources, opportunities, and constraints certainly differ between men and women. Given these differences, women can be expected to differ from men in their processes of innovation and business formation even in the same economic environment. While much of the literature focuses on the defining characteristics of the entrepreneur herself, such as her management experience, risk preferences, and network disadvantages, there are relatively few studies of women entrepreneurs in regional science. Yet regional economies are particularly useful units of analyses for understanding the local contextual factors, many of which may be significantly influenced by policy. These milieus and policies could uniquely affect entrepreneurial outcomes for women, as well as their ability to generate tangible economic impacts, particularly in terms of job creation.

Compared to relatively stable trends in male business ownership, the trends in female business ownership have evolved rapidly. Perhaps the most distinct among these trends is the accelerated growth in the number of women-owned businesses. Based on their growth in the 1990s, women-owned businesses were expected to equal male-owned business in number within a decade. Interestingly, the growth trend observed in the late twentieth century appears to have changed course, slowing dramatically since 2000, with women-owned business still far outnumbered and outperformed in terms of sales and employment growth by their male-owned counterparts.

The dramatic changes in the growth trajectory of women-owned business during the last 70 years seem to indicate that there are broad socio-cultural shifts that are continually shaping and reshaping rights and roles for women, including women entrepreneurs. For example, women have gone from a relative minority on college campuses to earning the majority of college degrees, reducing the education barrier to entrepreneurship but only to deepen the gender segregation across fields of study. Similarly, women's access to institutional financing has changed over time, relaxing an important constraint on their entrepreneurial potential. However, the relatively low share of women entrepreneurs actually using institutional finance may be indicative of the lower risk preferences among women, a less obvious issue before credit was widely available to women. These examples suggest that the factors driving and constraining the growth of women's entrepreneurship may continue to evolve as women entrepreneurs advance toward gender equality. Perhaps the most insightful future research will treat women's entrepreneurship as an evolving phenomenon that partly results from unique institutional circumstances that continue to change and influence processes of women's business formation. Considering women entrepreneurs over time and by cohort may be especially useful for identifying and understanding the factors that drive trends in women-owned business and shape the innovative capacity of women entrepreneurs.

There are likely regional determinants of entrepreneurship that are unique to women, such as child care availability, but even the well-established determinants of entrepreneurship could vary by gender. Human capital has proven an important factor driving entrepreneurial outcomes across U.S. regions, but it also has an important role in explaining the gender differences in those outcomes. Educational attainment at the regional level varies by gender, but recent work suggests that even if educational attainment were the same between men and women, there are important behavioral differences that would still result in wide gender gaps (Conroy and Weiler 2015). Thus distilling behavioral differences will be critical to future assessments of the gender differentials in entrepreneurship and innovation. Understanding these gendered entrepreneurial behaviors will also be essential to transitioning from gender-blind policies based predominately on the male experience to more informed policies better suited to equitably enhancing entrepreneurship for both men and women.

One of the most promising aspects of gendered entrepreneurship research in regional science, both theoretically and empirically, lies in the personal networks of entrepreneurs. The entrepreneur's social network is important because it can be a source of moral and financial support, but perhaps most importantly, networks are a source of information that could be useful in starting her business. Networks as an information transmission mechanism may be especially critical for women who have historically operated on the economic periphery and consequently have less access to mainstream market information. Yet their networks may also compound the information disadvantage for women entrepreneurs. Generally, men and women have networks with members primarily of the same gender. Given the historically and currently low rates of entrepreneurship among women, it is less likely that women will have entrepreneurial peers and role models in their networks. Consequently, women may find it all the more difficult to access market information and integrate themselves into the business community.

The informational disadvantage for women entrepreneurs may be especially relevant to their financing strategies. Potentially, the information deficit among women entrepreneurs also plays out as bankers and financiers try to evaluate the credit-worthiness of female-led ventures. With relatively few comparable projects to reference, women-led ventures seem relatively risky, and thus less attractive candidates for financing. Relationship lending can mitigate the potential information asymmetries between bankers and entrepreneurs regarding the quality of their businesses. However, with little information based on female-led projects and no personal network connecting women entrepreneurs to decision-makers in the financial community, their projects are less likely to get funding, leading to undercapitalization and lower chances of success. The existing literature focuses on possible gender discrimination in the credit market as an explanation for their small share of women using institutional finance, but as of yet there is little focus on the role of information in the credit market in relation to gender and entrepreneurs. These relationships between information, finance, and women's entrepreneurial outcomes are an exciting area for future research, particularly in the wake of the most recent recession, which was coupled with a financial crises and tight credit markets.

Considering the social aspect of entrepreneurship might help explain why some places are much more entrepreneurial than others, as well as why men appear much more entrepreneurial than women. According to some social models, utility maximizing agents benefit from a behavior of others (a spillover), and more importantly, benefit from behaving similarly (Brock and Durlauf 2001, 2007). Thinking of entrepreneurship in such a theoretical framework can help us understand highly entrepreneurial places as those settings where, not only do individuals benefit from being surrounded by other entrepreneurs via knowledge spillovers, but they also gain from mimicking the (successful) entrepreneurial behaviors of their peers. Empirically, accounting for social networks has a natural analog in spatial econometrics. Much like we often expect two places located near each other in geographic space to influence each other, we would expect two entrepreneurs who are near each other in "social space" to influence each other. The closer the relationship between two agents, the stronger the behavioral influences between them. With ideal data, it would be possible to set up a matrix that defines the strength of the social influence between any two agents much like a spatial weights matrix does for any two places. In the spirit of Kilkenny and Nalbarte (2000), describing and comparing the structure of the social influences between agents in any region could be insightful, let alone the possible lessons from a fully tractable econometric model.

Gendered networks and information asymmetries are one explanation for the strong regional employment effects from women-owned businesses. The same information disadvantage that creates an obstacle to the success of women entrepreneurs also makes their projects especially valuable. In many ways women entrepreneurs are exploring relatively unknown market terrain, generating valuable information about the market including the viability of new projects and consumer demand. Women have to be particularly innovative as they start their businesses and bring unique products, services, and management styles to the market. As a recent example, a woman in tech who designs new applications specifically for women is innovative in several ways. First, in her sector there are few women available to her as mentors to give advice and share market insights making her role as a business leader in a male-dominated sector a valuable example to those that may follow. Second, given that most apps are developed by men who likely know men's preferences better than they know and understand women's preferences, her product is testing the viability of a new idea. The consumer response and success of such new ideas will be a valuable indicator for related projects.

Given recent evidence of decreasing returns relationship between information and regional employment in both regional and gendered networks (Conroy and Weiler 2016), we can expect relatively large payoffs from incremental increases in female entrepreneurship. The persistently low rate of female entrepreneurship compounded by their network disadvantages implies that women entrepreneurs enter a relatively thin market. In the context of the information payoff function, the marginal female-led entrepreneurial project lies to the left of the curve and results in relatively large incremental increases in employment. In contrast, entrepreneurial men enter a thick market with detailed information available based on a long history of past projects led by other male entrepreneurs across all sectors of the economy. With ample market information already available, the marginal male-owned projects would lie much further to the right of the diminishing-returns function and generate small employment effects.

Empirically, the gendered market information hypothesis holds, suggesting that the innovative capabilities of women entrepreneurs, both in terms forging the way for women that may follow as well as in identifying new market niches, can lead to significant job creation. As they become business leaders and grow in economic presence and significance, women entrepreneurs are also building foundational network ties in the business community that will benefit themselves as well as their peers and those they mentor. These new and growing networks can mitigate the information disadvantage for women, allowing them to better evaluate their prospects for success and have an impact on the local economy.

Identifying the specific aspects of women's entrepreneurship that lead employment growth will be important for future research. For example, their sectoral concentration may partly explain the large employment effects. Women tend to be concentrated in sectors, especially service sectors such as healthcare and education that are inherently local in their activities. Properly assessing such localized markets places an especially high premium on information flows, where growth can be stymied by limited numbers and thus networks among women business-owners. But this situation also implies potentially very high returns for entrepreneurial women pioneers.

In addition to the local aspect, such service sectors are already undergoing major shifts brought on my new policies and rapid technological change. These aspects of women-owned businesses could explain the strong employment effects in the context of the information hypothesis. Future studies that account for industry sector and more precisely identify the contributions of women-owned firms will be valuable extensions of the regional entrepreneurship literature.

The most challenging aspect of furthering gendered research on entrepreneurship is the data limitations. The gendered data that are currently available publicly are typically limited in geographic scope or frequency. The Survey of Business Owners, for example, is available only in 5-year increments, specifically years ending in "2" and "7" which can make it difficult to use with conventional decennial sources. The Current Population Survey is only representative down to the state level, making it difficult to study small geographic areas. To our knowledge, there are no publicly available gendered data on business dynamics (openings, closings, expansions, and contractions), which are often preferable for entrepreneurial research. Until gendered data on entrepreneurial and business outcomes are more readily available, these data limitations will constrain the progress of research in the field.

The importance of collecting and publishing high quality gendered data is even more critical when considered as a component of informed policy. Advances in research on women entrepreneurs and other demographic groups are essential to policy designs that successfully and equitably enhance entrepreneurship. Designing policy based on current research, most of which is at least implicitly based on the male experience, may be ineffective for enhancing entrepreneurship among women business leaders and minority groups more generally. Hence, data improvements together with regional studies, which often lend themselves naturally to informing community and regional development strategies, focused on women entrepreneurs are an important step towards effectively tapping into the job creation potential of women-owned business and spurring regional economic growth.

5.4 Pioneers of the Twenty-first Century

Our understanding of how regions grow—and equally importantly why regions decline—has evolved rapidly in the last 50 years. Regional science has provided illumination on these core questions of its own discipline, but along the way its regions have also been productive higher-resolution lenses through which to understand aggregate growth (Krugman 1991). Most recently, entrepreneurship has been confirmed as a major driver of job growth, putting yet further into question economic developers' remarkably myopic focus on "primary" industries, among other implications.

Yet we need to extend this understanding to better inform decision-makers—public, private, and non-profit—about how such an entrepreneurship-innovation-growth nexus can best enhance a region's opportunities for quality job growth. A major first and obvious step is better utilization of well-educated resources already in place, namely potential women entrepreneurs—and the expanded market they may be uniquely able to access. The emerging stagnation of high-growth firms and their job creation is a troubling trend, interestingly one that is occurring alongside the noted flattening of new women-owned businesses. Both regions and nations would do well to consider more carefully the neglected half of the entrepreneurial-growth puzzle, as jobs, incomes, and quality-of-life trajectories may depend on women even more than we realized.

References

Brock W, Durlauf S (2001) Discrete choice with social interactions. Rev Econ Stud 68(2):235–260

Brock W, Durlauf S (2007) Identification of binary choice models with social interactions. J Econ 140(1):52–75

Bunten D, Weiler S, Thompson E, Zahran S (2015) Entrepreneurship, information, and growth. J Reg Sci 55(4):560–584. doi:10.1111/jors.12157

Conroy T, Low S, Weiler S (2016) Fueling job engines: impacts of small business loans on establishment births in metropolitan and nonmetro counties. Working paper

Conroy T, Weiler S (2015) Where are the women entrepreneurs? Business ownership growth by gender across the American urban landscape. Econ Inq 53(4):1872–1892. doi:10.1111/ecin.12224

Conroy T, Weiler S (2016) It's who and what you know: women entrepreneurs, information asymmetries, and economic growth. Working paper

Decker R, Haltiwanger J, Jarmin R, Miranda J (2014) The role of entrepreneurship in US job creation and economic dynamism. J Econ Perspect 28(3):3–24. doi:10.1257/jep.28.3.3

Decker R, Haltiwanger J, Jarmin R, Miranda J (2015) Where has all the skewness gone? The decline in high-growth (young) firms in the U.S. NBER (no 21776)

Haltiwanger J, Jarmin R, Miranda J (2013) Who creates jobs? Small versus large versus young. Rev Econ Stat XCV(2):347–361

Kilkenny M, Nalbarte L (2000) Network analysis of a community. In: The web book of regional science. Regional Research Institute, WVU. http://www.rri.wvu.edu/WebBook/Kilkenny/editedkeystone.htm

Krugman P (1991) Increasing returns and economic geography. J Polit Econ 99(3):483–499. doi:10.3386/w3275

U.S. Department of Commerce (2010) Women-owned businesses in the 21st century. http://www.esa.doc.gov/reports/women-owned-businesses-21st-century

Weiler S (2000a) Pioneers and settlers in Lo-Do Denver: private risk and public benefits in urban redevelopment. Urban Stud 37(1):167–179. doi:10.1080/0042098002348

Weiler S (2000b) Information and market failure in local economic development: a new role for universities? Econ Dev Q 14(2):194–203

Weiler S, Conroy T (2014) Colorado innovation report: the state of innovation. Colorado State University, Fort Collins, CO. http://innovation.colostate.edu/

Weiler S, Giltner C, Platt T, Totten G, Yeadon M (2012) Colorado innovation report: reaching our innovation summit. Colorado Innovation Network, Denver. http://innovation.colostate.edu/

Weiler S, Hoag D, Fan C-M (2006) Prospecting for economic returns to research: adding informational value at the market fringe. J Reg Sci 46(2):289–311. doi:10.1111/j.0022-4146.2006.00442.x

Tessa Conroy is an Economic Development Specialist in the Department of Agricultural and Applied Economics and Center for Community and Economic Development at the University of Wisconsin-Madison/Extension. Her primary research interests are regional economic development, business dynamics, and entrepreneurship. She has a particular interest in women entrepreneurs and understanding the gender discrepancy in business ownership and performance. Dr. Conroy earned the Ph.D. in economics from Colorado State University in 2014.

Stephan Weiler is William E. Morgan Professor of Economics at Colorado State University. His current work focuses on regional economic growth and development, particularly in rural and inner-city areas, combining theoretical, empirical, and policy analyses on topics such as information, innovation, industrial restructuring, land use, public/private partnerships, immigration, entrepreneurship, and the environment. He was previously a research faculty member at the Regional Research Institute and Assistant Vice President and Economist at the Federal Reserve's Center for the Study of Rural America. He received his BA (Honors) in Economics and MA in Development Economics from Stanford University in 1988, and his Economics Ph.D. from UC-Berkeley in 1994.

Part II
Regional Growth, Regional Forecasts, and Policy

Chapter 6
Agglomeration and Automation in the Twenty-First Century: Prospects for Regional Research

Doug Woodward

6.1 Introduction

There are few absolute certainties, but we can be reasonably sure that the world's population will grow and continually concentrate in space. By 2050, total population is estimated to be 9.6 billion, with the share residing in urban regions rising to 67%; it was just 30% in 1950 (United Nations 2014).

For regional science, there is ample evidence and widespread agreement that spatial concentration will not only persist, but also improve the world's economic welfare. Scholars widely concur that urban agglomeration raises regional productivity. In *Economics of Agglomeration*—a definitive work on urban and regional growth—the authors conclude: "Consequently, the wealth or poverty of nations or regions seems to be more and more related to the development of prosperous and competitive clusters of industries as well as to the existence of large and diversified metro areas" (Fujita and Thisse 2002, p. 4).

While the world's population will certainly enlarge and remain highly concentrated in the twenty-first century, technological change will fundamentally alter productivity, employment, and competitive regional clusters. Brynjolfsson and McAfee (2014) document how we are entering a second machine age, with potentially exponential growth in output based on pervasive, interconnected computing power and massive digital information. The first machine age was essentially mass production that began in the late nineteenth century. In the first machine age, productivity, employment and income increased together, along with urban and industry agglomeration. The second machine age—artificial intelligence, cloud computing, machine learning, autonomous vehicles, drones, and robotics—heralds

D. Woodward
Department of Economics, Darla Moore School of Business, University of South Carolina, Columbia, SC, USA
e-mail: woodward@moore.sc.edu

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_6

a new era of accelerating automation and rising productivity. Nevertheless, there is the potential for massive labor contraction across many industries. Can we expect regional economic concentration, as in previous eras?

In this paper, we consider the implications of the second machine age on agglomeration and related regional research. We begin with background on the second machine age, exploring the implications for employment and regional development. Automation will substantially raise productivity across numerous industries. Consequently, technological change will alter industry and urban development. Then we turn to an overview of what regional science currently knows about industry and urban agglomeration. Agglomeration is the spatial concentration of firms and workers beyond expected clustering by natural advantages or randomness. In regions with strong agglomeration economies, positive externalities generated by spatial proximity lead to increasing returns to scale and higher productivity. A substantial body of research has demonstrated that agglomeration is a major determinant of productivity. Yet most of the findings about productivity predate the second machine age and largely focus on manufacturing. Will future agglomeration research overturn our understanding of regional (urban) growth and productivity?

Next, we delve deeper into the three main sources of industry agglomeration: (1) better labor skill matching; (2) enhanced local learning and knowledge flows; and (3) more efficient input sharing, along with deeper and more specialized supplier relationships. Will local labor matching/pooling diminish as an agglomeration driver, given projected declines in jobs from automation, robots, machine learning and artificial intelligence? How will local knowledge spillovers change? Will supplier development and clustering remain crucial to agglomeration?

We then consider measures of agglomeration. Measuring agglomeration accurately is central to regional science. Despite major improvements in spatial concentration indices, many are based on employment as a gauge of scale. Moreover, they still apply to individual industries, but not clusters or more complex interindustry interactions. Therefore, it is imperative that we press on with the current work on clusters and agglomeration indices.

Essentially this chapter contemplates how accelerating automation will challenge our understanding of agglomeration, one of the most established tenets of regional science. We are not just entering a new era of automation, but of urban and industry agglomeration. One does not have to portend "the rise of the robots and the threat of a jobless future" (Ford 2015) to accept that by 2050 automation will profoundly affect the world's 9.6 billion people and the regions where they reside.

6.2 The New Age of Automation

The second machine age of the twenty-first century will bring substantial economic benefits, along with disruption for decades to come (Brynjolfsson and McAfee 2014). Pratt (2015) argues that the new technological era will be revolutionary, rather than evolutionary, spurred by interactive "machine intelligence." Contemporary examples of machine intelligence include cloud computing, in which machines learn from one another; and deep learning, in which machines analyze enormous amounts of data to magnify their competencies. Following recent breakthroughs, robots with sensors can scan their surroundings and execute complex physical tasks much better than humans perform, even for skilled work. Autonomous drones, self-driving vehicles, 3-D printing, and globally interconnected devices (the Internet of Things) will revolutionize the economy by performing tasks far exceeding human capabilities.

As a result, automation and artificial intelligence will upend numerous industries on a scale and scope the world has never before witnessed. One highly cited study predicts that by 2033, 47% of U.S. jobs face displacement by robotics, machine learning, and other technological advances (Frey and Osborne 2013). The estimates show the probability of employment loss by automation for 702 occupations based on an original Gaussian process classifier. To give a regional perspective, Walden (2015) used the occupations identified Frey and Osborne as having a high probability of displacement in North Carolina. He found the state could lose 1.9 million jobs over the next 40 years (nonfarm wage and salary employment was 4.3 million in 2015).

This disruption will not be confined to developed economies. Frey and Osborne (2016) project that 85% of employment in developing countries, including India and China, could be displaced by automation. In the future machine productivity may likely outstrip human productivity in many activities. In turn, productivity gains may be divorced from tasks carried out by humans. The implications for regions across the world could be momentous, especially if employment diminishes, while population continues to enlarge.

With growing value added and declining employment over the long run, then the economy would not generate the income needed to buy back the fruits of this productivity. Generally U.S. productivity has constantly increased since the beginning of the twenty-first century, but both private employment and median family income have stagnated (Brynjolfsson and McAfee 2015). Labor compensation since the 1970s has lagged far behind productivity growth in the United States (Bivens and Mishel 2015).

Countering the present-day prophets of the coming technological revolution, some prominent economists are pessimistic about the *rate* of productivity growth, which has fallen since the 1990s. Gordon (2015) predicts an enduring productivity growth slowdown for the United States. He claims that recent inventions have concentrated in entertainment and communication devices that do not significantly raise productivity or living standards. Innovations enabled by the Internet (social media apps and online banking, for example) are trivial compared with previous innovations like indoor plumbing, the telephone, and the internal combustion engine. For the decade from 2004–2014, which includes the Great Recession, U.S. labor productivity and total-factor productivity (TFP) actually decelerated when compared with 1995–2004 (Byrne et al. 2016). Accordingly, it appears that recent technology benefits show up only in consumer products and services, not production gains. Consumers use new devices to provide services that they apparently value, but mostly for leisure time. Smartphone communications, online searches, and social media interactions—and now virtual reality—are largely nonmarket, consumer activities. So far, then, there is no measurable productivity gain from the applications of Internet communications.

This "productivity paradox" perspective looks backward, however. It eschews the enormous potential production from automation and accelerating technological change. As the most significant applications of digital technology move from leisure to work activities, economic output will surge with robotics, 3-D printing, artificial intelligence, and machine learning, in which computers perform complex tasks for which they were not explicitly programmed.

Today, we find evidence for these changes in many regions and a broad array of sectors: manufacturing, services, finance, insurance, retail trade, distribution and logistics, health care, education, the military, construction, agriculture, education, energy, and mining. Across the world, select regional economies are restructuring around robotics and smart machines. Beyond the vaunted Silicon Valley, a prime example is Shenzhen, China. The city, along with the greater Pearl River Delta region, represents one of the most remarkable cases of urban development in history, having risen as a globally competitive manufacturing and innovation hub in just 30 years. Starting as a special economic zone following the economic liberalization in the early 1980s, this thriving metropolitan region attracted millions of migrants from rural areas to work in labor-intensive industries. Today Shenzhen's factories produce the world's iPhones and many other global products of the twenty-first century. Yet the regimented, grueling pace of work in the region's factories led to widespread discontent, demands for higher wages, and high turnover among employees. Shenzhen suffered with a reputation as a "suicide" cluster (Barbosa 2010). Now Shenzhen's manufacturers are replacing workers with robots (Chan 2015), even as they establish a cutting-edge urban cluster of robotic design and assembly. Taiwan-based Foxconn, which assembles iPhones in Shenzhen and hires 1.2 million workers in China, has invested heavily in robotics. Management's explicit goal is to deploy one million robots and leave only human labor for engineers and technicians (Kan 2013).

History suggests that technological revolutions provide opportunities for new employment and spawn new occupations. Typically, the argument invokes Joseph Schumpeter's concept of creative destruction. No one disputes that productivity grew while jobs were created during previous technological revolutions. There were three major technological eras before the present: the first era was the industrial revolution (steam power and mechanization) that began in the late 1700s; the second age commencing in 1870 (mass production and electrical energy); and a third era of computerization that started in the late 1960s. In each period, aggregate employment expanded, even as technology eliminated jobs. Clearly, the number of jobs is not fixed; there is no "lump of labor."

Economists are only beginning to analyze the labor impact of new technologies in the twenty-first century (Autor and Dorn 2013). To estimate the effects of automation on jobs, we need better parameters, including precise estimates of the elasticity of substitution between human- and non-human capital. Further, we need to know much more about factor-specific productivity and the extent to which technological change favors non-human capital over human capital.

Labor economists are attempting to identify tasks in which labor retains a comparative advantage in the second machine age. Some analysts posit unskilled labor and capital as substitutes, but skilled labor and capital are complements. Autor (2015) suggests that the shifting relationship between machine and human comparative advantages will lead machines (hardware and software) to substitute for labor in routinized tasks, yet boost the comparative advantage of non-routine, creative, and adaptable workers. In short, machine learning and robotics will both *substitute* and *complement* employment.

However, the scaling of aggregate employment may depart from previous waves of technological change in the emerging technological era. In the nineteenth and twentieth centuries, manufacturing employment expanded significantly, offsetting displacement as agricultural productivity led to structural job losses. Services then rose with the decline in manufacturing jobs. It is still an open question as to which industries or occupations will be augmented substantially to compensate for growing structural unemployment. Consider the restaurant business, a major service sector employer. Jobs in “eating and drinking establishments” scaled up for decades, as manufacturing employment contracted. These low-wage, low-skill “McJobs” are often disparaged as a poor substitute for factory employment, but even “hamburger-flippers” may be vulnerable to technological displacement.

Hence, despite complementary opportunities for skilled labor, significant labor-intensive jobs may shift to machines that are more efficient. These ever-cheaper machines and software programs, interconnected through cloud computing, do not bargain for wages or require retirement and health care benefits. They are never absent or sick. They are able to work through the night and on holidays. They do not file workers’ compensation claims.

Stagnating aggregate employment relative to productivity has deep implications for regions. Technology could unsettle urbanization, as output depends more on the “cloud”-based, rather than “crowd”-based advantages. Unlike smart machines, humans spatially cluster to learn and improve. In the past, regional scholars have shown that the concentration of employment and population is a source of productivity gains, as cogently articulated in *Economics of Agglomeration* (Fujita and Thisse 2002) and backed up by a large corpus of regional research. Agglomeration persisted throughout all previous technology-driven periods of economic growth. Now there could be a massive regional realignment as productivity gains shift away from human to machine interactions.

Despite accelerating technological advances, regions will respond and transform slowly. To capture the productivity gains from automation, regional institutions and organizations will have to adapt and innovate as well (Brynjolfsson and McAfee 2014). This process will likely transpire over decades, given institutional resistance and organizational challenges. History suggests that in the face of accelerating technology, organizations and institutions at first resist change. Famously, David (1990) discovered that the invention and spread of electrification in U.S. manufacturing did not initially increase productivity when it substituted for steam power. At first, it was unprofitable to introduce new electric dynamos, given that factories had well-functioning production technology from the existing water and steam-driven mechanical power. Further, productivity gains from factory electrification only climbed after manufacturers substantially reorganized industrial processes, which took place over decades.

Thus, while it may make take time, by 2050 we should see substantial changes from automation that will eventually transform regional development. According to Amara's Law, "We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run" (quoted in Mokyr et al. 2015, p. 48).

6.3 Agglomeration

To set the stage for our discussion of automation and agglomeration, in this section we briefly review what regional science and urban economics means by agglomeration. Broadly speaking, agglomeration is the spatial concentration of firms and workers beyond expected clustering by natural advantages or randomness. In regions with strong agglomeration economies, positive externalities generated by spatial proximity lead to increasing returns to scale and higher productivity. In spatial theory, these positive regional externalities of agglomeration offset congestion, pollution and other negative regional externalities (Henderson 1974).

The concept encompasses three types of external returns from spatial concentration, beyond the well-known gains of individual firms with internal economies of scale:

1. *Industry localization*, which are spillovers that are external to local business establishments in a particular industry, but internal to the regional industry.
2. *Clusters*, which cross individual industries with mutual benefits for local networks of firms from specialized suppliers and services, a common pool of skilled workers, shared infrastructure, and supporting institutions like higher education.
3. *Urban agglomeration*, which are more general spillovers that are external to individual industries and clusters, but still captured internally within the region.

Literally and figuratively, agglomeration is a core principle of regional science, economic geography, and urban economics. From a myriad of perspectives and methods, analysts have extensively investigated industry-, cluster, and urban-level concentration. Duranton and Kerr (2015) provide an incisive review of theoretical and empirical developments. Besides the urban and regional literature, scholarly articles span many fields, including labor economics, industrial organization, international business, entrepreneurial studies, and economic history.

The new economic geography, in part centered on agglomeration, has had a far-reaching influence. Krugman's writing on the new economic geography helped the author win the Nobel Memorial Prize in Economics for insights about the role of economies of scale and increasing returns in determining trade and location patterns. *The Competitive Advantage of Nations*, the popularization of clustering by Porter (1990), garners more than 100,000 citations identified on Google Scholar. Agglomeration and cluster concepts are widely used by policy makers to shape regional development strategy across the world (TCI 2016).

6.4 Agglomeration and Productivity

Research on agglomeration is pertinent to technological change because of its findings about productivity. Since Adam Smith published *the Wealth of Nations* in 1776, scholars have recognized that productivity is a fundamental determinant of economic growth (Smith 1904).

The essential logic of agglomeration and productivity originates with Alfred Marshall's (1920) observations about England's specialized regions of the first machine age, exemplified by Sheffield (cutlery) and Birmingham (jewelry). Subsequent historical research corroborates Marshall's insights as agglomeration helped spread the productivity gains of the industrial revolution across Europe (Hohenberg and Lees 1985). During the mechanization and urbanization of the late nineteenth and early twentieth century U.S. economy, productivity rose dramatically as a result of Marshallian industry-specific agglomeration (Crafts and Klein 2015).

Overwhelming evidence confirms that regional scale economies improve productivity (Combes and Gobillon 2015). The size and density of economic activities increases the productivity of firms and workers. Consistently, scholars find that productivity increases with urban scale (population, firm and establishment counts, and employment). The empirical literature on agglomeration posits that doubling urban size increases productivity from 2 to 8% (Rosenthal and Strange 2004). The results hold up with controls for firm-worker sorting and selection. Yet the responsiveness of productivity to urban scale depends on the nature of the industry and region.

In testing this agglomeration hypothesis, many researchers prefer total factor productivity (TFP) to labor productivity (output per worker), following Moomaw (1981). TFP represents the residual output not accounted for by inputs (physical and human capital) and captures the efficiency effects on output of technology (along with ignorance of other factors with positive effects on production). Early work based U.S. manufacturing from 1959–1973 detected no association between urbanization and total factor productivity growth (Beeson 1987). Yet, Henderson (2003) analyzed plant-level data in U.S. high technology and machinery industries and uncovered evidence for localization (own industry) effects on productivity for high technology, but not for machinery. Urbanization economies do not affect productivity in these industries. Researchers continue to investigate the TFP-agglomeration association with more sophisticated controls and instruments (Puga 2010).

Future regional studies of total factor productivity and agglomeration could look specifically at different forms of automation like robotics as an input to the production function, along with skilled and unskilled labor. Yet attempts to test the determinants of TFP at the regional level are limited by data and the difficulties posed by defining and measuring regional automation. How do we capture capital (and knowledge) embodied in spatially dispersed, but linked devices and sensors, investments in cloud computing, or many other forms of emerging technology?

Beyond attempts to measure total factor productivity, a large research stream investigates productivity gains from agglomeration with wage regressions (Wheaton and Lewis 2002; Combes et al. 2008; Freedman 2008; Mion and Naticchioni 2009; Matano and Naticchioni 2012; Figueiredo et al. 2014). The premise is that higher relative wages in regional clusters reflect the productivity benefits resulting from industry-specific external economies. Indeed, higher wages would lead firms to relocate elsewhere unless there were some significant compensating productivity advantages. In addition to the industry-specific externalities (localization), the wage premium may pick up productivity gains associated with urbanization—economic gains that accrue from the concentration of general economic activity, rather than a particular industry.

Most empirical researchers base their agglomeration and wage premium (productivity) conclusions on assessments of developed economies. A lack of reliable data impedes research in many developing economies. While we base most of what we know about agglomeration on studies in Europe and North America, there is a notable rise of Chinese, Korean and Indian research, along with incipient work in Latin American countries. Based on wage regressions, the elasticities of wages with respect to market size and density vary, but researchers detect higher results for China and India than those found in developed economies (Duranton 2014). In a comprehensive review of the literature, Combes and Gobillon (2015) find that these larger elasticities in developed countries range from 0.10 to 0.12, approximately three times the estimates for developed countries.

Overall, the association between agglomeration and productivity is remarkably robust. Thus, existing research endorses the fundamental regional principle that prosperity depends on "…competitive clusters of industries as well as to the existence of large and diversified metro areas" (Fujita and Thisse 2002, p. 4).

Before we enshrine the agglomeration-drives-productivity thesis with the status of "known knowns" in regional science (Jackson 2011), we should remember that the relationship is not axiomatic. It is a falsifiable hypothesis.

Conceivably, agglomeration advantages may diminish in the second machine age. As argued in the Introduction, we need to question the advantages of larger, denser regions in the era of artificial intelligence and machine learning. As productivity gains spread across sectors from automation, many regions will require fewer workers to produce the same level of output. Recall that robotics, machine learning, and other technological advances threaten to supplant 47% of U.S. jobs (Frey and Osborne 2013). Unlike human spatial concentration, smart machines do not need spatial spillovers and social networks to raise productivity, even exponentially. As artificial intelligence and robotics propel productivity, the new era may lead to relatively smaller gains from urban and regional external scale economies. By 2050, the agglomeration benefits for productivity may not disappear, but they may diminish in many sectors of the economy.

Nascent evidence indicates the extent to which robots increase aggregate productivity and affect the labor market. A cross-country study looked at the introduction of industrial robots in 17 countries from 1993 to 2007 and found statistically significant gains for total factor productivity (Graetz and Michaels 2015). Interestingly, increasing the density of robots led to progressively smaller productivity increases, suggesting diminishing marginal returns. Note that this congestion result was found for early-stage industrial robots, not other innovations of the second machine age. More importantly, Graetz and Michaels (2015) find that robots lead to less employment and falling wages for low- and medium-skilled workers. Apparently, there is no adverse impact on the employment of high-skilled workers.

There is no doubt, however, that smart machines will work in tandem with skilled, adaptable workers in regions. To some extent, urban areas with highly skilled workers and competitive firms will continue to benefit from urban size and density. There is widespread consensus that there will be continual sorting of smart people around other talented people, capital, and technology. Flexibility will be the key to regional success in the new technological age (UBS 2016). Regional economies with the most flexible labor markets, educational systems, infrastructure, and institutions will raise productivity.

Frey and Osborne (2016) find resilient urban areas least affected and most favored in the second machine age have a deep pool of talent and a developed technology sector. The San Francisco Bay Area, Boston, Seattle, Austin, and Raleigh, among other high-technology hubs in the United States, will continue to prosper, despite higher rents and living costs, drawing workers from across the world.

It should be emphasized that the scale of employment in high-technology firms could be modest, at least compared with the previous industrial era. It is still unclear that even competitive technology firms will need denser, larger population and industrial clusters as a source of productivity improvements.

Even in the favored technology regions, future urban and regional productivity may rely more on the *sorting* of skilled labor and competitive firms for productivity gains than on agglomeration The most productive workers and firms *select* regions and raise productivity, apart from the positive externalities of agglomeration (Behrens et al. 2014). High-skill workers and highly competitive firms will *choose* to locate in expensive cities, despite higher costs. Besides natural amenities, regional sources of capital draw the most productive workers and companies. Typically, productive inventors and entrepreneurs seek out “angel” investors and other private sources of income through local social networks. Venture capital is also surprisingly concentrated (Florida and King 2016), exemplified by the high-technology clusters like the San Francisco Bay Area, Boston, and India’s Bangalore region. Emerging companies and skilled labor (notably engineers) cluster around venture capital. Technology clusters prosper because of this sorting process, not necessarily because of agglomeration.

Future enquiries, then, need to unravel the effects of sorting, agglomeration, and productivity. There is reason to believe that regional science will advance our understanding. As Duranton and Kerr (2015) point out, the quality of empirical work has improved substantially with the availability of firm- and establishment-level employment data (see also the review by Combes and Gobillon 2015).

Andersson et al. (2014) exemplify the promising research stream on agglomeration and sorting. The paper studies the urban wage premium (productivity) in Sweden. The authors build on the work of Bacolod et al. (2009), which finds that the urban wage premium mostly applies to workers with cognitive and people skills. The Swedish study shows the virtues of having access to a longitudinal, matched employer–employee data base. Importantly, the authors ascertain that agglomeration gains are spread unevenly among specific types of worker skills. For example, they examine the wage premium among workers with non-routine and routine job tasks. They find no agglomeration gains for workers with routine tasks, yet agglomeration engenders particularly strong productivity benefits for non-routine skill sets. They conclude that spatial density creates productivity gains when problem solving and interaction matter. In addition, Andersson et al. (2014) point to spatial sorting as the foremost source of the urban wage premium. As skilled workers sort into regional clusters, the wage premium reflects workers' greater capabilities, not externalities from clustering.

This key finding about agglomeration's advantages for non-routine skills, along with sorting, is clearly relevant in the age of automation. To date, automation has mostly substituted for routine tasks, both cognitive and manual. Figure 6.1 depicts the distinctions among routine, non-routine, manual, and cognitive tasks. The lower two quadrants elucidate the prospects for non-routine work. Non-routine skills typically entail interactive problem solving, where the tasks do not follow strict rules. The requisite knowledge is often tacit, best explained by the Polanyi's paradox: "we know more than we can tell." Note that machines do not replace tasks, not jobs per se. Non-routine tasks based on tacit knowledge will be more difficult to automate and should thus provide employment in the future.

There has been growth in both categories of non-routine jobs in the United States: high-paying, cognitive skills (professional and technical occupations) and lower-paying, manual skills (assisting and caring for others). However, jobs are declining decisively for routine skills. The only category growing significantly in the United States is *cognitive, non-routine* employment (Dvorkin 2016). Traits such as interpersonal skills, adaptability, and problem solving define cognitive, non-routine skills and endow humans with comparative advantages over machines (Autor 2015).

Thus, as indicated in the lower right quadrant of Fig. 6.1, the future workforce will expand with occupations requiring non-routine, cognitive tasks. Instead of facing labor substitution and contraction, workers with these skills complement automation and job prospects should expand.

	Manual	Cognitive
Routine	Repetitive, codified, narrow task set Assembly-Line Distribution (picking, sorting) Future Task Substitution and Job Contraction through Automation	Repetitive, codified, narrow tasks set Clerical, Record Keeping, Customer Service Future Task Substitution and Job Contraction through Automation
Non-Routine	Broad, flexible, interpersonal task set Service (waiters) and Protection (security) Direct patient care Transport (truck drivers, taxi drivers) Future Employment: Limited Expansion with Tasks Displaced through Robotics and Autonomous Vehicles	Complex task set: problem-solving, creativity, and persuasion Professional Services, Scientific, Technical Managerial, Engineering, Law, Health, Education, Design, Marketing, Entertainment Future Employment: Expansion with Some Tasks Displaced through Artificial Intelligence and Deep Learning

Fig. 6.1 Future employment prospects

This complementary process has already unfolded in the increasingly automated automotive sector. Automakers are the largest users of industrial robots, with 1.5 million used in production in 2014 and an additional 1.3 million expected within 2 years (Gibbes 2016). Productivity has soared with rising robotics.

Robots possess comparative advantages over humans in strength and precision. Even so, innovative automakers like Mercedes-Benz and Toyota have reversed the automation process to some degree, substituting more humans for machines. The head of production at Mercedes-Benz claimed that future increases in productivity required more flexibility and a need to insert human labor back into the industrial process (Gibbes 2016). Following years of robotic expansion, automotive production moved from large-scale to smaller, more agile systems in which humans work in tandem with robots. Robot-human interaction is known as collaborative robotics or cobotics. This collaboration between non-routine human skills and smart machines is necessary, given the ever-changing variety and options on the assembly line, making it difficult for machines alone to raise productivity (Gibbes 2016). The recent cobotic machines are equipped with sensors to enable them to detect and react to humans. Cobots complement the automotive factory workers' comparative advantages: the capacity to observe, reason, and adjust to new situations.

Concerning the large-scale growth of non-routine, cognitive employment, however, it should be emphasized that machines can learn and improve in a range of complex tasks, as was demonstrated when Google's DeepMind beat the world champion in the strategy game Go (Brynjolfsson and McAfee 2016). Deep learning programs sift through enormous amounts of data to get machines to adapt and learn more like humans, without explicit instructions. Amelia (created by IPsoft) and Watson (IBM) are but a few of the current examples of artificial intelligence programs that have potential to replace non-routine, cognitive jobs. While future work will largely involve non-routine mental abilities, some of these jobs will be replaced by technology.

To sum up, productivity and agglomeration will continue to play a central role in urban and regional development, and accordingly, in research. It will be interesting to see if the strong effects of agglomeration on productivity, now widely accepted as gospel, will hold up under much more intensified automation. The size and stability of agglomeration benefits, then, remain open questions, a known unknown in regional science (Jackson 2011), requiring further inquiry.

There are at least one strong reason that agglomeration—urbanization, localization and clustering—will diminish as a determinant of productivity. Through substitution, labor may be a less significant input in producing an increasingly more abundant array of goods and services.

At the same time, there is also reason to believe that agglomeration will remain a determinant of productivity, at least for advanced regions. Research so far suggests there are agglomeration gains for workers in non-routine tasks, the kinds of skills apparently needed in the second machine age. Even so, research now points to labor and firm sorting as a more significant determinant of productivity than agglomeration, even for non-routine labor.

Therefore, future research will have to treat the long-standing hypothesis about agglomeration and productivity as refutable. The findings from wage regressions will need to be tested across many more countries and regions. This ongoing exploration will require reliable microeconomic data at the regional level, notably establishment-level data, with matched employee–employer data sets linked to skills. As in the past, researchers have to rely mostly on official government data. Data restrictions confine much of the demanding empirical work to European studies (where microeconomic data are disclosed and available at the necessary level of detail) and to a lesser extent, the United States, Asia, and Latin America.

6.5 Determinants of Agglomeration

As discussed in the last section, an abundance of empirical evidence points to the size and extent of agglomeration externalities. Still, there is much we do not know about the specific factors that underlie agglomeration effects on productivity (Combes and Gobillon 2015; Duranton and Kerr 2015). In this section, we examine more closely the three major determinants of agglomeration:

1. Deeper labor pools and better labor skill matching;
2. Enhanced localized learning spillovers and knowledge flows;
3. More efficient input sharing, including infrastructure, and deeper, more specialized supplier relationships, and vertical disintegration among firms.

These potential determinants contribute to industry localization along the lines first suggested by Alfred Marshall (1920). For each, the hypothesis is that they boost regional productivity with the increasing scale of an industry at a location. In theory, they apply to urbanization and clustering as well as industries. In turn, labor pooling/matching, knowledge spillovers, and sharing/specialization spawn empirical research (Rosenthal and Strange 2004; Duranton and Puga 2005).

6.6 Labor Pooling and Matching

Given projected declines in jobs from automation, robotics, machine learning, and artificial intelligence, regional labor pooling and firm-worker matching may seem likely to diminish as a regional agglomeration driver. Many occupations are projected to involve non-specific, non-routine skills that do not require precise industry matching between firm and the worker: interpersonal skills, flexibility, and problem solving. Automation will raise the comparative advantage of adaptable workers, not the skills that have strict job classifications.

Labor pooling is an advantage because there is a thick, reliable regional labor market for skilled workers. One rationale for the gains from large labor pools is that these regions can withstand shocks by easing labor transfer from low to high productivity establishments (Puga 2010). Firm establishments with varying and volatile employment may find it beneficial to locate in places where a large number of workers possess the requisite skills. Puga and Overman (2010) test the labor-pooling hypothesis and show that those industries with more "idiosyncratic volatility" are more spatially concentrated. Employment expands and contracts with positive and negative shocks, so a dependable external labor pool anchors firms to regions. In the past, much of the volatility in employment was for routine manual tasks. The regional advantage of large labor pools may be a relic of the nineteenth and twentieth century. In the future, automation will shield firms from unpredictable labor demand.

Beyond the labor-pooling hypothesis, researchers will need to continue to probe the extent of firm-labor matching advantages of agglomeration. For studying matching externalities, the availability of detailed microeconomic data holds considerable promise for research (Duranton and Kerr 2015). For example, wage regressions testing for agglomeration need extensive controls for various sources of observed and non-observed heterogeneity. An example of the information required for studying matching is Portugal's *Quadros do Pessoal* database. This is a mandatory survey for every firm operating in Portugal for businesses with wage earners. The data set includes precise information on firm and establishment location, sector of activity, type of ownership, employment, earnings, and workforce characteristics such as gender, age, tenure, and years of schooling. A worker identification code allows for the tracking of workers over time.

Based on an analysis of the Portuguese data, Figueiredo et al. (2014) find little evidence that the quality of labor matching increases with firm concentration within the same industry. Freedman (2008) reached a similar conclusion about matching and agglomeration in the U.S. software industry. In other words, the empirical evidence for matching and agglomeration appears surprisingly tepid, even before the second machine age begins in earnest. Yet it is important to understand how we measure agglomeration in these studies. Typically, industry-specific externalities (localization) are proxied with regional employment in the industry (e.g., Wheaton and Lewis 2002; Combes et al. 2008; Freedman 2008; Mion and Naticchioni 2009). In these studies, regional employment by industry is often introduced in the regressions as a density measure (divided by the area of the region) or as a share (with total employment included to capture urbanization economies).

The problem for future research on labor matching will be to measure agglomeration effects without relying on employment data. Firm (establishment) counts offer an alternative measure of agglomeration, but the total number of firms in a region may not reflect the scale of an industry in the region. We need an appropriate mass variable besides employment or firm counts for agglomeration. Clearly, output measures of agglomeration would be preferable, but they generally are not available for regions at a detailed industry level.

Despite the dearth of evidence for industry agglomeration and matching, the literature suggests that there is a productivity gain (or a large and significant wage premium) for industry clustering. In wage regressions, concentration within the same industry matters, even when matching does not figure prominently (Figueiredo et al. 2014). Thus, besides labor matching, other sources of industry-specific external economies may be at work; namely, learning/knowledge spillovers and input sharing and supplier specialization.

6.7 Learning and Knowledge Spillovers

The learning benefits of urban regions would seem to be crucial in the second machine age, given the need for flexible, adaptable, and non-routine labor skills. The city, our "greatest invention," enables people to share ideas and to interact and collaborate with each other at close proximity (Glaeser 2011). Extensive knowledge networks develop with urban scale and density, thereby raising productivity. Learning in cities facilitates the genesis and spread of technology, along with organizational innovation. The related Marshallian hypothesis is that spatial clustering by firms within a given industry increases intra-regional knowledge transfer through close interactions.

Distance from knowledge generation may influence the extent of regional knowledge spillovers, apart from the effect of industry clustering. Proximity to invention and innovation may positively affect knowledge transfer, notwithstanding the acceleration of information technology and communication through the Internet. The hypothesis is that even outside of industry clusters, being near knowledge generation is a major learning advantage for firms and individuals. Caragliu and Nijkamp (2015) provide a comprehensive review of proximity and knowledge.

Contemporary regional science has investigated knowledge externalities and spatial concentration, along with proximity effects. Still, scholars have only an inchoate understanding of learning in cities (Puga 2010). The empirical literature finds varying evidence on the role of proximity as a determinant of knowledge spillovers. Mostly, the studies use patent citations to measure local knowledge spillovers, following Jaffe et al. (1993). There is also evidence to support the agglomeration hypothesis, again using patent citations to assess local knowledge spillovers. One study of U.S. patent citations found that industry agglomeration compensates for the friction of distance (Figueiredo et al. 2015). In this case, a higher concentration of an industry's employment (or establishments) in a U.S. county has a positive impact on the number of patent citations in that county and industry, after controlling for distance from patent generation (the proximity effect).

Patenting records form the basis for studies of regional learning and agglomeration, in large part because the data are available. Patents represent a form of explicit, codified knowledge (along with scientific papers) that can be tracked across space. Yet the research on patent generation and diffusion does not directly measure productivity effects from knowledge spillovers.

Tacit, non-codified knowledge also spreads through agglomeration and may be subject to distance decay, but it remains difficult to assess since, unlike patents, it leaves no "paper trail." Much of tacit knowledge transfer is firm-specific, affecting productivity through in a variety of channels, including training, communication, and observation (Argote et al. 2000). Researchers in international business use surveys to assess knowledge transfer through organizations and across space (McDermott and Corredoira 2010). Devoted mainly to sophisticated econometric analysis, regional science has produced little survey-based analyses of tacit knowledge transfer.

Machines, unlike humans, do not depend on developing cognitive skills from proximity and close connections with other humans. An important question for future research on agglomeration is the extent to which learning externalities exist as machines perform non-routine, cognitive tasks. Carbon-based life forms will certainly learn and become more productive along with intelligent machines. Yet the learning advantages of human concentration in cities and industry clusters may erode.

6.8 Input Sharing and Specialization

The original insights into agglomeration viewed the size (or extent) of the market as allowing for firm specialization (Smith 1904). With a larger market reach, there would be demand for more differentiated products. Following Marshall (1920), an industry's regional size and density lead to the emergence of more efficient, specialized suppliers not found elsewhere. Essentially, agglomeration leads to vertical disintegration among suppliers, which may raise productivity. Firms in the local industry can share and benefit from this intermediate supplier network, along with other supporting institutions and infrastructure in the regional cluster.

Empirically, studies at first found little evidence for input-sharing hypothesis, compared with other sources of agglomeration (Rosenthal and Strange 2001). In more recent research, input sharing emerges as a strong determinant in agglomeration (Overman and Puga 2010).

The vertical disintegration argument for regional agglomeration gains remains a subject of regional research. Holmes (1999) represents the first systematic empirical study addressing the hypothesis (see Rosenthal and Strange 2004). In the U.S. manufacturing sector, there appears to be a positive association between localization of an industry and vertical disintegration among firms.

More recently, a study used Portuguese data to evaluate the proposition that the vertical disintegration of firms should be greater in areas where an industry concentrates (Guimarães et al. 2010). The approach addresses several problems associated with Holmes's (1999) paper. With access to detailed data for all regions and industries, the authors were able to use the plant as the unit of observation. After controlling for firm size and sector, the estimates indicated that vertical disintegration by firms is about 3% higher in regions where industries agglomerate.

Plausibly, the benefits of having an efficient base of specialized suppliers and other shared inputs will persevere, despite automation and potential large-scale job displacement. Firms agglomerate in regions, not just workers. Indeed, shared inputs facilitate an interaction between suppliers and buyers, but do not require a mass of labor.

Thus, supplier specialization and development within clusters should continue to be considered as a source of agglomeration gains. More broadly, research will have to account for more complex interaction across firms and industries. Input sharing, along with cooperation and collaboration within clusters, is an evolving process.

Twenty-first century firm agglomeration may also be affected by the trend toward functional specialization and the concentration of innovative activities (Carlino and Kerr 2015). Conceivably, some regions will specialize in goods production, while others specialize in innovative functions (Duranton and Puga 2005). One recent survey of leading global companies revealed that that firm sorting by functional specialization remains a trend in contemporary value chains (World Economic Forum 2016).

6.9 Agglomeration Measures

Finding a theoretically grounded measure of agglomeration remains an essential objective for regional science. Over the past few decades, researchers have refined measures of industry localization, notably the Ellison and Glaeser (1997) index, an approach that has the virtue of clear microeconomic foundations (random utility maximization). However, to date researchers have not settled on a single measure that estimates the strength of industry agglomeration. A significant contribution is the continuous distance measures of geographic concentration of Duranton and Overman (2002). These measures have proven to be useful because they exploit available microeconomic data on employment and establishments by industry and region. Ideally, spatial concentration is measured using information on the precise location of each business unit. The data are, unfortunately, rarely available. Moreover, most agglomeration indices base the analysis on employment to assess the relative regional scale of an industry. This approach masks regional industry concentrations in high-value added, but relatively low employment sectors like biotechnology in Boston–Cambridge, Massachusetts. Value added, or output measures would be preferable, but the regional data are not available or disclosed to researchers.

Further, despite major improvements in spatial concentration indices, they still apply to individual industries. This limits the scope of potential agglomeration and our understanding of the far-reaching effects of automation. In innovative regions that are likely to benefit from the creative side of creative destruction, broad technology clusters supplant industry agglomeration. In other words, firms and workers do not concentrate within the industries typically defined in government statistics. The classic case is the Silicon Valley, which is not only an agglomeration of firm headquarters in the semiconductor industry, but of information technology companies that span hardware, software, suppliers, specialized services, along with supporting academic and government institutions. Regional clusters, a broader form of agglomeration, revolve around knowledge, skills, inputs, local demand, and other regional linkages that lead to productivity enhancing economic activities. The "algorithm" for defining clusters is still in an experimental stage, based on inter-industry, occupational, and co-location linkages (Delgado et al. 2016). If an accepted procedure emerges, then cluster definitions could supplant industry-based approaches.

With satisfactory measures, researchers will then be in a much better position to discern how clusters evolve. As Jackson (2015) avers, one of the fundamental challenges for future research will be to comprehend the tension between industry concentration and diversification. His "cluster assessment diversification strategy" offers a promising method for assessing the ongoing structural changes we will face over the next 50 years.

6.10 Conclusion

Extensive research on agglomeration agrees that competitive regional clusters and large metropolitan regions raise productivity. External scale economies allow firms and workers to generate more output with the same inputs in larger, denser regions.

While agglomeration is a bedrock principle of regional science, the ground is shifting. Agglomeration research will have to address ongoing structural changes induced by the new era of smart, interacting machines. The scale of regional employment could fall with the substitution of routine human tasks by machines. We are already finding declining jobs for routine skills in the United States (Dvorkin 2016). At the same time, there has been growth in non-routine jobs: high-paying, cognitive skills. In the second machine age, we could find agglomeration effects for non-routine, high-skill, jobs, which endure by complementing automation (Autor 2015). Local labor pooling/matching and knowledge spillovers, first recognized in the late nineteenth century by Alfred Marshall may last as an agglomeration driver in the twenty-first century.

Thus, agglomeration economies should survive the twenty-first century, unless machines substitute for non-routine, cognitive employment. Even today's productive jobs are not permanently productive (Duranton 2014). Regional development must be upgraded through innovation, entrepreneurial activity, and employment expansion in high-growth firms. The intricacies of this dynamic process merit future study.

We will also need a deeper understanding of how the benefits from agglomeration spread unevenly across regions, favoring a few elite regions that attract and retain high-skill workers. We know that productive workers and firms choose regions and thus raise productivity, apart from agglomeration (Behrens et al. 2014). We will need to evaluate firm and worker sorting relative to agglomeration. This research requires detailed microeconomic data, especially matched employer-employee matched linked to skill levels.

Finding a tractable agglomeration measure remains essential for regional science. We also need to assess how regional clusters boost productivity through knowledge networks, skills, inputs, local demand, and other regional linkages. With better measures of agglomeration, future research will potentially probe deeper into the three main sources of regional productivity from agglomeration: better labor skill matching and local pooling; more efficient input sharing, along deeper and more specialized supplier relationships; and enhanced localized learning and knowledge flows.

Ideally, to study agglomeration, academics would have access to the "big data" amassing on worker-machine interactions, as well as the huge range of other information that is now proprietary and held closely by private companies. Only these firms can access and analyze the data and thus better understand human behavior and the economy. There have been calls for an equally potent, publicly accessible data infrastructure—"big data for the masses" (Nielsen 2015). Unlocking these data for academic research would open up a treasure trove for regional science.

References

Andersson M, Klaesson J, Larsson JP (2014) The sources of the urban wage premium by worker skills: spatial sorting or agglomeration economies? Pap Reg Sci 93(4):727–747

Argote L, Ingram P, Levine JM, Moreland RL (2000) Knowledge transfer in organizations: learning from the experience of others. Organ Behav Hum Decis Process 82(1):1–8

Autor DH (2015) Why are there still so many jobs? The history and future of workplace automation. J Econ Perspect 29(3):3–30. Summer

Autor DH, Dorn D (2013) The growth of low-skill service jobs and the polarization of the U.S. labor market. Am Econ Rev 103:1553–1597

Bacolod M, Blum BS, Strange W (2009) Skills in the city. J Urban Econ 65:136–153

Barbosa D (2010) After suicides, scrutiny of China's grim factories. New York Times, 6 June

Beeson P (1987) Total factor productivity growth and agglomeration economies in manufacturing, 1959–1973. J Reg Sci 27(2):183–199

Behrens K, Duranton G, Robert-Nicoud F (2014) Productive cities: sorting, selection, and agglomeration. J Polit Econ 122(3):507–553

Bivens J, Mishel L (2015) Understanding the historic divergence between productivity and a typical worker's pay. Economic Policy Institute, 2 Sept

Brynjolfsson E, McAfee A (2014) The second machine age: work, progress, and prosperity in a time of brilliant technologies. WW Norton, New York

Brynjolfsson E, McAfee A (2015) The great decoupling. Harvard Business Review, June

Brynjolfsson E, McAfee A (2016) Where computers defeat humans, and where they can't. New York Times, 16 Mar

Byrne DM, Fernald JG, Reinsdorf MB (2016) Does the United States have a productivity slowdown or a measurement problem? Brookings Papers on Economic Activity, 1 Mar

Caragliu A, Nijkamp P (2015) Space and knowledge spillovers in european regions: the impact of different forms of proximity on spatial knowledge diffusion. J Econ Geogr:1–26. doi:10.1093/jeg/lbv042

Carlino G, Kerr W (2015) Agglomeration and innovation. In: Duranton G, Henderson V, Strange W (eds) Handbook of regional and urban economics, vol 5. Elsevier, Amsterdam, pp 349–404

Chan K (2015) Robot revolution sweeps China's factory floors. AP, 23 Sept

Combes P-P, Gobillon L (2015) The empirics of agglomeration economies. In: Duranton G, Henderson V, Strange W (eds) Handbook of urban and regional economics, vol 5. Elsevier, Amsterdam, pp 247–348

Combes P-P, Duranton G, Gobillon L (2008) Spatial wage disparities: sorting matters! J Urban Econ 63:723–742

Crafts N, Klein A (2015) Agglomeration economies and productivity growth: U.S. cities, 1880–1930. CEPR discussion papers 10673

David PA (1990) The dynamo and the computer: an historical perspective on the modern productivity paradox. AEA papers and proceedings, May, pp 355–361

Delgado M, Porter M, Stern S (2016) Defining clusters of related industries. J Econ Geogr 16(1):1–38

Duranton G (2014) Agglomeration and jobs. In: Fisher M, Nijkamp P (eds) Handbook of regional science. Springer, Berlin, pp 631–648

Duranton G, Kerr W (2015) The logic of agglomeration. Harvard Business School working paper 16-037

Duranton G, Overman HG (2002) Testing for localization using micro-geographic data. Rev Econ Stud 72(4):1077–1106

Duranton G, Puga D (2005) From sectoral to functional urban specialization. J Urban Econ 57:343–370

Dvorkin, Maximiliano, 2016. Jobs involving routine tasks aren't growing. On the Economy, Federal Reserve Bank of St Louis, 4 Jan

Ellison G, Glaeser EL (1997) Geographic concentration in U.S. manufacturing industries: a dartboard approach. J Polit Econ 105(5):889–927

Figueiredo O, Guimarães P, Woodward D (2014) Firm-worker matching in industrial clusters. J Econ Geogr 14:1–19

Figueiredo O, Guimarães P, Woodward D (2015) Industry agglomeration, distance decay, and knowledge spillovers: following the patent paper trail. J Urban Econ 89:21–31

Florida R, King KM (2016) Spiky venture capital: the geography of venture capital investment by metro and zip code. Martin Prosperity Institute, University of Toronto, Rotman School of Management

Ford M (2015) The rise of the robots: technology and the threat of a jobless future. Basic Books, New York

Freedman ML (2008) Job hopping, earnings dynamics, and industrial agglomeration in the software publishing industry. J Urban Econ 64:590–600

Frey CB, Osborne MA (2013) The future of employment: how susceptible are jobs to computerization? Oxford Martin School, Oxford

Frey CB, Osborne MA (2016) Technology at work v2.0: the future is not what it used to be. Oxford Martin School and Citi, Oxford

Fujita M, Thisse J (2002) Economics of agglomeration: cities, industrial location, and regional growth. Cambridge University Press, Cambridge

Gibbes S (2016) Mercedes-Benz swaps robots for people on its assembly lines. Financial Times, 26 Feb

Glaeser E (2011) Triumph of the city: how our greatest invention makes us richer, smarter, greener, and happier. MacMillan, New York

Gordon RJ (2015) The rise and fall of American growth: the U.S. standard of living since the civil war. Princeton University Press, Princeton

Graetz G, Michaels G (2015) Robots at work. CEP discussion paper no 1335. Centre for Economic Performance, The London School of Economics and Political Science, March

Guimarães P, Figueiredo O, Woodward D (2010) Vertical disintegration in Marshallian industrial districts. Reg Sci Urban Econ 40(1):73–78

Henderson JV (1974) The size and types of cities. Am Econ Rev 64(4):640–656

Henderson V (2003) The urbanization process and economic growth: the so-what question. J Econ Growth 8(1):47–71

Hohenberg P, Lees LH (1985) The making of urban Europe (1000–1950). Harvard University Press, Cambridge, MA

Holmes T (1999) Localization of industry and vertical disintegration. Rev Econ Stat 81(2):314–325

Jaffe A, Trajtenberg M, Henderson R (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. Q J Econ 108(3):577–598

Jackson RW (2011) Knowns, unknowns, and impacts. Rev Reg Stud 41:5–11

Jackson RW (2015) Are industrial clusters and diversity strange bedfellows? Rev Reg Stud 45:113–129

Kan M (2013) Foxconn to speed up 'robot army' deployment. PCWorld, 26 June

Marshall A (1920) Principles of economics, 8th edn. Macmillan, London

Matano A, Naticchioni P (2012) Wage distribution and the spatial sorting of workers. J Econ Geogr 12:379–408

McDermott GA, Corredoira RA (2010) Network composition, collaborative ties, and upgrading in emerging-market firms: lessons from the Argentine autoparts sector. J Int Bus Stud 41(2):308–329

Mion G, Naticchioni P (2009) The spatial sorting and matching of skills and firms. Can J Econ 42:28–55

Mokyr J, Vickers C, Ziebarth NL (2015) The history of technological anxiety and the future of economic growth: is this time different? J Econ Perspect 29(3):31–50

Moomaw R (1981) Productivity and city size: a critique of the evidence. Q J Econ 96(4):675–688

Nielsen M (2015) Who owns big data? MIT Technology Review, 5 Jan

Overman HG, Puga D (2010) Labour pooling as a source of agglomeration: an empirical investigation. In: Glaeser EL (ed) Agglomeration economics. University of Chicago Press, Chicago, IL

Porter M (1990) The competitive advantage of nations. The Free Press, New York

Pratt GA (2015) Is a Cambrian explosion coming for robotics? J Econ Perspect 29(3):51–60

Puga D (2010) The magnitude and causes of agglomeration economies. J Reg Sci 50(1):203–219

Puga D, Overman H (2010) Labour pooling as a source of agglomeration: an empirical investigation. In: Glaeser E (ed) Agglomeration economics. University of Chicago Press, Chicago, pp 133–150

Rosenthal SS, Strange W (2001) The determinants of agglomeration. J Urban Econ 50(2):191–229

Rosenthal SS, Strange W (2004) Evidence on the nature and sources of agglomeration economies. In: Henderson V, Thisse J-F (eds) Handbook of regional and urban economics, vol 4. North-Holland, Amsterdam, pp 2119–2171

Smith A (1904) An inquiry into the nature and causes of the wealth of nations, 5th edn. Methuen, London

Sprague S (2014) What can labor productivity tell us about the U.S. economy? U.S. Department of Labor, Bureau of Labor Statistics, Beyond the numbers, vol 3, No 12, May

TCI (2016) The global practitioners network for competitiveness, clusters, and innovation. http://wwwtci-networkorg/. Accessed 15 Mar 2016

United Nations, Department of Economic and Social Affairs, Population Division (2014) World urbanization prospects: the 2014 revision, highlights (ST/ESA/SER.A/352)

Walden ML (2015) North Carolina in the connected age: challenges and opportunities in a globalizing economy. The University of North Carolina Press, Chapel Hill, NC

Wheaton WC, Lewis MJ (2002) Urban wages and labor market agglomeration. J Urban Econ 51:542–562

World Economic Forum (2016) The future of jobs: employment, skills, and workforce strategy for the fourth industrial revolution. Global challenge insight report, January

Doug Woodward is professor of economics and the director of Division of Research at the Darla Moore School of Business, the University of South Carolina. His primary research interests are regional economic development, firm location, urban and industrial clustering, and foreign direct investment. Dr. Woodward earned his Ph.D. at the University of Texas at Austin in 1986.

Chapter 7
Designing Policies to Spur Economic Growth: How Regional Scientists Can Contribute to Future Policy Development and Evaluation

Carlianne Patrick, Amanda Ross, and Heather Stephens

7.1 Introduction

Policymakers at all levels of government design policies with a goal of increasing economic growth. Lawmakers frequently design programs to attract new businesses with the hope that these enterprises will drive future growth within their jurisdiction (Neumark et al. 2005). However, such policies often have heterogeneous effects and induce unintended outcomes. It is important to understand fully the direct and indirect impact of these policies on regional economic development. Researchers in regional science have the unique ability to aid policymakers in the design of local economic development policies so that they can achieve desired goals. Ideally, regional science research can also provide guidance for future policies—helping to avoid some of the unintended consequences that may occur when designing policy.

Research in regional science can contribute to policy evaluation. However, previous policy evaluation research has largely ignored the unintended consequences of local economic development policy on regional growth. Since there is a limited amount of land (Roback 1982), local economic development comes at a cost, especially since most policies are designed to attract businesses of a specific type. Whether that goal is specifically stated, or if the policy is written in such a way that

C. Patrick (✉)
Department of Economics, Georgia State University, Atlanta, GA, USA
e-mail: cpatrick@gsu.edu

A. Ross
Department of Economics, Finance, and Legal Studies, University of Alabama, Tuscaloosa, AL, USA
e-mail: aross@cba.ua.edu

H. Stephens
Resource Economics and Management, West Virginia University, Morgantown, WV, USA
e-mail: heather.stephens@mail.wvu.edu

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_7

a specific type of business is more likely to be enticed by it, attracting one type of business frequently comes at the cost of other businesses. We start our discussion by highlighting previous research that has examined these heterogeneous effects of policies on local economic development. Specifically, we look at the research regarding sorting across different types of industries in response to different policies, as well as differences in sorting based on the size and corporate structure of the firm. However, this literature needs to be expanded; not only in terms of research on these issues but also through more and better data.

More work also needs to be done to distinguish the effects of these policies on attracting existing businesses from other areas versus helping to grow or start new entrepreneurial enterprises. While policymakers at the local level might not care if they are attracting businesses from other regions, state and federal policymakers might be concerned with taxpayer funds being used simply to redistribute businesses. Local entrepreneurs also may have stronger ties to the region and may be less likely to move once a program ends, increasing the economic effects of helping these businesses. Finally, there needs to be increased attention on how the effects of policies may vary across regions. While, on average, programs may have benefits, the benefits of the programs may not be shared equally across all eligible areas.

Regional scientists, and almost all researchers in the social sciences, frequently face issues with the availability and quality of data. Numerous quality data sets are compiled by federal, state, and local governments across the world. However, in times of tight budgets, policymakers frequently look to save on costs by cutting the collection and maintenance of these data. Unfortunately, policymakers often underappreciate the importance of data to research in the social sciences and effective program design, including economic development. We discuss the importance of data availability both now and in the future to researchers in regional science. One area that needs to continue to improve is the availability of administrative data or data collected as part of program operations and record keeping. While survey data are readily available, there are increasing issues with response rates and accurate reporting. At the same time, a challenge with administrative data is that they are frequently gathered with no consideration for their use to researchers. We discuss in detail below the pros and cons of each type of data as well as the importance and possibility of combining the two.

Finally, we discuss the importance of policy evaluation methodologies in the field of regional science. It is important that analysts credibly determine a causal effect of an existing or hypothetical policy. However, the non-random nature of implementation of most policies creates challenges for estimating causal effects. We have seen an emergence of experimental economics, where the laboratory environment allows researchers to control for various factors that could bias estimates, but this type of research frequently faces concerns regarding external validity. This has led to growth in research using quasi-experimental methodologies, such as difference-in-differences and regression discontinuity methods, that attempt to reproduce the results that could be obtained from an experiment with observational data. While these methods are improvements over simple ordinary least squares methods, there are still shortcomings and limitations to these types of analysis.

We conclude by discussing what we view as the future of regional policy analysis and the role of regional scientists in the development of policy. In England some researchers are working hand-in-hand with policymakers to set up policy evaluation programs, but these efforts are woefully underemployed across the world. In the United States, there are times when the government does consult with regional scientists and economists to evaluate policy, but frequently this is done too late in the process. Therefore, it is important that we as regional scientists work with policymakers, in order to aid the design of optimal policy.

7.2 Heterogeneous Effects of Government Policies

Policymakers aim to implement policies that create economic growth, especially job creation. Some policies affect an entire country, while other programs are focused on specific areas. Furthermore, some policies are enacted at the local or state level, while others are national programs aimed at helping disadvantaged or struggling[1] areas. Whether or not policies to attract businesses to a specific region are effective in supporting growth partially depends on whether or not the standard spatial equilibrium of regional economics holds. The idea behind the spatial equilibrium model is that, when there is perfectly mobile labor, workers will move until utility is equalized across space, and firms will move until profits are equalized across space (Roback 1982). While the concept of a spatial equilibrium may seem logical, there are numerous studies questioning whether or not it actually exists due to incomplete models or persistent frictions (Bartik 1991, 1993; Kline and Moretti 2013; Moretti 2011). Thus, policies that are directed at people and firms who can relocate may not be effective in helping struggling regions.

If a spatial equilibrium will not occur on its own, then many argue that location-based policies, such as those designed to attract businesses to a specific region, may be the best way to achieve locational equilibrium and help struggling regions (Partridge et al. 2015). Partridge et al. (2015) lay out three main considerations for such interventions: (1) whether new jobs lead to employment of the local unemployed and reductions in poverty, (2) if new jobs are sustainable over the long term, and (3) if the long-term benefits outweigh the costs. Much of the previous research has focused on examining the employment and poverty effects. Very little research thus far has considered the sustainability of new jobs or conducted the more difficult full cost-benefit analysis of such policies; these are questions for future research.

Job creation is frequently a goal of economic development policies because there is evidence that policies that support job growth lead to reductions in poverty and

[1] An area can be considered struggling for many reasons, typically criteria such as high poverty rates or high unemployment rates. Other criteria may also be used, such as areas with high population loss or out-migration.

other positive outcomes for regions. Analysis of policies that focus on job growth have found that they are also effective at reducing poverty (Partridge and Rickman 2007) and increasing labor force participation (Bartik 1991).[2] Bartik (1991) also found evidence of earnings gains for existing residents—with about half of the increase due to higher local employment rates and the other half due to moving up to higher paying occupations. However, his results also suggest that at least some of the new jobs are going to new residents. This research also found that these effects are similar across regions—and across areas with different underlying economic conditions. However, in more recent work, Bartik (2005) suggests that the benefits from a job creating policy will be higher if the policy addresses a market failure or if there is public infrastructure that is underutilized—such as in distressed areas that have lost population and employment.

Policymakers have various means for supporting job growth in specific regions and must decide which type of policies are most effective. Because previous research has suggested that new businesses tend to create more new jobs than existing businesses (Haltiwanger et al. 2013), policymakers frequently create location-based incentive programs, such as taxes or subsidies, to attract businesses to a specific area. Existing research on the success of these programs has found mixed results, with some research suggesting these programs have positive effects (Billings 2008; Busso et al. 2013; Krupka and Noonan 2009; Papke 1994) while others find that the policies have no effect (Boarnet and Bogart 1996; Bondonio and Engberg 2000; Bondonio and Greenbaum 2007; Greenbaum and Engberg 2004; Hanson 2009; Kolko and Neumark 2010). Neumark and Simpson (2014) provide a review of the existing literature on these location-based policies and the current state of knowledge regarding these programs.

Future research should explore why there does not appear to be an overall effect from these policies. Are these programs creating distortions among different types of businesses? Are these programs leading to sorting across space of firms in different industries or with different types of structures (corporations versus sole proprietorships)? Are they simply causing business turnover? Some recent research has begun to look at where these policies may be creating distortions.

The argument behind differential effects across industries was formalized by Hanson and Rohlin (2011), who developed a model to explain differential impacts across industries from the federal Empowerment Zone (EZ) tax credit. The EZ program provides a tax credit to firms that locate in specific areas if they also hire workers from that area. Therefore, the EZ program can be considered a labor tax credit, as it reduces the cost of hiring workers with no direct effect on the cost of capital. The authors estimate the ratio of capital cost to total cost of production to determine which industries are more capital intensive and which are more labor intensive. Hanson and Rohlin (2011) show that industries that are more capital intensive, such as manufacturing and mining, have higher values of the ratio of

[2] Goetz et al. (2010) review the landscape and literature as it pertains to rural entrepreneurship programs.

capital cost to total cost of production than more labor intensive industries, such as retail and services. Given the focus on the EZ program on reducing the cost of labor, they demonstrate that the EZ program attracts more labor intensive businesses, such as retail, at the cost of more capital intensive businesses, such as manufacturing.

Patrick (2016b) applies a similar type of analysis using a measure of capital subsidies versus labor tax credits. Using the Incentives Environmental Index, which was constructed based on state constitutional provisions that structure the ability of government entities to aid private enterprise, the author looks at how capital subsidies affect the capital-labor substitution in industries.[3] Findings indicate that increasing the number of capital subsidy tools that can be utilized by state and local governments causes a change in the local industry mix, particularly attracting more firms in capital intensive industries.

Other research has taken a similar approach to examining this type of sorting across industries from location-based tax policies. Freedman (2015) and Harger and Ross (2016) apply a similar type of analysis to the New Markets Tax Credit (NMTC) program, a program operated by the Community Development Investment Financial (CDFI) Fund. While the EZ program was effectively a labor tax credit and attracted labor-intensive businesses, the NMTC program attracted more firms in the industries that tended to receive the tax credit, including manufacturing and retail. Both papers find that the NMTC program created a sorting across industries in eligible versus ineligible areas, consistent with the arguments put forth by Hanson and Rohlin (2011).

Harger et al. (2015) consider both the NMTC and the CDFI program, both of which operate through the CDFI Fund. They also find evidence of sorting across industries. In addition, they consider whether this new business activity comes at the cost of existing businesses by examining firm deaths. Their results do not support the argument that the program simply creates business turnover. This conclusion is consistent with the work of Rohlin and Ross (2016), who consider the effect of bankruptcy law on business turnover and found that more generous bankruptcy exemptions increase the number of existing businesses, likely because more generous protections in bankruptcy encourage entrepreneurs to incur the risk of opening their own business.

Rohlin et al. (2014) also consider how state tax policy differences may cause a sorting of different types of business activity across state boundaries. Overall, they find no effect from the differences in state tax policies, however, the results suggest that this is because there is offsetting activity from other state policies. For example, the presence of a reciprocal agreement, which determines if you pay income tax in the state where you work or where you live, affects the sorting of industries. In particular, labor intensive industries are less sensitive to state income tax rates when individuals pay taxes based on where they live rather than if individuals pay personal income taxes in the state where they work. Because the authors are looking at places close to the state border, it is easy for individuals to work in the high

[3]For more information on the creation of the Incentives Environment Index, see Patrick (2014).

personal income tax state but live across the border in the low personal income tax state. They also find that corporations are more sensitive to the corporate income tax rates, while sole proprietorships are more sensitive to the personal income tax rates, consistent with expectations because corporations are more likely to be subject to the corporate income tax while sole proprietorships are more likely to be subject to the personal income tax.[4]

Overall, this growing literature is finding that while there may be no effect overall of various programs, these policies are creating a sorting of different types of businesses. As this research continues to expand, and as the quality of data available improves, these issues should be considered further. Future research could include additional exploration of the impacts on industrial composition and other dimensions of business activity, such as the impact of policies on business activity based on firm size and industrial structure.

While recent research has found a link between either small businesses or proprietors and job creation (Deller 2010; Shrestha et al. 2007; Stephens and Partridge 2011; Stephens et al. 2013), another challenge for regional science research in the future is understanding the regional variation in the impact of policies to support job creation. For example, Shrestha et al. (2007) find that the strongest impact is in metropolitan areas. However, Stephens et al. (2013) show that even in struggling regions, like the Appalachian region of the United States, having more self-employment results in higher employment growth.[5] While location-based relocation packages are going to outside firms by design, many of the programs (such as those run by the CDFI) could be used to help expand existing businesses; and future research should examine this further.

Even if it appears that a program, on average, is creating jobs, the question remains whether it is effective in all cases. To explore the heterogeneity in the drivers of job growth across the U.S., Partridge et al. (2008) use a geographically-weighted regression (GWR) approach which allows these factors to vary across regions. They find significant heterogeneity—suggesting that policymakers should be wary of estimates that provide a global average effect from a policy as that average may be masking very different impacts across local areas. Deller (2010), also finds evidence of differences in the impact of small businesses on economic growth—where the differences vary based on industry and region of the country. If federal programs affect different regions in different ways, it is important to understand these differences. With advances in computing power, access to microdata, and new methodologies, this is an opportunity for future researchers.

---

[4]Borchers et al. (2016) also looked at the relationship between various types of state taxes and found that most of the impact of policy is on small businesses, versus larger establishments.

[5]Michelacci and Silva (2007) find that firms are larger and more capital intensive when owners are from the region. This positive relationship holds even though data limitations sometimes make it necessary to use measures of entrepreneurship that include entrepreneurs of necessity who start businesses when there are no other opportunities—and thus are not likely to contribute to economic growth (Stephens and Partridge 2011).

7.3 Issues Regarding Current and Future Data

Progress on empirical policy evaluation is aided by improvements in the data available to researchers and methodologies used for analysis. The recent proliferation of data, particularly microdata, and advances in policy evaluation methods provide opportunities to answer some important unresolved questions discussed earlier; yet, there are significant challenges that may hamper progress.

In order to answer relevant policy questions, economic development researchers need both microdata and aggregate data at fine levels of geography. Microdata allow researchers to investigate individual and firm level policy effects, potentially uncovering the mechanism through which the policy acts. However, microdata can miss general equilibrium effects that are revealed through analyzing aggregate data at varying spatial scales. For example, microdata on establishments benefiting from a policy might reveal that recipients create more jobs than non-recipients, while analysis of the same policy using aggregate data indicates no job creation. As is the case with some of the programs discussed above, further investigation reveals these seemingly incongruent results arise from the sorting of firms across industries and regions. In this way, the full effects of the policy can only be understood using both microdata and aggregate data at varying spatial and sectoral scales.

There are more detailed micro data available to researchers today than ever before, and as more and more information is gathered electronically, this trend is likely to continue (Einav and Levin 2013; Varian 2014). In addition to transforming data obtained from traditional sources, the Internet houses a wealth of data available for researchers to collect. Through "scraping" websites for data compiled for other purposes, researchers can collect detail on local prices, jobs, etc. (Edelman 2012). More detailed data allow economic development policy researchers to take advantage of new methods as well as apply existing methods previously underutilized because of data limitations. The sheer volume of the data may present a challenge in this environment, but shared computing resources and new database structures provide researchers with practical solutions to storage and computational limits (see Varian 2014 for an overview).

Data proliferation also generates situations in which the number of available predictors exceeds observations (high-dimensional data) and thus requires tools for variable selection (Belloni et al. 2014; Varian 2014). Belloni et al. (2014) demonstrate that these high-dimensional data methods substantially improve causal estimation of policy effects.

Although data are generally becoming more available, there are serious threats to the continued availability and quality of some data essential to economic development policy research. For example, many researchers in the U.S. were opposed to the removal of the long-form decennial census and its replacement with the annual American Communities Survey (ACS), as many researchers do not believe the ACS has the same amount of information at small levels of geography. In Canada there have been suggestions about making the census voluntary, raising obvious problems regarding selection and use of this important data set.

Publicly available aggregate data at fine levels of geography are typically created through survey and administrative data programs at government agencies. Recently, government data programs have been targeted as a seemingly innocuous way to reduce government spending. Data programs facing budget cuts must make hard choices about which data to collect and which data to make available, sometimes leading to the elimination of state and local data programs. In addition to the replacement of the census long form, the 2013 U.S. federal sequestration mandated budget cuts at the Bureau of Economic Analysis (BEA). The result of this was the elimination of the Local Area Personal Income (LAPI) program that provides annual county level data on economic activity by sector and income by source. As argued by Partridge et al. (2013), these data programs are necessary for the evaluation of local economic development policies. Fortunately, regional scientists and other public data users successfully lobbied for the restoration of part of the LAPI, but other state and local data programs remained victims of budgetary cutbacks.

On the positive side, microdata have become more accessible through publicly available de-identified government survey datasets [such as the Current Population Survey, (CPS)], Internet sources, and increased access to confidential survey and administrative data. The latter is aided by growth in the number of secured access locations, such as the Census Research Data Centers, and through on-line remote portals to restricted data, such as the Data Enclave through the Michigan Center on the Demography of the Aging. The same budgetary pressures discussed above threaten the resources supporting increased (public or secure) data access. Cuts to survey programs like the ACS have become a perennial topic for debate. This is particularly concerning given declining survey quality. Meyer et al. (2015) argue that there are three main sources contributing to the decline in survey quality—increased unit non-response, increased item non-response, and measurement error (inaccurate reporting)—in important datasets for regional policy analysis such as the ACS, CPS, and Survey of Income and Program Participation (SIPP). In particular, there is evidence of misreporting self-employment income, earnings, type of college attended, transfer program participation, and other key variables for evaluating economic development policy. Such misreporting can lead to substantial bias in estimated policy effects, although there is little evidence on the extent of that bias.

The decline in the quality of survey data coupled with the increase in access to administrative data will likely bolster the current trend towards administrative microdata research (and away from survey microdata) documented in Chetty (2012). However, as Meyer et al. (2015) note, administrative data are not collected with research in mind. These data vary substantially in quality, accessibility, and coverage—making replication of results and generalizable conclusions more difficult. Administrative data also tend to contain a limited set of characteristics. Some of these limitations, however, may be overcome by linking survey and administrative data. An emerging literature uses linked survey and administrative data, but no research so far (of which we are aware) has been by regional scientists evaluating economic development policies.

Limited survey and administrative data on economic development incentives has significantly hampered evaluation of these policies. The recent studies on

Empowerment Zones and New Market Tax Credits employ federal data. Similar data are available for some state and local programs; however, comprehensive data on state and local government incentives continue to prove elusive (Patrick 2014; Thomas 2011; Buss 2001). Future research will benefit from recent developments in state and local government accounting standards requiring uniform reporting of economic development tax expenditures (i.e., tax abatements and credits) (Sharma and Davidoff 2015). The next step towards comprehensive administrative data is similar reporting requirements for non-tax incentives (i.e., cash, grants, land, and low-interest financing), which research suggests comprise the bulk of state and local resources devoted to economic development (Patrick 2014, 2016b; Bartik et al. 2003).

7.4 Considerations of Methodologies Used in Regional Science

Regardless of data, the field now requires that researchers establish the extent to which their particular combination of data and methods: (i) produces stylized facts from correlations and patterns; (ii) highlights causal estimates of policy effects (and key parameters), and/or (iii) generates reliable out-of-sample predictions. The descriptive approach to empirical studies provides stylized facts without any causal or predictive claims. There is an increasing expectation that policy research go further by either convincingly attributing outcome changes to the policy, credibly predicting the change in outcomes from potential policy changes, or both. This is particularly difficult for economic development policy because policies are not randomly implemented across locations and interact with locational attributes to produce heterogeneous effects.

In recent years, causal and predictive research has been categorized as following either the structural or the experimental approach. Rust (2014) describes the structural approach as “empirical work that takes theory seriously” (p. 821), whereas work following the experimental approach avoids tight integration of theory and focuses instead on causal inference with few assumptions. Angrist and Pischke (2010) proclaimed a credibility revolution in empirical work associated with the experimental emphasis on quality research design. Although there are significant supporters, not all empiricists agree. A series of articles published alongside Angrist and Pischke (2010)—Keane (2010), Leamer (2010), Nevo and Whinston (2010), and Stock (2010)—summarize the relative merits and drawback of the structural and experimental approaches. In general, the greatest strength of the experimental approach is the focus on credibly exogenous sources of identifying variation. On the other hand, experimental results may have limited external validity, provide little basis for out-of-sample predictions, and, therefore, cannot be used to analyze potential effects of policies not-yet-implemented. The latter is the primary benefit of structural work; whereas identification may be more problematic in structural models.

In a 2010 assessment of research in regional economics, Holmes (2010) suggests that most work in the field follows the descriptive approach. He also notes that the experimental approach is gaining traction, but very little work is using the structural approach outside of research that overlaps with empirical sorting models in public finance. Since Holmes' writing, economic development policy evaluation has increasingly employed experimentalist methods. Although economic development policy is not generally the subject of randomized controlled trails (RCTs) and lab experiments, there is a mature literature on experimental methods using natural experiments, quasi-experiments, difference-in-differences, matching, instrumental variables, discontinuity designs, and other methods. These empirical strategies now appear more frequently in the literature. For example, the cross state-border double-differencing methodology in Rohlin et al. (2014) compares changes in similar locations that credibly serve as the alternatives from which a particular business owner is choosing—a twenty-mile wedge at state borders. By removing time-invariant factors specific to the business owners' location choice set, they are able to obtain a causal estimate of tax policies. We expect future research will continue to employ these types of methods, with increased scrutiny of identification strategies.

Leamer expressed concern that the focus on experimental methods was leading researchers to "apply push-button methodologies without sufficient thought regarding their applicability and shortcomings" (Leamer 2010). These methods require researchers to think deeply about the assumptions of their empirical strategy (i.e., institutional settings for natural experiments and the exogeneity case for instrumental variables). Blundell and Costas (2009) and Imbens and Wooldridge (2009) provide practical guidance and outline the identifying assumptions for the most commonly employed experimental strategies in policy analysis. Even within a quasi-experimental framework, violations of key identifying assumptions can lead to incorrect policy effect estimates.

As an example, Patrick (2016a) compares the local economic development effects of successfully attracting a large new manufacturing plant obtained from two quasi-experimental empirical strategies: difference-in-differences estimation combined with either a natural experiment (in which counterfactuals are revealed in a magazine) or geographically-proximate matching control by design. The two identification strategies yield different estimates of the effects of large plants on economic activity and fiscal surplus. Patrick (2016a) presents distributional and placebo tests indicating that the natural experiment results may be driven by systematic differences in counterfactual county outcomes, while the geographically-proximate matching strategy does not appear to suffer from this threat to identification. These tests suggest geographically-proximate matching is the preferred strategy, which is in line with recent findings that combining the difference-in-differences estimator with methods for pre-processing data produces results closer to those from an ideal experiment (Ferraro and Miranda 2014). The placebo test exercise in Patrick (2016a) follows a recent trend of requiring sensitivity analysis and falsification tests for experimental studies. We expect this to become standard practice for future economic development policy studies employing these methods.

While experimental methods may generate credible estimates of policy effects, Rust (2014) critiques empirical work that is not grounded in theory because it places limits on the ability to conduct counterfactual policy analysis. Keane (2010) provides an example of this in the labor literature by using an experimental approach to document large improvements in student outcomes from reductions in class size. He argues that this is not a ceteris paribus effect and that because “we don’t have estimates of the structural parameters of the cognitive ability production function, or the decision rules that parents and schools use to determine other inputs, we cannot determine if reduced class size would be a more cost effective way to improve student achievement than, say, higher teacher salaries or better nutrition and health care in utero (p. 49).”

Structural empirical studies set out to model such decision rules and uncover structural parameters in a way that allows researchers to use past changes to predict responses to other not-yet-observed changes (Nevo and Whinston 2010). However, there are still very few structural studies in regional science. A few studies of household sorting and migration have used the structural approach, but there is virtually no economic development policy analysis using a structural approach outside of input-output modeling. However, as Nevo and Whinston point out, structural analysis does not obviate the need for credible inference.

In general, we agree with Rust (2014) that there is an unnecessary divide between structural and experimental work on policy evaluation. Future empirical economic development policy evaluation would benefit from more theory in studies using the experimental approach and more structural work incorporating the identification insights from experimental work. Using theory to generate hypotheses will allow regional scientists to use experimental methods to distinguish between competing theories and uncover the mechanisms driving changes in economic development outcomes resulting from different policies. The studies by Hanson and Rohlin (2011), Rohlin et al. (2014), and Patrick (2016b) represent a step in this direction. Structural models that benefit from the identification from experimental work will be especially useful in further exploring the firm sorting behavior in response to policy interventions.

7.5 Conclusions and Recommendations for the Future

Regional scientists bring a unique set of skills to the analysis of economic development policies, enabling researchers to provide recommendations and guidance for policymakers on how effectively to design policy. However, it is important that researchers and policymakers work together early in the process. It is also important to acknowledge the limitations and possible unintended consequences of policies which were designed with good intentions. Finally, in order for regional scientists to conduct thorough, careful analyses, we must ensure they have the proper tools—especially the right data.

We have discussed how policies can have unintended consequences, for example, attracting one industry at the cost of another or attracting small businesses at the cost of larger establishments. This should not be too surprising since such sorting of businesses is grounded in economic theory. However, more work needs to be done to distinguish the mechanisms of these effects. Furthermore, policymakers need to be cognizant and aware of the possibly negative spillovers of well-intended policy. Land will always be limited to some degree; therefore, attracting businesses of one type will always come at some cost that must be known and understood. At the same time, because of the interaction between policies and the other differences between regions, even policies that are effective in some places will not be effective everywhere.

Policymakers have a lot to learn from regional scientists and other social scientists about the economic development effects of policy. However, in order for researchers to provide a proper analysis, they need quality data. The quality and amount of microdata has exploded as computing power and the Internet have grown, but we must make sure that these data are maintained. The shift from the long form in the decennial census to the ACS may have saved money, but it also may have come at a cost to researchers that relied on quality census data. In addition, all parties need to find ways to ensure that the quality of survey data does not continue to decline. No matter how good and readily available administrative data becomes, we will still have questions that need to be asked to individuals and firms through a survey.

Finally, researchers need to continue to develop quality methodologies to assess the impact of policy on local development. Randomized control trials of economic development policy are unlikely to occur in most situations. However, policymakers and regional scientists may be able to work together in some situations to conduct small scale social experiments. There will always be concerns regarding external validity of such efforts, but that does not mean insights cannot be gained from such endeavors.

Over the last 50 years, we have made great strides in the field of regional science. As we move forward into the next 50 years, one goal of researchers in the field should be to provide more guidance and insights regarding the effect of policy on local economic growth. We discussed in this chapter the current state of this research and what we view are some of the key areas that future analysis should consider. We also stressed the importance of quality data and continuing to improve our methodologies. Hopefully, over the next 50 years, regional scientists will become one of the key sources of information for all levels of government regarding how to design effective policies that will promote local economic development.

References

Angrist JD, Pischke J-S (2010) The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. J Econ Perspect 24(2):3–30

Bartik TJ (1991) Who benefits from state and local economic development policies? Upjohn Press, Kalamazoo, MI

Bartik TJ (1993) Who benefits from local job growth: migrants or the original residents. Reg Stud 27(4):297–311

Bartik TJ (2005) Solving the problems of economic development incentives. Growth Chang 36:139–166. doi:10.1111/j.1468-2257.2005.00272.x

Bartik TJ, Erickcek G, Eisinger P (2003) Economic development policy in Michigan. In: Ballard C, Courant P, Drake D, Fisher R, Gerber E (eds) Michigan at the Millennium. Michigan State University Press, East Lansing, MI, pp 279–297

Belloni A, Chernozhukov V, Hansen C (2014) High-dimensional methods and inference on structural and treatment effects. J Econ Perspect 28(2):29–50

Billings S (2008) Do enterprise zones work? An analysis at the borders. Public Finance Rev 37:68–93

Blundell R, Dias MC (2009) Alternative approaches to evaluation in empirical microeconomics. J Hum Resour 44(3):565–640

Boarnet M, Bogart W (1996) Enterprise zones and employment: evidence from New Jersey. J Urban Econ 40:198–215

Bondonio D, Engberg J (2000) Enterprise zones and local employment: evidence from the states' programs. Reg Sci Urban Econ 30:519–49

Bondonio D, Greenbaum R (2007) Do tax incentives effect local economic growth? What mean impacts miss in the analysis of enterprise zone policies. Reg Sci Urban Econ 37:121–136

Borchers E, Deskins J, Ross A (2016) Can state tax policies be used to grow small and large businesses? Contemp Econ Policy 34(2):312–335

Buss TF (2001) The effect of state tax incentives on economic growth and firm location decisions: an overview of the literature. Econ Dev Q 15 (1): 90–105

Busso M, Gregory J, Kline P (2013) Assessing the incidence and efficiency of a prominent place based policy. Am Econ Rev 103(2):897–947

Chetty R (2012) Time trends in the use of administrative data for empirical research. NBER Summer Institute presentation. Available at the author's website. http://www.rajchetty.com/chettyfiles/admin_data_trends.pdf

Deller SC (2010) Spatial variations in the role of microenterprises in economic growth. Rev Reg Stud 40:71–97

Edelman B (2012) Using internet data for economic research. J Econ Perspect 26(2):189–206

Einav L, Levin JD (2013) The data revolution and economic analysis. NBER Working Paper 19035

Ferraro PJ, Miranda JJ (2014) The performance of non-experimental designs in the evaluation of environmental programs: a design-replication study using a large-scale randomized experiment as a benchmark. J Econ Behav Organ 107:344–365

Freedman M (2015) Place-based programs and the geographic dispersion of employment. Reg Sci Urban Econ 53:1–19

Goetz SJ, Partridge MD, Deller SC, Fleming DA (2010) Evaluating rural entrepreneurship policy. J Reg Anal Policy 40:20–33

Haltiwanger J, Jarmin RS, Miranda J (2013) Who creates jobs? Small versus large versus young. Rev Econ Stat 95(2):347–361

Hanson A (2009) Local employment, poverty, and property value effects of geographically-targeted tax incentives: an instrumental variables approach. Reg Sci Urban Econ 39:721–731

Hanson A, Rohlin S (2011) The effect of location-based tax incentives on establishment location and employment across industry sectors. Public Finance Rev 39(2):195–225

Hanson A, Rohlin S (2013) Do spatially targeted redevelopment programs spillover? Reg Sci Urban Econ 43(1):86–100

Harger KR, Ross A (2016) Do capital tax incentives attract new businesses? Evidence across industries from the new markets tax credit. J Reg Sci. doi:10.1111/jors.12286

Harger KR, Ross A, Stephens HM (2015) What matters more for economic development, the amount of funding or the number of projects funded? Evidence from the community development financial investment fund. WVU Working paper

Holmes TJ (2010) Structural, experimentalist, and descriptive approaches to empirical work in regional economics. J Reg Sci 50(1):5–22

Imbens GW, Wooldridge JM (2009) Recent development in the econometrics of program evaluation. J Econ Lit 47(1):5–86

Keane MP (2010) A structural perspective on the experimentalist school. J Econ Perspect 24(2):47–58

Kline P, Moretti E (2013) Place based policies with unemployment. No. w18758. National Bureau of Economic Research

Kolko J, Neumark D (2010) Do enterprise zones create jobs? Evidence from California's enterprise zone program. J Urban Econ 68:1–19

Krupka D, Noonan D (2009) Empowerment zones, neighborhood change and owner occupied housing. Reg Sci Urban Econ 39:386–396

Leamer EE (2010) Tantalus on the road to asymptopia. J Econ Perspect 24(2):31–46

Meyer BD, Mok WKC, Sullivan JX (2015) Household surveys in crisis. J Econ Perspect 29(4):199–226

Michelacci C, Silva O (2007) Why so many local entrepreneurs? Rev Econ Stat 89(4):615–633

Moretti E (2011) Local labor markets. Handb Labor Econ 4:1237–1313

Neumark D, Simpson H (2014) Place-based policies. No. w20049. National Bureau of Economic Research

Neumark D, Zhang J, Wall B (2005) Employment dynamics and business relocation: new evidence from the National Establishment Time Series. No. w11647. National Bureau of Economic Research

Nevo A, Whinston MD (2010) Taking the dogma out of econometrics: structural modeling and credible inference. J Econ Perspect 24(2):69–82

Papke L (1994) Tax policy and urban development: evidence from the Indiana enterprise zone program. J Public Econ 54:37–49

Partridge MD, Rickman DS (2007) Persistent pockets of extreme American poverty and job growth: is there a placed-based policy role? Agric Resour Econ Rev 32(1): 201–224

Partridge MD, Rickman DS, Ali K, Rose Olfert M (2008) The geographic diversity of U.S. Nonmetropolitan growth dynamics: a geographically weighted regression approach. Land Econ 84(2):241–266

Partridge MD, Rickman DS, Rose Olfert M, Tan Y (2015) When spatial equilibrium fails: is place-based policy second best? Reg Stud 49(8):1–23

Partridge MD, Goetz SJ, Kilkenny MR (2013) Why we need federal statistical data for states and counties. Choices. Available online http://choicesmagazine.org/choices-magazine/submitted-articles/why-we-need-federal-statistical-data-for-states-and-counties

Patrick C (2014) Does increasing available non-tax economic development incentives result in more jobs? Nat Tax J 67(2):351–386

Patrick C (2016a) Identifying the local economic development effects of million dollar facilities. Econ Inq. doi:10.1111/ecin.12339

Patrick CE (2016b) Jobless capital? The role of capital subsidies. Reg Sci Urban Econ. doi:10.1016/j.regsciurbeco.2016.07.009

Roback J (1982) Wages, rent and the quality of life. J Polit Econ 90(6):1257–1278

Rohlin SM, Ross A (2016) Does bankruptcy law affect business turnover? Evidence from new and existing business. Econ Inq 54(1):361–374

Rohlin S, Rosenthal SS, Ross A (2014) Tax avoidance and business location in a state border model. J Urban Econ 83:34–49

Rust J (2014) The limits of inference with theory: a review of Wolpin (2013). J Econ Lit 52(3):820–830

Sharma A, Davidoff K (2015) New rules will improve understanding of costs, benefits of business tax incentives. Pew Charitable Trusts Stateline. Available online http://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2015/10/16/new-rules-will-improve-understanding-of-costs-benefits-of-business-tax-incentives

Shrestha SS, Goetz SJ, Rupasingha A (2007) Proprietorship formations and U.S. Job growth. Rev Reg Stud 37:146–168

Stephens HM, Partridge MD (2011) Do Entrepreneurs enhance economic growth in lagging regions? Growth Chang 42:431–465

Stephens HM, Partridge MD, Faggian A (2013) Innovation, entrepreneurship and economic growth in lagging regions. J Reg Sci 53(5):778–812

Stock JH (2010) The other transformation in econometric practice: robust tools for inference. J Econ Perspect 24(2):83–94

Thomas KP (2011) Investment incentives and the global competition for capital. Palgrave Macmillan, New York

Varian HR (2014) Big data: new tricks for econometrics. J Econ Perspect 28(2):3–28

Carlianne Patrick assistant professor, Department of Economics, Georgia State University (GSU). Her primary research interests are determinants of the spatial distribution of economic activity, economic development policies, agglomeration externalities, and local public good provision. She is a recipient of the Miernyk Medal for early career scholars in regional science, the Andrew Young School of Policy Studies Dean's Early Career Award, Charles M. Tiebout Prize in Regional Science, Barry M. Moriarty Prize, W.E. Upjohn Foundation Early Career Award, and Regional Science Association International Dissertation Award. She received her Ph.D. from The Ohio State University in 2012.

Amanda Ross is an assistant professor in the Department of Economics, Finance, and Legal Studies at the University of Alabama. Her primary research interests are urban economics, public economics, real estate economics, entrepreneurship, and the economics of crime. Previous faculty positions were at West Virginia University, where she was a faculty research associate at the Regional Research Institute and the Center for Free Enterprise. Dr. Ross earned the Ph.D. in economics from Syracuse University in 2011.

Heather Stephens is an assistant professor, Division of Resource Economics and Management, and a faculty research associate, Regional Research Institute at West Virginia University (WVU). Her primary research interests are in regional economic development and the differences in policies across regions, examining questions related to entrepreneurship, taxation, energy, and environmental quality. She previously was an assistant professor in the Department of Economics and the director of the Office of Economic Research at California State University, Long Beach. She also has prior experience in the public and private sectors. Dr. Stephens earned a Ph.D. from The Ohio State University in 2012.

Chapter 8
Regional Science Research and the Practice of Regional Economic Forecasting: Less Is Not More

Dan S. Rickman

8.1 Introduction

Despite the prevalence of regional forecasting in practice, recent research on regional forecasting has been limited. A search through four conference programs (2011–2014) of the North American Regional Science Association Meetings revealed an average of only two papers each conference on forecasting aggregate regional economies. Often, regional economic forecasting is done by those who do not conduct academic research on regional economies. Forecasts also often are based on anecdotal evidence and explanations offered for patterns in key economic indicators often are not supported by the evidence in the regional science literature. Analytic techniques used in regional forecasting typically do not reflect the state of knowledge in the literature.

Accurately forecasting regional economies has always posed challenges. The forces that affect regional economies continually alter recent trends in the data, often in ways that are not well-understood. In addition, de-industrialization, globalization, skill-biased technical change, climate change, and aging of the population are among a number of factors that are presenting new challenges for regional economic forecasting. The modern era of communications also presents new challenges and opportunities in disseminating economic forecasts. Information on regional economic indicators is easily obtained from the internet, reducing the need for forecasters to present this information. What increasingly is needed is research-based information about the regional economy that is forecasted.

In this essay, I argue that more research is needed on regional forecasting methods and that regional forecasting in practice should be better integrated with

D.S. Rickman
Oklahoma State University, 338 Business Building, Stillwater, OK, 74078, USA
e-mail: dan.rickman@okstate.edu

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_8

the regional science literature. In the next section, I discuss six areas where more research is needed on forecast methodology: (1) dependencies between sectors, (2) spatial interdependencies, (3) population growth/change, (4) treatment of labor market(s), (5) role of housing, and (6) employment by occupation. In Sect. 8.3, I also discuss three complementary research areas: (1) benchmarking to other regions, (2) sub-state analysis, and (3) industry trends. The discussion and examples presented draw on my experiences in regional economic forecasting in the U.S., particularly for Oklahoma.

8.2 Research on Forecasting Methods

8.2.1 Dependence Between Sectors

Regional econometric forecasting models typically have followed in the tradition of export base theory in distinguishing sectors that are export-oriented and dependent on external economic conditions from those more dependent on local economic activity (e.g., Bell 1967). Commonly, manufacturing, mining and agriculture are considered as export-oriented. Right-hand-side variables for export-oriented sectors in regional econometric forecasting models then include measures of corresponding national industry quantities and/or prices.

As an energy-producing state, for Oklahoma a primary external economic driver is the mining sector and in my econometric model of Oklahoma, national energy sector measures drive mining employment. Mining employment and/or other energy sector measures are then used as explanatory variables in employment forecast equations for several other sectors: Administrative and Support Services and Waste Management and Remediation Services; Construction; Fabricated Metals Manufacturing; Machinery Manufacturing; Professional, Scientific and Technical Services; Real Estate Services; Retail Trade; State Government; Transportation and Public Utilities; and Wholesale Trade. The equation specifications are based on perceptions of inter-industry linkages and model specification exercises. The mining sector employment multiplier from the model is consistent with those in the regional science literature (Brown 2014; Weinstein 2014; Munasib and Rickman 2015), but is dramatically lower than those reported by others that relied on anecdotal evidence or by arbitrarily assigning sector growth that is caused by mining sector growth. Similar exercises could be performed for regions dependent on manufacturing or agriculture.

Increasingly, however, because of trends such as the shift to services, bifurcation of sectors into export versus local activity has become more complex and is not straightforward for all sectors. While estimated coefficients on national and local variables in a regional sector employment equation will determine the relative sensitivity of the sector forecasts to local versus national economic conditions, collinearity between the right-hand-side variables may make it difficult to separate local from national effects. National growth across sectors may be positively correlated, which can produce positive collinearity between national activity in the sector and other local economic activity. If the collinearity is stable over time, forecast accuracy is not affected, though interpretation is problematic. If the correlation between sectors changes at the national level, such as from a recessionary shock, then the collinearity will not be stable and having precise estimates for national versus local activity will be critical for obtaining accurate regional forecasts.

One approach to obtaining more precise estimates is to use mixed estimation, which combines prior information with regional data. Mixed estimation overcomes problems of collinearity and over parameterization. The most common use of mixed estimation has been in Bayesian Vector Autoregression (BVAR) models (e.g., LeSage and Magura 1991; Partridge and Rickman 1998). Mixed estimation also has been used in regional econometric models in constructing right-hand-side variables for the employment block of equations (Rickman 2002). Mixed estimation methods heavily rely, however, on finding accurate prior information on inter-industry relationships.

A common source of prior information on inter-industry relationships is regional input-output models. A number of studies have reported improved regional forecast accuracy from the use of input-output information (e.g., LeSage and Magura 1991; Partridge and Rickman 1998; Rickman 2002).[1] Input-output models contain well-known implicit assumptions though, and most available regional input-output models are constructed by regionalizing national input-output models based on methods for assessing regional trade patterns.

An additional, relatively unexplored, avenue of research regarding inter-industry specification worth pursuing is the use of a structural vector autoregression (SVAR) model. A reduced-form VAR could be estimated with the regional variable of interest, the corresponding national variable, and the local conditions variable in each equation, including lags. Identification for an SVAR could be obtained by assuming an ordering of innovations. The national variable would be ordered first, which assumes it is the most exogenous variable. The local conditions variable could be second in the order, with the variable of interest third. Impulse response functions then could be used to assess the relative impacts of the national and local conditions variables, and checked against the coefficients of the estimated econometric equation in the model. Energy states, such as Oklahoma, and states dominated by exported agricultural or manufactured goods, naturally fit such an ordering. Generalized impulse responses also could be used to avoid imposing an ordering (Pesaran and Shin 1998) for states without an obvious exogenous external driving force. Impulse response functions could be used to cross-check estimated

[1] Rickman and Miller (2002) also compared the accuracy of using input-output information in mixed estimation against using the information to specify which sectors to use in other sector forecasting equations in Bayesian Model Averaging. Rey (1998) discusses strategies for integrating input-output information into regional econometric models.

multipliers produced from the econometric model, would aid in interpretation of the forecasts, and could be used as prior restrictions in mixed estimation models.

In short, incorporating economic dependencies between sectors may not only increase forecast accuracy but also allows for calculation of employment multipliers. Multipliers help users of forecasts understand the influence of key industries. Other methods for calculating multipliers for policy analysis such as using input-output or computable general equilibrium models depend heavily on assumptions that typically have not been empirically verified for the region of study (Partridge and Rickman 2010).

8.2.2 Spatial Interdependencies

Not only may interdependence exist between sectors within a region, but regions may be economically-interdependent, whereby, forecast accuracy may be improved by accounting for regional interdependencies. Regional differences in industry composition cause national industry shocks to have unequal effects across regions. There also may be regionally-idiosyncratic shocks related to policy or other factors. Because of distance effects in information and relocation, households and firms may weigh economic conditions in nearby regions more heavily than in those further away. Interindustry relationships also may exist between regions, where the costs of distance may cause there to be more inter-industry relationships between regions of close proximity. These considerations cause a regional economy to not only be dependent on the national economy, but also on the economic performance of nearby regions relative to that of the nation. Accurately capturing spatial interdependencies in a regional forecasting model is challenging, but potentially worthwhile.

Granger causality tests can be used to establish whether data from one or more nearby regions are helpful in forecasting the performance of another. But regional economies are dependent on the national economy and interdependence between nearby regions also can be bi-directional. Collinearity between regions can vary over the business cycle, as can the relationship between the region and the nation. Correlation in industry shocks across regions one period may not persist into the forecast period. The challenge is to separate spurious correlations from spatial spillovers.

Prior information on expected spatial spillovers such as distance and relative sizes of the regional economies (Rickman et al. 2009), or spatial contiguity (LeSage and Krivelyova 1999) can be used in formulating priors in BVAR forecasting. Chang and Coulson (2001) implemented an SVAR model to assess whether there were spillover effects between central cities and suburbs of four large metropolitan areas. Identification came from assuming an ordering of shocks among the nation, central city, suburbs and industries of each metropolitan area. Impulse responses from an SVAR could be used in specifying BVAR priors, akin to the approach suggested in inter-industry forecasting above, or at a minimum, in checking the predictions of the regional forecasting model.

8.2.3 The Roles of Supply Versus Demand in Population Growth

Regional population growth consists of three broad components. The first source of growth is natural population growth, which results from the difference between the numbers of births and deaths. The other two components are the balances of individuals moving between countries (immigrants) and between areas within the nation (internal migrants). In the U.S., for periods of time typically of interest in regional forecasting, natural population growth and immigration generally are stable in comparison to internal migration. Based on U.S. Bureau of Census intercensal population estimates from 2001–2009, the ratio of the standard deviation to the absolute value of the mean over the period in the annual change in U.S. state population attributable to internal migration exceeded those for immigration and natural population growth for all states. The median values of the ratios for natural population growth, immigration, and internal migation are 0.09, 0.16 and 0.89, respectively. Thus, internal migration was the most important factor for forecasting short-run fluctuations in state population growth over the period.

Internal migration can either be a response to shifts in labor demand or represent shifts in labor supply. Exogenous shifts in industry labor demand have been shown to be a strong determinant of internal migration in the U.S., though the connection appeared to weaken post-2000 (Partridge et al. 2012). Using a structural vector autoregression approach (Keating 1992), Partridge and Rickman (2006) demonstrate the dominance of labor demand shifts in explaining the variation in internal migration on average across states.

Thus, for many states, and particularly for Oklahoma, it is appropriate to regress population growth on job growth in an econometric model. To be sure, a simple population annual growth equation containing a constant term to capture the stable influence of immigration and natural population growth, and a 1-year lag of Oklahoma employment growth relative to the nation, accurately explains Oklahoma population growth since 1999. Tests of the lag length revealed a strong 1-year lagged relationship. Where employment growth is not dominated by a single exogenous sector, to address concerns with endogeneity, the sectors most identified as exogenous to local conditions could be used as instruments for overall employment growth, or an industry mix measure from the shift-share model could be used as an instrument, in estimating the linkage between population and employment growth.

Population growth will not be as dominantly explained by labor demand shifts in all states or regions and may in many places be explained more by amenity-based migration. Although amenity-based migration in the U.S. has occurred for several decades, a number of factors can cause amenity migration to vary over shorter periods of time. To be sure, Partridge and Rickman (2006) found migration in Sunbelt states to be much better explained by shocks to migration, representing labor supply shifts, than by labor demand shifts. Shocks to internal amenity migration can occur from catastrophic events such as hurricanes, earthquakes, droughts or fires or from reduced quality of life that occurred with previous growth (Gabriel et al. 2003). In this case, it would not be appropriate to consider jobs as labor demand and regress population growth on job growth. Both job growth and population growth are outcomes of shifts in labor supply (Rickman 2010). Some accounting for the changing amenity-attractiveness of the area would need to be reflected in the econometric model. This poses a significant future challenge for forecasting population, considering the expected increased frequency of catastrophic events associated with global climate change. SVAR approaches akin to Partridge and Rickman (2006) could be used to identify the labor supply shocks and assess their role in internal migration in the region.

Both labor demand and amenity-based internal migration also may vary over the business cycle. Saks and Wozniak (2011) found U.S. internal migration to be pro-cyclical, dropping dramatically during the Great Recession. Rickman and Guettabi (2015) found amenity demand dramatically dropping off during the Great Recession, reducing population and employment growth in natural amenity-rich areas. Areas hard hit by manufacturing job losses experienced significant net out-migration. Rickman and Guettabi also found amenity-migration related to the buildup and bursting of regional housing bubbles prior to the Great Recession. Thus, population growth equations should be tested for shifts during different phases of the national business cycle.

8.2.4 Labor Market Closures

Predicted policy impacts and regional forecasts are affected by labor market closures. With few exceptions, regional econometric models are demand driven and ignore potential supply constraints (Beaumont 1990). While this is often argued to be appropriate for short-run analysis, this depends on whether there is excess capacity. Where regional economies face short-term supply constraints that ease over time, a demand-driven model would be more appropriate in the long run. Alternatively, an economy with slack also would be appropriate to forecast with a demand-driven model, but in the short run.

In a BVAR employment forecasting approach for several metropolitan areas in Oklahoma and surrounding states, Rickman et al. (2009) concluded that assuming perfectly elastic supply was more accurate than assuming perfectly inelastic supply. However, in an out-of-sample forecast analysis of employment and wage rates for all fifty states plus Washington D.C. for 1981–1988, Rickman and Treyz (1993) found that specifying imperfectly elastic labor supply was on average more accurate than either of the extreme specifications. Imperfectly mobile labor was captured in the study by an econometrically-estimated net migration equation, in which net migration only fully-responded to labor demand shocks in the very long run.

The current movement of the U.S. economy back towards full employment and ageing of the labor force will limit U.S. national and regional growth for years to come. Use of top-down regional forecasting models will partly transmit the supply constraints to the regional level if they are incorporated into the national models. Yet, if a region has a high composition of industries that are expected to grow faster, demand-side models may project employment growth that implies unprecedented and unattainably high labor force participation rates and/or low unemployment rates. Regional forecasters across the country will then need to incorporate supply constraints to avoid forecasting overly optimistic employment growth.

This might be done ad hoc by specifying a ceiling for the regional labor force participation rate and floor for the regional unemployment rate based on historical precedents. Because of changing demographic and economic structure over time, this is unlikely to be accurate. What is needed is incorporating equations that feed supply constraints back to regional industry employment forecasts. Equations capable of distinguishing labor demand from labor supply in employment growth would be needed; i.e., the problem of identifying demand versus supply in population growth above also applies here. Such equations may be part of CGE, econometric or VAR models (Rickman 2010). Estimating supply constraints may require including a demographic component in a regional forecasting model, where estimates of unemployment and labor force by age and gender would be forecasted.

8.2.5 The Role of Housing

Despite the importance of housing in regional economies, regional housing prices seldom are forecasted. The importance of forecasting housing prices extends beyond housing-related sectors. Housing prices affect the attractiveness of a region to households and can reflect the general health of the regional economy.

A lack of housing can greatly limit regional growth, at least in part through high housing prices. But low housing prices do not necessarily indicate an economic advantage for a region. Lower housing prices in Oklahoma, for example, routinely have been argued to be a harbinger of future strong growth in the state. The traditional spatial equilibrium model, however, predicts lower land and housing prices in locations where there is lower quality of life and firm productivity. More recent formulations of the spatial equilibrium model provide new insights into the role of housing supply in regional growth.

Using a spatial equilibrium growth model, Glaeser and Tobio (2008) argue that a favorable housing regulatory environment underpinned strong growth in the South near the end of the twentieth century, not increased demand for the weather in the region. Strong growth in population and wages relative to housing prices revealed a favorable regulatory environment for housing in the South. In fact, they found rising relative real wages in the South, which is an indicator of declining relative household amenity demand. There also was some evidence of strong productivity growth, revealed by increased population and wages.

Recent trends in housing prices, wages and population (or labor force) then can be used to understand the sources of recent growth and forecasted growth. Returning to the example of Oklahoma, it is believed to have a relatively favorable housing regulatory environment (Gyourko et al. 2008). Using housing prices from the Federal Finance Housing Authority (all transaction indices), Bureau of Economic Analysis nonfarm wage rates and U.S. Census Bureau population estimates for Oklahoma and the U.S., a Glaeser and Tobio (2008) growth decomposition suggests that the change in amenity demand averaged 1% less per year in Oklahoma relative to the nation from 1999–2007, improving to only about 0.5% less from 2010–2014. Productivity growth in Oklahoma slightly exceeded that of the nation over the respective two periods (0.1 and 0.5%). In terms of housing, there was an absence of a price bubble during 1999–2007, and a 1% per year disadvantage over 2010–2014 after taking into account firm productivity and household amenity effects. Overall, Oklahoma slightly out-performed the nation in productivity over the latter period likely because of the energy industry, where housing and quality-of-life were slightly below the national average. So, population growth in Oklahoma will likely fall below that of the nation in the future unless the energy industry is strong; lower housing prices in Oklahoma in general are not indicative of a competitive advantage, but rather they result from lower amenity demand.

This decomposition can be routinely produced in regional econometric models as all needed economic indicators can be forecast, giving forecasters a tool for interpreting the causes of relative regional growth. Explanatory variables for regional housing prices can include mortgage rates, national housing price inflation, and relative regional population growth. Consistent patterns in the decomposition across years reveal the relative strengths and weaknesses of the region.

8.2.6 Forecasting Employment by Occupation

Industry level data are sufficient for regional forecasting purposes to the extent fluctuations in employment are driven by shifts in industry demand. Yet, Partridge et al. (2012) suggest that shifts in regional labor demand post-2000 increasingly became associated with national occupational employment shifts relative to industry employment shifts. The authors surmise that occupations became less tied to certain industries, such as computer programmers that are demanded by many industries. National shifts in occupations also may be supply-driven; e.g., educational funding and immigration policies may affect the supply of labor in high-skilled occupations. Wage rates may be set in the labor market more by occupation and skill; e.g., high-skilled labor is paid relatively more regardless of the industry. Fluctuations in regional industry employment may be influenced by occupational supply constraints. Therefore, accurately forecasting regional employment and wages may require integrating occupational detail.

Regional employment data are routinely provided at the industry level but not at the occupational level. The U.S. Bureau of Labor Statistics (BLS) releases occupational employment and wage statistics once a year based on a semi-annual survey of nonfarm establishments in May and November. Data are available for states, metropolitan areas, and aggregated nonmetropolitan portions of states. Occupation links to industries are only provided at the national level.

One approach to integrating occupational detail is to use national industry-occupation linkages provided by BLS (Treyz et al. 1992). An industry-occupation bridge matrix could be constructed based on national data and used to translate regional employment at the industry level to the occupational level. Changes in BLS regional occupational wage rates could shed light on whether relative excess demand or supply existed across occupations, which suggests whether demand or supply primarily underlies occupational growth and growth in key industries. Not only could this increase forecast accuracy of regional wages and income but it also could help policy makers decide whether to focus on stimulating labor demand versus labor supply to increase regional economic growth.

8.3 Research to Supplement Regional Model Forecasts

8.3.1 Benchmarking Against Other Regions

Because different phases of the national business cycle can have distinct regional patterns, forecasts for a region can be compared to those of other regions. The comparisons can be used both to assess the accuracy of recent employment statistics and forecasts, while also aiding in interpreting the economic trends in the state. The patterns surrounding the recent Great Recession provide an instructive example.

During the 2001–2007 expansion, Rickman and Guettabi (2015) found that employment and population grew faster in high natural amenity states, while these states also experienced housing market bubbles. The natural amenity pattern disappeared during the recession, and only began to re-emerge at the beginning of the current expansion. Farm states also grew faster during the 2001–2007 period, while there were no significant differences for manufacturing and energy states (Rickman 2012). During 2007–2010, energy and farm states grew significantly faster, while manufacturing and high natural amenity states grew slower (Rickman 2012). The 2007–2010 patterns mostly were evident during the 2009–2011 period, though the negative effect for manufacturing disappeared (Rickman 2013).

The patterns above helped me establish the reasons for the relatively stronger performance for Oklahoma during the Great Recession. The importance of natural amenities for employment growth and housing market bubbles elsewhere helped explain Oklahoma's average employment growth and stability in the housing market during 2001–2007 given Oklahoma's slightly below levels of natural amenities. Oklahoma was then spared the ravages of the bursting of regional housing market bubbles, and along with rising energy prices, performed stronger during the Great Recession.

It also is tempting to ascribe differences in state economic growth to state policy differences. Although the regional economics literature is replete with policy studies of regional growth differences, the evidence is often imprecise or inconclusive, giving policy makers little guidance (Bartik 2012). Nevertheless, additional simple analysis can be performed that is informative.

Because of its generally stronger employment and population growth, Texas often is held up as the policy model for Oklahoma to follow. However, in comparing Oklahoma county growth to that of neighboring states for 2000–2007, controlling for industry dependence, natural amenities, the rate of immigration in the previous decade and the degree of urbanization, I found that Texas did not enjoy any growth advantages over Oklahoma (Rickman 2013). To be sure, in comparing counties along the Oklahoma-Texas border, I found lower per capita income growth in Texas, presumably related to its higher rate of immigration. Using the spatial hedonic model, Wang (2016) reports an absence of capitalized policy advantages in Texas.

In short, recent and forecasted regional economic growth should be put in perspective. Is there a pattern across regions that helps explain your region's performance? Are longer-term growth differences easily explained by exogenous factors rather than by policy differences? Not only may answers to these questions improve the accuracy of forecasts by assessing the accuracy of recent data and model specification, getting the story correct may be as, or more, important than producing accurate forecasts.

8.3.2 Sub-state Analysis

Along with state-level forecasts, sub-state areas are routinely forecasted. In Oklahoma, growth typically has been relatively strong in its two largest metropolitan areas, Oklahoma City and Tulsa. Growth in nonmetropolitan areas also may vary because of their relative remoteness. Not only do the differences across areas within the state need to be addressed but so do the differences for the areas in Oklahoma compared to similar areas in other states.

Numerous studies have documented the *ceteris paribus* stronger growth of (all but the largest) U.S. metropolitan areas in recent decades and slower growth in rural areas, where the farther an area was located from larger urban areas the slower has been its growth. Using a spatial equilibrium growth decomposition approach, Partridge et al. (2010) conclude that productivity disadvantages primarily underlie the lower growth of remote areas. *Ceteris paribus* factors, however, include natural amenity attractiveness of the area, presence of natural resources, and location on an interstate highway or waterway.

I have often been asked why metropolitan areas grow faster. Is it that firms prefer to locate in them? Or is because of household location preferences? The literature is wide ranging and inconclusive on the subject.

Partridge et al. (2009) suggest that rather than New Economic Geography agglomeration economies, it is the position of a metropolitan area as a Central Place in the urban hierarchy that makes them attractive. But Central Places can be attractive to both firms and households. Glaeser et al. (2001) argue that large cities offer attractive urban amenities to households, which is confirmed by the empirical evidence of Lee (2010). The results of Moretti (2013) suggest that the advantages are driven by labor demand.

Economic forecasters do not necessarily need to engage in sophisticated academic research to address this issue in their forecast presentations. Forecasters can examine the panoply of publicly available economic indicators for their states and sub-state areas and deduce which theory best fits the pattern. Strong employment growth, lower unemployment, higher labor force participation, positive wage growth and lower poverty suggest that labor demand shifts dominated labor supply shifts (Partridge and Rickman 1999). At the very least, consistent with the discussion above on comparisons to other states, regional economic forecasts should be presented as part of a larger research-based narrative for the region rather than a sequence of forecasts for individual economic indicators.

8.3.3 Industry Trends

Forecasters commonly discuss industry composition in presenting and explaining their forecasts. The location of some industry activities is determined by the presence of resources such as energy or agriculture. Forecasters then commonly become knowledgeable about trends in these industries to forecast and discuss overall economic activity in the state.

For many industries that are not resource-based, national and international competitiveness factor heavily into industry performance in the state. In recent years, the presence of industry clusters has commonly been argued to increase economic competitiveness. Although some industry clusters are long-standing and may be attributable to a single entrepreneur, such as Henry Ford and automobile manufacturing in Detroit, others may be more recent, and believed to be related to government policy such as the emergence of the automotive cluster in the southeastern region of the U.S. (Jacobs 2010). The successes of regions that contain clusters such as in automotive manufacturing or in high-tech sectors have garnered interest among policy makers. I have been asked many times during forecast presentations which industry clusters the state or city should be attempting to develop.

Policy makers often believe that if they use incentives to attract a major firm in an industry, this can evolve into a self-sustaining cluster that propels the state forwards economically (Porter 2000). However, Duranton (2011, p. 45) argues that clusters are a "... complex second-order issue that wrongly receives first-order attention." He questions the benefits of clusters, the ability of policy makers to identify the market failures that require correction for the cluster to develop, and the ability to design policies without being influenced by special interests. Defining and measuring clusters also is problematic. Yu and Jackson (2011, p. 112) identify reasons why regional industrial clusters likely cannot be "successfully identified, mapped and promoted by regional governments". The empirical literature on the economic benefits of clusters appears to support the views of skeptics of regional industry cluster policy.[2]

Therefore, regional forecasters should be careful not to confuse strong growth in a sector from national or other local demand sources with the emergence of a cluster that can be enhanced with government policy. A well-specified block of sector equations can help inform the sources of growth in key sectors. Does the sector grow because of other sectors in the economy? Is there a trend in the size of the sector employment share relative to that of the nation, all else equal? But even if such growth can be identified, the above discussion from the regional science literature cautions against government cluster policies.

8.4 Summary and Conclusions

Following national and international trends, regional economies continue to evolve, posing new challenges to regional forecasters. Regional science researchers have offered insights into the complex workings of modern regional economies. Yet, regional forecasting has not kept pace with these developments. Regional forecasting often is anecdotal, and where models are used, they bear striking resemblance to early models. The quantity of research on regional forecasting models likewise has diminished. I argue in this essay that more regional science research is needed in regional economic forecasting. I identified several areas where more research is needed on forecast methodology and where research supplements regional forecasting efforts.

For one, I suggest greater use of the structural vector autoregression (SVAR) approach (Keating 1992). Current methods for capturing dependencies between sectors suffer from data limitations and the methods are better able establish correlation between sectors rather than causality. This is becoming more important as bifurcation of a regional economy into export versus local activity is becoming more problematic with the shift to a more service-based economy. Likewise, I suggest use of the SVAR approach for capturing spatial interdependencies, as current methods better capture correlation between regions than causal relationships between them. The SVAR also is one of the approaches I suggest for forecasting and understanding both employment and population growth. De-industrialization and the rise of amenity-based migration have raised the importance of labor supply factors in U.S. regional growth. Because population is not labor supply and employment is not labor demand– both are influenced by labor demand and supply–the SVAR method is one approach for disentangling the relative roles of labor demand and supply and increasing forecast accuracy. The SVAR approach also is one of several approaches I suggest should be explored in specifying labor market closures in regional econometric models. In particular, I note the slowing of labor force growth nationally, constraining supply, which may cause demand-based

[2] See Fallah et al. (2014) for a review of the studies and for additional evidence.

regional econometric models to be inaccurate. In addition to the SVAR approach, other structural modeling approaches in macroeconomics also are promising for use in regional forecasting and analysis (Rickman 2010).

I also note that housing has become more prominent in national and regional business cycles and discuss how to include housing in regional econometric models. The possibility of regional employment shifts increasingly arising more from national occupational shifts than national industry employment shifts labor markets is discussed; I suggest how forecasters might overcome limited regional data on occupational employment. Finally, I discuss areas of research to complement regional forecasting to increase knowledge of the forecasted regional economy.

In summary, regional forecasters need to be more knowledgeable of regional forecasting methods and regional science research. Increased integration of regional science research in regional forecasting models will increase both forecast accuracy and foster greater knowledge of the regional economy. The challenges ahead posed by recent economic trends arising from climate change, de-industrialization, globalization and skilled-bias technical change, will necessitate more research, not less, in the practice of regional economic forecasting.

References

Bartik TJ (2012) The future of state and local economic development policy: what research is needed. Growth Chang 43(4):545–562

Beaumont PM (1990) Supply and demand interaction in integrated econometric input-output models. Int Reg Sci Rev 13:167–181

Bell FW (1967) An econometric forecasting model of a region. J Reg Sci 7(2):109–128

Brown JP (2014) Production of natural gas from shale in local economies: a resource curse or blessing? Fed Reserve Bank Kansas City Econ Rev 99(1):5–33

Chang SW, Coulson NE (2001) Sources of sectoral employment fluctuations in central cities and suburbs: evidence from four eastern U.S. cities. J Urban Econ 49(2):199–218

Duranton G (2011) California dreamin': the feeble case for cluster policies. Rev Econ Anal 3(1):3–45

Fallah B, Partridge MD, Rickman DS (2014) Geography and high-tech employment growth in U.S. counties. J Econ Geogr 14(4):683–720

Gabriel SA, Mattey JP, Wascher WL (2003) Compensating differentials and evolution in the quality-of-life among U.S. states. Reg Sci Urban Econ 33:619–643

Glaeser EL, Tobio K (2008) The rise of the Sunbelt. South Econ J 74(3):610–643

Glaeser EL, Kolko J, Saiz A (2001) Consumer city. J Econ Geogr 1:27–50

Gyourko J, Saiz A, Summers A (2008) A new measure of the local regulatory environment for housing markets: the Wharton residential land use regulatory index. Urban Stud 45(3):693–729

Jacobs AJ (2010) Collaborative regionalism and foreign direct investment: the case of the southeast automotive core and the 'new domestics'. Econ Dev Q 26(3):199–219

Keating JW (1992) Structural approaches to vector autoregressions. Fed Reserve Bank St Louis Rev 74(September/October):37–57

Lee S (2010) Ability sorting and consumer city. J Urban Econ 68:20–33

LeSage JP, Krivelyova A (1999) A spatial prior for Bayesian vector autoregressive models. J Reg Sci 39(2):297–317

LeSage JP, Magura M (1991) Using interindustry input-output relations as a Bayesian prior in employment forecasting models. Int J Forecast 7(2):231–238

Moretti E (2013) Real wage inequality. Am Econ J Appl Econ 5(1):65–103

Munasib A, Rickman DS (2015) Regional economic impacts of the shale gas and tight oil boom: a synthetic control analysis. Reg Sci Urban Econ 50:1–17

Partridge MD, Rickman DS (1998) Generalizing the Bayesian vector autoregression approach for regional interindustry employment forecasting. J Bus Econ Stat 16(1):62–72

Partridge MD, Rickman DS (1999) Which comes first, jobs or people? An analysis of the recent stylized facts. Econ Lett 64(1):117–123

Partridge MD, Rickman DS (2006) Fluctuations in aggregate U.S. migration flows and regional labor market flexibility. South Econ J 72(4):958–980

Partridge MD, Rickman DS (2010) CGE modelling for regional economic development analysis. Reg Stud 44(10):1311–1328

Partridge MD, Rickman DS, Olfert MR, Ali K (2009) Agglomeration spillovers and wage and housing cost gradients across the urban hierarchy. J Int Econ 78(1):126–140

Partridge MD, Rickman DS, Olfert MR, Ali K (2010) Recent spatial growth dynamics in wages and housing costs: proximity to urban production externalities and consumer amenities. Reg Sci Urban Econ 40(6):440–452

Partridge MD, Rickman DS, Olfert MR, Ali K (2012) Dwindling U.S. internal migration: evidence of spatial equilibrium or structural shifts in local labor markets? Reg Sci Urban Econ 42(1–2):375–388

Pesaran HH, Shin Y (1998) Generalized impulse response analysis in linear multivariate models. Econ Lett 58(1):17–29

Porter ME (2000) Location, competition, and economic development: local clusters in a global economy. Econ Dev Q 14:15–34

Rey SJ (1998) The performance of alternative integration strategies for combining regional econometric and input-output models. Int Reg Sci Rev 21:1–36

Rickman DS (2002) A Bayesian forecasting approach to constructing regional input-output based employment multipliers. Pap Reg Sci 81(4):483–498

Rickman DS (2010) Modern macroeconomics and regional economic modeling. J Reg Sci 50(1):23–41

Rickman DS (2012) Looking backwards and forwards. http://economy.okstate.edu/caer/files/state-of-oklahoma-j12.pdf

Rickman DS (2013) 2013 economic outlook: National, state and local forecasts. http://economy.okstate.edu/caer/files/Oklahoma-Economic-Forecast-Dan-Rickman.pdf

Rickman DS, Guettabi M (2015) The great recession and nonmetropolitan America. J Reg Sci 55(1):93–112

Rickman DS, Miller SR (2002) An evaluation of alternative strategies for incorporating interindustry relationships into a regional employment forecasting model. Rev Reg Stud 32(1):133–147

Rickman DS, Treyz GI (1993) Alternative labor market closures in a regional model. Growth Chang 24:34–50

Rickman DS, Miller SR, McKenzie R (2009) Spatial and sectoral linkages in regional models: a Bayesian vector autoregression forecast evaluation. Pap Reg Sci 88(1):29–41

Saks RW, Wozniak A (2011) Labor reallocation over the business cycle: new evidence from internal migration. J Labor Econ 29(4):697–739

Treyz GI, Rickman DS, Shao G (1992) The REMI economic-demographic forecasting and simulation model. Int Reg Sci Rev 14(3):221–253

Wang H (2016) The Texas economic model, miracle, or mirage? A spatial hedonic analysis. Ann Reg Sci 56(2):393–417

Weinstein A (2014) Local labor market restructuring in the shale boom. J Reg Anal Policy 44(1):71–92

Yu J, Jackson R (2011) Regional innovation clusters: a critical review. Growth and Chang 42(2):111–124

Dan S. Rickman is Regents Professor, Department of Economics and Legal Studies in Business, Oklahoma State University. He also serves as the director of the Center for Applied Economic Research. His primary research interests are regional economic forecasting and modeling; regional economic development, growth and poverty; and state and local public finance. Previous faculty and visiting positions were at Colorado State University, Georgia Southern University, University of Nevada-Las Vegas and the University of Saskatchewan. Dr. Rickman earned a Ph.D. in economics from the University of Wyoming in 1987.

Chapter 9
Energy for Regional Development

Paulo Henrique de Mello Santana

9.1 Introduction

As civilizations advanced in history, the evolution of conversion and the different utilization of energy were synonymous with progress. In the eighteenth century, the steam engine was a milestone of the Industrial Revolution, moving the world into a new paradigm. Coal was used at this time to generate steam for locomotives, factories and heating buildings. In 1880, coal powered a steam engine attached to the world's first electric generator, launching an electricity industry that would change the energy industry and the welfare of mankind. By the late 1800s, the oil market was developed, and this fuel has been the most important source of energy since the second half of the twentieth century. Technical development of exploration, production, transport and end use equipment was the main driving force of the energy industry. All of these areas were usually market driven until researchers and the government found that energy was an inelastic and essential good. Furthermore, the electricity and the natural gas industry gave rise to natural monopolies in transport and distribution, triggering the emergence of energy regulatory agencies. The characteristics of inelasticity, essentiality, natural monopoly and the time period to build a power facility distinguish energy as a good that is and should be viewed by society quite differently than other goods.

We may say that the field of energy economics research came into its own worldwide after the 1973 oil crisis, since the market couldn't answer some of the emerging questions. Energy price and supply were probably the most important driving forces for regional research on energy economics at this time. However, in the last two decades the environment entered energy research and policy as a major

P.H. de Mello Santana
Engineering, Modelling and Applied Social Sciences Center, ABC Federal University, Sao Paulo, Brazil
e-mail: paulo.santana@ufabc.edu.br

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_9

additional concern. Although some authors question the veracity of anthropogenic influences on climate change, the first part of the Intergovernmental Panel on Climate Change (IPCC's) fifth assessment report says scientists are 95% certain that humans have been the "dominant cause" of global warming since the 1950s. CO_2 emissions from fossil fuel combustion and industrial processes contributed about 78% of the total GHG emission increase from 1970 to 2010, with a similar percentage contribution for the period 2000–2010 (IPCC 2014).

We may say further that regional energy economics was already born in an interdisciplinary environment, since energy economics is a scientific area that includes aspects of economics (microeconomics, macroeconomics, econometrics and environmental economics), engineering, geology, environmental, political and the social sciences.

Section 9.2 describes the existing regional energy research topics considering energy exploration, production, transport and end use. Section 9.3 organizes and classifies some energy policy mechanisms and develops a new approach to help policymakers better design and deploy those mechanisms. Section 9.4 details the most challenging question, cutting edge topics and methods that policymakers will deal with in the next 50 years in this field. Finally, Sect. 9.5 presents a summary and conclusions.

9.2 Energy Research

The literature up until the late 1980s shows that disciplinary issues of energy supply, mainly related to engineering techniques and economics, dominated publications in regional energy economics. Energy economists, geologists, thermal, electrical, chemical and environmental engineers worked to deepen the knowledge in their specific field of energy research. The main driving forces were techno-economic and strategic; the main focus was in the expansion, optimization and relation of variables to primary energy source exploration (mainly fossil fuels), conversion, primary and secondary energy transport and use. Initiatives and research to develop renewable energy were mostly focused on decreasing oil dependence, not environmental aspects.

Because of the environmental movements around the world, new issues in regional energy research emerged, increasing the importance of public policy in energy development. The environmental and social aspects of energy introduced a new horizon on regional energy economics. The interdependence of technical, economical, strategic, social and environmental aspects of energy is changing the way that energy research is being conducted. Specific issues still exist, but now in a broader context, with increased financing from and participation of policymakers. Furthermore, interdisciplinary research is increasing worldwide, trying to connect different aspects of energy into a major objective. It will be increasingly important to design and deploy energy policies that cover all these aspects. Before describing the most challenging questions and cutting edge research issues that policymakers will deal with in the next 50 years, Sect. 9.3 organizes some energy policy mechanisms and develops a new approach to help policymakers better design and deploy those mechanisms.

9.3 Energy Policy Mechanisms: A New Approach for Policymakers

Different policy mechanisms were used by policymakers to promote energy technologies worldwide in recent years, considering both technology-neutral and technology-specific approaches. In the technology-neutral approach, policymakers set up the goal of a policy without picking a specific technology to promote. Cap-and-trade policy is a good example, where governments set up emission limits rather than dictate technologies. Technology-specific approaches pick one or more technologies to promote though specific policy mechanisms. The Wind Production Tax Credit (PTC), a part of the Energy Policy Act of 1992 in the U.S., is an example of a technology specific approach.

A technology-push perspective indicates the key role that science and technology play in developing technological innovations and adapting to the changing characteristics of the industry structure. A demand-pull approach identifies a broader set of market features, including characteristics of the end market (particularly, the users) and the economy as a whole, that affects the performance of innovation (Stefanoa et al. 2012).

The literature also classifies energy policy mechanisms in terms of command-and-control or market incentive policies. Command-and-control is direct regulation in which governments set a mandatory regulation that the market must follow. Although those policies typically have a high impact effect, they are usually costly because of regulatory enforcement and market opposition. Market incentives are policies designed to incentivize the market to develop or provide a good or service. Changes are not mandatory, which makes these policies more easily accepted by the market, but these policies are normally less effective than a command-and-control policy portfolio. Demand-pull policies target increases in private profits, while technology-push policies reduce private costs (Nemet 2009). Command-and-control policies contain mandates, while market incentives are voluntary. Unfortunately, the literature doesn't combine these mechanisms.

Figure 9.1 classifies the mechanisms into four categories, using the four criteria of technology-push, demand-pull, market incentives and command-and-control. This section aims to aggregate these concepts to classify and clarify these policy mechanisms to help policymakers better design and deploy energy policy mechanisms. The first category is called technology-control, where policies are technology-push and command-and-control type because they have the potential to reduce private costs and are mandatory. The second is called market-control policy, where policies are demand-pull and command-and-control type because they

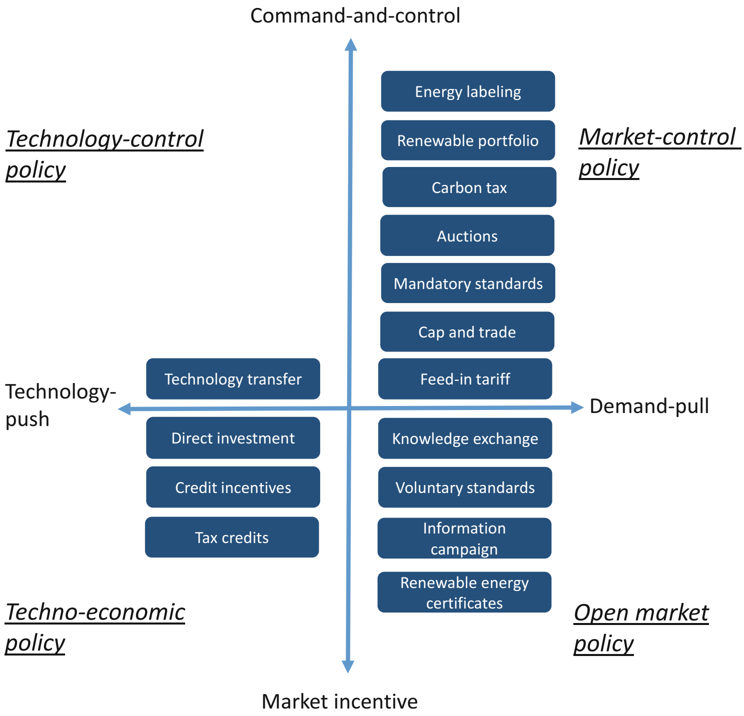


Fig. 9.1 Classification of energy policy mechanisms

have the potential to increase private profits and are mandatory. The third is called open market policies, where policies are demand-pull and market-incentive type because they have the potential to increase private profits and are not mandatory. The last category is called techno-economic, and they are technology-push and market-incentive because they have the potential to reduce private costs and are not mandatory.

The lower-left quadrant of Fig. 9.1 shows the technology-push and market incentive category, called techno-economic policy in this book. Direct government investments, credit incentives and tax credits are the most used policy mechanisms to promote clean energy technologies in this case. These are technology-push policy mechanisms because they reduce private costs, and they also are market incentive driven since they are not mandatory. Direct government investments provide grants to demonstrate or deploy clean energy technologies. Full grants are usually provided to demonstration plants, and cost sharing is used to promote commercial deployment facilities. In the commercial deployment case, the share amount is usually given so the investment is feasible to investors. The effectiveness of this policy mechanism is usually high because it can make a technology economically feasible, depending only on the amount of investment provided. Credit incentives are lower than market interest rate loans that are provided by state banks or other government agencies.

Because initial investment is a large percentage of clean energy technology life cycle costs, credit incentives are very effective. Tax credits are also used to promote clean energy technologies around the world. They are less effective than credit incentives and direct investments, but they are also useful, especially if used together with other policy mechanisms.

The upper-left quadrant of Fig. 9.1 shows the technology-push and command and control category, in which the most important policy mechanism is technology transfer. These are called technology-control policies in this book. These are technology-push policy mechanisms because they reduce private costs since production in other countries may lead to lower costs. They are command-and-control mechanism because they are mandatory. Technology transfer alone normally can't promote the development of an energy technology, but it's useful when combined with other mechanisms.

The upper-right quadrant of Fig. 9.1 shows the demand-pull and command and control category, called market-control policy in this book. Renewable portfolio, carbon tax, auctions, mandatory standards, cap-and-trade, feed-in tariffs and energy labeling are the most used policy mechanisms to promote clean energy technologies in this case. These are demand-pull policy mechanisms because they increase private profits; moreover, they are also command-and-control mechanisms because they are mandatory. These policy mechanisms usually have a high potential to be effective in promoting clean energy technology. They usually create stable market conditions, decreasing risk perception. The renewable portfolio standard is a regulatory mandate to increase production of energy from renewable and/or clean sources at a given rate, percentage or amount. A carbon tax is a form of carbon pricing for emissions that are released into the atmosphere to promote clean energy sources. Auctions are used when a regulatory authority announces that it wishes to install a certain capacity of a given technology or suite of technologies. Project developers then apply to build the project and name the price at which they are willing to develop it. The bidder with the lowest offer is selected and can go ahead with the project. Usually the parties sign a long-term contract (power purchasing agreement). Auctions are frequently used to meet government-set quotas in systems where there is no trading of certificates (International Energy Agency 2011). Auctions are similar to feed-in tariffs, which also offer long-term contracts for energy producers and the costs are shared by customers. Feed-in tariffs are subsidies per MWh generated, paid in the form of guaranteed premium prices, and combined with a purchase obligation by the utilities (Río and Cerda 2014). The main difference between the two is that in auctions the price paid is set by the market depending on the amount of energy traded, and for feed-in tariffs the price is set by government, most commonly based on production costs. Mandatory standards can be effective mechanisms for limiting or promoting specified energy technologies. For example, government might prohibit technologies that emit more than a designated rate, or they might conduct a specific auction to promote a specific technology. Cap-and-trade sets a mandatory cap on emissions in any designated region, which may be a city or even the world. Lastly, energy labeling is the information about energy-efficient equipment, and it is used to inform the customers though mandatory labels.

The lower-right quadrant of Fig. 9.1 shows the demand-pull and market incentive category, called open market policy in this book. Knowledge exchange, voluntary standards, information campaigns and renewable energy certificates (RECs) are the most widely-used energy policy mechanisms in this category. These are demand-pull policy mechanisms because they increase private profits; moreover, they are also market incentive because they are not mandatory. Knowledge exchange, according to the Economic and Social Research Council, is a two-way process where social scientists and individuals or organizations share learning, ideas and experiences. Voluntary standards are non-mandatory standards used to incentivize the market to adopt best energy technologies and practices. Information campaigns are used to inform the market about any existing energy policy mechanism.

A renewable energy certificate (RECs), also known as a green certificate, is a guarantee that one megawatt hour (MWh) of renewable energy was produced and delivered to the power grid. When customers buy RECs, they are promoting renewable electricity generation. REC may be considered the most effective open market policy, with a higher effectiveness than knowledge exchange, voluntary standards and information campaigns. Those policies are the most accepted by the market because they are market incentives that are not mandatory.

The choice of which mechanisms to use depends on the technology chosen. Some mechanisms work better on a regional level while others work better on a national level. Some energy policy mechanisms are also harder to monitor than others. Some may be technology-neutral, technology specific or both. A higher level of regulatory enforcement is required to monitor direct government investments and mandatory standards when these policy mechanisms are set up. Once governments design and deploy mandatory standards like minimum efficiency levels or emission level, they can be costly to monitor. Open market policies don't need any enforcement, but they are usually less effective.

Having established this organization and classification of some energy policy mechanisms to help policymakers better design and deploy these mechanisms, this chapter describes the next 50 years in regional energy economics research.

9.4 The Next Fifty Years in Regional Energy Economics Research

Interdisciplinary studies and the connection among policymakers, market and academia in techno-economic, strategic, social and environmental aspects of energy will likely be the main driving force in the future. Regional research in this field will probably follow this path in an interdisciplinary environment; disciplinary studies will be part of energy research in a broader sense. Figure 9.2 shows a bigger picture of energy, illustrating how policymakers and regulators have to deal with different aspects of energy, the main subjects, and their connections to interdisciplinary research.

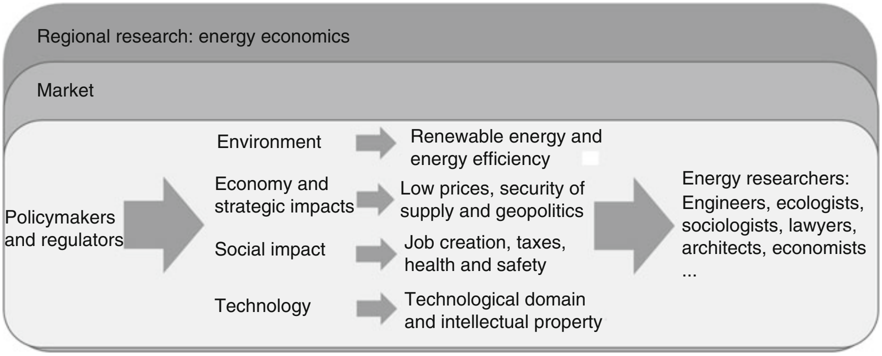


Fig. 9.2 Interdisciplinary research in energy: market and policymaker environment

Despite recent efforts to mitigate greenhouse gas (GHG) emissions worldwide, annual GHG emissions have grown between 2000 and 2010. During this period, total anthropogenic GHG emissions were the largest in the history of mankind. Global investments in the energy sector are key to sustainable development in a context of a low carbon economy. Energy efficiency, renewable energy sources, carbon capture and storage, moreover, decentralized energy generation researches are considered crucial to the success of environmental strategies worldwide.

Regional researchers, policymakers and regulators have questions about environmental, economic, strategic, social and technological impacts of energy. Most of these questions are interdisciplinary and interrelated.

Energy resources, exploration, production, conversion, transportation and end-use are the mainstream of the energy chain. Security of supply and low energy prices are usually conflicting variables, since oversupply or capacity surplus is often necessary to avoid shortages. Short term energy supply is also a source of preoccupation for researchers, and the objective of secure energy is often in conflict with the goal of providing cheap energy.

Economic and strategic variables, in both the short and long term, are also important in regional economics. Low energy prices and security of supply are usually conflicting variables in the energy industry, since an oversupply is usually associated with higher energy costs. In the long term, energy geopolitics and indicative energy planning are fields of interest of researchers and policymakers for the next 50 years.

Social impacts regarding job creation, taxes, health and safety are among the research fields that regional energy economists will work on. Technological advances and its interrelation with basic and applied R&D, commercial deployment and market diffusion are also an important field of research, together with intellectual property issues. The interrelationships among environmental, technological, social, economic and strategic variables will be the main driving forces for energy research and policymakers in the next 50 years.

This chapter shows the likely direction of energy economics research in the next 50 years. Some of these questions and research methods are highlighted below. While it would be impossible to identify all relevant questions and methods, we try to show the direction of regional energy economics research through some important problems that all involved agents will have to deal with.

Problem-driven questions have an increasingly importance in the last two decades. It doesn't mean that method-driven in energy is less important. However, method-driven research in energy economics will likely be inserted into problem-driven questions in a broader sense. An overarching general question might be, "How can safe energy expansion at a low cost, with security of supply, minimal environmental impact and maximum social welfare in the short and long term be promoted?" Questions like this, which span a wide range of qualitative and quantitative dimensions, can be approached using methods like multiple-criteria decision analysis to achieve robust solutions. Questions of a more uni-dimensional character would include:

1. How to promote sustainable energy in a given region?
2. How to promote renewable energy with low cost and security of supply?
3. How to develop competition and security of supply?
4. How to promote energy efficiency?
5. How to promote technological innovation and diffusion of emerging clean energy technologies?
6. How to project energy demand and price in a given region?
7. How to mitigate GHG emissions in energy systems?

Some of the methods that likely will be used to address these more specific questions are listed below. All of these methods might be tied to the organization and classification of energy policy analysis detailed in Sect. 9.3. Policies in the short and long term face different challenges, and all of them deal with techno-economic, strategic, social and environmental aspects of energy.

1. Engineering: technical development of renewables, energy efficiency, carbon capture, energy storage, decentralized generation technologies and smart grids;
2. Architecture: sustainable houses and cities;
3. Life cycle assessment;
4. Public policy: Cost-benefit analysis;
5. Multi-objective and multi-model optimization methods;
6. Qualitative methods: surveys, Delphi and case studies;
7. Multiple-criteria decision analysis: multi-criteria evaluation and design, including spatial multi-criteria;
8. Forecasting techniques in both short and long term: time series, econometrics, spatial econometrics and end use modeling.

Promoting investments in energy efficiency, renewables sources, carbon capture and carbon storage, and decentralized energy generation are considered crucial to the success of international negotiations on climate change mitigation. Sustainable houses and cities, life cycle assessment and public policy cost-benefit analysis researches will help society to achieve this objective. Methods like multi-objective, multi-model optimization, quantitative methods, multi-criteria, spatial multi-criteria and forecasting techniques will help researchers and policymakers to deal with conflicting energy variables.

The eight methods listed above can be used effectively to deal with the future problem-driven and method-driven questions listed in this section. Promoting safe energy expansion at a low cost, with security of supply, while minimizing environmental impact and maximizing social welfare in the short and long term will be a challenging issue for researchers, regulators and policymakers. The methods that might answer these questions lie within a continuum from qualitative to quantitative.

9.5 Conclusion

The promotion of energy for sustainable development is a challenge for governments and regulators. Regional energy economics will have to deal with techno-economic, strategic, social and environmental aspects of energy production and consumption to achieve interrelated objectives. Society wants cheap, clean, safe and secure energy. Questions emerge because those objectives sometimes conflict. In the next 50 years, regional energy economics research will have to deal with these challenging issues, identifying the nature and extent of the tradeoffs involved. Energy research will likely become increasingly interdisciplinary, addressing interrelated disciplinary research and problem domains to achieve a hybrid solution in a broader sense. It will be increasingly important for policymakers to design and deploy energy policies that cover multiple objectives. The connection among policymakers, market and academia in techno-economic, strategic, social and environmental aspects of energy will be the main driving force for energy research in the future.

References

IEA (2011) Deploying renewables 2011: best and future policy practice. OECD/IEA, Paris

IPCC (2014) Climate change 2014: mitigation of climate change. Intergovernmental Panel of Climate Change, Geneve

Nemet GF (2009) Demand-pull, technology-push and government-led incentives for non-incremental technical change. Res Policy 38(5):700–709

Río PD, Cerda E (2014) The policy implications of the different interpretations of the cost-effectiveness of renewable electricity support. Energy Policy 60:364–372

Stefanoa GD, Gambardellab A, Veronab G (2012) Technology push and demand pull perspectives in innovation studies: current findings and future research directions. Res Policy 41(8):1283–1295

Paulo Henrique de Mello Santana is professor, Engineering, Modeling and Applied Social Sciences Center, ABC Federal University, Sao Paulo, Brazil. His primary research interests are energy policy; climate change, energy efficiency, energy technologies and energy market. Previous faculty positions were at the Interdisciplinary Center of Energy Planning (NIPE)—Unicamp, and 1 year as research scholar at the West Virginia University (WVU). Prof. Santana earned the Ph.D. in Energy Planning from the State University of Campinas (UNICAMP) in 2009.

Chapter 10
Regional Perspectives on Public Health

Sara McLafferty and Alan T. Murray

10.1 Introduction

The impacts of local and neighborhood contexts on population health have attracted enormous research attention in the past two decades. Researchers are investigating how diverse place characteristics, such as environmental pollution; residential segregation; and access to recreational, health, and social services and employment opportunities, affect people's health and well-being, as well as how health in turn shapes people's lives and livelihoods. Much of the focus is on the local scale, while the regional scale—the primary focus of regional scientists—has attracted less attention. At the same time, regional science work has not generally addressed health and health care, despite the importance of these issues to regional economies, populations, and economic and social development processes. Health is crucial for sustaining a productive and innovative workforce, providing a key cornerstone for productive and reproductive activities that are embedded within regional economies. Regional development processes affect population health through complex linkages to incomes, occupations, infrastructure developments, and myriad other socioeconomic and contextual factors. In addition, medical services and other health-related industries have emerged in the past several decades as critical components of local economies, providing employment, income, innovation, and other outputs that ripple through local and regional economies. In the U.S., health care accounts for 17% of GDP, making it one of the largest economic sectors.

S. McLafferty (✉)
Department of Geography & GIScience, University of Illinois at Urbana-Champaign, Champaign, IL, USA
e-mail: smclaff@illinois.edu

A.T. Murray
Department of Geography, University of California at Santa Barbara, Santa Barbara, CA, USA
e-mail: amurray@ucsb.edu

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_10

For these and other reasons, we see much potential for regional scientists to increase attention on health and health care as topics of study and to investigate health-related patterns and processes, especially at the regional scale. Although there is no hard and fast boundary between local and regional scales, a regional focus—one that fits well within regional science—involves understanding and modeling population-level processes that operate within interconnected regional settings. Regional investigations emphasize the economic development and demographic processes that drive changes in health and health care within and among regions. Planning of public health interventions to improve population well-being takes place at the regional and local scales, providing opportunities to evaluate and model access to services and to design more effective service delivery systems.

This chapter identifies key themes in regional science research related to public health. We discuss the importance of regional science perspectives and methods in seven main areas of public health research: inequalities; determinants; development processes; mobility; detection, surveillance and monitoring; provision and access; and location modeling. Opportunities for closer collaboration between health researchers and regional scientists are identified. Because the literatures on these topics are vast, we cite only a small subset of relevant literature to give an overall sense of research contributions and directions.

Our chapter adopts a socio-ecological definition of health. We view health as a positive quality that encompasses social, physical, and emotional health and not merely the absence of disease. Health care comprises diverse activities—personal, social, and institutional—for maintaining and promoting health. Our discussion of health care focuses primarily on formal health care—health care provided by trained professionals in organized settings such as hospitals and clinics—although it is important to acknowledge that much health care is provided in informal settings.

10.2 Inequalities

Health outcomes vary widely among places and population groups. These *health inequalities* are evident at all spatial scales, from global to regional to local. At the global scale, we observe large and persistent differences in the most basic indicators of life and death, including infant mortality and life expectancy, as well as sharp disparities in the incidence and prevalence of specific health concerns, such as malaria, HIV/AIDS, diabetes, violence, various cancers, and many others. Global health inequalities are strongly linked to differences in economic and social development, with low-income countries typically having higher rates of morbidity and mortality than more economically advanced countries. However, notable exceptions exist, such as Cuba and Costa Rica, where investments in sanitation, water supply, nutrition, and public health services result in very positive population health indicators despite low or moderate levels of economic development.

The epidemiological transition offers a conceptual framework for understanding the relationships between health disparities and economic development and modernization at the global scale. This framework characterizes low-income countries as having high mortality rates due to undernutrition, poor sanitation, and high rates of communicable diseases, such as malaria, tuberculosis, and diarrheal diseases. As economic development takes place, nutrition and food supply improve, and investments in public health-related infrastructure and medical services take place, resulting in a decline in mortality and a life expectancy increase. During this transition, the causes of death shift from malnutrition and communicable diseases to degenerative diseases, such as heart disease, cancer, and stroke. While the growing importance of degenerative diseases is partly linked to increases in life expectancy that give time for degenerative conditions to develop, such diseases also emerge in economically advanced places due to increases in health-damaging behaviors, like smoking and substance abuse, and increased exposure to environmental and occupational hazards that are by-products of industrialization. As development and modernization continue, incidence of certain degenerative diseases declines as wealthy countries invest in environmental regulations and public health interventions and as polluting industries move to low-wage countries (Martens 2002).

Recent evidence suggests that health transitions reflect distinctive contexts and causal mechanisms. Non-communicable diseases, like diabetes, are increasing across the globe, even in places where economic development is slow and among populations of low socioeconomic status. Long-term declines in mortality that occur as economies develop may also “reverse” in some cases, as economic shocks and downturns lead to rising morbidity, mortality, and/or infectious disease outbreaks. Such a reversal took place during the Eastern European health crisis of the 1990s, when life expectancy fell in countries like Russia and Ukraine as they switched abruptly to a market economy. Despite these complexities, the health transition framework is important for emphasizing the close ties of population health to social and economic development processes. Regional scientists’ knowledge of such processes can help to improve our understanding of important national and global health transitions.

Wide health inequalities also exist at the regional and local scales. For the U.S., regional and county-scale disparities in life expectancy approach those observed between developed and developing countries. Geographic disparities in life expectancy in the U.S. are especially pronounced for low-income individuals (Chetty et al. 2016). Other countries, like Canada, Italy, and the United Kingdom, also show large regional variations in indicators of population health (e.g., Ferrara and Nisticò 2015). In some instances, regional variations in health follow urban-rural divides, or align with north-south gradients in regional economic development as is the case in Italy.

Uneven geographic variation in health outcomes has also been widely documented at the local scale where detailed data reveal patterns of health (dis)advantage that reflect population and place characteristics (Krieger et al. 2005). In Chicago, low birthweight, the percentage of infants born weighing less than 2500 g, is four times higher in some neighborhoods than in others—a gap that is as wide as that between corresponding percentages for Sweden and Bangladesh. Diseases like HIV/AIDS, cancer, and asthma also vary greatly across urban metropolitan areas, with the highest rates often occurring in areas of socioeconomic disadvantage. In the 1980s and 1990s, areas of health disadvantage were typically located in inner city neighborhoods close to the city center; however, urban development processes like gentrification and the closing of public housing projects are shifting vulnerable populations to peripheral urban and suburban locations where new concentrations of health disadvantage are emerging.

10.3 Determinants

Determinants of regional and local health inequalities can be divided into two broad categories—composition and context (Macintyre et al. 2002). Composition emphasizes characteristics of people within an area that affect their risk of ill-health, such as age, gender, and race/ethnicity. Context describes characteristics of the places in which people live and interact that affect their health risks and health-related behaviors. Contexts exist at multiple and overlapping scales, from the personal spaces of daily life to the broader regional and global contexts that shape and constrain these everyday spaces. Contextual factors include environmental hazards; parks and green spaces; transportation, retail, employment, and health care opportunities; opportunities for social interaction and social capital; residential segregation and housing; national, state and local policies; and many others. The contextual and compositional factors affecting health inequalities vary greatly depending on health outcome and study area of interest.

Obesity represents an illustrative example of how composition and context influence health outcomes. Age, gender, and genetic factors are known to affect obesity risk, so areas in which population composition is skewed towards high risk groups tend to have higher rates of obesity. However, research shows that geographic variation in obesity goes well beyond the effects of population composition. People's food intake levels have been linked to the availability and prices of healthy and unhealthy foods as reflected in the local mix of food retail outlets and people's knowledge of and willingness and ability to access them. Food deserts—places in which people lack access to affordable and healthy food—present a particularly constrained environment for healthy eating (Walker et al. 2010). In addition to food intake, obesity is influenced by physical activity, which is rooted in environmental contexts. Research shows that diverse contextual factors affect physical activity, including urban sprawl, neighborhood street patterns, ease and safety of walking, transportation access, crime and safety, and access to parks and green spaces.

Selection bias results from people actively selecting particular kinds of residential contexts, and is a critical issue in studies of contextual effects on health. We cannot be sure if an observed association between neighborhood context and health represents a causal effect or a preference that motivates people to choose to live in that neighborhood context. For example, do people engage in physical activity because they live near a park, or do they choose to live near a park because they enjoy engaging in physical activity? The gold standard for establishing causal effects is the controlled experiment, but it is nearly impossible to conduct controlled experiments to analyze relationships between contextual variables and health. In addition, the segregated residential landscapes that result from residential sorting processes mean that some population groups do not experience a diverse spectrum of contextual effects (Messer et al. 2010). One way to address these concerns is through natural experiments associated with health policy changes—for example, analyzing the effects on health behaviors of changes in smoking regulations, marijuana legalization, or construction of a new supermarket in a food desert.

While acknowledging that context and composition are both important, contemporary work emphasizes interdependences between them (Cummins et al. 2007). Contexts affect people's health-related behaviors and experiences, but people and their social and political institutions create, choose, and change place contexts. The issue of selection bias illustrates well the mutual interactions between people and contexts. Understanding how these interactions unfold and their implications for population health calls for dynamic approaches that consider feedbacks and changing trajectories. The dynamic approaches of regional science have a lot to offer for analyzing how processes like regional and urban economic development and migration relate to population health.

10.4 Development Processes

Processes associated with development are related to health in complex ways through effects on incomes, employment, occupations, prices, land uses, and many others. The health effects of such economic variates have been widely studied and generally show that people who have low incomes, are un- or under-employed, and/or who work in risky occupations have poorer health outcomes than those who are more affluent. However, fewer studies have drilled down to the regional scale, where place-specific combinations of economic opportunities, population groups, health and social services, infrastructure, and other dimensions of regional context result in varying health outcomes. Associations that are strong at the national scale, like the inverse association between health outcomes and income inequality, sometimes disappear when data are analyzed at the regional and urban scales.

Regional development processes influence both population composition and contextual environments in ways that are relevant for health. The types of people who live in an area, and their patterns of residential segregation, are profoundly shaped by regional economic and political trends. Even more complex are the effects of development processes on health-related contexts. Regional economic growth leads to investments in water, sanitation, and transportation infrastructure, education, health, and social services, and retail activities that can have both beneficial and adverse effects on health. Growth can also lead to increased air and noise pollution, and long commuting times due to traffic congestion, all of which have been associated with stress and other adverse health effects. Changes in urban form, such as sprawl, segregation, and gentrification, unevenly impact opportunities to engage in healthy (and unhealthy) behaviors and social interactions. These health-related multiplier effects depend greatly on the mix of industries in a region and their interindustry linkages and workforce and input requirements, interconnected economic systems that regional scientists understand well.

Health impacts are also important as intended or unintended consequences of regional development processes. From the outbreak of Schistosomiasis that followed construction of the Aswan Dam to the increased spread of Lyme disease in the northeast U.S. linked to patchy suburban development in second-growth forests, examples abound of development projects heightening risk of communicable diseases (Mayer 2000). Emergence and spread of non-communicable diseases can also be traced to contextual changes associated with development processes, although the pathways are highly complex. For regional science, an opportunity exists to extend input-output modeling frameworks to estimate health impacts using methods similar to those employed in modeling environmental impacts such as the single- and multi-region input-output models used in estimating changes in CO_2 emissions.

Health also feeds into regional and urban development processes, generating economic and social benefits that lead to wider societal gains. Increased productivity, income, employment, education, and social capital are among the many benefits of improvements in health. There is great interest in quantifying and forecasting these benefits to better represent the benefits of public health investments and interventions. A recent study estimated that in developing countries, spending $5 more per person for woman and child would generate approximately nine times that value in social and economic benefits (Stenberg et al. 2014). Such benefits are likely to vary regionally depending on regional context and population. With their detailed knowledge of interconnected population and economic processes, regional scientists can provide methods and perspectives needed to estimate benefits of investments in public health infrastructure and services within specific regional and local settings.

10.5 Mobility

Time-space contexts reflect people's movements from place to place, and are critically important for representing the environments people experience daily and over the life course. Health-related exposures and opportunities occur within these ever-changing spaces. Kwan (2012) highlights the fact that contexts shift over time and space, reflecting individual choices and constraints within particular settings. As a result, using fixed geographic units like census tracts or counties to represent contexts may lead to inaccurate and biased exposure estimates. To address this issue, researchers have used commuting flow data to devise contextual measures based on the spaces people traverse in their daily routines. A growing number of studies use time-space activity patterns for examining contextual influences on health. Beyond effects on contextual exposures, daily mobility itself can influence health: people with restricted activity spaces may experience social isolation, while those with expansive activity spaces experience stress and time constraints. These health impacts have not been widely studied, although they are becoming increasingly relevant in our hyper-mobile societies.

Mobility also occurs at longer time scales through migration processes that induce changes in population composition and expose people to new environments. Migrants are often healthier than populations left behind, so migration can exacerbate health inequalities if flows among regions are uneven. Many regions in which poor health outcomes persist are also regions of strong net out-migration. The impacts of migration on individual migrant health have been widely studied in relation to processes of settlement and acculturation. Migration imposes a significant social and economic burden on migrants, severing social ties and connections to places. Forced migrants face huge challenges that may lead to diminished social and physical health. Researchers have uncovered mixed support for the "healthy migrant" thesis as immigrant health experiences vary from one group to another and evolve over time as migrants cope with and adjust to new environments (Newbold 2005). Neighborhood contexts play a role in migrant health by providing opportunities for social interaction and access to jobs, resources, and services. For some migrants, living in an ethnic enclave with other members of one's cultural or ethnic group appears to have protective effects on certain health outcomes, although for other migrants, living in a dense ethnic concentration may be detrimental (McLafferty et al. 2012). The interactions between immigrant labor markets; settlement patterns; and access to social, health, and cultural resources are important topics where regional scientists' understanding of social and economic interlinkages and space-time dynamics is important.

10.6 Detection, Surveillance and Monitoring

The significance of public health is undeniable. Consider, for example, human immunodeficiency virus (HIV), whether in a small rural town in Indiana or among working age adults of sub-Saharan Africa or the growing number of children across the globe living with the virus, HIV impacts all of us in one way or another. As communicable disease like HIV spread across the globe, and non-communicable diseases grow and decline unevenly in different places, surveillance and monitoring efforts to detect and/or track these public health concerns are paramount. Detection and surveillance requires an ability to identify issues at local, regional and global scales, to monitor where they are occurring, and to predict when and where the issues will emerge in the future. Spatio-temporal data are plentiful for supporting public health surveillance efforts. Many spatial analytical methods have been developed to aid in the detection of clusters, identification of hot spots and understanding patterns.

Much progress has been made in the development of methods to support health detection, surveillance and monitoring, although not necessarily in regional science. A range of supporting software, tools and technologies have been and could be brought to bear on questions concerning public health (Cromley and McLafferty 2011; Yao and Murray 2014). Sensing technologies. including GIS (geographic information systems), GPS (global positioning system), and remote sensing are widely available, facilitating spatial data creation, mapping and geovisualization, and analysis. GIS is a collection of software, hardware and procedures that enable spatial information to be created, managed, manipulated, displayed and analyzed (Church and Murray 2009; Cromley and McLafferty 2011). This may involve digitization and geocoding of incident locations but also various sorts of manipulation, spatial and aspatial. Beyond this, GIS provides the capacity for mapping and geovisualization of geographic data.

Geospatial technologies can and could be helpful for public health policy development, planning response and assessment. GPS is the constellation of satellites orbiting the earth along with base stations and individual receivers, all combining to enable determination of location (and time) essentially anywhere on the surface of the earth. Of significance is that most cellular phones have GPS, so movement, behavior and interactions can be tracked directly or indirectly for individuals or groups of people. Because people often carry diseases with them as they move from place to place, these movement data can be used to anticipate where diseases are most likely to spread. There are also many different remote sensing platforms, including those tracking shopping purchases (and where), cameras mounted on drones, sensors on aircraft, and satellites recording observed conditions from space, facilitating observation as well. What this means is that we can detect from afar that countries like Zimbabwe are and will be experiencing food shortages due to local/regional drought conditions, as an example.

These technologies are currently being used for health surveillance to record and map outbreaks of disease globally and regionally. Systems like HealthMap collect data from news reports and public health agencies and map the information as reports come in. In fact, early reports of the 2009 H1N1 (Swine Flu) epidemic appeared in HealthMap's reporting system (McLafferty 2010). (H1N1 is an Influenza A virus subtype that spread globally in 2009 resulting in hospitalizations and human deaths.) Other geographic health surveillance systems process data from web searches, drug store sales, and social media to reveal when and where outbreaks of diseases like flu are taking place.

There are a range of spatial analytical methods that support detection, surveillance and monitoring. This may involve local and global measures of spatial autocorrelation to detect hot/cold spots, or a host of approaches for identifying clusters and patterns (Murray et al. 2014). Oftentimes, one is interested in areas that exhibit concentrated levels of disease, either high or low. On the spatial autocorrelation side, Moran's I and G* are popular approaches. Exploring spatial pattern may involve using k functions. Cluster detection methods include hierarchical and non-hierarchical approaches, such as kernel density, spatial scan, and other explicitly spatial optimization methods. A range of confirmatory methods have been utilized in public health analysis, including regression (e.g., OLS, spatial autoregressive, spatial error, geographically weighted regression, logistic regression, etc.) as well as kriging, multi-level and Bayesian hierarchical models, among others. In the end, the interests may varying depending upon context, but research efforts have certainly focused on monitoring change over time, identifying disease outbreaks, and predictive modeling of disease spread. Regional scientists have made valuable contributions to spatial analytics methods development and refinement, and clearly there is opportunity for this to continue in addressing public health issues.

10.7 Provision and Access

Health care provision and the availability of doctors, nurses, monitoring mechanisms, testing capabilities, diagnostic and preventive services have proven to be vitally important for keeping people healthy. To be effective, services must be available, provide high quality and appropriate care, and be utilized by people in need. It is not always the case, unfortunately, that people and communities have access to appropriate and high quality health services. As a result, we have and will continue to observe much regional variability in health and well-being. For example, the Appalachian region of the U.S. has a high prevalence of diabetes attributable to obesity due to sedentary lifestyle, diet, etc. This coupled with, poverty, low educational attainment and a general lack of medical services has led to a health situation of great concern (Grubesic et al. 2014).

The implications are clear. Access to health care services is vital. Although many factors influence access, both the local supply of health services and their proximity to the population in need are critical (Wang 2012). The types, quality, and costs of health services vary spatially, and as a result, impact individual access. Proximity is highly correlated with health services utilization (Mayhew and Leonardi 1982) and is also important for saving lives (Cudnik et al. 2012). Measures and metrics associated with spatial access abound. Discussion and reviews can be found in Cromley and McLafferty (2011) and Wang (2012). Perhaps the most widespread access measures are average distance, maximum distance and distance decay (see Bennett et al. 1982; Rushton 1984; Kumar 2004; Murray and Grubesic 2016). Average distance is derived through the measurement of proximity between an individual (home residence) and their closest healthcare service center/facility. It is used as a summary indicator of the level of potential service an area or region has. An alternative is the maximum distance (or travel time) standard—a maximum distance beyond which access is assumed to be poor. Using GIS, we can determine the percent of population within this standard, or alternatively, who is not within the standard. Other spatial access measures, like the 2-step floating catchment method, capture the supply of health services relative to demand and can incorporate distance decay (a drop-off in availability with distance), differences in transportation access, and other complicating factors (Wang 2012). Gravity models have also been used to analyze health care access and to predict changes in service use after service facilities close or are added.

In the U.S., the Dartmouth Atlas of Health Care project documents wide regional disparities in utilization of specific medical procedures and in the costs of those procedures, even after factors affecting patients' need for services are controlled (Wennberg et al. 2002). Some of these differences stem from practice variations—local and regional differences in health provider behavior and norms. Regional differences in health services also point to the significance of health care to local and regional economies. In many areas, health care providers and related industries account for a large share of employment and income, making them central to economic development. Regional science can contribute to better understanding the economic significance of health care industries and the impacts on quality, utilization and costs.

10.8 Location Modeling

Policy formulation, evaluation and assessment having to do with public health is extensive, but not without varying levels of controversy. Of course, there is less controversy when policy is supported by rigorous methods in order to gain insights. To this end, a variety of location model (spatial optimization) based approaches have been utilized to evaluate existing healthcare systems, plan for new service networks, and/or extend an existing service system in some manner. Discussion and reviews of modeling in this area can be found in Rushton (1984), Cromley and McLafferty (2011), Daskin and Dean (2004) and Murray and Grubesic (2016) (see Church and Murray 2009 for broad classes of location models). Effectively, different location modeling approaches rely on the different measures of access and accessibility, including average distance or travel time from population to facilities (see Oppong and Hodgson 1994) or the percent of population within a maximum distance standard (see Bennett et al. 1982). These kinds of locational objectives have been used to identify optimal locations for new healthcare providers, to identify health facilities that are no longer needed, and to identify service assignments/areas. The models also may be used to evaluate the spatial efficiency of an existing system, as suggested in Rushton (1984), Kumar (2004) and Yao and Murray (2014).

Researchers have proposed more complex location models to emphasize the likely health outcomes associated with specific numbers and locations of health facilities. These models reflect the fact that patient outcomes are typically better in health facilities that treat more patients, known as the patient volume effect, which favors large, centralized facilities. This advantage can be traded off with long travel times associated with a centralized system (McLafferty and Broe 1990). Locating health facilities often involves multiple objectives, which can be expressed using location modeling. While many examples of modeling along these lines can be found, worth noting is the work of Osleeb and McLafferty (1992) and more recent work by Cocking et al. (2012) and Leira et al. (2012) in support of public health decision making. A final point is that health services are hierarchically organized and interconnected. Murray and Grubesic (2016) review the hierarchical and non-hierarchical contexts of healthcare service systems, where various types of referral mechanisms and relationships are observed. Regional scientists have been particularly active in this area of public health, and there remains much opportunity for continued efforts and contributions.

10.9 Conclusion

Health and health care are pressing social concerns that cut across the theoretical frameworks and methods developed and used by regional scientists. Health is deeply embedded in regional economic and social systems: healthy populations are crucial for regional growth and sustainability; regional development processes impact population health both directly and indirectly through effects on contextual and compositional health determinants; and health services have an increasing role in regional economies, affecting not only access to care, but also local employment, workforce requirements, and incomes. Regional scientists are uniquely poised to address these important issues.

This chapter has provided an overview of themes in health research while pointing to opportunities for closer collaboration between health researchers and regional scientists. Health outcomes vary greatly at the regional and local scales, and the methods and perspectives of regional sciences can be employed to better understand such variation. Regional economic and demographic processes are central to health outcomes, affecting factors like incomes, poverty, occupations, and population demographics that are well-established health determinants. Modelling how these processes affect population health, and how health and health services, in turn, affect development processes represents an important area for future investigation. Methods used by regional scientists can also be brought to bear on critical issues that underpin efforts to improve public health, including health detection, surveillance and monitoring; evaluating spatial access to health services; and improving the locations and deployment of services to populations in need. Spatial and space-time "big data" from sensing systems and GPS-enabled devices will be central to these efforts. Multidisciplinary approaches, in which regional scientists collaborate with health researchers, are also important, given the complexity and multifactorial nature of health issues. Looking ahead, improving health and health care access require a fuller understanding of the roles of health in regional systems based on accurate and timely geospatial information, topics that present exciting opportunities along the research frontiers of regional science.

References

Bennett VL, Eaton DJ, Church RL (1982) Selecting sites for rural health workers. Soc Sci Med 16:63–72

Chetty R, Stepner M, Abraham S et al (2016) The association between income and life expectancy in the United States, 2001–2014. JAMA. doi:10.1001/jama.2016.4226

Church R, Murray AT (2009) Business site selection, location analysis, and GIS. Wiley, New York

Cocking C, Flessa S, Reinelt G (2012) Improving access to health facilities in Nouna district, Burkina Faso. Soc Econ Plan Sci 46(2):164–172

Cromley EK, McLafferty SL (2011) GIS and public health, 2nd edn. Guilford, New York

Cudnik MT, Yao J, Zive D, Newgard C, Murray AT (2012) Surrogate markers of transport distance for out-of-hospital cardiac arrest patients. Prehosp Emerg Care 16(2):266–272

Cummins S, Curtis S, Diez-Roux AV, Macintyre S (2007) Understanding and representing 'place' in health research: a relational approach. Soc Sci Med 65(9):825–1838

Daskin MS, Dean LK (2004) Location of health care facilities. In: Brandeau M, Sainfort F, Pierskalla W (eds) Operations research and health care: a handbook of methods and applications. Kluwer Academic, Boston, pp 43–76

Ferrara A, Nisticò R (2015) Regional well-being indicators and dispersion from a multidimensional perspective: evidence from Italy. Ann Regional Sci 55(2/3):373–420. doi:10.1007/s00168-015-0704-y

Grubesic TH, Miller JA, Murray AT (2014) Geospatial and geodemographic insights for diabetes in the United States. Appl Geogr 55:117–126

Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV (2005) Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the public health disparities geocoding project. Am J Public Health 95(2):312–323

Kumar N (2004) Changing geographic access to and locational efficiency of health services in two Indian districts between 1981 and 1996. Soc Sci Med 58:2045–2067

Kwan M-P (2012) The uncertain geographic context problem. Ann Assoc Am Geogr. doi:10.1080/00045608.2012.687349

Leira EC, Fairchild G, Segre AM, Rushton G, Froehler MT, Polgreen PM (2012) Primary stroke centers should be located using maximal coverage models for optimal access. Stroke 43:2417–2422

Macintyre S, Ellaway A, Cummins S (2002) Place effects on health: How can we conceptualise, operationalise and measure them? Soc Sci Med 55(1):125–139

Martens P (2002) Health transitions in a globalizing world: towards more sustained health? Futures 34(7):635–648

Mayer JD (2000) Geography, ecology and emerging infectious diseases. Soc Sci Med 50(7–8):937–952

Mayhew JD, Leonardi G (1982) Equity, efficiency, and accessibility in urban and regional health-care systems. Environ Plann A 14:1479–1507

McLafferty S (2010) Placing pandemics: Geographical dimensions of vulnerability and spread. Eurasian Geogr Econ 51(2):143–166

McLafferty S, Broe D (1990) Patient outcomes and regional planning of coronary care services: a location-allocation approach. Soc Sci Med 20(3):97–304

McLafferty S, Widener M, Chakrabarti R, Grady S (2012) Ethnic density and maternal and infant health inequalities: Bangladeshi immigrant women in New York city in the 1990s. Ann Assoc Am Geogr 102(5):893–903

Messer LC, Oakes JM, Mason S (2010) Effects of socioeconomic and racial residential segregation on preterm births: a cautionary tale of structural confounding. Am J Epidemiol 171:664–673

Murray AT, Grubesic TH (2016) Locational planning of health care facilities. In: Kanaroglou P, Delmelle E, Paez A (eds) Spatial analysis in health geography. Ashgate, New York, pp 243–259

Murray AT, Grubesic TH, Wei R (2014) Spatially significant cluster detection. Spat Stat 10:103–116

Newbold KB (2005) Self-rated health within the Canadian immigrant population: risk and the healthy immigrant effect. Soc Sci Med 60:1359–1370

Oppong JR, Hodgson MJ (1994) Spatial accessibility to health care facilities in Suhum District, Ghana. Prof Geogr 46(2):199–209

Osleeb J, McLafferty S (1992) A weighted maximal covering model to aid in dracunculiasis eradication. Pap Reg Sci 71(3):243–257

Rushton G (1984) Use of location-allocation models for improving the geographical accessibility of rural services in developing countries. Int Reg Sci Rev 9:217–240

Stenberg K et al (2014) Advancing social and economic development by investing in women's and children's health: a new global investment framework. Lancet 383:333–354

Walker RE, Keane CR, Burke JG (2010) Disparities and access to healthy food in the United States: a review of food deserts literature. Health Place 16:876–884

Wang F (2012) Measurement, optimization, and impact of health care accessibility: a methodological review. Ann Assoc Am Geogr 102:1104–1112

Wennberg JE, Fisher ES, Skinner JS (2002) Geography and the debate over medicare reform. Health Aff. doi:10.1377/hlthaff.w2.96

Yao J, Murray AT (2014) Locational effectiveness of clinics providing sexual and reproductive health services to women in rural Mozambique. Int Regional Sci Rev 37(2):172–193

Sara McLafferty is professor, Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign. Her research interests include health inequalities and access to health care in U.S. cities, GIS and spatial analysis, and spatial access to employment for women and minorities. Previous faculty positions were at Columbia University and Hunter College, City University of New York. Prof. McLafferty earned the Ph.D. in geography from the University of Iowa in 1979.

Alan T. Murray is professor, Department of Geography, University of California at Santa Barbara (UCSB). His primary research interests are geographic information science; spatial optimization; health informatics; urban growth and development; land use planning; urban, regional, and natural resource planning and development; and, infrastructure and transportation systems. He previously held academic appointments at Drexel University, Arizona State University and Ohio State University. Previous faculty positions were at Drexel University, Arizona State University and Ohio State University. Dr. Murray earned a Ph.D. in geography from the University of California at Santa Barbara in 1995.

Chapter 11
New Approaches to Gender in Regional Science

Katherine Chalmers and Walter Schwarm

While regional science argues that space matters, it has long ignored the important context of gender, specifically how gender inhabits space. Central to this discussion is the notion that gender implies a broader understanding than merely biological sex. Nelson (2003) argues for a humanist approach that widens the scope of analysis away from the binary to an inclusive feminist perspective. Rather than thinking merely in terms of masculine-good/feminine-bad, Nelson (1996) argues that economics, and by extension regional science, should consider both the positive and negative attributes of each gender. This means that regional science must do more than include sex as an explanatory variable in its analysis by examining how gender constructs change not only potential outcomes but indeed how spatial relationships are perceived by acknowledging gender and space's own social constructions. Doing so will help regional science to become more *relevant* and its policy prescriptions more prescient.

What does it mean to broaden regional science to include gender considerations? It means doing more than paying lip service to differences by expanding our thinking beyond simply a male/female explanatory variable. To include gender means changing the nature of the model, not merely tacking sex onto the variable list. How men and women move through space and organize their lives is distinctly and significantly influenced by their assigned genders. For example, both genders embrace different social and familial responsibilities that cause them to encounter space in distinctly different ways. Traditionally, women have borne the burden of childcare and the bulk of household production, and these responsibilities constrain women's ability to move through space in a way that men do not necessarily

K. Chalmers (✉)
California State University, Sacramento, CA, 95819, USA
e-mail: chalmers@csus.edu

W. Schwarm
Demographic Research Unit, California Department of Finance, Sacramento, CA, USA

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_11

recognize. Furthermore, these constraints may have unintended consequences as gender responsibilities and societal norms change. Therefore, regional science should embrace these differences and create models that help to explain why these differences exist, and craft appropriate policies that take them into account.

Regional Science, as with economics, relegated the consideration of gender for a long time to the marginal fringes of the discipline. While Becker (1973, 1974) notably provided the first step to understanding how gender might influence economic decisions with conceptual modeling, Akerlof and Yellen (1988), in developing their theory of efficiency wages, have shown how economics can more broadly expand beyond its traditionally masculine viewpoint. As Nelson (1995) points out, by incorporating notions of "fairness" into their theory of efficiency wages, Akerlof and Yellen have moved beyond the rigid boundaries of economics. Unfortunately, to date, regional science largely ignores gender and continues to see the issue along more traditional analytical lines. This narrow focus will hamper the relevancy of regional science because issues that will arise are likely to be infused with underlying currents that invoke a consideration of gender issues. Indeed, as the world moves further along in its development, regional differences continue to exert their influences at both ends of the size scale. These differences arise, as we know, from many differing spatial characteristics but, we must also be cognizant that those different spatial characteristics include gendered differences.

While there have been notable exceptions, regional science has largely considered gender only in the context of migration and the labor market issues under localized wage differentials and commuting. While these are important issues and will remain so in the future, they do not fully capture the breadth of analysis that is possible or indeed warranted when one examines the greater role of gender in the underlying fields of economics and geography.

11.1 Areas of Likely Future Focus

11.1.1 Gender, Income and Inequality

When considering models of urban development, the price of land and housing decreases as one moves farther from either the center or spatial clusters. Firms that elect to locate away from these areas can pay their workers lower wages because workers do not bear the cost of commuting. These intra-urban wage differentials have been credited with contributing to spatial unevenness in wages across areas of cities Carlson and Persky (1999) and regions Steven and Riker (1999). These monocentric and clustering models yield associated wage differentials that may be incomplete because spatial constraints for certain classes of workers fail to be adequately addressed. Urban/suburban wage differentials might exist only for certain kinds of workers and those who are more limited spatially in their commute, such as second-earner women. In this case, women's wage rates by job location would be much more distance-sensitive than would men's.

One of the markers of economic development has been the increasing status of women. With development, women assume productive employment, which gives them a stronger voice in household economic decisions. Therefore, models that speak to women's concerns are more likely to be effective, unencumbered by unintended consequences. For example, human capital theory ignores the societal expectations attached to gender roles that may influence women to perform more traditional gender specific activities so that women's failure to achieve gender parity in labor markets is not merely indicative of women's lower opportunity costs in performing household tasks but rather an indication of the failure of economic theory to consider the situational gender aspects of the labor decision Barrett (1982). This is especially true in participatory democracies in which women form a significant voting bloc and politicians may seek to satisfy their demands. As women's political voices become louder, laws that promote gender equality must consider how women live and the responsibilities they bear. For example, the rate of part-time work for women in the Netherlands is significantly higher than elsewhere in the world and is a growing concern to politicians there when coupled with an aging population Noback et al. (2015); however, programs designed to increase full time work that fail to account for the drivers of part time work such as child-care and increasing elder-care responsibilities will face potential failure. Thus, a consideration of the intertwined role between political influence, regional growth, and the employment decision motivate research at this level and at this time.

As traditional gender roles undergo change, consideration must be given to the interactions between gender, wages, amenities and schools within a larger regional or spatial construct. Traditional suburban housing that requires automotive transport to successfully fulfill household responsibilities means that completing those tasks necessarily requires commuting time, time that cannot be spent doing other tasks. Given that suburbanization results in lower density living, and in the U.S. typically consists of traditional, single-family homes with minimum lot sizes enforced by zoning restrictions, and that household tasks have historically fallen on women, this combination can present a serious logistical constraint upon the female labor force participation rate. As Jacobsen (2007) notes, while this situation can be argued to be an expression of free choice, that assertion ignores the extent to which zoning laws and pre-existing housing stock influence household decisions. Furthermore, the traditional (American) school calendar of approximately 180 days of school, with each day consisting of less than 8 h of schooling with a 3-month summer break is not conducive to a dual-earner household, particularly when one gender bears primary responsibility for the transportation of children to/from school. And shortened school days are not a uniquely American problem, but a common issue faced by those raising children in many countries. Indeed the more common nature of homeschooling in the United States may be attributable, at least in part, to the difficulty that women face in juggling competing responsibilities within an unmanageable distance Miller (2006).

Clearly there is important work being done examining the spatial and regional characteristics that generate differential outcomes in wages, income and opportunities. These range across the entire spectrum of sophistication, from the most basic narrative approach to those using the most sophisticated tools of the art. It is interesting to note, however, how many either fail to examine or assume away the problem of gender. While we have argued herein that the mere inclusion of a gender dummy is insufficient, it at least begins to address the potential differing interactions. As an example in the Florida and Mellander (2013) working paper examining the difference and determinants of wage and income inequality across U.S. Metros, no mention is given to gender as a component of their analysis, either as a determinant nor as a conditioning variable. The lack of such consideration is unusual but not uncommon, and unless we are prepared to declare the equality of opportunity and outcomes across gender, such oversight or simplified assumptions are, in our opinion, unrealistic and prone to making grave errors.

11.1.2 Household Location, Composition, and Influences

In the regional science literature, the household is usually treated no differently than an individual would be. Typically, the household has a single objective function that it wishes to maximize, having several variables of adjustment at its disposal to do so (e.g., location, home size, commute time), and subject to various constraints. This corresponds to a unitary vision of the household. The influence of Mincer's (1978) seminal work in the differential weights of migration decisions within a household have only made moderate inroads in regional science. The reality is that we should consider not just commuting decisions, but also labor supply, and indeed almost all actions a household undertakes as compound events. The relative impact of moving, labor supply, and residential location are not uniform distributed across members of the household, and those economic impacts, as well as psychological and environmental ones are apt to be unequally borne, and might fall heaviest on those in Mincer's "tied mover" category McCollum (1990). As Donni underlines, "there is increasing agreement, however, that economists cannot ignore the fact that most households are composed of several individuals who take part in the decision process" Donni (2008), particularly as more data is now available. There are various reasons why household members, in particular a husband and his wife, might have diverging views regarding residential choices. We can at least mention three main kinds of situations:

- Mincer's "tied mover" vs. "tied stayer" where the decision to move might be beneficial to one spouse more than the other, in particular if it is based on professional considerations.
- The decision to either own or rent has usually a major impact on the household consumption and savings behavior for the following years. Rules of the marriage contract regarding wealth distribution in case of divorce are thus likely to influence the views of each member regarding this decision, and potentially generate conflicts.

- The location choice is a typical example of diverging objectives within the household, as each member wishes to minimize commuting time (subject to other constraints in terms of quality of the neighborhood, home size, and so on).

The existing literature studying residential choices using collective models is quite scarce, and it generally focuses solely on the location choice, not on the within household interactions. Empirical works are even rarer. It would seem to us that the analysis of residential choices through the prism of collective decision-making remains to be done in its largest part.

A regional approach to labor and household location issues forms a direction that has also not yet been adequately considered. While there are models and previous research that examine the motivations and economic driver, there is still considerable room for deepening our understanding. The question that most clearly must be addressed is how the locational decision takes place. What forms the basis of the spatial outcome? The underlying spatial determinants of the migration/moving locational decision needs to be further examined. While labor pull effects have been examined in the literature before, there is more that can be done to examine these issues considering the household as a joint decision system with its own outcomes. The joint household decision literature characterized by Chiappori et al. (2002), for example, could be operationalized within a more regional specific framework and begin to provide a considerable drive towards generating such an understanding.

The household decision framework presented above propels us to consider further research areas that move regional science forward. While Chiappori is clearly examining the effect of the marriage market and its outcomes on the labor supply decision of the member of a household, we can also see that such a decision framework should be influenced by the potential pool of alternative mates and labor market opportunities. There has been some thought given to the effect of assortative mating on labor market characteristics and the role that location specific factors may influence the outcomes, Turney and Harknett (2010) and Harknett and Hartnett (2011, 2014), but with the growing imbalance in educational attainment arising from sex ratio differences in higher education, there is the opportunity to investigate further. Indeed, perhaps the job skills match model from urban economics can be adapted to model desirable matching qualities in mates? A potential direction of research might also be found by examining the labor pool sharing model to examine not only matching mate choices but the eventual household labor and locational choices that result from the marriage.

11.1.3 Composition of Neighborhoods and Social Change

Janice Madden's contributions to this sub-area warrant special acknowledgment. Her continuing research has done much to lay out the existing state of gender in regional science. Likewise, her research contains the suggestions of the future directions we lay out, one that is inclusive, but quantitative based on highlighting and teasing the spatial distribution of challenges and opportunities for the residents of regions.

While Madden and Ruther (2014) begin a movement in the direction that encompasses not only gender, but also alternative household composition, many more questions are raised than answered. A consideration of the greater effect of household formation and its spatial components would incorporate the arguments of Badgett (1995) and expand the economic household models to include LBGT households and the location decisions that accompany their formation. This would further allow the discipline to focus on the more quantitatively realized results that argue against the more qualitative arguments that fail to be realized, particularly when one examines the nature of neighborhoods pre-and-post social change.

11.1.4 Innovation and Its Regional Environment

While regional science has considered such marginalized groups as homosexuals [see Madden and Ruther (2014), Florida (2002), and Florida et al. (2008)], to date, women have not been a specific research focus as a stimulus for innovation. Overlooking women may lie in their relative ubiquity as well as the implied assumptions of masculine universality. One of the challenges researchers and policymakers face is to expand their concepts of regional and urban development beyond those processes associated with technologically defined and growth-oriented originality, such that notions of local development may enhance the social well-being of places and be more gender inclusive. Gender needs to be specifically targeted to emphasize the ways networks are embedded in place-based social, economic, cultural, and political structures, which shape entrepreneurs' identities and affect access to resources. This extends to understanding the importance of investigating the impacts of social identity on entrepreneurs' networks if the relationship is between entrepreneurship and place.

Regional scientists have a keen interest in innovation because of its connection to regional economic advantage. Except for a very few papers Hanson and Blake (2009) innovation is understood as predominantly technological and product driven rather than considering the social spatial network that fosters it. The context, both social and geographical, of an innovation is elementary to its identification as innovative. There are many instances of innovation that occur in economic sectors and by agents that are typically ignored or undervalued by current research and policy. One of the challenges researchers and policymakers face is to expand their concepts of regional and urban development beyond those processes associated with technologically defined and growth-oriented originality, such that notions of local development may enhance the social well-being of places and be more gender inclusive.

11.2 Conclusions

Topics that will continue to move regional science forward include the relationships between economic structure, growth regimes and gender inequities and the mechanisms and the extent to which unpaid work constrains women's participation in paid work and access to economic opportunities. Forward looking studies must consider the complete implications of women's labor market participation on their well-being and for intra household allocation of time. There should be consideration of the structural, microeconomic aspects of women's employment in the informal sector and formulation and analysis of gender-aware policy interventions and frameworks for integrating the role of unpaid work in measures of regional well-being. These steps, combined with an understanding of the various underlying motivations and structures that drive innovation and locational decisions, serve to set regional science on a path that would bring it to the forefront of quantitatively-based, gender-informed research.

References

Akerlof G, Yellen J (1988) Fairness and unemployment. Am Econ Rev 78(2):44–49

Badgett ML (1995) The wage effects of sexual orientation discrimination. Ind Labor Relat Rev 48(4):726–739

Barrett NS (1982) The impact of public policy programs on the status of women. AEA Pap Proc 71(2):160–165

Becker G (1973) A theory of marriage: part I. J Polit Econ 81(4):813–846

Becker G (1974) A theory of marriage. In: Schultz T (ed) Economics of the family: marriage, children and human capita. University of Chicago Press, Chicago

Carlson VL, Persky JJ (1999) Gender and suburban wages. Econ Geogr 75(3):237–253

Chiappori P, Fortin B, Lacroix G (2002) Marriage market, divorce legislation, and household labor supply. J Polit Econ 110(1):37–72

Donni O (2008) Labor supply, home production, and welfare comparisons. J Public Econ 92(7):1720–1737

Florida RL (2002) The rise of the creative class: and how it's transforming work, leisure, community and everyday life. Basic Books, New York, NY

Florida R, Mellander C (2013) The geography of inequality: difference and determinants of wage and income inequality across US metros. Working paper series in economics and institutions of innovation. Royal Institute of Technology, No 304. CESIS – Centre of Excellence for Science and Innovation Studies

Florida R, Mellander C, Stolarick K (2008) Inside the black box of regional development: human capital, the creative class and tolerance. J Econ Geogr 8(5):615–649

Hanson S, Blake M (2009) Gender and entrepreneurial networks. Reg Stud 43(1):135–149

Harknett K, Hartnett C (2011) Who lacks support and why? An examination of mother's personal safety nets. J Marriage Fam 73(4):861–875

Harknett K, Hartnett C (2014) The gap between births intended and births achieved in 22 European countries, 2004–2007. Popul Stud 68(3):265–282

Jacobsen J (2007) The economics of gender, 3rd edn. Blackwell, Malden, MA

Madden J, Ruther M (2014) Gayborhoods: the economics and demographics of the concentration of gays within large American cities. In: Rose A, Nijkamp P, Kourtit K (eds) Regional science matters – studies dedicated to Walter Isard. Springer, Berlin

McCollum AT (1990) The trauma of moving: psychological issues for women. Sage library of social research, vol 182. Sage, Thousand Oaks, CA

Miller HJ (2006) Social exclusion in space and time. In: Axhusen KW (ed) Moving through nets: the physical and social dimensions of travel. Selected papers from the 10th International conference of travel behaviour research, Elsevier, pp 353–380

Mincer J (1978) Family migration decisions. J Polit Econ 86:749–773

Nelson J (1995) Feminism and economics. J Econ Perspect 9(2):131–148

Nelson J (1996) Feminism, objectivity, and economics. Routledge, London

Nelson J (2003) How did the 'moral' get split from the 'economic'? In: Barker D, Kuiper E (eds) Toward a feminist philosophy of economics. Routledge, London

Noback I, Lourens B, van Dijk J (2015) Gender-specific dynamics in hours worked: exploring the potential for increasing hours worked in an ageing society. In: Karlsson C, Andersson M, Norman T (eds) Handbook of research methods and applications in economic geography. Edward Elgar, Northampton, MA

Steven R, Riker DA (1999) Geographic mobility, race, and wage differentials. J Urban Econ 45(1):17–46

Turney K, Harknett K (2010) Neighborhood socioeconomic disadvantage perceptions of social support among new parents. J Fam Issues 31(4):499–524

Katherine Chalmers is associate professor, Department of Economics, California State University at Sacramento (CSUS). Her primary research interests are in the intersection of gender, public policy, and regional economic development. Additionally, she specializes in the evolution of historical institutions, particularly with respect to sub-Roman and Anglo-Saxon economic development in the face of international migration and religious conversion. Previous faculty positions were at Bowling Green State University in Ohio and Central College in Pella, Iowa. Dr. Chalmers earned the Ph.D. in economics from the Colorado State University in 2004.

Walter Schwarm is senior demographer, Demographic Research Unit, State of California Department of Finance. His primary research interests are in the intersection of household fertility decisions, location dynamics, and household formation. Additionally, he serves as adjunct faculty in the Department of Economics at California State University at Sacramento (CSUS). Previous faculty positions include Coe College in Cedar Rapids, Iowa and a post-doctoral fellowship at the Regional Research Institute at West Virginia University (WVU). Dr. Schwarm earned the Ph.D. in economics from Colorado State University in 2001.

Chapter 12
Identifying Sleeping Beauties in the Lore of Regional Science

Rajendra Kulkarni and Roger R. Stough

12.1 Introduction

The Scientific world has seen an unparalleled growth in new knowledge creation over the last several decades. Growth rates from 1% just a couple of centuries ago to over 8 and 9% per year by 2012 have been estimated (Bomann and Mutz 2014). Compared to natural sciences, the field of regional science is relatively young, and yet the new knowledge creation rates are comparable to other scientific disciplines. Given this extraordinary growth, some researchers have been investigating a not-so-rare phenomenon of what Van Raan (2004) refers to as "sleeping beauty" for research that goes unnoticed for a long duration.

If "sleeping beauties" are not uncommon in the scientific literature, then, regional science is likely to have its own share of such research. In this paper, we present a novel method that combines network analysis with text and bibliometric analytics and that does not explicitly depend on citation patterns to help identify "sleeping beauties" in regional science. Additionally, we explore how changes in published scholarly work may change over the next few decades and how some of those changes are likely to affect future sleeping beauties.

R. Kulkarni (✉)
George Mason University, Fairfax, VA, USA
e-mail: rkulkarn@gmu.edu

R.R. Stough
Schar School of Policy and Government, George Mason University, Fairfax, VA, USA
e-mail: rstough@gmu.edu

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_12

12.2 Evolution of Scientific Literature and the Concept of Sleeping Beauties

Collective scientific knowledge—when measured in terms of publications, appears to be incremental, i.e., new scientific ideas are built on top of the existing knowledge base, which in turn traces its roots to older research. Mostly a linear view of knowledge building is accepted as the norm with rare exceptions where truly path breaking research emerges that takes the discipline or the field of research by storm and leads the field in a different direction. And then there are those—apparently not-so-rare—examples of research that lay dormant or were ignored by the scientific community at the time they were published but were discovered much later for their true significance and contribution to the field of research. Such works of published scientific work with delayed recognition have been referred to as "sleeping beauties" of science (Crassey 2015).

This delayed recognition phenomenon raises such questions as: Why does this happen?; Was the research premature for its time?; and How to quantify research with delayed recognition. Since the mid-1970s, there has been growing interest in the fields of science of sciences, scientometrics, infometrics, etc. that is devoted to researching all aspects of scientific research, including citation and co-citation analysis, bibliometric coupling, co-occurrence and co-authorship network analysis. Garfield (1970) and Price (1976) tried to answer the questions about why it happens and pointed to possible causes such as poor communication skills on the part of authors. In the past, Garfield (1980) and others have discussed issues related to premature research, or research before its time.

As mentioned in the introduction section, van Raan in 2004 developed a methodology to quantify "occurrence of sleeping beauties" in the science literature (Raan 2004). Using the bibliometric approach, he identified different phases of the phenomenon, such as deep sleep, duration of sleep and awakening duration and was able to identify case of sleeping beauty and awakening prince. Qing K et al. (2015) analyzed over 22 million pieces of published literature and developed a "sleeping beauty" coefficient that can define and help identify such publications. The main thrust of both these studies was about gathering quantifiable information for very large number of published materials and analyzing these based on temporal citation distributions. In other words, if a published scientific work has no citations, then it would not be part of sleeping beauty analyses, which filters out very large portions of published knowledge. As the number of avenues for publishing has increased, the citation rates have not, according to Larivière and Gingras (2015).

A vast majority of published research receives fewer than five citations and often no citations. Remler, in her blog (2014), observes that depending on the field or discipline, the number of articles that get zero citations varies from 12% in medicine, 27% in natural science, 60% in social science and a whopping 82% in humanities. Thus, on average almost 45% of the published articles are never cited. In that case, any analyses of identifying sleeping beauties based only on citation frequencies is likely to miss a large number of sleeping beauties.

In the following sections, we explore a novel method that does not depend on citation counts and can be used to identify potential sleeping beauties within a field or discipline using just the bibliometric data for that field or discipline. Specifically, we will explore how one could find potential sleeping beauties in the field of regional science.

In the next section, we briefly describe bibliometric data used for this exploratory analysis, followed by a section on the specifics of the methodology, which will be followed by couple of case studies that help identify sleeping beauties in the field of regional science.

12.3 Citation Indexing Databases and Data Extraction

The data for the analysis was extracted from Thomson Reuter's Web of Science (WoS). The initial attempt to conduct a topic (key word or phrase) search using "regional science" and "regional analysis" had to be abandoned since we discovered that those phrases are not unique to the field of regional science but have been used by life sciences; especially in biomedical fields such as radiology, nuclear medicine, neuroscience, neurology, MRI analysis, image analysis, etc.

Instead of topic searches we generated a list of journals that publish regional science research and extracted articles based on journal names. (See Table 12.1 for a sample list of journals.) This process generated a total of over 17,000 records (WoS July 17, 2015). Of these, about 300 records were discarded due to missing information in some of the fields. Of the remaining 16,700, just over 55% (9200) of records representing published scientific research had one or more citations since the publication date (Fig. 12.1). The flip side of this is that nearly 45% of articles (7495) had zero citations (Fig. 12.2). In theory, one could use Google Scholar as an alternate source, and in some fields Google Scholar does have better citation counts

Table 12.1 Some of the journals used to extract data from Web of Science

Regional Science Journals	
Papers in Regional Science	International Regional Science Review
Journal of Regional Science	Journal of Regional Analysis and Policy
Regional Science Policy and Practice	Journal of Urban Economics
Applied Spatial Analysis and Policy	Papers in Regional Science
The Annals of Regional Science	Regional Science and Urban Economics
Regional Studies, Regional Science	Regional Studies
Spatial Economic Analysis	Review of Regional Studies
Annals of Regional Science	Growth and Change
Canadian Journal of Regional Science	Review of Urban and Regional Development Studies

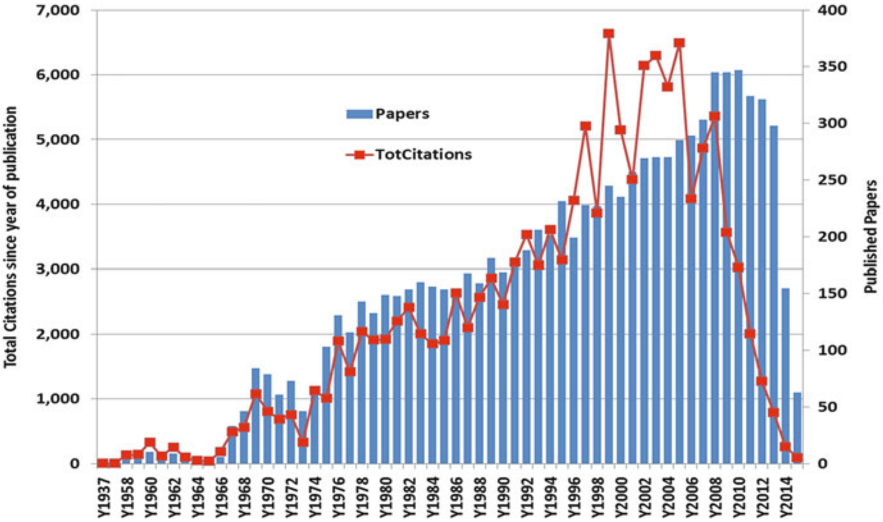


Fig. 12.1 Regional science papers from WoS: Articles and total citations by year of publication

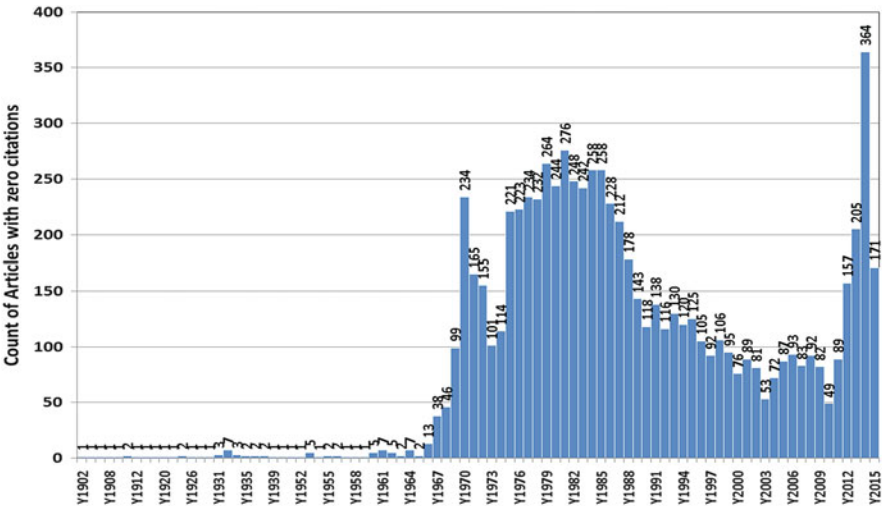


Fig. 12.2 Papers with no citations by year of publication

(Anne-Wil 2016); however, the percent of published papers with zero citations is likely to remain high.

In the following section, we briefly discuss the burst detection analysis that may be useful to identify sleeping beauties.

12.3.1 Using Burst Analysis to Identify Potential Sleeping Beauties

Kleinberg (2002) developed a burst detection and analysis algorithm to "extract meaningful structure from document streams that arrive continuously over time". By analogy, one could think of citations received by published articles over time as streams of information and apply burst analysis to identify structural changes in the citation patterns, such as low or no citations to a sudden burst of citations for a time period.

One could also apply a burst algorithm to detect sudden fad or popularity of certain keywords/terms associated with scientific articles. Both of these methods were used to detect bursty articles as well as bursty terms by applying the burst detection algorithm to the regional science articles. This particular analysis raises two issues:

- Use of the burstiness property of citation patterns by definition ignores nearly half (45%) of published articles that have zero or no citations. By definition, burst is a sudden change from few citations to more citations one may safely assume that all articles that received between 1 and 3 citations won't be considered for the burst analysis either. In other words, nearly, two-thirds (62%) of articles are ignored by citation burst analysis and, therefore, detection of sleeping beauties is limited only to articles that have more than a certain number of citations.
- The burstiness property of terms identifies sudden popularity, or a fad, of using certain terms over time, however, that does not address the main objective of identifying specific articles that contributed to a term with high burstiness. In theory, one could develop another set of analyses to identify the contribution of each article towards burstiness and then figure out which article started that trend. That part of the analysis will be explored in future work.

12.3.2 Use of Multimode Network and Text Analytics to Detect Sleeping Beauties

Typically, each full record W_i, where i is the index into number of records, extracted from WoS, has four text attributes (fields): Title Ti, Abstract Ab (if present), a set of Key words KW (if present) and scientific classification Ci (one or more). These four elements of text fields are converted into a vector X_i of N-gram tokens of length 1–4, such that $X_i = [W_i\ (t1), W_i\ (t2), W_i\ (t3), \ldots, W_i(tn)]$. A simplistic example will illustrate how a title of a paper such as "Industry and Firm Location Analysis by Regions," is turned into N-gram tokens. Each word in the title except common words such as "and" and "by," etc. is called a uni-gram or 1 gram. A sequence of two words would be bi-gram or 2-gram such as "Industry Firm", "Firm Location", etc. In the same vein, a 3-gram consists of 3 words and 4 gram consists of 4-words.

Each full record W_i also has other attributes, such as cited references Rf, year of publication Py, times cited Tc, authors Au, affiliation Af, journal details Jd, unique Id, etc. For an article W_i, let Y_i refer to a vector whose members are the published article and all of the cited references associated with that article. We used software package Citespace (2015) to build a multi-mode,[1] mixed direction network consisting of co-cited articles and references as well as co-occurrence N-gram tokens. The co-cited connections and co-occurrence N-gram are not directed, while the N-gram tokens associated with articles are connected to articles by directed edges from N-gram tokens to articles.

After generating a network consisting of P articles, R references and T n-gram tokens, we computed network statistics such as node degree, weighted node degree and node centrality measure called "betweenness" index. The node degree refers to number of connections to its immediate neighbors while weighted node degree refers to weighted sum of connections to its immediate neighbors. The main difference between a degree and weighted degree of a node is that if node "A" has two connections to node "B" and one connection to node "C," then the weighted degree of node "A" is 3, while degree of node "A" is 2. The node betweenness index refers to how often a node is visited for an abstract set of paths from each node to every other node. Thus, a node with high betweenness index is more central to the network than a peripheral (or leaf) node that is either a "start" or an "end" node for a path from (to) that node to (from) other nodes in the network.

12.4 Network Analysis and Sleeping Beauties

In theory, a network consisting of 16,700 WoS articles (P) with over 300 K reference records (R) and nearly 150 K tokens from terms T forms a humongously large multi-mode mixed direction network consisting of nearly a half million nodes and

[1]When a network consists of just one type of node and edges connecting these nodes, then such a network is referred to as a uni-modal network. If a network consists of two types of nodes, then it is referred to bi-modal or multi-modal network. Depending on whether edges connecting nodes have directions or a combination of directional and non-directional edges, the network would be called as having directional and mixed directional connectivity respectively.

at least many magnitudes more of edges between these nodes. However, due to computing limitations, we built a network consisting of 33 K nodes and almost 350 K edges with a diameter of 12 hops and an average path length 3.5; and over 600 million shortest paths. The diameter of 12 refers to the longest path from over 600 million shortest paths; in other words, the longest shortest path of 12 refers to 12 connections (also called as “hops”) between the farthest nodes in the network, while on average the shortest path (hops) is of 3.5 connections. Next, we computed a weighted node degree and node centrality betweenness index. The following is a short list of token nodes with the highest centrality betweenness index: *Growth, Regional Study, Regional Policy, Regional Development, Economic Development, Cities, Location, Metropolitan Area, Economic Growth, Regional Science, Employment, Urban Area, Regional Growth, Migration, Agglomeration, Unemployment, Demand, Trade, Local Governments, Competition, Increasing Return,* etc.

Tokens with a high betweenness index are central to the network consisting of articles and tokens. For the current findings, we randomly chose a few of these high betweenness index token nodes and built their 1 to 2 “hops” ego-networks where the token of interest and its immediate 1 and 2 hop neighbors form an ego-network. The concept of ego-networks comes from the field of social network analysis where the ego-network of node A consists of all the nodes that are directly connected to node A and that direct connection indicates that they have a close social relationship with node A. This idea of close relationships among nodes with ego-node will be used to identify potential sleeping beauty candidates.

When a set of articles that are part of a token’s ego-network are sorted by the date (year) of publication, then the article with the earliest publication date is likely to be a sleeping beauty provided that, several years had lapsed since its publication and that:

1. It has been not been cited
2. Or cited infrequently
3. Or lay dormant for a long duration and then received large number of citations for next several years.

As mentioned before, for a few of these token nodes with very high betweenness centrality; their ego networks were constructed, i.e., the token node of interest and all the nodes that are its immediate neighbors and then searched for a node that happened to have the earliest year of publication (Fig. 12.3).

The Hotelling article is not part of the published papers data (P), however, it has been part of cited references (R). We were able to find citation analysis for the article from the WoS Core database. The chart below shows the output. It is clear

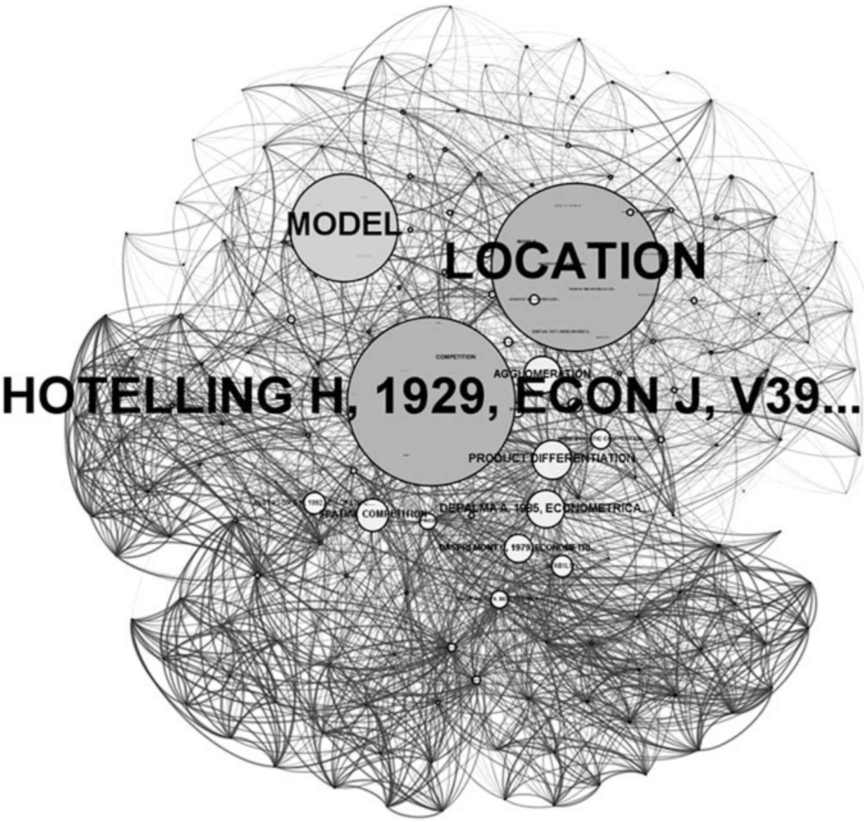


Fig. 12.3 Ego-network of token "Location" and its sleeping beauty, "Hotelling, 1929, Stability in Competition, *Journal of Economics*, V39 N33, DOI 10.2307/2224214"

from the citation analysis shown in Fig. 12.4 that Hotelling's article was hardly cited for a few dozen years but then garnered a large number of citations much after its publication. This citation pattern fits our definition of sleeping beauty (Figs. 12.5, 12.6, 12.7, 12.8, and 12.9).

We present ego-network analysis for a couple more terms and show associated sleeping beauties. In each of these cases, the sleeping beauties identified were not part of the published paper dataset (P) but were inferred from the ego-network analysis since these were part of the cited reference set (R).

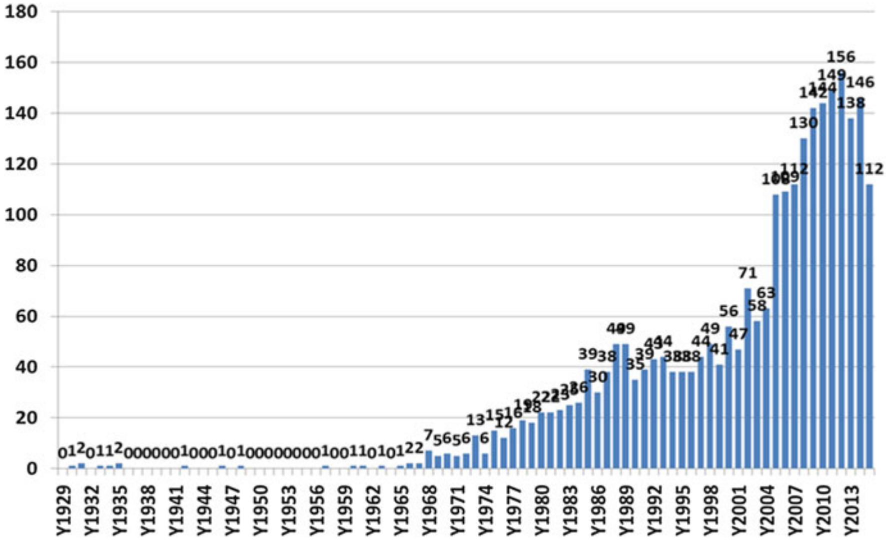


Fig. 12.4 Citation patterns for the Hotelling, 1929, Stability in Competition, *Journal of Economics*, V39 N33, DOI 10.2307/2224214 article

The published article dataset (P) that was used for the analysis does not have a record for Alfred Marshall's work, however, it appears to be in cited references (R) for quite a few articles. Google scholar search reveals an uneven citation history and it appears that the work remained relatively uncited during early 1900s. Even though, this work appears to be part of ego-network of "Growth," conventional wisdom says that regional science researchers do not associate Alfred Marshall's work with the literature on "Growth." It may be that this spurious association between a token term and scholarly work is result of co-citation by scholarly works that are related to "Growth." In future, we will have to analyze number of different data samples to determine whether an association exists consistently across different samples.

12.5 Conclusions

Based on our current analysis of over 15,000 published papers from 1902 to 2015 in the field of regional science; a multi-modal network analysis combined with extracted ego networks of nodes with high betweenness centrality index scores can be useful to identify potential sleeping beauties in the field. After constructing a multi-modal network from co-cited published papers (P), references (R) and co-occurrence of tokens (T), a nodal betweenness centrality measure is computed. As stated before, the betweenness centrality of a node is a proxy for gauging how

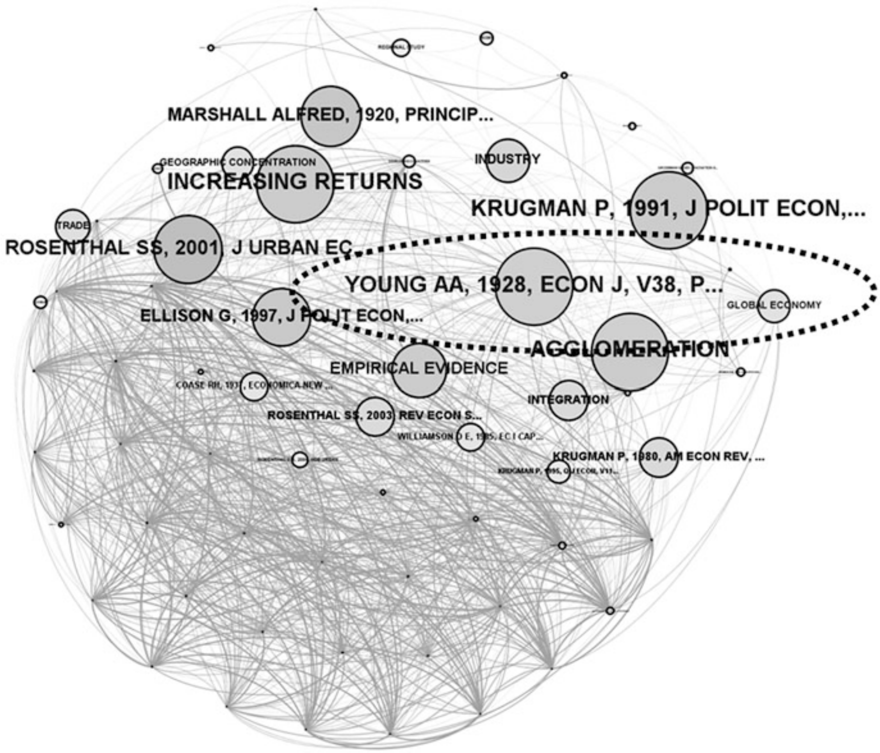


Fig. 12.5 Ego-network of "Agglomeration" and associated sleeping beauty, "Young, Allyn A., Increasing Returns and Economic Progress, *Economic Journal* DEC 1928; DOI: 10.2307/2224097"

central that node is to the network. In terms of a social network, centrality of an actor (node) indicates how powerful or crucial that actor (node) is for relations between all other actors (nodes). For a set of tokens with high betweenness centrality scores, one then can extract their ego-networks to discover the oldest published works associated with that token. If such a published work has zero or very few citations, then that research is a potential candidate for sleeping beauty in that research field. On the other hand, if the published paper does have citations, graphically determine the citation patterns. If the citation pattern shows dormancy of a time period exceeding five or more years, then that paper is a potential candidate for sleeping beauty.

The advantage of using multi-modal network with betweenness centrality score and ego network approach is that, this methodology is agnostic about citation analysis and yet is able to find potential sleeping beauties. One of the limitations of the current analysis is that the cited references may not have article titles (TI), and thus these references do not contribute to the set of tokens (T) whose ego-networks help identify potential sleeping beauties. Another issue that needs to dealt in future

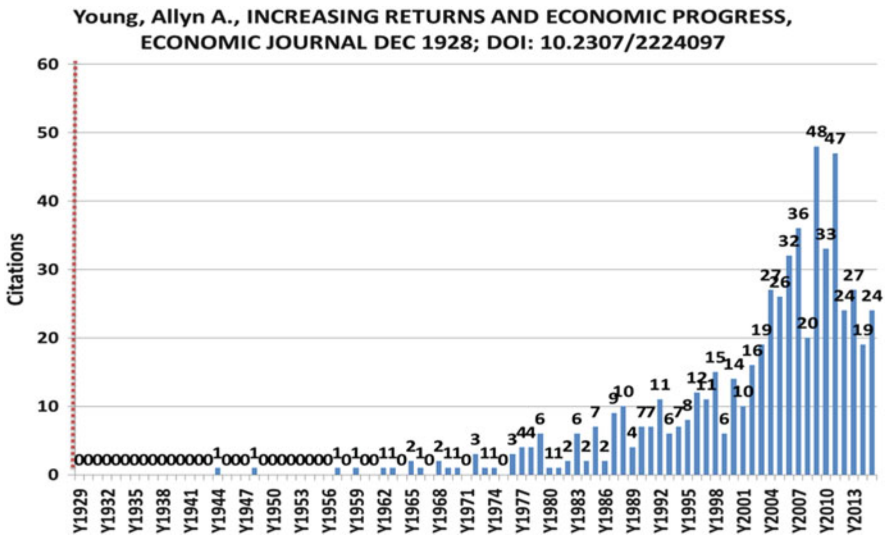


Fig. 12.6 The bar chart shows the citation pattern for the article. Clearly for first few decades this article garnered hardly any citations

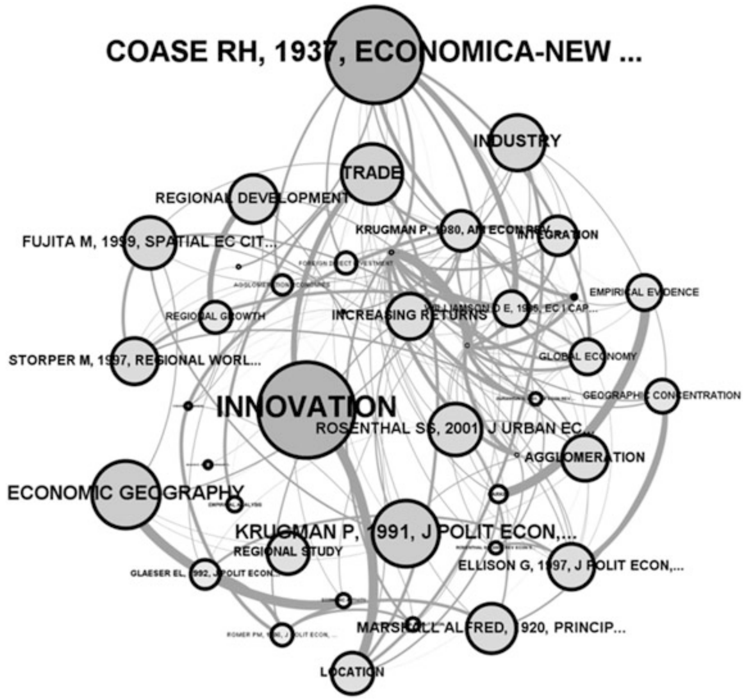


Fig. 12.7 Ego-network for "Innovation" and associated sleeping beauty, "Coase, R., 1937, The Nature of Firm, *Economica-New Series*, V4, N16, DOI:10.1111/j.1468-0335.1937.tb00002.x"

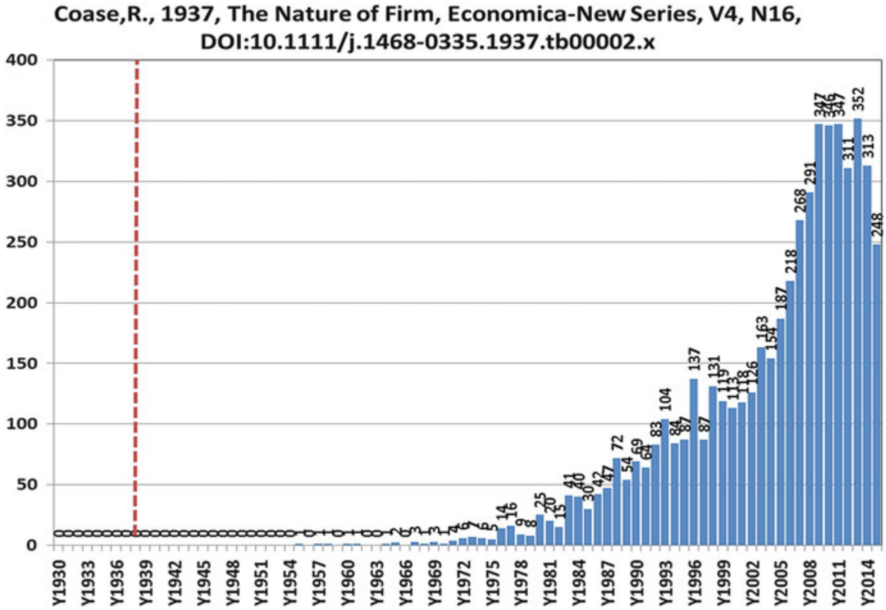


Fig. 12.8 The bar chart shows citation pattern for the article. Clearly for first few decades this article garnered hardly any citations

analysis is to check for consistency of output by using different sets of data. This is essential to filter out spurious association between a scholarly work and a field. In other words if some associations appear randomly between different data sets then those associations will have to filtered out from the analysis. For our future research, we would like to include reference titles which in turn could result in very large networks that will require big data analytics and substantial increases in the required computational resources. However, benefits of adding reference titles could help to convert the multi-mode mixed direction networks into bi-partite networks consisting of articles and tokens. Such bi-partite networks offer new avenues for analysis and identification of sleeping beauties. For the future analysis, we would also like to identify the “waking princes,” those papers that rediscover previously published work, adding a crucial piece of knowledge to the existing knowledge base. The current analysis does show in rudimentary fashion how some of the sleeping beauty articles multiply and are strongly connected to more than one token, and that these multiple tokens are indicators of different disciplines. In our future research, we would like to explicitly identify how certain sleeping beauties may have helped start new research avenues, thereby helping establish new disciplines.

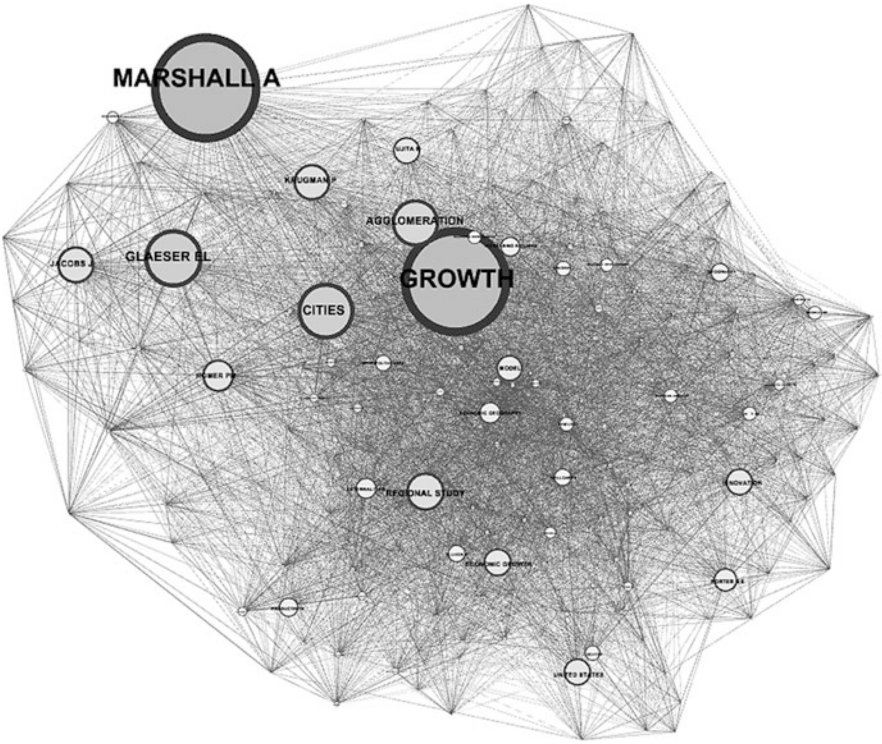


Fig. 12.9 Ego-network of token "Growth" and a potential sleeping beauty article, "Marshall, A, 1890, *Principles of Economics*"

12.6 Future of Research in Regional Science

As mentioned in Sect. 12.1, the pace at which new scholarly work is contributing to a near doubling of scientific knowledge every 9 years. At this rate, total scientific knowledge is likely to be more than 32 times the current levels in 50 years. Assume that the field of regional science is growing at just a fraction of that rate of growth, the total volume of knowledge associated with the field of regional science will be many magnitudes more than the current level over next 50 years. Given the increasingly rapid growth rate in scholarly work over the next five decades, the likelihood or odds of sleeping beauties will become much greater. For example, much of this growth in the future will likely be due to emergence of new technologies which themselves will be occurring at increasingly more rapid rates. This means that the odds will increase that earlier work that was not recognized as important will, with new technologies, become important and, thus, the likelihood that the number of sleeping beauties will rise.

Further, related disciplines are likely to emerge as well as regional science addresses new questions and new issues associated with spatio-temporal analyses of regions all over the globe. For example, one can envision in the near future, regional scientists and analysts trying to find new answers and solutions to resolve issues related to transportation, warehousing operations of commercial drones and small goods deliveries as fleets of these drones and/or automated vehicles start home and business deliveries over short hauls. Ever changing consumer taste for newer products and services along with demand for faster and efficient deliveries are likely to have ripple effects in regional manufacturing and services sectors, commercial surface transportation sector, labor productivity, regional employment and incomes. A contemporary example is how the local passenger transportation sector is affected by the rise of app-based taxi business (Uber) or how the hospitality sector may have to adjust to the rise of the private homes as temporary rental places (Airbnb). These contemporary examples of new technologies are offering competing or substitutional business opportunities that affect local and regional economies to different degrees, thereby giving researchers new avenues for policy analyses.

Discovering sleeping beauties within huge volumes of scholarly work may not be the most challenging issue, since we already have outlined in this paper an exploratory methodology that can be scaled up to big data. Part of the difficulty in the analysis may be with the compilation of data (published scholarly work) since much of that is accessible only with paid subscription. As it is now, free or public access to published literature is limited to open access journals. Regional science does not have a pubmed (http://www.ncbi.nlm.nih.gov/pubmed) equivalent repository of published literature. Google Scholar (https://scholar.google.com/) may be the closest to a publically available server of published scholarly work. In theory, one could make use of pre-print servers such as SSRN (https://www.ssrn.com); however, the full access is still limited. There may be a few free access servers such as Sci-Hub on the dark internet (https://en.wikipedia.org/wiki/Sci-Hub), but any analyses based on such data source will not be accepted.

Maybe all is not lost. The U.S. Office of Science & Technology Policy (OSTP), via its memo, has put forth proposals that would make access to any published literature free if the underlying research has been supported by U.S. tax payer dollars (OSTP 2013). In response to OSTP's plans, American Publishers Association has embarked upon a public-private partnership initiative called "Clearinghouse for the Open Research of the United States" (CHORUS 2013), and it has been tracking how U.S. agencies are implementing open access to published research (US Agencies—Public Access Plans Details 2016). Recently, European Union's Competitive Council during its May 27–30 meeting adopted conclusions that state, "Member states agreed to common goals on open science and to pursue concerted actions together with the Commission and stakeholders. Delegations committed to open access to scientific publications as the option by default by 2020 and to the best possible reuse of research data as a way to accelerate the transition towards an open science system" (EC Competitive Council 2016; Science 2016). And yet the task of data compilation will not be easy since published scientific works in future are likely to be in the multi-media domain that include blogs, v-blogs (video blogs), self-publishing, microblogging, nanopublishing, simulated participatory games and in the same vein publications authored by the public at large where different parts of the research are carried out by different teams at different times. Some of these publications would be in the form of living-documents such that the results obtained and conclusions reached are episodic in nature. Not that far out in the future, questions will arise about how to attribute idea origins when the underlying scholarly work is "crowd-sourced" or "crowd-created." Or it may have been a joint effort by humans and machines—Artificial Intelligence (A.I.)

Who gets credit for original research ideas if they are either the outcome of AI or they are a living-document with episodic output or a multi-team, multi-media effort? Additionally, with fast moving innovative technologies giving rise to variety of new disciplines that involve inter- , cross- and multi-disciplinary research, how does one identify origins of research and ideas? We are seeing beginnings of addressing some of these issues in "altmetrics" (http://altmetrics.org/manifesto/ 2010).

Assuming some or all of the above of how regional science research will unfold in the future, then the scientific work of discovering sleeping beauties associated with origins of ideas or new disciplines is going to be quite challenging. We think that some variation of our current approach based on a combination of text mining/analytics and network analysis will play a key role in identifying sleeping beauties in the future of regional science.

Acknowledgement We would like to express our special thanks to GMU's Office of Research Computing (orc.gmu.edu) for computational resources to conduct largescale network analysis.

References

Anne-Wil H (2016) Google Scholar as a new data source for citation analysis. http://www.harzing.com/publications/white-papers/google-scholar-a-new-data-source-for-citation-analysis. Accessed 3 June 2016

Bomann L, Mutz R (2014) Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. http://arxiv.org/abs/1402.4578. Accessed 3 June 2016

Chen C (2015) How to use CiteSpace. https://leanpub.com/howtousecitespace

CHORUS (2013) Clearinghouse for the open research of the United States. http://publishers.org/news/publishers-introduce-chorus. Accessed 31 May 2016

Competitiveness Council (2016) Research and innovation: open science. http://www.consilium.europa.eu/en/meetings/compet/2016/05/st09357_en16_pdf/ence. Accessed 31 May 2016

Crassey D (2015) http://www.nature.com/news/sleeping-beauty-papers-slumber-for-decades-1.17615

SSRN (2016) https://www.ssrn.com. Accessed 3 June 2016

Garfield E (1970) Would Mendel's work have been ignored if the 'Science citation index' was available 100 years ago? Curr Contents 2:69–70

Garfiled E (1980) Premature discovery or delayed recognition available via essays of an information scientist. http://www.garfield.library.upenn.edu/delayedrecognition.html. Accessed 3 June 2016

Google Scholar (2016) https://scholar.google.com/. Accessed 3 June 2016

Ke Q, Ferrara E, Radicchi F, Flammini A (2015) Defining and identifying Sleeping Beauties in science. Proc Nat Acad Sci U S A 112(24):7426–7431. www.pnas.org/cgi/doi/10.1074/pnas.1424329112. Accessed 3 June 2016

Kleinberg J (2002) Bursty and hierarchical structure in streams. Proceedings of the 8th ACM SIGKDD international conference on knowledge discovery and data mining. https://www.cs.cornell.edu/home/kleinber/bhs.pdf. Accessed 3 June 2016

Larivière V, Gingras Y (2015) The decline in the concentration of citations, 1900–2007. http://arxiv.org/ftp/arxiv/papers/0809/0809.5250.pdf. Accessed 3 June 2016

OSTP Memo (2013) Increasing access to the results of federally funded scientific research. https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf. Accessed 31 May 2016

Price DDS (1976) A general theory of bibliometric and other cumulative advantage processes. J Am Soc Inf Sci 27:292–306. doi:10.1002/asi.4630270505

Priem J, Taraborelli D, Groth P, Neylon C (2010) Altmetrics: a manifesto. http://altmetrics.org/manifesto. Accessed 31 May 2016

Remler D (2014) Are 90% of academic papers really never cited? Reviewing the literature on academic citations, London School of Economics. http://blogs.lse.ac.uk/impactofsocialsciences/2014/04/23/academic-papers-citation-rates-remler/. Accessed 4 June 2016

Science (2016) http://www.sciencemag.org/news/2016/05/dramatic-statement-european-leaders-call-immediate-open-access-all-scientific-papers. Accessed 31 May 2016

Thompson Reuter's Web of Science (2015) https://en.wikipedia.org/wiki/Web_of_Science. Accessed 17 July 2015

U.S. Agencies Public Access Plans Details (2016) http://www.chorusaccess.org/resources/us-agency-public-access-plans-details/. Accessed 31 May 2016

van Raan AFJ (2004) Sleeping Beauties in Science. Scientometrics 59(3):461–466

Rajendra Kulkarni is researcher/analyst at George Mason University (GMU) and affiliated with the Office of Research and School of Public Government and International Affairs. His research interests include nonlinear dynamical systems and their application to social sciences, complexity theory, network analysis, GIS, Transportation and Regional Science. He has BTech in Engineering and MS in Computer Science with extensive experience working with industry, business, government and academia. He holds a patent and has published in variety of disciplines with main focus on interdisciplinary model and algorithm development.

Roger R. Stough is university professor, the Schar School of Policy and Government, Vice President for Research and Economic Development (2008–2013) and currently Vice President, of the Regional Science Academy, an affiliate of the Regional Science International Association. His primary research interests are regional economic development and policy, transport policy, and entrepreneurship and innovation policy. He holds an honorary doctorate of philosophy, University of Jonkoping, Sweden.

Chapter 13
Regional Policy and Fiscal Competition

Santiago M. Pinto

13.1 Introduction

Almost everywhere around the world, governments implement policies aimed at developing certain geographic areas or regions within their respective countries. While in some places the responsibility of designing and implementing such policies falls entirely on the federal or central government, in others the decisions are made by sub-national, regional, or local authorities. The decentralization of these responsibilities has its benefits. Local and regional authorities have an informational advantage over the central authority about the region and local economy, so they are better positioned to understand the needs and demands of local residents. However, a decentralized decision making process may potentially lead to large inefficiencies and a lower overall welfare level.

The design of regional policies in an integrated system of regions presents multiple challenges. First, to the extent that households, factors of production, and goods are mobile, regional policies may only induce a potentially wasteful relocation of resources. Second, the policies can be counterproductive if they simply attempt to offset the effects of other policies implemented in competing regions. And third, regional policies can be harmful for the entire economy if regions engage in a fiscal game to attract resources to their regions at the expense of other regions.

The views expressed are those of the author and do not necessarily represent official positions of the Federal Reserve Bank of Richmond, or of the Federal Reserve System.

S.M. Pinto
Federal Reserve Bank of Richmond, Richmond, VA, USA
e-mail: Santiago.Pinto@rich.frb.org

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_13

Different theories explain the outcome resulting from the spatial interaction between regional governments.[1] Early work considered the case in which regional governments compete in capital tax rates to attract mobile firms to their regions. The main conclusion from this literature is that competition leads to a "race to the bottom", in which regions choose inefficiently low levels of tax rates, and, consequently, low levels of local public goods.[2]

Regions, however, compete in multiple dimensions, not only tax rates. Some recent models of fiscal competition extend the previous analysis and examine more generally how mobile resources (individuals, firms) respond to different types of regional policies. For instance, in order to attract businesses and increase production, local authorities may decide to invest in local transportation networks, devote resources to reduce local crime, or offer different types of public services and regulatory policies. Other models argue that the presence of informational externalities may explain why regional policies in one region may affect other regions, even when factors of production are immobile. Such models, generally referred to as "yardstick competition" models, assume that local voters obtain information from outcomes observed in neighboring regions and use that information as a benchmark against which they can evaluate the performance of local government officials. As a result, when deciding their own policies, local officials are forced to consider what happens in those other regions. Regional competition in richer institutional settings may lead to completely different conclusions than those predicted by the basic tax competition model.[3]

A wide range of questions arises concerning the design and implementation of regional policies in a context as the one described above. For instance, what exactly is the role of regional governments in promoting local economic growth? Under what conditions do regional policies contribute not only to the development of a particular region, but also to the development of the nation as a whole? To what extent is it desirable to allow a complete decentralization of the decision making process as opposed to one where decisions are made in a coordinated or centralized way? Considering that the location decisions of productive factors are affected by multiple causes, are such regional policies really effective? How should regional economic development policies be designed, for instance, in the presence of agglomeration economies? This chapter attempts to examine how recent work on regional policies and fiscal competition has addressed some of these questions.

[1]The literature on fiscal competition, both theoretical and empirical, is very extensive. For survey papers in this area, see Brueckner (2006), Wildasin (2003), Wildasin (2006), or Zodrow (2010).

[2]Wilson (1986) and Zodrow and Mieszkowski (1986) were the first papers to formalize this idea.

[3]Even though this chapter examines exclusively models of fiscal competition involving regional governments (also known as models of horizontal fiscal competition), it is important to acknowledge that a strand of the literature has focused on the outcomes arising from the strategic interaction between different levels of governments (vertical fiscal competition). In the latter case, strategic behavior may be triggered by the fact that the regional and federal governments levy taxes on the same tax base.

At the same time, it intends to identify issues that require further study and provide guidance on the direction of future research in the area.

The main idea stressed in this chapter is that further advancements in the literature on fiscal competition would require the use and development of methodological innovations, including the integration of quantitative techniques within a more structural approach. From a policy standpoint, it is not only important to determine whether fiscal decisions in one region influence other regions, but also to quantify such effects. While the empirical work has thus far provided strong evidence indicating that governments do behave strategically, this approach cannot entirely assess the magnitude and significance of such effect. Moreover, the empirical line of research is subject to numerous identification challenges. Structural methods, which analyses data through the lens of a well-specified parametric theoretical framework, may potentially offer new insights regarding the importance of fiscal competition. However, this approach is by no means straightforward and may have its drawbacks. First, the estimation and identification of the underlying structural parameters (in other words, those parameters that are truly exogenous and invariant to policies), requires an extensive use and availability of detailed micro level data. The lack of appropriate data may lead to imprecise parameter estimates, and, consequently, to inaccurate predictions concerning the impact of different policies. Second, computational requirements for this approach are generally quite large as well. Theoretical models become very complicated, so the implementation of the structural methods is performed on simpler model specifications. This process entails introducing several assumptions (for instance, about functional forms, stochastic distribution of shocks, and so on), which are generally non-testable.

Structural methods are already widely used in other fields in economics. The research in fiscal competition can be developed further by adopting some of these techniques; many other areas in regional economics would benefit as well. Such an approach, however, should be viewed as complementary to the other current research efforts that employ clever identification strategies based on experimental or quasi-experimental data.

The chapter is organized as follows. Section 13.2 offers a brief description of the kind of regional policies considered in the present chapter and Sect. 13.3 defines fiscal competition. Section 13.4 provides a brief critical summary of the main findings and challenges faced by the current literature on fiscal competition. Finally, Sect. 13.5 provides guidelines for future research.

13.2 Regional Policies

Economic development is not evenly distributed across space. Moreover, disparities between some regions tend to persist over time. Regional policies are typically justified on the basis that they intend to reduce the gap between the development of regions, integrate lagging regions, reduce economic and social disparities, or revitalize certain geographic areas. Policies aimed at achieving these goals are generally classified into two broad categories: place-based and people-based programs.[4] People-based programs are designed to assist low-income households regardless of their place of residence. The main objective is to improve low-income households' access to education and job opportunities, and ultimately facilitate their migration to other areas offering better prospects. Examples include income support programs (such as the Earned Income Tax Credit in the USA), education assistance programs, housing assistance programs based on housing vouchers, and job training programs. Place-based policies are usually divided into pure place-based interventions and place-based people strategies. Both interventions target specific distressed areas. However, while pure place-based policies are intended to develop or revitalize a specific region, place-based people strategies are directed to benefit a specific group of households residing in the area. Programs that seek the revitalization of the central business district, for example, would qualify as pure place-based policies, and enterprise zone programs aimed at attracting jobs to areas where poor households reside and job opportunities are weak, are examples of place-based policies targeted to people.

In practice, federal governments are generally responsible for implementing broader people-based policies. However, regional governments occasionally rely on similar policies to help some people move out of disadvantaged neighborhoods and locate in places that offer better opportunities within the same region. The implementation and design of place-based programs fall under the sphere of federal, sub-national, and local governments.

Even though institutional arrangements differ across countries, multiple levels of governments are generally engaged in delivering regional policies, which clearly introduces economic and organizational challenges. In the first place, the academic literature does not seem to offer clear cut rules on the optimal allocation of responsibilities across different levels of governments concerning regional policies. The public economics literature, and in particular the research work on fiscal federalism, has examined closely the trade-offs associated with alternative allocation of responsibilities across different levels of governments.[5] While there is some agreement that certain responsibilities should be maintained under the orbit of the central government, there are also economic reasons to decentralize some decisions to lower levels of governments, a process that has been taking place worldwide.[6] The rationale behind the benefits of fiscal decentralization can be traced back to the seminal work by Tiebout (1956). Tiebout suggests that a decentralized provision of public goods leads to an efficient outcome. The basic idea is that communities choose their levels of public goods, and mobile consumers sort themselves

[4]See Ladd (1994) for a detailed classification of policies aimed at dealing with distressed areas.

[5]See Epple and Nechyba (2004) for a survey on fiscal decentralization.

[6]For example, it is generally believed that regional governments are limited in their capabilities of redistributing income because household mobility reduces the effectiveness of these policies. As a result, the federal government should undertake this responsibility.

according to their preferences for those goods.[7] Decentralization, however, may entail potential costs, especially when households and factors of production are mobile and regional governments only have access to distortionary taxation. In this environment, regional policies aimed at attracting factors of production are potentially damaging because they would simply encourage a wasteful relocation of resources.[8]

In the second place, a decision-making process as the one described above, which combines the participation of multiple levels of governments, requires a significant amount of coordination, both between regions and the federal government and across regions. Failure to articulate such efforts would lead to other kinds of inefficiencies, associated with the organization and adminstration of regional policies, which may result in an overlap of programs and duplication of efforts and, even worse, the possibility that efforts at one level negate efforts at another level of government.

A large strand of the literature on regional economics studies the effectiveness of regional policies in achieving their intended objectives. The present chapter focuses precisely on one aspect that may condition their effectiveness: the existence of fiscal competition. The theoretical literature supports the idea that the implementation of uncoordinated regional policies may encourage regions to behave strategically. It is claimed that this kind of behavior initiates a process of fiscal competition, which eventually leads to less favorable outcomes for the region and the nation as a whole. If this conclusion holds, the ability of regional policies to help develop a region or revitalize a specific area would be severely constrained. This chapter does not really focus on people-based policies since, as stated earlier, they are largely implemented by central governments.[9] The policies under consideration are those that target regions, including place-based and place-based people strategies. As explained earlier, these policies are commonly thought to trigger the kind of fiscal competition explained above. The next section revisits the conditions for this to happen.

13.3 Fiscal Competition

The work by Tiebout mentioned earlier highlights some of the beneficial aspects of decentralization and fiscal competition among regions. The Tiebout model concludes that under certain assumptions, the provision of local public goods is

---

[7] We will revisit the assumptions required for this conclusion to hold in Sect. 13.4.

[8] See Brueckner (2004) for a detailed discussion on the benefits and costs of centralization versus decentralization. Wildasin (1991) focuses on income redistributive policies and explores what happens when different levels of government are responsible for undertaking such polices.

[9] All policies should aim at improving the well-being of people. The difference between policies rests on the specific way in which they are structured to achieve the intend goal. To some extent, the discussion about which policy is better has been exaggerated. Empirical evidence is far from conclusive, suggesting that no policy clearly dominates another.

efficient.[10] This result deviates from the traditional public economics literature, which states that the private provision of public goods is inefficient due to the well-known preference revelation problem. The mechanism proposed by Tiebout claims to solve this issue by asserting that in an economy with many communities, consumers would reveal their preferences by "voting with their feet" and sorting themselves into communities that offer their most preferred level of local public goods. The ability of consumers to costlessly move across communities disciplines local governments in the sense that they are forced to compete with other regions. As a result, regional governments end up providing local services at the lowest cost and offering goods and services that closely match consumers' preferences.

Fiscal competition within the Tiebout framework is viewed as something positive. However, for such a mechanism to lead to an efficient outcome, the provision of local public goods should not, among other things, generate spillover effects across jurisdictions. In principle, two conditions should be met for regional governments to engage in a game of fiscal competition. First, policies in one jurisdiction should affect other jurisdictions, and second, regions should choose their policies in a decentralized and uncoordinated way. The key issue is that when choosing such policies, regional policymakers behave strategically and do not internalize the consequences of their decision on other locations.

The traditional models of fiscal competition focus on the negative side of decentralization. In those models, policymakers compete for mobile resources (factors of production, goods, etc.) by choosing different regional policies.[11] Such policies, however, generate external effects on other regions because they ultimately encourage the spatial relocation of resources. For instance, consider a local policymaker who in an effort to revitalize local economic activity offers tax benefits and other advantages to businesses locating there. The policies could either target businesses already operating in other jurisdictions, or new businesses considering where to setup their factories or stores. Because it is expected that policymakers in other regions will behave in a similar way, the final outcome would typically be sub-optimal.[12]

---

[10]The complete list of assumptions is the following: (i) consumers are perfectly mobile; (ii) there is full information; (iii) the number of communities is large; (iv) income is exogenous; (v) there are no spillovers across communities; (vi) the average cost curve for the provision of local public goods is U-shaped (i.e. there exists a cost minimizing population size); and (vii) communities with population sizes below (above) the cost minimizing size seek to expand (contract).

[11]The basic tax competition model, for example, focuses on the strategic determination of capital tax rates. Regional governments, however, compete on many other dimensions, so fiscal competition takes place through multiple channels.

[12]In the basic tax competition literature, the outcome is characterized by a lower than optimal spending on local public goods. The fact that the size of the local public sector becomes smaller under tax competition is not perceived as something necessarily bad by the public choice literature. This literature (see, for example, Brennan and Buchanan 1980) questions the assumption that governments are benevolent and choose the tax rates that maximize the welfare of their citizens. The primary objective of self-interested, rent-seeking governments, according to this view, is to

Competition between regional governments can be welfare-enhancing, though. Regions may compete to become more attractive, and this process may benefit everyone, but for different reasons than those pointed out by Tiebout. For instance, by attracting firms, regions may exploit the benefits of agglomeration economies. This point will be developed further in the next section.

Another strand of the literature argues that in the context of a political equilibrium model, regional policies in one region may also affect other regions even when factors of production are immobile.[13] Consider the case of a local authority seeking re-election. To the extent that local voters derive information from neighboring jurisdictions to evaluate the performance and select their respective governments officials, local authorities will have incentives to mimic the behavior observed in other regions. This theory, generally referred to as "yardstick competition", suggests that policies chosen in other regions contain valuable information and generate informational externalities. Local voters eventually use this information to constrain the behavior or "discipline" local authorities.[14]

13.4 Comments on the Current State of the Literature

This section begins by introducing a simple model of fiscal competition which helps illustrate some of the most relevant findings of the theoretical and empirical literature. It also highlights some of the topics that require deeper understanding and will drive the empirical and theoretical research in this area.

13.4.1 A Simple Fiscal Competition Model

Consider an economy with two regions, i and j.[15] The payoffs or utility of a representative agent in region i, $V^i(\mathbf{z})$, depends on policies chosen in i and j,

maximize tax revenues and use those resources for unproductive purposes. So anything that would limit such behavior would be an improvement for the society.

[13]See, for example, Besley and Case (1995).

[14]Decisions by regional governments may generate other types of spillovers, such as direct geographic spillovers, which do not affect the relocation of mobile factors of production. For instance, consider spending on pollution abatement. When a region devotes more resources to reduce pollution, it creates a positive externality on neighboring jurisdictions. The latter does not include the case in which regions strategically choose regulatory policies to attract firms. Also, parks and other local amenities provided by one jurisdiction may be enjoyed and visited by residents from other locations. The decentralized solution in these two examples, as in the previous cases of fiscal competition, will also be sub-optimal, and a system of inter-regional transfers would be required to restore efficiency. This chapter will not focus on these external effects, however, since in principle they do not generate fiscal competition among regions.

[15]The model presented here follows closely Brueckner (2006).

where $\mathbf{z} = (z^i, z^j)$.[16] As in the basic models of fiscal competition, we assume the representative agent is completely immobile. Region i chooses the value of z^i that maximizes $V^i(\mathbf{z})$ taking z^j as given. The first-order condition $\partial V^i(\mathbf{z})/\partial z^i \equiv V^i_{z^i}(\mathbf{z}) = 0$ implicitly determines z^i as a function of z^j.[17] Jurisdiction j chooses z^j in a similar way.[18] The Nash equilibrium policies are given by a combination $\mathbf{z} = (z^i, z^j)$ that satisfies the first-order conditions for each region.

The Nash equilibrium is generally defined in terms of best-responses. From the first-order conditions, we obtain the best-response functions for i and j, $b^i(z^j)$ and $b^j(z^i)$, respectively.[19] The intersection of the best responses determines the Nash equilibrium $\mathbf{z} = (z^i, z^j)$.

Following the game theory literature, strategies are classified as strategic complements or strategic substitutes. This distinction becomes useful to characterize the possible outcomes of the fiscal competition game. If the best response of a region i is to increase z^i when region j increases z^j, then strategies are strategic complements and the game supermodular. In the model above, z^i and z^j are strategic complements when $b^{i\prime}(z^j) > 0$ and $b^{j\prime}(z^i) > 0$. If, alternatively, $b^{i\prime}(z^j) < 0$ and $b^{j\prime}(z^i) < 0$, then z^i and z^j are strategic substitutes and the game is submodular.

Note that the slope of i's best-response is $\partial z^i/\partial z^j \equiv b^{i\prime}(z^j) = -V^i_{z^i z^j}(\mathbf{z})/V^i_{z^i z^i}(\mathbf{z})$. Since by the second-order condition $V^i_{z^i z^i}(\mathbf{z}) < 0$, then $b^{i\prime}(z^j)$ and $V^i_{z^i z^j}(\mathbf{z})$ should take the same sign. This means that strategic complementarity and substitutability can be defined in terms of the impact of a policy on the marginal payoffs of the other region. For example, if the effect of z^i [z^j] on $V^j_{z^j}(\mathbf{z})$ [$V^i_{z^i}(\mathbf{z})$] is positive, then z^i and z^j are strategic complements; if instead, the effect is negative, then they are strategic substitutes.

The next step generally entails comparing this Nash equilibrium to a benchmark case in which policies are decided in a centralized way.[20] Specifically, consider a central authority that chooses the policy vector $\mathbf{z}$ with the objective of maximizing joint welfare $W \equiv V^i(\mathbf{z}) + V^j(\mathbf{z})$. The value of z^i in this case would satisfy $W_{z^i} \equiv V^i_{z^i}(\mathbf{z}) + V^j_{z^i}(\mathbf{z}) = 0$. The first term captures the direct impact of z^i on jurisdiction i and the second term accounts for the external effect of policy z^i on jurisdiction j. If $V^j_{z^i}(\mathbf{z}) > 0$, policy z^i has a positive external effect on j or generates positive

[16] Note that in the case of R regions, $\mathbf{z}$ would simply represent an $(R \times 1)$ vector of policies.

[17] Assuming that the second-order condition for a maximum $\partial^2 V^i(\mathbf{z})/\partial(z^i)^2 \equiv V^i_{z^i z^i}(\mathbf{z}) < 0$ holds.

[18] It is generally assumed that decisions are made simultaneously. Among other things, this setup considers a situation in which no jurisdiction have the advantage of moving first.

[19] It should be noted that in general region i's best response to z^j may consist of a set of values of z^i. This means that $b^i(z^j)$ is actually a correspondence rather than a function. When not necessary, we will stay away from such technicalities and refer to $b^i(z^j)$ simply as a function.

[20] While some papers derive the optimal allocations from the central planner's problem and compares the optimal allocations to the decentralized solution, most of the literature simply compares the outcomes arising in the non-cooperative (decentralized) and cooperative (centralized) cases. In the latter case, it is assumed that governments have access to the same set of policy instruments.

spillovers, and if $V^j_{z^i}(\mathbf{z}) < 0$, it has a negative external effect on j or generates negative spillovers.[21] Similar expressions can be obtained for z^j.

The rules determining the policies in the decentralized and centralized cases do not generally coincide. This can be seen by evaluating the decentralized first-order condition for z^i, $V^i_{z^i}(\mathbf{z}) = 0$, at the corresponding centralized condition, which gives $W_{z^i} \equiv V^j_{z^i}(\mathbf{z})$. To the extent that $V^j_{z^i}(\mathbf{z})$ is different from zero, the two solutions would clearly differ. Moreover, the sign of $V^j_{z^i}(\mathbf{z})$ reveals whether there is under- or over-provision of z^i in the decentralized case relative to the centralized solution.

13.4.2 Assumptions and Characterization of Results

There is no fiscal competition model that captures all possible spatial interactions in a single analytical framework. Currently available models highlight only one or just a few factors affecting the strategic behavior of regional governments. Also, models are generally built on different assumptions, depending on the specific problem under study, reaching different, and sometimes, conflicting outcomes. Below, the chapter briefly reviews some of the approaches frequently taken in the literature and examines the implications of adopting those specific directions.

13.4.2.1 Strategic Complementarity

One of the most compelling results in the traditional tax competition literature is that strategic interaction between regional governments leads to a "race to the bottom". In other words, the Nash equilibrium of the game is characterized by capital tax rates that are too low, which leads to an underprovision of local public goods. The "race to the bottom" result holds under at least two conditions. First, a higher level of z^i [z^j] creates a positive spillover effect on region j [i], or $V^j_{z^i}(\mathbf{z}) > 0$ [$V^i_{z^j}(\mathbf{z}) > 0$] (policies are plain complements). Second, regional policies are strategic complements, or the best-response functions have positive slopes, i.e., $b^{i\prime}(z^j) > 0$ and $b^{j\prime}(z^i) > 0$. In other words, the tax competition model is supermodular.

Showing strategic complementarity is convenient because additional insights can be obtained from the literature on supermodular games. For instance, it follows that under strategic complementarity the coordination of tax policies would lead to an efficient outcome. It also results that as the number of competing regions increases, tax rates and tax revenues decline further, and the return on capital increases.

[21] Note that this is different to the effect of policy z^i on the marginal payoffs V_{z^j} defined earlier, or $V^j_{z^j z^i}(\mathbf{z})$. Eaton (2004) classifies policies depending on their effects on the payoffs of the other region. Specifically, policies are called plain complements when the spillover effects are positive, and plain substitutes when they are negative.

However, proving that tax rates, or more broadly regional policies, are strategic and plain complements is not straightforward. The fiscal competition literature is very extensive and several modelling approaches have been followed, making it difficult to establish a general characterization of those properties. A few papers derive conditions under which $V^i_{z^j}(\mathbf{z}) > 0$, $V^i_{z^i z^j}(\mathbf{z}) > 0$ (and similarly for region j), but only for certain types of games.[22]

In general, the theoretical work on fiscal competition has not kept pace with the developments of the literature on supermodularity: the results emerging from this literature have not been fully exploited. The latter is an example where significant progress can be made by relying on advancements achieved in fields outside of urban, regional, and public economics.

13.4.2.2 Strategic Substitutability

If, in contrast, z^i and z^j are strategic substitutes, so that the best-response functions have negative slopes, and the game is submodular, then tax coordination is not necessarily efficient. In fact, either under- or over-provision of the local public good may take place under these conditions. There are numerous models of fiscal competition in which regional policies are strategic substitutes. For example, Wildasin (1988) shows that tax rates become strategic substitutes (i.e., there is a negative relationship between tax rates) when governments strategically choose the level of public goods, and tax rates residually adjust to finance the cost of providing those goods. The rationale is that when other jurisdictions decrease their tax rates, a given region experiences a decline in its tax base due to capital outflows. In order to compensate for such loss and collect enough tax revenue to finance the local public goods, the region is then obligated to raise its own tax rate.

13.4.2.3 Homogeneous or Heterogeneous Agents

Some fiscal competition models assume that regional governments are benevolent and choose the policies that maximize the welfare of local residents. Following this approach requires making several modelling choices which eventually affect the conclusions. For instance, households residing in the region may be homogeneous (in terms of endowments, preferences, etc.), in which case a benevolent regional government would choose the policies that maximize the welfare of a representative agent. Alternatively, the regional policy could have conflicting effects if local residents are heterogeneous. Capital movements, for example, may affect the return of local factors of production in different and maybe conflicting ways. Within this environment, to the extent that a specific regional policy induces the relocation of

[22] For example, Rota-Graziosi (2015) provides conditions to establish the plain and strategic complementarity of capital tax rates in the traditional tax competition.

capital, the policy would be favored by some residents, and opposed by others. Pinto and Pinto (2008) consider a political economy model in which a domestic government imposes a tax on foreign direct investment (FDI). The paper shows that the level of the tax rate will depend on the technological relationship between FDI and domestic factors of production.

13.4.2.4 Ownership of Capital

Analyzing the effects of fiscal competition in a general equilibrium model requires an assumption on the distribution of capital endowments or ownership structure. Most models assumes that ownership is uniformly distributed across immobile households in the entire federation. If capital is not employed where households reside, capital returns should then be expatriated to their respective owners, generating financial flows from capital importing regions to capital exporters. The latter effects complicates the analysis further, and conclusions generally become ambiguous. To avoid such complications most models focus on symmetric equilibria at which capital does not move.[23]

13.4.2.5 Relationship Between Private and Public Goods

In the tax competition models, for example, the specific relationship between local public goods and private goods in consumption, specifically the degree of substitutability or complementarity among these goods, is critical for determining the sign of $V^i_{z^i z^j}$. The literature has worked around this problem by making further restrictive assumptions to restore strategic complementarity. These approaches include assuming that regional governments are tax revenue maximizers (instead of utility maximizers), or that private consumption is always valued more than public good consumption (for example, by using a linear utility function with constant marginal rate of substitution).[24]

13.4.2.6 Multidimensional Policy Space

While most of the literature examines the implications of fiscal competition when regions compete on one specific policy instrument, in reality, the policy space is multidimensional. Regional governments do not only choose tax rates, as in the basic tax competition model, but they decide on the structure of the entire tax

[23] Another complication arises when households are also mobile, in which case it is necessary to keep track of where capital is employed and where households end up residing.

[24] See, for instance, Brueckner and Saavedra (2001), or Vrijburg and Mooij (2016), for further discussion.

system (including who and what to tax and how much to tax), types of regional goods to offer, the structure of the regulatory environment, and so on.[25] Additionally, those policies may affect the localization decisions of multiple mobile resources in different ways.

An implication of the previous statement is that fiscal competition may not necessarily lead to an inefficient outcome after accounting for all the externalities generated by regional policies.[26] In terms of the model, this means that $V^j_{z^i}(\mathbf{z})$ could be zero under fiscal competition after accounting for the net impact of different policies with potentially conflicting effects. For example, suppose that the production function in region i is defined by $F^i(L^i, K^i, G^i)$, where L^i denotes employment, K^i capital, and G^i a regional public good (such as infrastructure). Moreover, region i finances G^i with a tax on capital, and $F^i_{K^iG^i} > 0$, so that a higher level of G^i increases the marginal productivity of capital. When region i increases the capital rate, G^i increases as well, which eventually raises the marginal productivity of capital. If this effect offsets the additional cost of a higher capital tax rate, then the change in the regional policy would not produce an interjurisdictional spillover.

Another related example is explained in Pinto (2007b). In that paper, local authorities finance local law enforcement policies using a tax on capital. While a higher tax rate induces capital to relocate to other jurisdictions, generating a positive externality, a higher level of local law enforcement expenditures shifts crime to neighboring areas, producing an externality on the opposite direction.[27]

13.4.2.7 Symmetric Equilibria

The literature has generally focused on symmetric equilibria. The reason is that the prospects of obtaining unambiguous results in the asymmetric cases are very limited. Although appealing from an analytical point of view, it should be acknowledged that this simplistic approach cannot provide a complete understanding of the implications of fiscal competition. However, the analysis of other types of

[25] The theory on supermodular games has been used until now simply to characterize the solution of the basic model of fiscal competition. However, to some extent, this theory has been underexploited in the literature. This theory is a lot more general, and can be applied to more complicated games in which regions have access to multiple policy instruments.

[26] For instance, the effects may compensate each other.

[27] Pinto (2007a) considers another model in which regional governments choose more than one policy instrument. Specifically, the paper studies the strategic determination of the structure of the corporate profit tax system when firms operate in multiple regions and regions use an apportionment formula to calculate the proportion of the firms' income subject to regional taxation. The formula, as used in the USA, assigns weights to the proportion of the firm's total sales, property, and payroll in a state. In the model, regions choose both the corporate income tax rate and the type of formula.

equilibria, in more complicated settings, would require the use of various numerical and computational techniques.

13.4.2.8 Agglomeration Economies

Models built on the new economic geography literature provide additional insights. These models study fiscal competition in the presence of agglomeration economies. The main conclusion from this literature is that regions offering greater agglomeration advantages are able to sustain higher tax rates. The intuition is as follows: when the firm produces in that region, it benefits from the positive effect of the agglomerative forces, and the resulting higher profits would partially offset the higher taxation costs. It has been argued that in this setting, economic integration may lead to a "race-to-the-top" rather than a "race-to-the-bottom".[28]

13.4.2.9 Other Assumptions

Many other reasons influence the outcome arising in fiscal competition models that will be too long to list in this chapter. We just briefly mention two of them. First, the supply of capital is generally assumed fixed in supply. Some of the conclusions of the basic fiscal competition models change when this assumption is relaxed. In dynamic models of tax competition that incorporate savings decisions, the total amount of capital becomes endogenous depending, among other things, on the entire profile of regional tax rates. Second, another important factor is that models typically assume that fiscal policies are chosen after localization decisions have been made. A model in which policies and agents' localization decisions are simultaneously decided introduces important theoretical and empirical challenges. Such complications arise, for example, when policies are decided through voting, and voters are perfectly mobile.

13.4.3 Summary of Findings

Decisions made in a decentralized way generate many different types of external effects. There is no single framework that examines all possible consequences of the externalities in a unified way. Existing models typically focus on one or a few aspects involving fiscal competition. The discussion thus far underscores, among other things, the ambiguity of the theoretical predictions when considering more sophisticated models. In brief, the inconclusiveness arises for several reasons. First, policy decisions are clearly made in a rich environment. Simplifying assumptions

[28] See, for example, Forslid (2005) and Baldwin et al. (2005), for a review of the literature.

are generally introduced to derive interesting results, but at the cost of realism. Second, regions have access to multiple policy instruments. While some regional policies affect neighboring regions in a specific way, others have the complete opposite impact. To evaluate the outcome of the fiscal competition game, it would be necessary to account for all possible externalities. Finally, and related to the previous point, multiple mobile factors are affected by regional policies, and occasionally, in conflicting ways.

Empirical results are also inconclusive. Early empirical work shows evidence of a "race to the bottom". However, more recent work reaches different conclusions. Spatial econometric models and techniques have been used to study the extent to which policies decided in a given region affect policies chosen in other regions. They provide empirical evidence of the degree of interdependency between regions. The precise identification of the effect of specific policies in such an environment is, however, extremely challenging.[29] Moreover, as emphasized earlier, the fact that fiscal competition across regions actually takes place in a much richer setting, makes the identification problem even more complicated. Finally, econometric models can provide a sense of the degree of interdependency across regions, but they cannot really quantify the effects nor evaluate the welfare implications of fiscal competition.

13.5 Guidelines for Future Research

The previous discussion reveals several important areas in the literature that deserve further study and that will drive much of the future research work on fiscal competition. While the theoretical literature offers different explanations on the possible channels through which fiscal competition may limit the effectiveness of regional policies, and the empirical work presents evidence supporting some degree of regional interdependency, the literature falls short of addressing some fundamental questions. Some of the questions that still remain unanswered include:

1. What is the magnitude of the effects generated by fiscal competition.
2. To what extent do currently observed regional policies differ from those considered "optimal"?
3. How strong, if any, are the incentives for regions to deviate from a centralized solution?
4. How large are the welfare losses/gains generated by fiscal competition? Are those effects strong enough to justify centralized intervention?
5. How are the welfare losses/gains distributed across regions?

To shed light on these issues, it will be necessary to adopt and develop novel approaches and techniques. A number of methods, such as quantitative techniques

[29]These challenges have been pointed out by Brueckner (2006) or Isen (2014), for example.

based on structural models, are already available and used in other areas in economics. The original models of fiscal competition expose in a very elegant and simplified way the consequences of making decisions in a decentralized and uncoordinated setting. The "race to the bottom" argument is both straightforward and striking. However, more complicated models are required to address the questions raised above and to fully understand the real implications of fiscal competition at both the regional and national levels.

Recent developments in quantitative models that combine calibration and structural estimation techniques with data may help establish such connection, though. There has been in the last few years an increase in the use of these models and techniques in the fields of urban and regional economics.[30] An analysis based on quantitative models offers several advantages. First, since this approach is founded on theoretical models, it becomes easier to identify the causal effects of policies. Second, quantitative models can be used to characterize, quantify and compare policy outcomes arising in different scenarios (for example, compare policy outcomes arising in the decentralized and cooperative cases), and construct other counterfactual policy evaluation exercises. And third, they can be used to perform a welfare analysis of regional policies, quantifying their impact on the entire economy and on each individual region.[31]

Arguably, this approach may have some drawbacks. For instance, a less simplistic approach would not produce a closed-form solution or equation that can be easily estimated empirically. Moreover, relating such more complicated models to data becomes challenging. Current fiscal competition models should be adapted and become more flexible in order to fit the data. The latter typically involves choosing parameterized versions of the models, which are generally non-testable, reducing the generality of the conclusions. Finally, quantitative models are data and computationally intensive. The number of underlying parameters to estimate rapidly increases with the complexity of the model, requiring a large amount of micro-level data, not always readily available.

Overall, the literature on fiscal competition can greatly benefit from the application of these methods and techniques. However, quantitative models in the area of fiscal competition should not preclude the use of other approaches. It is necessary, as usual, to validate the conclusions from a specific approach against alternatives. In fact, further development in quantitative analysis should not substitute, but rather complement similar advancements in theoretical and empirical work.

[30]It should be emphasized that the development of these techniques is far from uniform across different fields in economics. See, for example, Nevo and Whinston (2010) and Heckman and Vytlacil (2007) for general discussions on structural methods, and Epple and Nechyba (2004) or Holmes and Sieg (2015) for surveys on different applications of structural estimation to urban equilibrium and local public finance models.

[31]Some recent work has already taken such direction. For instance, Ossa (2015) uses a quantitative economic geography model to study the motives underlying the decision to subsidize the relocation of firms by regional governments.

References

Baldwin R, Forslid R, Martin P (2005) Economic geography and public policy. Princeton University Press, Princeton

Besley T, Case A (1995) Incumbent behavior: vote seeking, tax setting and yardstick competition. Am Econ Rev 85(1):25–45.

Brennan G, Buchanan JM (1980) The power to tax: analytic foundations of a fiscal constitution. Cambridge University Press, Cambridge

Brueckner JK (2004) Fiscal decentralization with distortionary taxation: Tiebout vs. tax competition. Int Tax Public Financ 11(2):133–153

Brueckner JK (2006) Strategic interaction among governments. In: Arnott RJ, McMillen DP (eds) A companion to urban economics. Blackwell Publishing Ltd, Malden, pp 332–347

Brueckner JK, Saavedra LA (2001) Do local governments engage in strategic property–tax competition? Natl Tax J 54(2):203–229

Eaton CB (2004) The elementary economics of social dilemmas. Can J Econ 37(4), 805–829

Epple D, Nechyba T (2004) Fiscal decentralization. In: Handbook of regional and urban economics, vol 4. Elsevier, Amsterdam, Chap. 55, pp 2423–2480

Forslid R (2005) Tax competition and agglomeration: main effects and empirical implications. Swed Econ Policy Rev 12(1):113

Heckman JJ, Vytlacil EJ (2007) Econometric evaluation of social programs, Part I: causal models, structural models and econometric policy evaluation. In: Heckman JJ, Leamer EE (eds) Handbook of econometrics, Part B, vol. 6, Elsevier, Amsterdam, Chap. 70, pp 4779–4874

Holmes TJ, Sieg H (2015) Structural estimation in urban economics. In: Duranton G, Henderson JV, Strange WC (eds) Handbook of regional and urban economics, vol 5. Elsevier, Amsterdam, Chap. 2, pp 69–114

Isen A (2014) Do local government fiscal spillovers exist? evidence from counties, municipalities, and school districts. J Public Econ 110(C):57–73

Ladd HF (1994) Spatially targeted economic development strategies: do they work? Cityscape 1(1):193–218

Nevo A, Whinston MD (2010) Taking the dogma out of econometrics: Structural modeling and credible inference. J Econ Perspect 24(2):69–81

Ossa R (2015) A quantitative analysis of subsidy competition in the US. NBER Working Paper 20975

Pinto SM (2007a) Corporate profit tax, capital mobility, and formula apportionment. J Urban Econ 62(1):76–102

Pinto SM (2007b) Tax competition in the presence of interjurisdictional externalities: The case of crime prevention. J Reg Sci 47(5):897–913

Pinto PM, Pinto SM (2008) The politics of investment partisanship and the sectoral allocation of foreign direct investment. Econ Polit 20(2):216–254

Rota-Graziosi G (2015) The supermodularity of the tax competition game. Working Paper

Tiebout CM (1956) A pure theory of local expenditures. J Polit Econ 64(5):416–424

Vrijburg H, de Mooij RA (2016) Tax rates as strategic substitutes. Int Tax Public Financ 23:1–23

Wilson JD (1986) A theory of interregional tax competition. J Urban Econ 19(3):296–315

Wildasin DE (1988) Nash equilibria in models of fiscal competition. J Public Econ 35(2): 229–240

Wildasin DE (1991) Income redistribution in a common labor market. Am Econ Rev 81(4), 757–774

Wildasin D (2003) Fiscal competition: an introduction. J Public Econ 5(2):169–176

Wildasin DE (2006) Fiscal competition. The Oxford handbook of political economy. Oxford University Press, Oxford, pp 502–520

Zodrow GR (2010) Capital mobility and capital tax competition. Natl Tax J 63(4):865–901

Zodrow GR, Mieszkowski P (1986) Pigou, Tiebout, property taxation, and the underprovision of local public goods. J Urban Econ 19(3):356–370

Santiago Pinto is a Senior Policy Economist at the Federal Reserve Bank of Richmond. His primary research interests are in the fields of urban and regional economics, public economics, and state and local public finance. His research focuses on regional policies targeted to the poor, fiscal competition, political determinants of foreign direct investment, and statistical foundations of diffusion indices. He joined the Federal Reserve Bank of Richmond in 2012. Previous faculty positions were at West Virginia University, Universidad Nacional de La Plata, Argentina, and one year as visiting professor at Syracuse University. Dr. Pinto earned his Ph.D. in Economics from the University of Illinois at Urbana-Champaign in 2002.

Chapter 14
Back to the Future: Lösch, Isard, and the Role of Money and Credit in the Space-Economy

David Bieri

> The geographic variations in interest rates are generally a mirror image of the spatial organization of the banking system and of regional differences in the economic structure of production.—Lösch (1940c, p.26, author's translation)

> It is invalid to take the position that price and monetary phenomena are merely surface manifestations and reflections of the more nearly basic and underlying relations and interactions of man with his physical environment. [. . .] Price systems and monetary institutions are in modern society an indispensable set of cultural tools, [. . .] which strongly shape the evolving organizational form and nature of man's economic and social activities. [. . .] To understand and anticipate the interaction of these forces, a knowledge of resources [. . .], and a knowledge of price, exchange control and monetary mechanisms [. . .] are each indispensable.—Isard (1956, p.6)

14.1 Introduction

The contemporary canon of regional economic theory has enshrined the 'classical dichotomy' in that it treats the spheres of money and production as analytically distinct. As such, regional theory upholds the neutrality of money in its most basic quantity-theory position that suggests it is only the absolute price level, not relative prices and interest rates, and hence real output, that is affected by

D. Bieri
School of Public & International Affairs, Virginia Tech, 140 Otey Street, Blacksburg, VA 24060, USA
e-mail: bieri@vt.edu

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_14

changes in the quantity of money, broadly conceived.[1] Regional researchers thus treat the monetary-financial system as the proverbial veil that renders money and financial interrelations a source for short-term frictions at best, but not relevant to the determination of regional market (dis)equilibria. Put differently, the dominant analytical approaches of mainstream regional economics exclusively focus on the real phenomena of a spatial barter economy in which money plays but a perfunctory role. In short, real factors determine real regional variables.[2]

Contrary to this view, I argue that the continued separation of monetary theory from price theory in regional thought represents a radical departure from the intellectual origins of the field of regional science, which has its roots in the pioneering work of August Lösch (1906–1945) and Walter Isard (1919–2010). In combining key elements of interregional trade theory and location theory, both Lösch's and Isard's treatment of monetary aspects of the 'space-economy' give rise to a *spatialized interpretation of the non-neutrality of money*—an area of research that has gained significant relevance once again in light of recent economic events.[3] Indeed, the disparate regional impacts of the financial dislocations during the crisis were a powerful reminder that the intersectoral flow of funds—always and everywhere—constitutes a local phenomenon with real effects across space.[4] In many ways, the muted post-crisis responses of monetary aggregates to the large-scale unconventional monetary policy experiments can be interpreted as long-overdue vindication of the critics of the quantity theory (cf. Minsky 1993; Marcuzzo 2002).

In a renewed engagement with regional aspects of money and credit, this chapter re-examines the monetary content in the foundational works of August Lösch and Walter Isard. The former a student of Joseph Schumpeter's and the latter a student of Alvin Hansen's, both Lösch and Isard represent important branches in the long lineage of twentieth century Continental and U.S. monetary thought, respectively. For our purposes here, we pay particular attention to Lösch's (1940a,b);

[1]See Patinkin and Steiger (1989) and Klausinger (1990) for complementary overviews on the origins of the term 'neutrality of money'.

[2]Recent attempts to infuse location theory with monetary analysis, such as Figueiredo and Crocco (2008) and Nogueira et al. (2015), represent rare exceptions in the otherwise languishing literature on money and its role in regional development. See Dow (1987) and Bieri and Schaeffer (2015) for comprehensive surveys of the literature on the treatment of money in regional economics.

[3]Isard first introduces the expression 'space-economy' in his *QJE* (1949) survey article wherein he defines the term as "concern[ing] itself with the local distribution of factors and resources as well as with local variations in prices, and thus with the immobilities and spatial inelasticities of factors and goods" (p. 478). Isard's usage of the term is clearly inspired by its German origin as *Raumwirtschaft* (cf. Weigmann 1931). While it has never found wide adoption in regional science beyond Isard, a variety of economic geographers with a political economy perspective continue to use the expression (e.g. Sheppard and Barnes 1990; Martin 1999).

[4]As an example of renewed interest in regional aspects of monetary phenomena, a string of recent empirical papers examines differences in regional price level dynamics in the wake of the crisis. See, e.g., Del Negro and Otrok (2007); Beckworth (2010); Fielding and Shields (2011); Beraja et al. (2015); Hurst et al. (2016).

Lösch (1940c); Lösch's (1940d, 1944; 1949) analysis of the spatial consequences of monetary-financial arrangements on the one hand and Isard's (1956, 1960) exploratory work on the flow of funds across space on the other hand. Above all, with regard to the view that there are important, neglected contributions in Lösch's and Isard's work far beyond their ordinarily acknowledged influence on location theory, this chapter echoes Ponsard (2007) in suggesting that both are "famous, but ignored economists".

Overall, then, this chapter has two closely related ambitions. First, it aims to document how the monetary content of August Lösch's and Walter Isard's spatial system completely disappeared from regional analysis. In fact, regional theorizing without money has its origins in the microfoundations-equilibrium transformation of the main corpus of orthodox economic theory, which now provides most of the epistemological and methodological underpinnings of contemporary regional science. Following Storper's (2013) terminology, I will refer to this dominant perspective in regional science as 'new neoclassical urban economics' (NNUE) whereby spatial heterogeneity of economic activity exclusively emerges from the optimal location choices of atomistic, representative agents (households, firms) and their respective interaction with the economies of agglomeration in equilibrium. As such, regional scientists' increasingly anaemic engagement with monetary issues during the discipline's first half-century is but a direct consequence of the axiomatic embedding of the neutrality of money in the NNUE framework. In the sense of Schumpeter's (1954) distinction between *real analysis* and *monetary analysis*, regional science today has thus completely turned its back on the latter, solely relying on the former apropos the assumption that the region as an exchange economy is fully described in terms of goods and services, and not monetary relations.

The second goal of this chapter is then to argue for the renewed importance of monetary analysis in regional science, placing the element of money "on the very ground floor of our analytic structure, abandoning the idea that all essential features of economic life can be represented by a barter-economy model" (Schumpeter 1954, p.278). In particular, I will outline a specific way for achieving this aim by re-considering the contemporary relevance of the Lösch-Isard approach to *interregional stock-flow analysis*—a key research program that once defined the core of the field (Isard 2004). As part of this argument, I contend that the intellectual roads not travelled in this regard hold the key to a promising research agenda in regional science.

The balance of this chapter is organised as follows. Section 14.2 sets the scene by retracing key intellectual developments that have induced 'monetary amnesia' in regional science. In Sect. 14.3, I present a brief genealogy of monetary thought in regional science, ascribing the central views on money, credit and banking in the work of Lösch and Isard to the monetary traditions of Schumpeter and Hansen. Section 14.4 presents elements of a future research agenda that reengages with regional aspects of money and credit, casting them as central pillars of a Lösch-Isard synthesis. Section 14.5 offers some concluding thoughts.

14.2 Crisis? What Crisis?

Propelled by his own gargantuan effort of transdisciplinary negotiation and discursive institution building, Isard's (1949, 1956) original grand vision for a 'general theory of the space-economy' at once aimed to be *unifying* and *pluralistic*. In essence, the idealized 'channels of synthesis' in the Isardian system readily called upon a broad spectrum of intellectual positions from diverse schools of economic thought, at least with regard to their stylized epistemological perspective, if not always their practical methodological approach (cf. Isard 1960). From inception, the ambition for regional science was to be a larger intellectual enterprise than the sum of its disciplinary parts.[5]

In this regard, or perhaps more precisely, in the epistemological pluralism that such a vision implies, the field may be facing its biggest crisis and may yet have to overcome its most enduring challenge. After more than sixty glorious years, regional science as a post-war scientific project has reached a historic cross-roads—a juncture that in many ways is characterized by an increasingly polarizing and siloed dialogue between regional economists and economic geographers (Barnes 2003). While the possible origins of 'regional science in crisis' and their jarring consequences were first identified some twenty years ago, (e.g., Isserman 1993; Lane 1993; Bailly and Coffey 1994; Pavlik 1995; Polèse 1995) little programmatic progress appears to have taken place since.

While the crisis in regional science (and with it the urgent need for new directions) refuses to go away two decades on, this chapter takes the view that the persistent rumours of the death of regional science are greatly exaggerated. Instead of more apocalyptic prognostication, I hope to delineate a constructive way out of what I see as mostly a self-imposed intellectual cul-de-sac. In particular, I propose a renewal from within that builds on an exegesis of the ideational and institutional foundations of the field. Rather than a plea for more transdisciplinary cross-fertilisation from related disciplines (e.g., Bailly et al. 1996), mine is thus a deliberately *narrow* stance vis-à-vis the field's original cast of characters, calling for a more conscious engagement with the history of regional science thought.

In order to contextualize my argument of a 'retrospective forecast' for the next 50 years in regional economic research, we need to briefly engage with the project's primordial epistemological and methodological roots. An important tenor among Isard and his early followers was the shared sentiment to guard against what Colander (2014) has recently criticised as the 'wrong type of pluralism' in the social sciences, i.e., pluralism that is paralyzed by cacophonic conversations and saturated

[5]Chapter 12 of Isard's (1960) magisterial *Methods of Regional Analysis*—at close to 200 pages by far the longest, but perhaps least remembered—identifies five 'channels of synthesis' that form the interdisciplinary core of regional science method: (i) Interregional input-output analysis; (ii) urban spatial structure; (iii) gravity modelling; (iv) values-social goals framework; and (v) the operational fusion of all previous channels. See also Schaeffer et al. (2012) for a recent discussion of regional science as an integrative social science project.

with language games, permitting little actual cross-fertilization of methods and approaches. Indeed, the early conceptualizations of multidisciplinary approaches to regional analysis during the early 1950s were already accompanied by these very concerns. Eventually, Isard's relentless efforts to overcome bounded disciplinarity, epistemological dissonance and methodological differences would culminate in the inaugural meeting of the Regional Science Association in Detroit in late 1954 (Isard 2003, chs. 2 and 3).

Recently, the intellectual upsets of the financial crisis have added new fuel to the original discussion about disciplinary pluralism in regional science; this time, much of the ideational soul searching is centered around rekindled tensions in the orthodoxy-heterodoxy nexus in economics proper (cf. Dobusch and Kapeller 2012; Skott 2014; Hands 2015). More specifically, much of this current wave of disciplinary introspection hinges upon an unsatisfactory treatment of the role of the monetary-financial system for the macroeconomy in mainstream economics. In many ways, the analytical integration of the real economy with the monetary-financial economy has emerged (once more) as the 'Holy Grail' of post-crisis economics (cf. Laidler 2011; Lavoie 2015).

Surprisingly, similar developments in regional science and urban economics are noticeable only by their absence. To the contrary, the neoclassical tenets of the NNUE orthodoxy seems to have emerged from the financial meltdown intact, as if vindicated, and any momentum for new post-crisis directions in urban and regional economic theory appear to have been lost. The triumph of *real* analysis over *monetary* analysis in regional science is perhaps best illustrated by the fact that even the solitary mention of 'money' in *Fifty Years of Regional Science* (2004), the predecessor to our two-volume effort here, is merely using the term as a synonym for informational frictions (McCann and Shefer 2004, p.183). The entire seventeen chapter tome does not contain a single reference to credit (in the financial sense).[6]

Above all, these conceptual lapses constitute missed opportunities to explore in more detail *how theories of money, credit and banking are brought to bear on the analysis of the space-economy*. In this sense, the disciplinary self-examination of the 'regional science in decline' debate of the past two decades has, paraphrasing Mirowski (2013), created more heat than light, and a perfectly good intellectual crisis may have been wasted in that no significant new research programs have emerged in the process.

To the extent that orthodoxy in economic thought has a tendency to emerge from heterodoxy (Davis 2008), the early multidisciplinary explorations about the scope of regional analysis present an important juncture for understanding the discipline's current orthodox intellectual core. In this sense, the research agenda outlined in this

[6]Monetary frictions that are consistent with the neoclassical dichotomy include the slow adjustment of nominal quantities, such as, for example, sticky prices, and money illusion. Importantly, this form of monetary non-neutrality would still be considered part of Schumpeter's real analysis as it predominantly concerns itself with the impact of the nominal money stock on real variables. In the same sense would Milton Friedman's monetarism also be considered as part of real analysis despite its "money does matter" maxim.

chapter describes a new heterodoxy that emerges from a return to the heterodox origins from which the regional science project arose. The next section situates the treatment of monetary phenomena in the work of Lösch and Isard within the larger pantheon of the history of economic ideas, thus laying the ideational foundations for a comprehensive research agenda on regional aspects of money and credit.

14.3 Lösch and Isard as Monetary Thinkers

Broadly speaking, monetary theory traditionally distinguishes between two separate approaches to money. The first, which includes 'metallism', develops monetary theory from the transactions, store-of-value and unit-of-account needs of a basic exchange economy with an exogenous amount of high-powered government money. The second approach, which includes 'chartalism', views money as a hierarchical form of credit which renders it essentially endogenous to the economic system.[7]

Rather than further emphasising their common roots in terms of location theory, one of the central aims to this chapter is to engage with Lösch and Isard in terms their monetary thought. More specifically, I will suggest that both Lösch and Isard can be viewed as important nodes in a long lineage of *chartalist* writings than span the monetary theories of Keynes and Schumpeter.

14.3.1 August Lösch's Schumpeterian Heritage

Lösch was a student of Schumpeter's at the Friedrich-Wilhelms-Universität Bonn, obtaining his doctorate in 1932, the final year of Schumpeter's tenure as department chair before taking a position at Harvard. It is precisely during this period that Schumpeter worked most intensively on his grand treatise on money, *Das Wesen des Geldes* (1970 [1943]), which, over the course of its forty year gestation period, experienced an inordinate amount of trials and misadventures and was only published posthumously.[8]

[7]The broad chartalism-metallism dual finds its earliest, modern systematization in von Mises (1917). Schumpeter (1954), arguably a 'chartallist' himself, classified Marx as a 'metallist' and Keynes as a 'chartalist'. See Trautwein (2000) and Arestis and Mihailov (2011) for more detailed overviews in terms of possible classifying the literature on monetary thought, including a good survey on the literature related to the 'credit view' of money.

[8]See Messori (1997) and Alvarado (2014) for a detailed chronology of Schumpeter's struggle with *Das Wesen*, the origins of which can be traced back to his *Das Wesen und der Hauptinhalt der theoretischen Nationalökonomie* (1908).

Examining the monetary content of the Löschian œuvre in more detail, I argue in Bieri (2016) that a hitherto overlooked aspect of his contribution is the development of a spatial theory of price level determination in a way that is consistent with credit theories of money, including the notion of monetary non-neutrality and money that is created endogenously. Indeed, Schumpeter's own monetary insights have shaped Lösch's thinking on spatial aspects of money and credit to a significant degree. Beyond Schumpeter's direct influence, Lösch's broader intellectual formation takes place during the waning years of the Weimar Republic, a period of intense monetary debate in Germany that—from Kahn, to Lautenbach and Neisser—was marked by a series of neglected contributions to a 'credit view' of money that has recently attracted renewed attention.[9] On these grounds alone, the lack of recognition of Lösch contributions to monetary theory, let alone his attempt to link the real and financial in a synthesis of location theory with modern credit theory represents a historical curiosity, if not a puzzle.[10]

Once in the New World, Schumpeter remained an important element in the development of Lösch's career and theorizing; it was not only with the help of his old mentor that Lösch was able to spend two extensive research stays in the U.S. on a Rockefeller Fellowship (1934–1935 and 1936–1937), but access to Schumpeter's own academic network—from Haberler, to Taussig and Hoover—became instrumental for much of the novel theorizing that shaped both the first and second editions of his path-breaking *Die räumliche Ordnung* (1940c; 1944).[11]

Figure 14.1 illustrates Lösch's rich lineage of monetary thought as a central node in a dense network of mentor-student relationships among a wide spectrum monetary theorists on both sides of the Atlantic, all of whom, to varying degrees, can be grouped as espousing a 'credit view of money' during the interwar period. Specifically, Lösch's (1940a,b,d) work on money, credit and financial markets acknowledges the importance of capital flows throughout the urban hierarchy, highlighting the spatial relationship between financial variables and institutional functions, such as interest rates or credit intermediation. Furthermore, Lösch (1949) recognizes that money and credit are fundamentally hierarchical in nature and that all money is credit money, even state money. The Löschian perspective on money and credit will be discussed in more detail in Sect. 14.4.

[9]Laidler and Stadler (1998), Klausinger (1999), and Laidler (2012) for a discussion of neglected contributions to monetary theory by German economists during the interwar period.

[10]See Bieri (2016) for more discussion of this point in particular.

[11]It is clear from Lösch's own records (partly published in Riegger 1971) that Schumpeter was more than an academic mentor, but also a personal inspiration and close friend with whom he resided several times in Cambridge, Mass. during his Rockefeller fellowship stays.

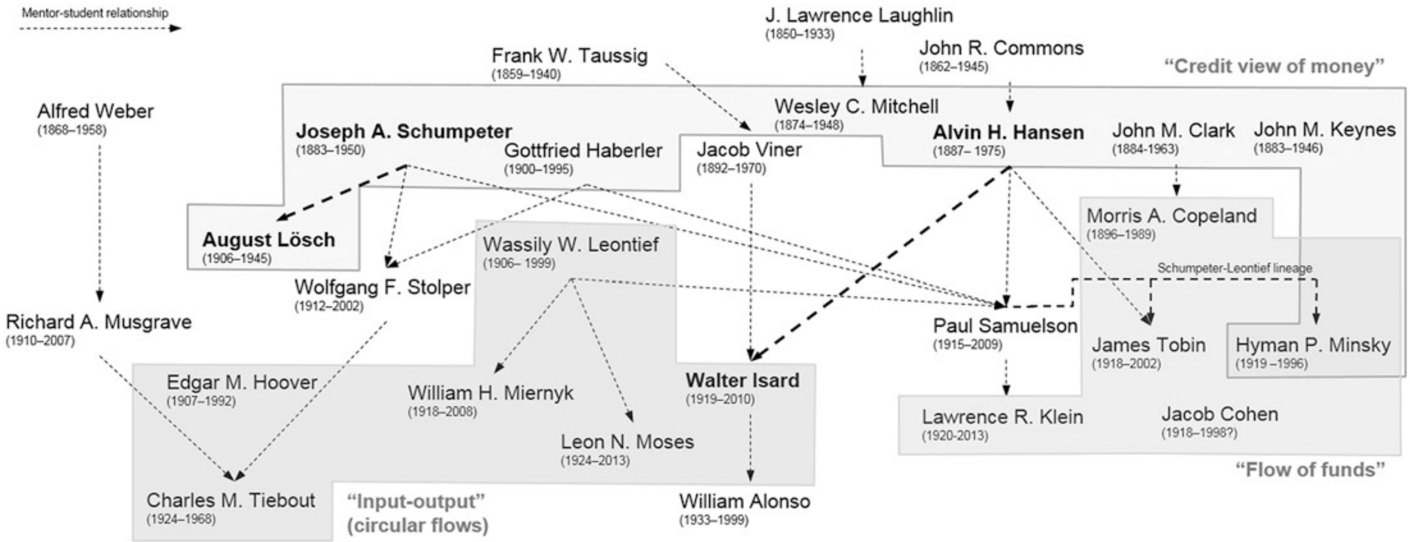


Fig. 14.1 Lösch-Isard lineage of monetary thought in regional science. *Notes*: The Lösch-Isard lineage of monetary thought is visualized as a mentor-student relationship, highlighting key areas of research in regional science ('input-output' and the 'flow-of-funds' analysis) where a 'credit view' of money perspective is instrumental to the integration of the spatial linkages between the real and financial sector. See main text for more details. *Source*: Author's illustration

14.3.2 *Walter Isard and the Influence of the 'American Keynes'*

In tracing Isard's monetary heritage, Alvin Hansen's influence stands out above anyone else. In his own account of Hansen's vital role during in his intellectual formation at Harvard, Isard refers to Hansen not only as the source for contemplating monetary factors as causes of the regional business cycle, but also as a "towering exception amid the widespread continued ignorance among Anglo-Saxon economists" with regard to the importance of location theory (Isard 2003, p.9).

At Harvard, Isard also came to study under Abbott Usher, who, in addition to his famous work on the transformational role of technology, was in the midst of a large project on the history of the early credit system in Europe (Usher 1943) when Isard arrived in Cambridge. Perhaps more importantly, Usher became, after the death of his European-trained colleague and mentor F. W. Taussig, something of a resident expert on the work of the German Historical School, particularly the work of Gustav Schmoller, who emphasized the effects of space on the trajectory of economic development (Molella 2005).

As with Lösch, a closer examination of Isard's main works reveals the clear intellectual imprinting of the mentors on the student's work—a fact that is best witnessed by the dedication of *Location and Space Economy* (1956) to both his teachers Hansen and Usher. It is Usher's influence that gave the impetus for Isard's foundational *QJE* (1949) article wherein he introduces an English-speaking general interest audience to the nuances of German location theory, including the work of Lösch. At the same time, however, Isard credits Hansen for kindling his interest in locational analysis and its relevance for to national policy (Isard 2003, p.8).

Although Hansen is mostly remembered for his 'Keynesian' stance in the context of post-war U.S. public policy, earning him the popularized moniker of the 'American Keynes' (cf. Breit and Ransom 1982), a central component of Hansen's work during the interwar period propounded a continental-style monetary theory of the business cycle—work that has regained prominence today in the context of a recent revival of his term 'secular stagnation'.[12] As a representative of the banking school tradition, Hansen played a pivotal role in the transformation of twentieth century monetary thought, advocating Keynesian fiscal activism and strong monetary restraint for economic stabilisation (Mehrling 1997, 1998). Indeed, Hansen's banking school position on the monetary transmission mechanism and credit creation is perhaps most clearly visible in Isard's own position regarding the importance and role of monetary institutions for the interregional flow of funds.

After taking courses at Harvard, Isard moved to Chicago to study for a Ph.D. where, in addition to Frank Knight and Oskar Lange, Jacob Viner soon became Isard's most important (monetary) point of reference (see also Fig. 14.1). And

[12] See Summers (2014a,b) for the contemporary revival and re-interpretation of Hansen's interwar idea of 'secular stagnation' in the context of the post-crisis limits of monetary policy to accomplish much more with interest rates at their lower bound.

perhaps in equal measure because of Viner's complex and contested role in defining the Chicago Monetary Tradition (e.g. Nerozzi 2009) and his own early exposure to Keynesian thinking at Harvard, Isard eventually positions himself against some of the Chicagoan tenets regarding "how money matters". For example, he rejects Viner's (1975 [1937]) assertion that there are "problems which fall within the domain of international trade and which distinguish it from domestic and intranational trade, particularly those associated with monetary phenomena." (Isard 1954, p.320*n*).

Little later, in his seminal *Location and Space-Economy* (1956), Isard takes an even stronger monetary stance, suggesting that "[it is] invalid to take the position that price and monetary phenomena are merely surface manifestations and reflections of the more nearly basic and underlying relations and interactions of man with his physical environment" (Isard 1956, p.6). By the time *Methods of Regional Analysis* (1960) is published as an explicit sequel to *Location and Space-Economy*, Isard has integrated his ideas on the regional role of money and credit into a ground-breaking treatment of the regional flow of funds, where linkages between the institutional evolution of money, credit and banking and the spatial structure of moneyflows form central pillars of the analysis.[13] As discussed more extensively below, Isard understood that the structure of regional economic activity is influenced by how institutional components of the monetary-financial system (financial instruments, financial markets, monetary and financial intermediaries) promote the interregional mobility of funds and, by extension, the mobility of funds among the various sectors of the space-economy.

14.4 Integrating Lösch and Isard: Elements of a Synthesis

In an extension of Schefold's (1997) characterisation of Schumpeter as a 'Walrasian Austrian' and Keynes as a 'Classical Marshallian', Lösch and Isard might each be viewed as 'Austrian' and 'Classical' with respect to their monetary ideas in general and their positions on the non-neutrality in particular.[14] In addition to details about Lösch's and Isard's intellectual lineage, Fig. 14.1 also provides a conceptual mapping of key areas of regional research where a monetary and credit perspective on the linkages between the real and financial sector offer significant promise. In particular, such work would focus on how structural linkages identified by input-

[13]Throughout, I adhere to Copeland's original terminology in his seminal *Study of Moneyflows in the United States* (1952) which uses 'moneyflows' as one word, rather than a hyphenated or two-word term.

[14]The mainstream claim about the original classical economists' adherence to the 'classical neutrality postulate', i.e., that money stock changes affect only the price level and not real output and employment, is subject to much debate (Humphrey 1991).

output analysis tie in with regional moneyflows studies within the larger context of a 'stock-flow consistent approaches' (SFCAs) to macroeconomic modelling.[15]

14.4.1 Monetary Hierarchy and Spatial Non-neutrality in the Löschian System

With regard to Lösch (1940c); Lösch's (1940d, 1949) pioneering analysis of the spatial consequences of monetary-financial arrangements, I document elsewhere (Bieri 2016, 2017) how this work contains hitherto neglected important theoretical insights for theorizing the flow of credit money across space. Specifically, I show that these lesser-known aspects of Lösch's work are broadly consistent with a spatialized version of (Post) Keynesian monetary theory, in particular with regard to some aspects of liquidity preference, the loan-to-deposit causality, and circuitist notions of the flow of funds (cf. Dow and Earl 1982; Arestis 1988, 1996; Chick and Tily 2014).[16]

At the same time, Post Keynesian monetary theory also implies what can be considered a 'hierarchy of monies' in that the modern monetary system is a hybrid, that is part public ('outside money', a net asset to the private sector) and part private ('inside money').[17] It has both public and private liabilities that circulate as money (Bell 2001; Mehrling 2013). Indeed, two specific aspects of Lösch's analysis of the spatial consequences of monetary-financial arrangements provide a useful lens for linking the hierarchy of money to the spatial structure of the financial system.

First, Lösch (1949, 1954) recognizes that money and credit are always and everywhere fundamentally hierarchical in nature and that all money is credit money, even state money. Table 14.1 illustrates the hierarchy of money in the Löschian system as a spatial monetary order where money and credit are created by different financial institutions at separate levels of the hierarchy. The Löschian monetary

[15] In their most general form, SFCAs to macroeconomic modelling are based on the strict discipline of social accounting matrices (SAM), relating all the flows and the stocks of an economy. SFCAs have their origins in the pioneering work of Copeland's (1947, 1952) flow-of-funds analysis and have recently re-gained prominence among, particularly among Post Keynesians, as a heterodox methodology for macroeconomic modelling based on stock-flow relationships, the flow of funds, interrelated sectoral balance sheets, and double-entry matrices. See Caverzasi and Godin (2015) for a survey of this literature.

[16] Throughout, I will use the convention of using the capitalized, non-hyphenated version of writing 'Post Keynesian', largely in keeping with the self-identification of the thinkers who use the label. See Davidson (1991), King (2002, pp.9–11), and Lavoie (2014, pp.42–45) for a discussion of the deep semantics behind the four different ways in which the term can be written (hyphenated or not and capitalized or not).

[17] The distinction between 'outside money' and 'inside money' goes back to seminal work of Gurley and Shaw (1960). In this context, 'outside money' is either of a fiat nature or backed by some asset that is not in zero net supply within the private sector, whereas 'inside money' is an asset backed by any form of private credit that circulates as a medium of exchange.

pyramid can be read both institutionally and, perhaps more importantly, in a functional manner, i.e., in terms of what constitutes money and credit as an accepted mean of settlement. In fact, with regard to the spatial propagation of changes in the price level, Lösch observes that the "shifting of the price level occurs only with credit creation; that is, with a hierarchy of different kinds of money, whereas in a region with a uniform currency, the price waves started by a shift in purchasing power necessarily suffice for transfer" (Lösch 1954, p.227).

A central feature of this monetary hierarchy is the fact that the distinctions between money and credit are not strict and largely depend on the specific vantage point from within each layer of the system. In this system, gold and deposits at the Bank for International Settlements are the ultimate money because they are the ultimate means of international payment. Currencies, both international money and national money, are deemed a form of credit insofar as they are promises to pay gold. Similarly, further down the hierarchy, bank deposits are viewed as a form of private credit money, effectively promises to pay currency on demand and thus twice removed from the promises to pay ultimate money. Private money in the form of debt obligations or securities is then a promise to pay currency or deposits over some specific time horizon.

Another crucial feature of this hierarchical view of money lies in the fact that at each layer the 'moneyness of credit' depends on the credibility of the promise by a given issuer to convert a specific form of credit into the next higher form of money. In other words, what counts as money and what counts as credit depends on the layer of the hierarchy under consideration, on what counts as ultimate means of settlement. The translated and augmented version of Lösch's original table in the bottom panel of Table 14.1 reveals that the Löschian monetary hierarchy maps directly into a Post Keynesian-Minskian perspective of monetary hybridity according to which the credit pyramid oscillates between a condition where money is 'scarce' and one where credit is 'elastic' (Wray 2009; Mehrling 2013).

Second, Lösch (1940c); Lösch's (1940d) work on financial markets acknowledges the importance of capital flows throughout the urban hierarchy, highlighting the spatial relationship between financial variables and institutional functions, such as financial regulation. Indeed, Post Keynesian monetary thinkers assign functional and institutional variation one of the most influential pathways for change in real-financial linkages (e.g. Dow 1982; Chick and Dow 1988, 1996). Another important, related perspective that is consistent with Lösch's work comes from Minsky's (1991, 1993) re-emphasis of Keynes's (1930) fundamental insight that the non-neutrality of money needs to be a "deep part of the system, not an afterthought in a capitalist economy" (Minsky 1996, p.78). Indeed, the similarities between Lösch's monetary thought and that of Minsky are far from coincidental: as Fig. 14.1 illustrates, both were students of Joseph Schumpeter's (Lösch at Friedrich-Wilhelms-Universität Bonn, and Minsky at Harvard).[18]

[18]The Lösch-Minsky relationship and its deep connection to the misadventures of Schumpeter's *Das Wesen des Geldes* (1970 [1943]) are discussed in more detail elsewhere (Bieri 2016).

Table 14.1 Hierarchical money in the Löschian system

1. Geld höchster Ordnung:	Weltgeld		(Bargeld: Gold; Buchgeld: BIZ)
2. Geld hoher Ordnung:	Teilgeld	Großraumgeld	(£, ℛℳ)
3. Geld mittlerer Ordnung:		Nationalgeld	(Banknoten, Zentralbankguthaben, mitunter entsprechendes Regionalgeld)
4. Geld unterer Ordnung:		Privatbuchgeld	(der Groß-, Regional-, Lokalbanken)
5. Geld unterster Ordnung:		Privatbargeld	(private oder fiskalische Schuldurkunden, besonders Wechsel)

Translated (and augmented) version:

Outside money[*]	1. Highest-order money:	Global money		(Currency: Gold; credit money: BIS[†])
	2. High-order money:	Regional money ('partial money')	International money[‡]	(£, Reichsmark)
	3. Mid-order money:		National money	High-powered money (currency, central bank reserves), occasionally equivalent regional money
Inside money	4. Lower-order money:		Private credit money	National commercial and retail banks, regional and local (community) banks
	5. Lowest-order money:		Private money	Private or fiscal debt obligations, in particular commercial paper

Source: Original table with monetary hierarchy in Lösch (1949, p.59). Author's translation and adaptation

Notes: This 'monetary order' links the hierarchy of money on the left hand side to the spatial structure of the financial system on the right-hand side

[a]'Outside money' is either of a fiat nature or backed by some asset that is in positive net supply within the private sector, whereas 'inside money' is an asset backed by any form of private liabilities (credit) that circulate as a medium of exchange, an analytical distinction first introduced by Gurley and Shaw (1960)

[b]Corresponds to both 'top currency' and 'patrician currency' in the terminology of Cohen's (1998, 2003) currency pyramid

[c]BIZ/BIS: Bank für Internationalen Zahlungsausgleich/Bank for International Settlements, Basel, Switzerland

14.4.2 The Flow-of-Funds Perspective and Isardian Monetary Space

In what follows, it will be useful to recall that Isard shared with Lösch the intellectual heritage of accessing location theory via the comparatively mature analytical apparatus of interregional trade theory.[19] Deeming to him the "most prominent living location theorist", Haberler suggests that Isard has succeeded "more than anyone else to combine trade and location theory in a comprehensive general equilibrium model comprising more than two countries and commodities as well as the space factor" (Haberler 1961, p.5*n*).

For the purposes of our discussion here, I want to highlight the conceptual importance of Isard's synthesis in terms of linking two separate but related strands of examining the sectoral structure of the regional economy. More specifically, Isard connects the structure of regional production with its corresponding moneyflows, and, in doing so, he aligns the flow-of-funds accounting pioneered by Copeland (1947, 1952) with Leontief's (1991 [1928]) conceptualization of the economy as a circular flow upon which all input-output methods, once the analytical workhorse of regional scientists, are based.

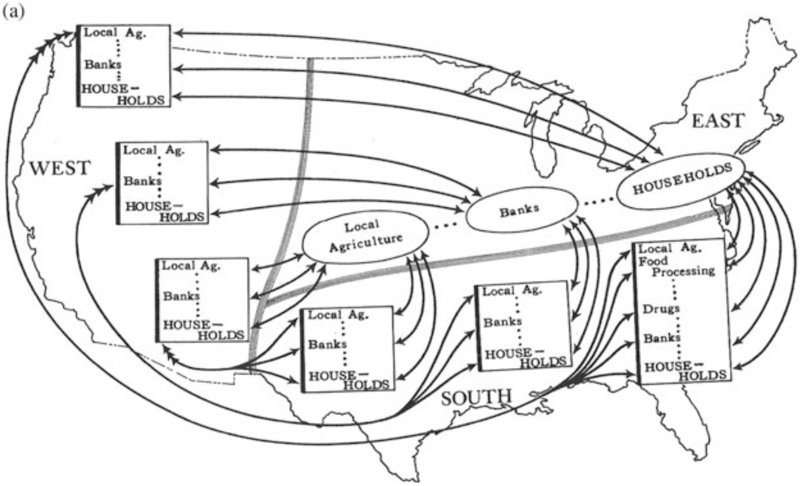
Figure 14.2 illustrates the core of this flow-based interregional system in the Isardian space-economy. Both graphics are reproduced from the first edition of Isard's seminal *Methods of Regional Analysis* (1960)—a fact that is worth highlighting here since the analysis of regional moneyflows had all but disappeared by the time the 7th and final edition of *Methods* (1998) was published.[20] The upper portion of the figure underlines the weblike connectivity of interregional moneyflows that arises from the balance-sheet relations of different sectors of the space-economy. The lower portion of the figure provides a regional money flow matrix representation of economic activity that is consistent with and expanded and refined by Cohen's (1968, 1972, 1987) work on the flow of funds.

The most important aspect of Isard's theoretical innovation lies in his vision to base interregional analysis on the implicit linkages of the three major national accounting statements, i.e., the national income and product accounts, the input-output tableaux, and the flow-of-funds accounts. In particular, his insight of complementing standard input-output relationships with monetary stock-flow data was well ahead of its time and essentially anticipates what was to become the hallmark of the rapidly expanding field of Post Keynesian stock-flow-consistent models—except for the fact that the latter have yet to develop a regional perspective! In bringing together the Isardian approach with the Löschian system, the next section now attempts to delineate the broad contours of a research agenda of the role of money and credit in the regional economy.

[19]In this regard, both Lösch's and Isard's foray into location theory can be viewed as the template for Krugman's (1998) discovery of space as the 'final frontier'.

[20]Perhaps mirroring much of the 'monetary amnesia' that befell regional science over the last half century, this development remained unremarked in all of the major reviews of the book (e.g. Rephann 2000; Stabler 2000).

(a)



(b)

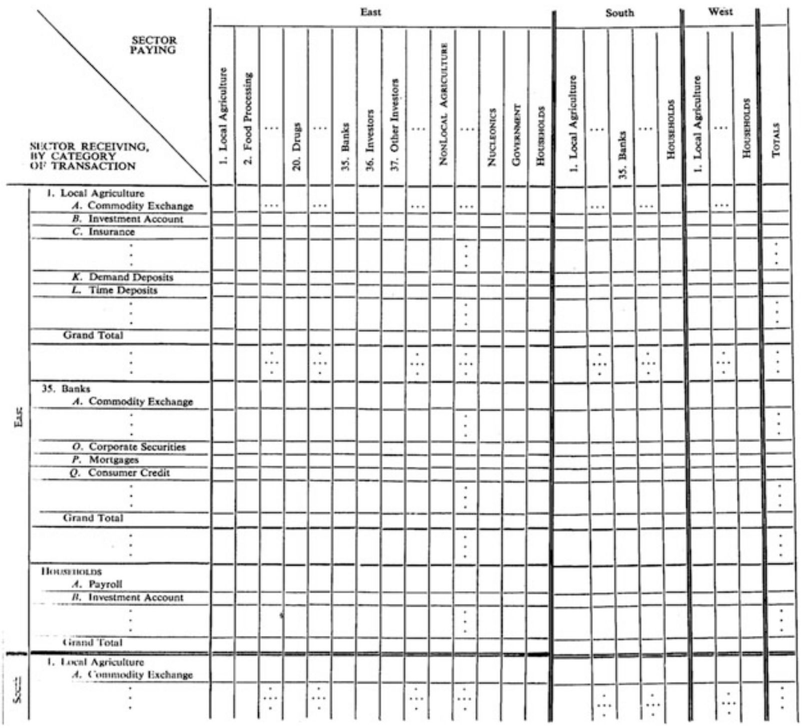


Fig. 14.2 Moneyflows across the space-economy. *Notes*: Panel (**a**) illustrates a set of hypothetical interregional moneyflows across different sector of the economy. Panel (**b**) presents a schematic representation of a corresponding flow-of-funds matrix across different sectors of the space-economy. *Source*: Isard (1960)

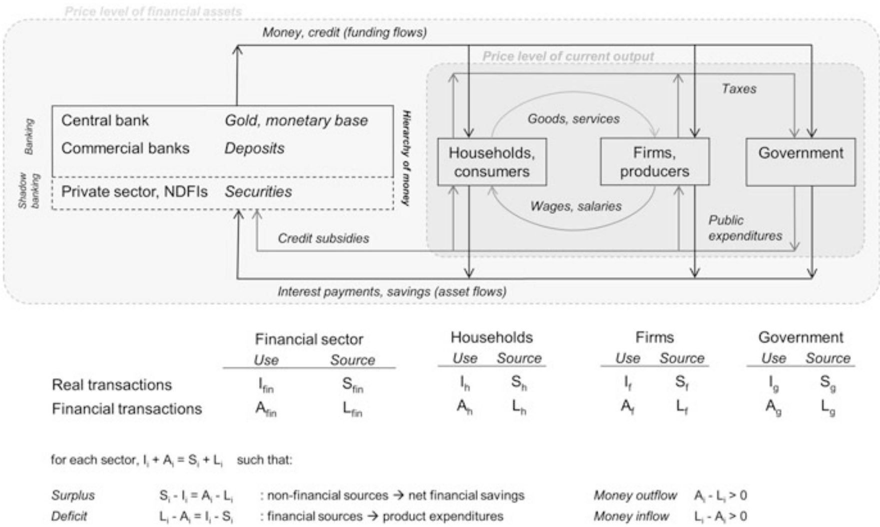


Fig. 14.3 The hierarchy of money and the flow of funds. *Notes*: Schematic representation of the flow of funds across different sectors of the economy, paying particular attention on the hierarchical relationships between different forms of money and credit. The lower portion of the figure presents a sectoral flow-of-funds table that is consistent with the money-flow accounting pioneered by Copeland (1947, 1952) and its extensions by Cohen's (1972, 1987). See main text for more details. *Source*: Adapted from Bieri (2017)

14.4.3 Linking Regional Moneyflows and the Hierarchy of Money

While the Löschian hierarchy of money provides the institutional and functional vector that underpin the spatial non-neutrality of money, Isard's regional flow-of-fund linkages form the accounting lens through which its outward appearance becomes empirically tractable. Figure 14.3 visualizes the key components of this Lösch-Isard system wherein the financial accounts follow funds as they move from sectors, such as households or firms that serve as sources of funds (net lenders), through intermediaries (financial institutions) or financial markets, to sectors that use the funds to acquire physical and financial assets.

In particular, the focus on the sources and uses of funds in the lower panel of Fig. 14.3 helps to emphasise the two key elements of the Löschian monetary system introduced above, namely, the hierarchical relationships between different forms of money and credit on the one hand and the (spatial) non-neutrality of money via the price level of output and the price level of financial assets on the other hand. In this setting, the non-neutrality of money arises from the simple fact that, for each sector i in region j, real transactions and financial transactions are closely linked as investment (I) and increases in financial assets (A) equal saving (S) and increases in financial liabilities (L) such that $I_{ij} + A_{ij} = S_{ij} + L_{ij}$.

Much of what both Lösch and Isard had originally envisioned by way of integrating the real and the financial for regional analysis took several decades before it was formalized by two Nobel Laureates, Leontief and Klein, in terms of an interface between input-output and flow-of-funds analysis (cf. Klein and Glickman 1977; Leontief and Brody 1993; Klein 2003). Despite these advances, it took several additional decades plus a major financial crisis before the importance of real-financial linkages, particularly financial flows and the composition of sectoral balance sheets, was more broadly recognized. A good six decades since it was originally conceived, flow-of-funds analysis is at long last experiencing veritable renaissance, propelled by flurry of academic and policy interests aimed at understanding central aspects of the financial crisis that the conventional equilibrium-based mainstream models were not able to capture by design (e.g. Palumbo and Parker 2009; Bezemer 2010; Winkler et al. 2013; Borio and Disyatat 2015).[21]

Table 14.2 summarizes our preceding discussion in terms of the most important conceptual differences between the orthodox view of money in regional science and its Lösch-Isard alternative. In particular, Table 14.2 compares these competing paradigms of monetary theorizing along key dimensions, namely, money, interest, prices, and the nature and structure of financial intermediation. Indeed, of the "continuing muddles in monetary theory", as Goodhart (2009) puts it, several are particularly relevant for the regional analysis of money because they are so deeply embedded in the theoretical fabric of the NNUE view of money. Above all, this includes the analysis of the monetary base multiplier of bank deposits, the current three-equation neoclassical consensus (which assumes perfect creditworthiness, and hence no need liquidity intermediation), and the analysis of the evolution of money. For each of these dimensions of monetary analysis, the last column of Table 14.2 outlines a few high-level areas of theoretical and empirically inquiry that are implied by the Lösch-Isard view. While too numerous to be elaborated in detail, I shall briefly discuss a few of the topics for expositional purposes.

For example, the financial crisis has reminded policy makers just how much the dynamics of regional cost of living adjustments depend on a clear understanding of house price movements, particularly in the U.S. where the recovery of house prices has shown substantial regional heterogeneity. Even in the absence of nominal exchange rate movements and trade barriers, some of the observed deviations from regional purchasing power party (PPP) are even more persistent than their international counterparts. Indeed, relative price levels among U.S. cities have historically shown mean reversions at an exceptionally slow rate, seemingly in contrast to recent evidence of falling transportation cost and the strong regional integration of the U.S. economy (e.g. Cecchetti et al. 2002; Chen et al. 2006). While non-traded local goods and services are one common real sector explanation for such deviations from PPP, the two-price level perspective of the Lösch-Isard view

[21] See Bieri (2017) for a discussion of the Leontief-Klein connection to Minsky's (1977, 2008) financial instability hypothesis. Cf. also the 'flow of funds' box in Fig. 14.1.

Table 14.2 Key dimensions of the monetary space-economy

	Orthodox view (NUUE-NEG)[a]	Lösch-Isard view	What are the (monetary) questions?
Nature of analysis	"Real"	"Monetary"	
Economic fluctuations	Business cycle[b]	Interaction between financial cycle, business cycle	(i) Finance-growth nexus of regional development; (ii) regional economic adjustment;
Money	Neutral, exogenous[c]	Non-neutral, endogenous	(iii) geography of money and inflation (e.g. regional money multiplier); (iv) optimal regional currency areas;
Interest	Natural interest rate[d]	Market interest rates	(v) regional interest rate differentials; (vi) regional capital market integration;
Prices	One price level (real output)	Two price levels (Financial assets, real assets/output)	(vii) regional cost of living differentials; (viii) spatial purchasing power parity, law of one price;
Financial intermediaries	Reduction of frictions, information asymmetries	Credit creation, transfer of purchasing power	(ix) regional transmission mechanism of monetary policy; (x) structure of financial intermediation (e.g. spatial disparities in credit creation by non-depository financial institutions); (xi) regulatory arbitrage across space;
Deposits	Sectoral endowments	Created by loans	(xii) regional deposit concentration; (xiii) spatial disparities in the 'moneyness' of deposits;
Source of investments	Savings	Financing flows	(xiv) regional discrepancies in liquidity preference; (xv) regional flows of finance vs. collateral; (xvi) spatial distribution of credit subsidies;
Flow of funds	Current account, net capital flows	Gross capital flows	(xvii) regional balance of payments (BoP); (xviii) classical 'transfer problem' vs. monetary approach to BoP; (xix) regional reserve flows.

Source: Author
Notes:
[a]"New neoclassical urban economics" (NNUE) and new economic geography (NEG)
[b]Real business cycle theory in the tradition of new classical macroeconomics
[c]Includes superneutrality of money, i.e., real variables are not only unaffected by the level of the money supply, but also by the rate of money supply growth
[d]The natural interest rate is unobservable, reflecting only real factors. Explanations for the source of divergences between the market and the natural rate differ between the Lösch-Isard view and the conventional view. See text for more detail

would suggest additional monetary phenomena, such as regional asset price inflation in the housing market, as an alternative causal pathway.

Similarly, discussions about regional differences in interest rates commonly assume that such divergences strictly reflect real factors, above all the balance between *ex ante* saving and *ex ante* investment which drive equilibrium in the goods market. Thus, in the standard view of real analysis, by construction, there is no difference between saving and financing (Borio and Disyatat 2011; Borio 2014). The monetary analysis of the Lösch-Isard view, by contrast, would highlight that such regional interest rate differentials could reflect a purely monetary phenomenon whereby variations in local credit conditions, not informational frictions, drive a wedge between the market rate and the (unobservable) natural rate.

14.5 The Future

Overall, a return to the foundational works of Lösch and Isard offers important opportunities for regional science's next half-century. Methodologically, the future of regional science thus lies in a rediscovery of the project's "macro-foundations", particularly in terms of its national accounting traditions that provide the intellectual undergirding for much of the work on input-output analysis and the flow of funds. Epistemologically, the future of regional science also ought to be more pluralistic by including heterodox approaches, such as Institutionalist, (Post) Keynesian, and socio-economic perspectives. With regard to the particular re-engagement of monetary analysis in Schumpeter's original sense, regional scientists could play an important role in this intellectual effort, not at least because these lesser-known aspects of Lösch's and of Isard's work are consistent with a spatialized version of central tenets of heterodox monetary theory.

In linking the structure of intersectoral money and credit flows with the structural relationships that govern the intersectoral flow of goods and services, the Lösch-Isard analytical framework outlined in this chapter aligns well with the renewed academic interest in modelling the pathways between financial markets and the macroeconomy.[22] Furthermore, this chapter has also provided the contours of a research agenda associated with the development of a spatial theory of money and credit as key research frontier for the next half-century of regional science. Specifically, I have argue that a re-engagement with the monetary foundations of the intellectual touchstones of regional science could yield a wide array of promising theoretical and empirical research for the future.

In addition to providing directions for the future, a return and re-interpretation of the foundational texts in regional science could help establish a much needed compass to anchor the enterprise more solidly in the intellectual space of its peers—something that has long been identified as being critical for reverting the discipline's

[22]See Morley (2016) for an overview of the recent literature on macro-finance linkages.

decline, or—perhaps more optimistically—for living up to its full promise and potential (e.g. Isserman 1993; Lane 1993; Isard 2004). While the reigning paradigms in regional science will not be changed overnight, it is my hope that this chapter marks a modest new beginning, not only in terms of the specific details of a revival of Lösch's and Isard's monetary tradition, but also in providing the impetus for a reconceptualization of a disciplinary grand narrative in regional science—in particular, a narrative that is faithful to the intellectual tradition of comparative political economy, including a growing '(regional) varieties of capitalism' literature and a (re)discovery of traditional business cycle and institutional thinking among urban and regional scholars.

In conclusion, then, the approach presented in this chapter argues for a return to the ideational roots of regional science and for a re-projection of its fundamental concepts whereby regional aspects of money and credit are re-cast as central pillars of a Lösch-Isard synthesis. As such, for spatial economists and economic geographers alike, regional science remains *the* systematic project tasked with advancing the theory and measurement of the *space-economy*.

Acknowledgements I thank the editors, Randy Jackson and Peter Schaeffer, for their detailed comments and encouragement on previous versions of this chapter. I am indebted to Scott Campbell for rich discussions of the intellectual history of regional science which were critical to my understanding of the Lösch-Schumpeter connection. I am also grateful for input from seminar participants at West Virginia University's Regional Research Institute and at Virginia Tech, and from session participants at the 2015 Regional Science Association International meetings (Portland, OR), the 2016 Association of American Geographers meetings (San Francisco, CA), and the 2016 History of Economic Society meetings (Durham, NC).

References

Alvarado R (2014) Introduction to the English edition. Schumpeter's treatise on money, Wordbridge Publishing, Aalten, pp xxxiii–xxxviii

Arestis P (ed) (1988) Post-Keynesian monetary economics: new approaches to financial modelling. New directions in modern economics, Edward Elgar, Aldershot

Arestis P (1996) Post-Keynesian economics: towards coherence. Camb J Econ 20(2):111–135

Arestis P, Mihailov A (2011) Classifying monetary economics: fields and methods from past to future. J Econ Surv 25(4):769–800

Bailly AS, Coffey WJ (1994) Regional science in crisis: a plea for a more open and relevant approach. Pap Reg Sci 73(1):3–14

Bailly AS, Coffey WJ, Gibson LJ (1996) Regional science: back to the future. Ann Reg Sci 30(2):153–163

Barnes TJ (2003) What's wrong with American regional science? a view from science studies. Can J Reg Sci 26(1):3–26

Beckworth D (2010) One nation under the Fed? the asymmetric effects of US monetary policy and its implications for the United States as an optimal currency area. J Macroecon 32(4):732–746

Bell S (2001) The role of the state and the hierarchy of money. Camb J Econ 25(2):149–163

Beraja M, Fuster A, Hurst E, Vavra JS (2015) Regional heterogeneity and monetary policy. Staff Report No. 731. Federal Reserve Bank of New York, New York

Bezemer DJ (2010) Understanding financial crisis through accounting models. Acc Organ Soc 35(7):676–688
Bieri DS (2016) August Lösch's monetary theory and its implications for the spatial neutrality of money. Global Forum on resilience working paper. Virginia Tech, Blacksburg
Bieri DS (2017) Regulatory space and the flow of funds across the hierarchy of money. In: Martin RL, Pollard J (eds) Handbook of the geographies of money and finance. Edward Elgar, Cheltenham, pp 373–414
Bieri DS, Schaeffer PV (2015) The treatment of money in regional economics: a reprise. Global forum on resilience working paper. Virginia Tech, Blacksburg
Borio CEV (2014) The financial cycle and macroeconomics: what have we learnt? J Bank Financ 45(1):182–198
Borio CEV, Disyatat P (2011) Global imbalances and the financial crisis: link or no link? Working Paper No. 346. Bank for International Settlements, Basel, Switzerland
Borio CEV, Disyatat P (2015) Capital flows and the current account: taking financing (more) seriously. Working Paper No. 525, Bank for International Settlements, Basel, Switzerland
Breit W, Ransom RL (1982) Alvin H. Hanson: the American Keynes. The academic scribblers, 2nd edn. Dryden Press, Chicago, pp 81–104
Caverzasi E, Godin A (2015) Post-Keynesian stock-flow-consistent modelling: a survey. Camb J Econ 39(1):157–187
Cecchetti SG, Mark NC, Sonora RJ (2002) Price index convergence among united states cities. Int Econ Rev 43(4):1081–1099
Chen LL, Choi S, Devereux J (2006) Accounting for U.S. regional real exchange rates. J Money Credit Bank 38(1):229–244
Chick V, Dow SC (1988) A post-Keynesian perspective on the relation between banking and regional development. Post-Keynesian monetary economics: new approaches to financial modelling. New directions in modern economics. Edward Elgar, Aldershot, pp 219–250.
Chick V, Dow SC (1996) Regulation and differences in financial institutions. J Econ Issues 30(2):535–541
Chick V, Tily G (2014) Whatever happened to Keynes's monetary theory? Camb J Econ 38(3):681–699
Cohen's J (1968) Integrating the real and financial via the linkage of financial flows. J Financ 23:1
Cohen's J (1972) Copeland's moneyflows after twenty-five years: a survey. J Econ Lit 10(1):1–25
Cohen's J (1987) The flow of funds in theory and practice. Financial and monetary policy studies, vol 15. Springer, Heidelberg
Cohen BJ (1998) Currency competition and hierarchy. The geography of money, 1st edn. Cornell University Press, Ithaca, pp 92–118
Cohen BJ (2003) The future of money. Princeton University Press, Princeton
Colander D (2014) The wrong type of pluralism: toward a transdisciplinary social science. Rev Polit Econ 26(4):516–525
Copeland MA (1947) Tracing money flows through the United States economy. Am Econ Rev 37(2):31–49
Copeland MA (1952) A study of moneyflows in the United States. National Bureau of Economic Research, Cambridge
Davidson P (1991) Controversies in post Keynesian economics. Edward Elgar, Aldershot
Davis JB (2008) The turn in recent economics and return of orthodoxy. Camb J Econ 32(3):349–366
Del Negro M, Otrok C (2007) 99 Luftballons: monetary policy and the house price boom across U.S. states. J Monet Econ 54(7):1962–1985
Dobusch L, Kapeller J (2012) Heterodox United vs. Mainstream City? Sketching a framework for interested pluralism in economics. J Econ Issues 46(4):1035–1057
Dow SC (1982) The regional composition of the money multiplier. Scott J Polit Econ 29(1):22–44
Dow SC (1987) The treatment of money in regional economics. J Reg Sci 23(1):13–24
Dow SC, Earl PE (1982) Money matters: a Keynesian approach to monetary economics. M. Robertson, London

Fielding D, Shields K (2011) Regional asymmetries in the impact of monetary policy shocks on prices: evidence from U.S. cities. Oxf Bull Econ Stat 73(1):79–103
Figueiredo ATL, Crocco MA (2008) The role of money in the location theory: a post-Keynesian approach. Revista Brasileira de Estudios Regionais e Urbanos 2(1):33–57
Florax RJGM, Plane DA (eds) (2004) Fifty years of regional science. Advances in spatial science. Springer, Berlin
Goodhart CAE (2009) The continuing muddles of monetary theory: a steadfast refusal to face facts. Economica 76(S1):821–830
Gurley JG, Shaw ES (1960) Money in a theory of finance. Brookings Institution Press, Washington
Haberler G (1961) A survey of international trade theory. Special papers in International Economics No. 1, International Finance Section, Department of Economics, Princeton University, Princeton
Hands DW (2015) Orthodox and heterodox economics in recent economic methodology. Erasmus J Philos Econ 8(1):61–81
Humphrey TM (1991) Nonneutrality of money in classical monetary thought. Fed Reserve Bank Richmond Econ Rev 18(2):3–15
Hurst E, Keys BJ, Seru A, Vavra, JS (2016) Regional redistribution through the U.S. mortgage market. Am Econ Rev 106(10):2982–3028
Isard W (1949) The general theory of location and space-economy. Q J Econ 63(4):476–506
Isard W (1954) Location theory and trade theory: short-run analysis. Q J Econ 68(2):305–320
Isard W (1956) Location and space-economy: a general theory relating to industrial location, market areas, land use, trade, and urban structure. Regional science studies series, vol 1, 1st edn. MIT Press, Cambridge
Isard W (1960) Methods of regional analysis: an introduction to regional science. Regional science studies series, vol 4, 1st edn. MIT Press, Cambridge
Isard W (2003) History of regional science and the regional science association international: the beginnings and early years. Springer, Berlin
Isard W (2004) The future (near and far) of regional science. Fifty years of regional science. Advances in spatial science. Springer, Berlin, pp 387–400
Isard W, Moses LN (1960) Interregional flow analysis and balance of payment statements. Methods of regional analysis. Regional science studies series, vol 4, 1st edn. MIT Press, Cambridge, pp 122–179
Isard W, Azis IJ, Drennan MP, Miller RD, Saltzman S, Thorbecke E (1998) Methods of interregional and regional analysis. Regional science studies series, 7th edn. Ashgate, Brookfield
Isserman A (1993) Lost in space? On the history, status and future of regional science. Rev Reg Stud 23(1):2–49
Keynes JM (1930) A treatise on money: the pure theory of money and the applied theory of money. Harcourt, Brace and Company, New York
King JE (2002) A history of post Keynesian economics since 1936. Edward Elgar, Cheltenham
Klausinger H (1990) The early use of the term 'veil of money' in Schumpeter's monetary writings: a comment on Patinkin and Steiger. Scand J Econ 92(4):617–621
Klausinger H (1999) German anticipation of the Keynesian revolution? the case of Lautenbach, Neisser and Röpke. Eur J Hist Econ Thought 6(3):378–403
Klein LR (2003) Some potential linkages for input-output analysis with flow-of-funds. Econ Syst Res 15(3):269–277
Klein LR, Glickman NJ (1977) Econometric model-building at regional level. Reg Sci Urban Econ 7(1–2):3–23
Krugman PR (1998) Space: the final frontier. J Econ Perspect 12(2):161–174
Laidler D (2011) The monetary economy and the economic crisis. Duke Department of Economics Working Paper No. 2011–04, Duke University Center for the History of Political Economy, Durham
Laidler D (2012) Competing monetary explanations of macroeconomic instability before 1936. Macroeconomics and the history of economic thought: Festschrift in honour of Harald Hagemann. Routledge, London, pp 96–117

Laidler D, Stadler G (1998) Monetary explanations of the Weimar Republic's hyperinflation: some neglected contributions in contemporary German literature. J Money Credit Bank 30(4):816–831
Lane T (1993) Is there a future for regional science in economics? Ann Reg Sci 27(2):285–293
Lavoie M (2014) Post-Keynesian economics: new foundations. Edward Elgar, Cheltenham
Lavoie M (2015) Teaching monetary theory and monetary policy implementation after the crisis. Eur J Econ Econ Policies Interven 12(2):220–228
Leontief W (1991 [1928]) The economy as a circular flow. Struct Chang Econ Dyn 2(1):181–212
Leontief W, Brody A (1993) Money-flow computations. Econ Syst Res 5(3):225–233
Lösch A (1940a) Die englischen Zwangskredite. Die Bank 33(32):567–569
Lösch A (1940b) Die räumliche Ordnung der Wirtschaft: Eine Untersuchung über Standort, Wirtschaftsgebiete und internationalen Handel, 1st edn. Gustav Fischer, Jena
Lösch A (1940c) Geographie des Zinses. Die Bank 33(2):24–28
Lösch A (1940d) Verrechnung und Goldwährung – ein Vergleich. Die Bank 33(34):603–606
Lösch A (1944) Die räumliche Ordnung der Wirtschaft: Eine Untersuchung über Standort, Wirtschaftsgebiete und internationalen Handel, 2nd edn. Gustav Fischer, Jena
Lösch A (1949) Theorie der Währung: Ein Fragment. Weltwirtschaftliches Arch 62:35–88
Lösch A (1954) The economics of location. Yale University Press, New Haven
Marcuzzo MC (2002) The demise of the quantity theory of money. Hist Econ Ideas 10(1):49–62
Martin RL (ed) (1999) Money and the space economy. Wiley, Chichester
McCann P, Shefer D (2004) Location, agglomeration and infrastructure. Fifty years of regional science. Advances in spatial science. Springer, Berlin, pp 177–196.
Mehrling PG (1997) Alvin Harvey Hansen (1887–1975). The money interest and the public interest: American monetary thought, 1920–1970. Harvard economic studies, vol 162. Harvard University Press, Cambridge, pp 85–158
Mehrling PG (1998) The money muddle: the transformation of American monetary thought, 1920–1970. Hist Polit Econ 30(Supplement):293–306
Mehrling PG (2013) The inherent hierarchy of money. Social fairness and economics: economic essays in the spirit of Duncan Foley. Routledge frontiers of political economy, Festschrift. Routledge, New York, pp 394–404
Messori M (1997) The trials and misadventures of Schumpeter's treatise on money. Hist Polit Econ 29(4):639–673
Minsky HP (1977) The financial instability hypothesis: an interpretation of Keynes and an alternative to "standard" theory. Challenge 20(1):20–27
Minsky HP (1991) The financial instability hypothesis: a clarification. The risk of economic crisis. University of Chicago Press, Chicago, pp 158–166
Minsky HP (1993) On the non-neutrality of money. Fed Reserve Bank NY Q Rev 18(1):77–82
Minsky HP (1996) The essential characteristics of post Keynesian economics. Money in motion: the post Keynesian and circulation approaches, The Jerome Levy economics institute series. St. Martin's Press, New York, pp 70–88
Minsky HP (2008) Stabilizing an unstable economy. McGraw Hill, New York
Mirowski P (2013) Never let a serious crisis go to waste: how neoliberalism survived the financial meltdown. Verso, London
Molella AP (2005) Review: the longue durée of Abbott Payson Usher: A. P. Usher, "A history of mechanical invention". Technol Cult 46(4):779–796
Morley JC (2016) Macro-finance linkages. J Econ Surv 30(4):698–711
Nerozzi S (2009) Jacob Viner and the Chicago monetary tradition. Hist Polit Econ 41(3):575–604
Nogueira M, Crocco AM, Figueiredo AT, Diniz G (2015) Financial hierarchy and banking strategies: a regional analysis for the Brazilian case. Camb J Econ 39(1):139–156
Palumbo MG, Parker JA (2009) The integrated financial and real system of national accounts for the United States: does it presage the financial crisis? Am Econ Rev 99(2):80–86
Patinkin D, Steiger O (1989) In search of the "veil of money" and the "neutrality of money": a note on the origin of terms. Scand J Econ 91(1):131–146
Pavlik CE (1995) Götterdämmerung or a brave new world? Int Reg Sci Rev 17(3):361–366

Polèse M (1995) On the cultural origins and future directions of regional science: a voice from the periphery. Int Reg Sci Rev 17(3):311–318
Ponsard C (2007) August Lösch: a famous, but ignored economist. Space-structure-economy: a tribute to August Lösch, Karlsruhe papers in economic policy research, vol 24, 2nd edn. Nomos Publishers, Baden-Baden, pp 151–162
Rephann T (2000) Review of 'methods of regional and interregional analysis' by Walter Isard et al. J Reg Sci 40(1):177–181
Riegger R (ed) (1971) August Lösch: in memoriam. Heidenheimer Schriften zur Regionalwissenschaft. Buchandlung Meuer, Heidenheim
Schaeffer PV, Jackson R, Bukenya JO (2012) Regional science reconsidered. Rev Reg Stud 41(3):161–177
Schefold B (1997) Schumpeter as a Walrasian Austrian and Keynes as a classical Marshallian. Normal prices, technical change and accumulation. Studies in political economy. Palgrave MacMillan, Basingstoke, pp 502–524
Schumpeter JA (1908) Das Wesen und der Hauptinhalt der theoretischen Nationalökonomie. Duncker und Humblot, Leipzig
Schumpeter JA (1954) History of economic analysis. Oxford University Press, New York
Schumpeter JA (1970 [1943]) Das Wesen des Geldes. Vandenhoeck und Ruprecht, Göttingen
Sheppard E, Barnes TJ (1990) The capitalist space economy: geographical analysis after Ricardo, Marx and Sraffa. Unwin Hyman, London
Skott P (2014) Pluralism, the lucas critique, and the integration of macroeconomics and microeconomics. Rev Polit Econ 26(4):503–515
Stabler JC (2000) Review of 'methods of regional and interregional analysis' by Walter Isard et al. Can J Reg Sci 23(2):365–367
Storper M (2013) Keys to the city: how economics, institutions, social interaction, and politics shape development. Princeton University Press, Princeton
Summers LH (2014a) Reflections on the 'new secular stagnation hypothesis'. Secular stagnation: facts, causes, and cures. VoxEU.org eBook. CEPR Press, London, pp 27–40
Summers LH (2014b) U.S. economic prospects: secular stagnation, hysteresis, and the zero lower bound. Bus Econ 49(2):65–73
Trautwein HM (2000) The credit view, old and new. J Econ Surv 14(2):155–190
Usher AP (1943) The early history of deposit banking in Mediterranean Europe. Harvard economic studies, vol 75. Harvard University Press, Cambridge
Viner J (1975 [1937]) A note on the scope and method of the theory of international trade. Studies in the theory of international trade. Augustus M. Kelley, Clifton, pp 594–601
von Mises LE (1917) Zur Klassifikation der Geldtheorien. In: Sombart W, Weber M, Schumpeter JA, Jaffé E (eds) Archiv für Sozialwissenschaft und Sozialpolitik, vol 44. J. C. B. Mohr Paul Siebeck, Tübingen, pp 198–213
Weigmann H (1931) Ideen zu einer Theorie der Raumwirtschaft: Ein Versuch zur Begründung einer realistischen Wirtschaftstheorie. Weltwirtschaftliches Arch 34:1–40
Winkler B, van Riet A, Bull P (eds) (2013) A flow-of-funds perspective on the financial crisis: money, credit and sectoral balance sheets. Palgrave studies in economics and banking, vol 1. Palgrave MacMillan, Basingstoke
Wray LR (2009) The rise and fall of money manager capitalism: a Minskian approach. Camb J Econ 33(4):807–828

David Bieri is Associate Professor in the School of Public & International Affairs at Virginia Tech (VT). He is a faculty research associate in VT's Global Forum on Urban and Regional Resilience. Dr. Bieri also holds an endowed appointment as Faculty Fellow in the VT Program in Real Estate. His current research examines spatial features of credit flows, regulatory aspects of global monetary governance and the history of economic thought. Previously, Bieri was a faculty member at the University of Michigan. He has held various positions in finance and central banking, most recently

as the adviser to the CEO at the Bank for International Settlements (BIS) in Switzerland. Dr. Bieri earned a Ph.D. in public policy from Virginia Tech, an M.Sc. in international finance from Durham University, and a B.Sc in economics from the London School of Economics.

Part III
Diasters and Resilience

Chapter 15
Economic Resilience in Regional Science: Research Needs and Future Applications

Adam Rose

15.1 Introduction

The frequency and magnitude of disasters are increasing throughout the world. This stems from a combination of new threats that have emerged, especially in the areas of terrorism and new technologies, and the fact that population and economic growth have generated more and larger targets.

With the exception of pandemics and armed conflicts, the direct effects of disasters are typically limited to the regional level in countries of significant geographic size. These disasters can threaten the survival of regional economies, as witnessed by the compound disaster of the nuclear reactor accident following the earthquake that befell the Fukushima Prefecture of Japan in 2011, the Wenchuan Province earthquake that killed 70,000 people in China in 2008, and the various towns and cities along the Gulf Coast of the U.S. still recovering more than 10 years after Hurricane Katrina.

Future predictions of disasters are even more dire, especially with regard to climate change. Although subject to political disagreement, there is near universal consensus in the scientific community of the reality of this existential threat (IPCC 2014). Evidence is irrefutable of increasing average global temperatures, melting of major glaciers, and rising sea levels manifesting themselves in such ways as extreme droughts and flooding. The former has already resulted in major population shifts, and the latter holds the potential to cause even more drastic shifts given the specter of large portions of major cities, ranging from Miami to Dacca, being underwater in 50 years.

A. Rose
Price School of Public Policy, and Center for Risk and Economic Analysis of Terrorism Events (CREATE), University of Southern California (USC), Los Angeles, CA, USA
e-mail: Adam.Rose@usc.edu

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_15

Resilience, broadly defined, refers to the ability to withstand and recover from such short-term and long-term threats. Unfortunately, resilience has become a popular "buzzword", and its meaning has been clouded by myriad of definitions, thereby generating confusion and undercutting progress on designing and implementing appropriate resilience tactics and strategies. In recent years, analysts and policy makers have narrowed the range of definitions and aided our understanding of the basic concept and some of its implications. The author has provided important distinctions among the definitions, formulated and successfully applied operational metrics of economic resilience, and measured its effectiveness (Rose and Liao 2005; Rose 2007; Rose et al. 2009; Rose and Krausmann 2013). Other regional scientists who have contributed to this field include: Chang (2009), Modica and Reggiani (2015), Kajitani and Tatano (2009), Okuyama (2015), and Cutler et al. (2015). Major contributions have been offered by scholars in related fields, such as Geography, Planning, Economics, Sociology, Organizational Behavior, Ecology, and Engineering (see, e.g., Adger et al. 2005; Martin and Sunley 2015; Cutter 2016 Godschalk 2003; Hallegatte and Dumas 2009; Tierney 2007; Hill and Paton 2005; Perrings 2006; Bruneau et al. 2003; Santos 2012).

Still, much work needs to be done to understand regional resilience, make it more relevant to regional research and practice, make its implementation more cost-effective, and make it applicable to a broader set of future disasters. This paper will focus on the following topics related to *regional economic resilience*:

- What are its key dimensions?
- What is its relationship to resilience in the various disciplines associated with regional science?
- What are its key spatial aspects?
- What are its key behavioral aspects?
- How does it differ between regional and national economies?
- How can the cost-effectiveness of alternative resilience tactics and strategies be measured?
- What is the relationship between resilience and adaptation to climate change?
- What are the key research challenges ahead?

Note that this chapter does not cover all topics under the heading of regional economic resilience. It focuses on disasters and recovery from the initial damages they incur. It does not deal with mitigation, which applies to actions taken in advance to reduce the frequency and/or magnitude of disasters. It also does not cover resilience to business cycles and more general economic phenomena. The reader is referred to the recent, comprehensive paper by Martin and Sunley (2015) for a critical review and conceptual framework for analysis of these more general events. This chapter also differs from the aforementioned in that the emphasis here is on resilience of individual firms and households, as well as on their interactions within the regional economy. This is because the author believes an understanding of the operation of fundamental units of analysis is key to identifying actionable resilience tactics and strategies that can promote economic recovery and that can be used to construct a useful resilience index to help individual decision-makers gauge their progress in advancing resilience.

15.2 Defining Economic Resilience

The etymology of resilience is the Latin term *resilio*, meaning to rebound. Although published accounts of its use date back to ancient Rome in Cicero's *Orations* (Alexander 2013) and some physicists and psychologists in the early twentieth century (Manyena 2006), ecologists were the first to embrace and make extensive use of the general concept of resilience more than 40 years ago (see, e.g., Holling 1973). Since then, it has been adapted or reinvented for the case of short-term disasters (see, e.g., Tierney 1997; Bruneau et al. 2003; Rose 2004) and long-term phenomena, such as climate change (see, e.g., Dovers and Handmer 1992; IPCC 2014). The analysis of resilience can benefit from a comparison of its definitions in ecology, engineering, organizational behavior, planning, psychology, sociology and economics over the past 40 years (Rose 2017). In the discussion below, we focus on points of agreement.

15.2.1 Basic Concepts

Resilience has four roles in the economics literature. Most generally, it is noted, typically in vague terms, as an attribute of the economy in studies of conventional (non-disaster) economic shocks (see, e.g., Dhawan and Jeske 2006). In ecological economics, it is a major focus of analysis as a key attribute necessary for sustainability (see, e.g., Perrings 2006; Folke 2006). Some initial attempts have been made to extend this research to the socioeconomic arena and have it overlap with the study of institutions (see, e.g., Levin et al. 1998). In the disaster literature, it has been an important dimension of hazard loss estimation and terrorist consequence analysis (Rose 2015a).

Following Rose (2004, 2007) we provide specific definitions of *Economic Resilience* to disasters by linking it to the common essence of resilience definitions common to other disciplines:

In general, Static Resilience refers to the ability of the system to maintain a high level of functioning when shocked (see, e.g., Holling 1973). *Static Economic Resilience* is: the efficient use of remaining resources at a given point in time. It refers to the core economic concept of coping with resource scarcity, which is exacerbated under disaster conditions.

In general, Dynamic Resilience refers to the ability and speed of the system to recover (see, e.g., Pimm 1984). *Dynamic Economic Resilience* is: the efficient use of resources over time for investment in repair and reconstruction. Investment is a time-related phenomena—the act of setting aside resources that could potentially be used for current consumption in order to reestablish productivity in the future. Static Economic Resilience does not restore damaged capacity and is, thus, not likely to lead to full recovery.

Note that the definitions are couched in terms of functionality, typically measured in economics as the *flow* of goods and services, such as individual business revenue, Gross Domestic Product (GDP) or broader measures of economic well-being, as opposed to property damage. It is not the property (capital *stock*) that directly contributes to economic welfare but rather the flows that emanate from these stocks. Two things should be kept in mind. First, while property damage takes place at a point in time, the reduced production flow, often referred to as business interruption (BI), begins at the time the disaster strikes but continues until the system has recovered or obtained a "new normal." Second, the recovery process, and hence the application of resilience, depends heavily on the behavior of economic decision-makers and on public policy. Of course, recovery is a multi-faceted activity. It is not as simple as, for example, just automatically rebuilding a factory, highway or school destroyed by an earthquake, hurricane, or terrorist attack.[1]

Another important distinction is between *inherent* and *adaptive* resilience. The former refers to aspects of resilience already built into the system, such as the availability of inventories, excess capacity, substitutability between inputs, and contingent contractual arrangements accessing suppliers of goods from outside the affected area (imports). Resilience capacity can be built up through these means and is then accessed after the disaster. Adaptive resilience arises out of improvisation under stress, such as Draconian conservation otherwise not thought possible (e.g., working many weeks without heat or air conditioning), changes in the way goods and services are produced, and new contracting arrangements that match customers who have lost their suppliers with suppliers who have lost their customers.

15.2.2 Resilience at Various Scales

At the microeconomic level, on the business supplier side, static economic resilience includes redundant systems, improved delivery logistics, and planning exercises. Many more options exist on the business customer side. Rose (2009, 2016a) has developed a framework for analysis that connects customer-side resilience to decisions captured by the economic production function with respect to input choices. Examples include conserving inputs, using excess inputs (capacity and inventories),

[1]Research on resilience is split into two camps. About half of the researchers view resilience as any action that can reduce losses from disaster, ranging from pre-disaster mitigation to post-disaster recovery. Not surprisingly, this group is dominated by engineers, whose work is primarily in the area of mitigation (see, e.g., Bruneau et al. 2003; Haimes 2006). The other camp focuses on resilience as actions following the onset of a disaster. Steps can be taken to enhance resilience before a disaster strikes, acknowledging that resilience is very much a process, but such actions are usually not implemented until afterward (e.g., stockpiling of critical materials, development of emergency plans). Recent events, such as the World Trade Center attacks and Hurricane Katrina, indicate that BI during disaster recovery can be as large as or larger than property damage. We focus on the second approach, noting that much of our analysis is applicable to mitigation as well without much modification.

substituting among them locally, importing them, and altering the locations and timing of their use. For example, all inputs (capital, labor, infrastructure services, and materials) can be conserved to some extent. The major obstacle is the necessity of the input in the production process. The major obstacles to importing needed goods are logistics and transport costs. Broadening the supply chain (see, e.g., Sheffi 2005) by expanding the range of current suppliers or others on a contingency basis is, especially outside the directly affected region, an increasingly popular option.[2] In essence, however, the micro level does not operate in isolation, but must also be evaluated in terms of its broader business ecology (Martin and Sunley 2015), at the meso and macro levels discussed below.

Tierney (2007) makes the point that many of the features that make businesses inherently resilient are related to those that make them less vulnerable to disaster in the first place. Vulnerability is related to such features as business size, ownership characteristics, and type (e.g., a small business with a minority group owner, located in an "Old Town" building and that sells a product with many substitutes, is more vulnerable and, hence, less inherently resilient than otherwise). Businesses with some of these characteristics are also likely to have a lower capacity for adaptation (Alesch et al. 2001).

At the mesoeconomic level, resilience can bolster an industry or market and include, for instance, industry pooling of resources and information, or introduction of innovative pricing mechanisms. What is often less appreciated is the inherent resilience of market prices that act as the "invisible hand" to guide resources to their best allocation in the aftermath of a disaster (see, e.g., Horwich 1995). Some pricing mechanisms have been established expressly to deal with such a situation, as in the case of non- interruptible service premia that enable customers to estimate the value of a continuous supply of electricity and to pay in advance for receiving priority service during an outage. The price mechanism is an inexpensive guide to redirecting goods and services. Price increases, to the extent that they do not reflect "gouging," serve the purpose of reflecting highest value use, even in the broader social setting. Moreover, if the reallocation violates principles of equity, the outcomes can be adjusted by income or material transfers to the disadvantaged.

Industry competitiveness also has an effect on inherent resilience and ability of both the firm and the industry to recover. Product characteristics and production technology have a major effect on the ability to adapt.

At the macroeconomic level, resilience is very much influenced by interdependencies between sectors. Consequently, macroeconomic resilience is not only a function of resilience measures implemented by single businesses but is also determined by the actions taken by all individual companies and markets, including their interaction. Of course, in most cases the whole is greater than the sum of the parts. Examples of resilience options at the macro-level would be economic diversity to buffer impacts on individual sectors and geographic proximity to

[2]Most of the resilience tactics associated with businesses are applicable to government and household operations as well, with some modification (see Rose 2009).

economies not affected by disaster to facilitate access to goods or aid. Others include fiscal (e.g., infrastructure spending to boost the affected economy) and monetary policy (e.g., keeping interest rates low to stimulate private sector reinvestment). The macro-level overlaps with the popular focus on "community resilience" and represents a more holistic picture (Norris et al. 2008).

Similar to Tierney, Martin and Sunley (2015), note that resilience is related to critical features of the regional economy. Other factors affecting resilience at the macro level include the extent to which the disaster affects its customers, and the demand for product. Locational and proximity aspects come into play ("hazardousness of place"), as does the size and level of economic development of the region. Agglomeration economies refer to advantages of large city size in reducing costs of production that can remain intact and keep the city competitive after a disaster (see, e.g., Chernick 2005). However, the close proximity of businesses can make them more vulnerable in the first place. Martin and Sunley (2015; p. 11) state that "The key contribution of the idea of resilience, in our view, is that it directs attention precisely to the impact shocks and their role in shaping the trajectories of regional growth and development."

The previous examples relate primarily to *Static Economic Resilience*. *Dynamic Economic Resilience* is applicable at all three levels as well in terms of expediting the recovery process and enhancing its outcome. At the micro-level, this can be promoted through rapid processing of insurance claims and arranging financing so as to facilitate repair and reconstruction. At the meso and macro levels, it includes accelerating and improving the economic effectiveness of the recovery process by improving logistics and coordinating recovery across sectors. Cross-cutting all three levels is adapting to changing conditions by promoting flexibility and translating short-run practices into sustainable ones through a continual learning process (see, e.g., Chang and Rose 2012; Zolli and Healy 2012; Rose 2014).

15.2.3 An Operational Metric

Following Rose (2004, 2009), we provide an admittedly crude but operational metric of resilience. *Direct* Static Economic Resilience (*DSER*) refers to the level of the individual firm or industry (micro and meso levels) and corresponds to what economists refer to as "partial equilibrium" analysis, or the operation of a business or household entity itself. *Total* Static Economic Resilience (*TSER*) refers to the economy as a whole (macro-level) and would ideally incorporate what is referred to as "general equilibrium" effects, which include all of the price and quantity interactions in the economy, macro- aggregate considerations, and the ramifications of fiscal, monetary and security policies related to the disaster.

An operational measure of *DSER* is the extent to which the estimated direct output reduction deviates from the likely maximum potential reduction given an external shock, such as the curtailment of some or all of a critical input. In essence, *DSER* is the percentage avoidance of the maximum economic disruption that a particular shock could bring about.

We illustrate the application of the definition with the following case study by Rose et al. (2009), who estimated the national and regional economic impact of the September 11, 2001, terrorist attack on the World Trade Center. The researchers refined available data indicating that more than 95% of the businesses and government offices operating in the WTC area survived by relocating, primarily to Midtown Manhattan or across the river in Northern New Jersey. Had all of these firms gone out of business, the potential direct economic loss in terms of GDP would have been $43 billion. However, relocation was not immediate, taking anywhere from a few days to as long as 8 months for the vast majority of firms. Rose et al. (2009) calculated this loss in GDP at $11 billion. They were then able to apply the resilience definition provided in this Section to estimate that the effectiveness of relocation as a resilience tactic in the aftermath of the 9/11 attacks was 72% ($43 minus $11, divided by $43).

Several studies have examined economic resilience in actual disasters or with the use of simulation studies. The major pioneer is Tierney (1997), who surveyed businesses in the aftermath of the Northridge Earthquake and Midwest Floods. Rose and Lim (2002) translated Tierney's findings into specific measures of resilience of the Los Angeles electricity system and found resilience reduced losses substantially. Several other simulation studies have been undertaken to estimate the effects of resilience on losses from disasters, using the metric presented in the previous section. Applying the metric presented above, Kajitani and Tatano (2009) used a survey to estimate the resilience of Japanese industries to various types of lifeline disruptions from disasters. Rose et al. (2007a, b) estimated the resilience of the Los Angeles water and power systems to a 2-week outage due to a terrorist attack and found that resilience could be as high as 90%, primarily due to production recapture. Rose and Wei (2013) examined such resilience tactics as excess capacity, inventories, and export diversion to reduce potential losses from a 90-day shutdown of a major U.S. seaport complex in a regional economy dominated by petrochemical production and estimated that these tactics could reduce GDP losses by more than 70%.

Unfortunately, the cost of implementing resilience has not been sufficiently studied. Anecdotal evidence indicates that it is relatively inexpensive in comparison to pre-event mitigation. For example, conservation often more than pays for itself, import substitution often just incurs additional transport costs, for inventories it is just the carrying and storage costs, and production rescheduling only incurs the payment of overtime wages. Most inherent resilience capabilities, such as excess capacity and the information that can be gleaned from the price system, are basically free.

15.3 Spatial Dimensions of Resilience

15.3.1 Business Location

Spatial considerations have important implications for resilience and disaster recovery. Several valuable analyses have been performed but typically in a relatively non-spatial manner. For example, Giesecke et al. (2012, 2015) in studies of radiological ("dirty bombs") and chlorine gas attacks on the Los Angeles (LA) financial district identified the contaminated area as the site of the impacts and translated the impacts into changes in factor availability and prices for firms located there. This affects firm competitiveness and, hence, imports and exports of the LA economy as a whole by averaging the direct impacts across all firms in its County. But several other spatial aspects were omitted, some of which relate to economic resilience.

Following Giesecke et al. (2015), we provide an overview of a systematic framework for the spatial analysis of the impacts of a terrorist attack, primarily with the objective of improving the accuracy of the estimates of economic consequences. We also indicate how the inclusion of a broader set of spatial dimensions would affect the outcome. We begin by noting two special features of a chlorine attack that distinguish it from other types of disasters. The first relates to fear and stigma effects (see, e.g., Slovic et al. 1981; Giesecke et al. 2012). Additionally, because of the uncertainty regarding the spread of chlorine gas or other insidious weapons whose dispersion is related to weather conditions and are difficult to detect, we should also consider that the fear/stigma will not halt abruptly at the financial district boundary. It is reasonable to consider a fringe zone where these behavioral considerations may spill over and have impacts, though likely less intensive than in the core area.

A behavioral consideration relates to the likelihood and pace of business relocation. The 9/11 example indicates that the response is likely to be rapid and not far from the original site. Both of these responses were conditioned somewhat upon broader aspects of resilience, in the form of demonstrating to terrorists that they cannot defeat their intended targets (Flynn 2008). There is every reason to believe that this "we will show them" attitude would prevail in LA as well.

A complication is that relocation may not be entirely out of the region for which the impact analysis was performed but may also take place within the region, as in the case of World Trade Center area firms moving to Midtown Manhattan. Also, business activity conducted in cyberspace and via tele-commuting have increased significantly in recent years, further blurring boundaries. Finally, there is the longstanding issue of the ready ability to shift economic activity among branch plants of the same company. Given all these considerations, a quarantine plus geographic averting behavior may not result in losses as great as initially predicted.

Below, we discuss in detail various aspects of the potential spatial realignment of economic activity in relation to the chlorine attack scenario:

Business Relocation Alternatives First, we must consider relocation out of the financial district. This would potentially include: (i) actual physical relocation; (ii) a shift of activity to other branches of the firm; and/or (iii) work primarily in cyberspace.

Physical relocation is likely to approach zero in a case where quarantine/decontamination lasts only a few days. Shift of activity to branch offices is a possibility, especially for firms involved in banking, finance and insurance. Work in cyber space, including telecommuting, deserves special attention here because it is becoming more prevalent, especially in the banking and finance sectors that are predominant in the area affected by the attack. This activity may not be affected in any significant way. Increasingly, businesses are backing-up their systems such that even if main computers are located in the financial district, relevant files can be accessed from elsewhere. Similarly, data are being increasingly stored on various types of "cloud." The prevalence of these cyber options warrants adjustment of direct impact estimates when data become available.

We must also distinguish shifts of locations within LA County from those going elsewhere. As in the post-9/11 shift to Midtown Manhattan, there are several advantages to relocating in close proximity to the original site. Doing so would not lead to any reduction in economic activity within LA County, all other things being equal. Of course, business relocation does have its costs. If this significantly increases the cost of doing business at the new location, this would have to be factored into the analysis and would lower the level of economic activity in LA County by affecting its competitiveness.

Economic Activity Shift Out of the Region This aspect does not pertain to the actual physical movement of firms, but rather to their activity levels in place. It relates to an increase in their cost of doing business due to increased wage demands and increased investor rates of return to compensate for increased risk. It also relates to their likely reduced profits as they have to provide customer discounts. In the case of the later, the base for the activity shifts decreases (fewer firms to which to apply what essentially declines in competiveness or product demand that lead to reduced sectoral output).

Temporal Dimension of Spatial Shifts It is important to distinguish between business relocation at various points in time. The implications of these decisions differ significantly for the consequence estimates between the initial contamination, clean-up, risk amplification, and stigma phases. While firms may not have time to physically relocate during the short-run response (decontamination), opportunities to do so increase over time, though the incentives to do so decline as well (see the discussion above of the decay rate for fear). Thus, some relocations decisions (including branch offices and cyber space activity) affect ordinary losses, and others affect behavioral losses.

One also needs to consider the potential for a reverse movement of businesses. Abadie and Dermisi (2011) investigated the movement of many WTC area firms who left Lower Manhattan right after 9/11 but returned. The same phenomenon could take place for the type of terrorist attack that we have analyzed, though it is less likely because of (misplaced) fear of lingering contamination. One would also need to consider the extent to which the attack site is seen as a prime target for future attacks, contributing to persistent stigma.[3]

15.3.2 Spatial Dimensions and People

We turn now to the special case of those geographically displaced, possibly permanently, by crises. This is becoming an increasingly frequent phenomenon due to climate change and armed conflict. Jeffrey Sachs (2012) has warned that many parts of the world will become "uninhabitable or at least uneconomic" as a result. Estimates of the impending number of refugees range from fifty million to one billion. Sachs points specifically to water (either too much or too little) as the main factor. This pertains not just to low-lying areas but also to semi-arid ones. It impacts greatly on human security and on the ultimate level of survival. Sachs points out that, for example, the violence in Darfur and in Somalia is caused by food and water insecurity.

Refugee crises are not necessarily limited to the world at large. The U.S. faced a significant refugee crisis in the aftermath of Hurricane Katrina, which stretched the capabilities of residents and governments in other regions. In addition to long-run climate change, short-run climate variability manifesting itself in more and larger hurricanes, for example, could mean that Katrina is not the exception of the likely rule in the future. Moreover, the specter of terrorist attacks utilizing insidious agents, such as biological, chemical, radiologic, and nuclear are serious matters of concern. For example, Dormady et al. (2014) simulated a large-scale anthrax attack on downtown Seattle and predicted 70,000 mortgage foreclosures and approximately 250,000 potential health migrants. All of these examples are exacerbated by potential inadequate planning and coordination by government agencies at all levels, as exemplified in the case of Katrina.

Internationally, borders are not necessarily open, and Piguet et al. (2011) point out that mobility is strongly affected by political decisions. Forced migrants are typically thought of as the neediest because of meager material resources and being dispossessed. But it is likely that at the outset they represent a broader cross-section of the population and, thus, have education and other skill levels closer to the average (McAdam 2011). Others pose the question of whether migration represents

[3] Problems of spillovers of disasters across political boundaries are beyond the scope of this volume. However, advances are being made in interjurisdictional cooperation to deal with them, including an application of resilience tactics (Rose and Kustra 2013).

an example of adaptation or the failure to adapt. The recent refugee crisis stemming from armed conflict in Syria and Iran is challenging this notion and has raised critical issues relating to motives of some refugees and the reception they receive.

Piguet et al. (2011) examined the governance framework in which migration flows take place. Those potentially forced to migrate as a result of climate change do not currently have official status as refugees. This issue is being hotly debated, where some argue that refugee status related to climate change would dilute the definition. This is, in part, due to a rigorous interpretation of the current concept of a refugee, but is not necessarily the best approach. It should also be noted that some oppose the extension of refugee status for fear of opening the floodgates due to migration. With respect to recent climate change treaties themselves, there has been slow progress on this issue (Hodgkinson et al. 2009). Refugee status matters greatly with respect to governments providing protection and assistance (Aminzadeh 2007; McAdam 2011). While these forms of assistance are strongly emphasized, there is inadequate attention to the role of resilience in assessing their need and effectiveness.

One of the features of climate change is that its slow onset could generate a new type of refugee, one with significant advanced time for planning. Advance plans are well-intended but might be potentially rigid. There is a great opportunity to build in flexibility in various ways to increase resilience capacity when things do not go as planned. Another factor is that this migration is likely to take place in waves, so that social connections, especially with respect to kindred, are likely to be the key.

As to resilience strategies, in addition to providing resources, it is very important to provide training on how to best use them in the new environment. It is also important to gauge the length of stay and transition to temporary status, rather than just visitation. It is important to maintain communities and to prepare a population for advancement on the journey, including the return to their homeland. Even in places where there have been efforts to ease the plight of refugees for decades, concerns over human security continues (see, e.g., UNDP 2013). Part of this is the dilemma of assistance, which can often turn into a dependency relationship.

15.4 Resilience, Adaptation and Sustainability

Several ecologists and ecological economists have linked resilience to the concept of *sustainability*, which refers to long-term progress and survival of a community without diminishing the quality of life for future generations (Adger et al. 2005). Others add the requirement: refers to society's ability to react effectively to a crisis and with minimal reliance on outside resources (see Mileti 1999; IFRC 2004). This is especially a challenge for sub-national areas.

What is the relationship between resilience and sustainability? Resilience is usually used in the context of responding to specific shocks and, thus, relates to short-run survival and recovery. This contributes to long-run survival, a key aspect of sustainability along with improving the quality of life and the environment.

Zolli and Healy (2012) have recently identified a major difference between resilience and sustainability that is especially important in relation to climate change. They see resilience focusing on disequilibrium situations and stability in contrast to sustainability's focus on equilibrium paths. They point to the need for a reorientation of infrastructure designs to be less brittle and more robust as well as overall more flexible so as to be able to rebound. Thus, many practices promoting sustainability do not necessarily promote resilience. A key example is new energy-efficient buildings, which include systems that promote longevity but not necessarily the ability to withstand or rebound from shocks.

In the context of longer-term disasters, such as climate change, Timmerman (1981) defined resilience as the measure of a system's capacity to absorb and recover from the occurrence of a hazardous event. In the climate change context, however, most researchers now refer to this as *adaptation* (see, e.g., IPCC 2014). Kates et al. (2012) argue that most of the ongoing adaptation to climate change is incremental but that this is likely to become increasingly insufficient, such that "transformational" adaptation (new to a region, larger and involving geographic shifts) will be needed.

Adaptation can be viewed as the complement to mitigation in the long run. For example, while mitigation of the causes of climate change is the preferred approach, the reality is that some amount of climate change is inevitable given the fact that greenhouse gases (GHGs) are "fund" pollutants (i.e., they have long residence times in the atmosphere and, hence, any emissions in a given year add to the existing concentrations). Thus, the second best response to climate change is adaptation—actions to minimize losses for the climate change that does occur (Mazmanian et al. 2010). Typically, adaptation is associated with long-term, or chronic, climate change as opposed to short-term climate variability, which sometimes manifests in natural hazards. There are as many adaptation strategies as there are resilience strategies, and many overlap. Examples of adaptation include the creation of drought-resistant crops, construction of seawalls, safeguards against wildfires, and population migration.

Thus, resilience can be thought of as a short-run version of adaptation, geared toward dealing with disasters related to climate change. However, some subtle differences arise. Building a levee or a seawall mitigates riverine floods or ocean storm surge, but is an adaptation strategy with respect to climate change. Also, if resilience refers to bouncing back, population migration is the antithesis, though there is an increasing realization that the optimal recovery from the disaster is not necessarily to return to prior population and economic levels if they are not sustainable (e.g., New Orleans following Hurricane Katrina). Still, many ways of translating resilience into sustainable practices can also extend resilience to adaptation practices as well (Rose 2014).

15.5 Future Research

Regional science is inherently an interdisciplinary topic given its holistic treatment of any given geographic area. Therefore, future research on resilience should continue to capitalize on cross-fertilization across disciplines. It should, on the one hand, seek to strengthen the foundations of the concept of resilience and fill in gaps in our knowledge. It should formulate key hypotheses and test them, and also perform more empirical analyses of resilience effectiveness. On the other hand, it should venture into new and often expanding areas of inquiry and develop new analytical tools to deal with them, as well as establishing a basis for improved data collection and more definitive empirical work. Below, I lay out a research agenda.

15.5.1 Solidifying the Foundations of Regional Economic Resilience

1. *Improving the fundamental understanding of resilience*. Without getting bogged down in the pedantry of definitions, it would be prudent to agree on some basic terminology and concepts within regional research and in relation to other disciplines. One concern is the parallel universe that currently exists between research in closely related fields, for example, members of the Regional Studies Association (e.g., Martin and Sunley 2015) and the Regional Science Association International (many of whom have been referenced in this chapter). These two threads of research complement each other nicely, as the former focuses primarily on regional macroeconomic considerations and the latter on micro and meso levels.
2. *Measuring Static and Dynamic Economic Resilience in Practice*. Very few studies have actually measured resilience in the aftermath of a disaster, and analysts and policy makers have been overly dependent on simulation analysis. This has likely resulted in an upward bias of the potential effectiveness of resilience because it does not deal with the many issues of implementation. This measurement requires well-formulated questions and well-designed survey instruments. Although there are likely several approaches, the author is currently engaged in research together with Noah Dormady and Heather Rosoff, utilizing a production theory approach and the operational metric presented above for static resilience. He is also involved in research with Dormady, Kathleen Tierney and Liesel Ritchie on an investment analysis approach to measure dynamic resilience. Both of these studies involve an application to Superstorm Sandy. Given the fact that so many aspects of resilience are firm, industry, and location specific, many studies will be needed in the future to differentiate the relative importance these features have on resilience.
3. *Identifying Obstacles to Resilience*. The simulation studies cited above are biased towards estimating resilience at its maximum effectiveness. This outcome is

unlikely to occur in practice due to the disarray accompanying most disasters, administrative obstacles, and personal failings. Moreover, Rose (2009) has analyzed the erosion of resilience during large disasters as inventories become depleted, extreme conservation becomes onerous, and opportunities for production recapture decline as customers abandon their traditional suppliers who are unable to deliver. Research is needed on the extent of these obstacles and to identify ways to overcome them.

4. *Evaluating Inherent Resilience Potential*. It is important to identify resilience that is inherent in the survival mechanisms of businesses and households from separate from those resilience tactics that require government policy assistance. Likewise, more research is needed on the ability of markets to provide adequate price signals for resource allocation in a crisis. Much research exists on the role of sectoral diversification in cushioning a region from ordinary shocks but does not necessarily deal with the idiosyncrasies of various types of disasters. If these separate types of resilience can be identified, future recovery efforts can better capitalize on existing capabilities and minimize duplication of government services. The focus of government can then be on facilitating inherent resilience by removing obstacles to private enterprise, reducing wait times for assistance, and more effectively targeting its own role.
5. *Costs of Resilience*. To make prudent resource management decisions, one must consider the cost of each resilience tactic as well as its effectiveness. There is a need to go beyond anecdotal evidence on resilience and perform analyses of all the direct and indirect costs (including any negative side effects) of individual resilience tactics. Some of the more challenging areas relate to transportation systems, where rerouting and modal substitution, as well as cost savings through trip-reduction via such options as telecommuting, should be explored. Once costs are specified, it is possible to compare them with the benefits reflected in resilience effectiveness to perform solid risk management valuations. Ultimately, post-disaster resilience strategies should be compared to pre-disaster mitigation strategies in a benefit-cost framework to strike a better balance than currently exists.
6. *Compiling Resilience Indices Based on Actionable Variables*. Recently, interest has shifted to identifying individual resilience indicators that can be aggregated into an overall index. This has emanated in part from the successful compilations of vulnerability indices (Cutter et al. 2003). Several well-intentioned examples of resilience indices include Cutter et al. (2010), Sherrieb et al. (2010) and ARUP and Rockefeller Foundation (2014) (see also the review by Cutter 2016), but many of their components are background conditions and many are not, in fact, important to the resilience of individual businesses or the economy as a whole during the crucial early stages of the recovery process. Specifically, resilience is not just the flip-side of vulnerability. A resilience index is not only useful to study the recovery process, but also to improve it. This speaks to the importance of actionable variables (Rose and Krausmann 2013). More research is needed to identify indicators that really matter to business decisions in the short-run.

7. *Complexity Analysis*. Shocks to the regional economy and its recovery involve many complex interactions. This includes impacts on and responses to various scales and the interconnections between them. Complexity analysis in general represents research on the frontiers of several disciplines, with major advances by regional scientists. This includes not only network analysis (see, e.g., Reggiani et al. 2002) but also such modeling approaches as input-output (Okuyama et al. 2004) and computable general equilibrium analysis (Kim et al. 2004; Rose et al. 2017. Further advances in this area that integrate GIS, spatial econometrics, and more recent advances in network theory are warranted.

15.5.2 Exploring New Areas

1. *Climate Change*. Although there is skepticism about climate change in the popular press and political circles, scientists are nearly unanimous in their concern over its realities, even current ones, and the likelihood that potential damages will increase without major steps to mitigate and sequester its GHG drivers (IPCC 2014). Moreover, given the centuries long residence time of GHGs in the atmosphere, even major mitigation efforts will not reduce the threat that currently exists and will remain as, at least, a baseline in the foreseeable future. Climate change is already manifesting itself in ways ranging from sea level rise to an increase in global temperatures, which have increased flooding and droughts, respectively. Many scientists are becoming convinced that climate change is already increasing the frequency and magnitude of tornadoes, hurricanes, and other severe storms. Resilience is a way of coping with this short-run climate variability. Moreover, lessons learned from it can carry over to improving long-run climate adaptation.
2. *Cyber Threats*. In an increasingly interconnected society, we are becoming more vulnerable to terrorist attacks, natural disasters, and technological failures related to cyber systems. Of course, these ripple outward to more than just person-to-person and business-to-business interactions, but also to interactions across these individual units and with respect to critical infrastructure and the rest of the region. As with other threats, we cannot eliminate all cyber risk and need to utilize remaining resources more efficiently and recover more quickly when they occur. Unique opportunities exist here with respect to clouds and satellite technology in general (Rose 2015b). The cyber area is also a major source of resilience capacity in relation to other threats as well and, thus, bears further exploration. For example, cyber storage of data (with back-ups) is less expensive and less vulnerable than most other options. Cyber capabilities are also better able to reduce losses through activities such as telecommuting.
3. *Incentives to Promote Resilience*. A major issue is how to increase resilience capacity and improve the effectiveness of its implementation. Businesses increase resilience capacity for the sake of their own profits, but resilience often yields spillover benefits in the form of positive externalities and public

goods that benefit others in the region. However, the standard market failure outcomes, primarily under-provision of resilience, are likely to take place under these conditions. Related to this is the fact that not all businesses are aware of the benefits to their own operations either. Efforts to promote greater awareness within the private sector of resilience potential are just beginning. One major theme relates to "Co-Benefits" of Disaster Risk Management, which is the attempt to quantify how improvements in resilience can reduce uncertainty and, thereby, produce improvements in the broader economic environment that will increase savings and investment in general (see, e.g., Tanner et al. 2015; Rose 2016b). More research is needed on how to induce decision-makers to address resilience from the broader perspective of society. This would relate to research on policy instruments, including traditional ones, such as taxes, subsidies, regulation and information campaigns, as well as innovative approaches relating to micro finance, insurance, and securitization.

4. *Equity Aspects and Environmental Justice*. Disasters typically have disproportionate impacts across socioeconomic groups, as well as economic sectors and regions. Disasters typically hurt poor and otherwise disadvantaged persons the most, because they are more vulnerable in the first place (see, e.g., Mileti 1999). But these groups have not been analyzed in terms of the effect that disaster resilience might have on them. It would be worthwhile to examine the impacts of resilience on lessening the burden on the weakest/poorest. These elements of society have fewer resources to build resilience capacity, but, at the same time, resilience is relatively less expensive than other tactics to reduce losses, so this may have a profound effect on narrowing the disparities of disaster impacts. Also, disadvantaged groups are likely to benefit differentially from the co-benefits of resilience.
5. *Displaced Persons and Refugees*. Two of the most critical issues for the future relate to refugees and displaced persons. A larger number of major disasters, such as Hurricane Katrina in the U.S. and the recent compound event of an earthquake/tsunami/nuclear reactor accident in Japan, resulted in massive population displacement. The specter of future disasters is even more ominous when we consider the potential of natural, technological, and man-made (mainly terrorism) initiators of large-scale disasters. Massive population adjustments cannot be left to market forces alone. Governments seem to be ill-prepared to deal with these crises, either in terms of operational capabilities or political will. More research is needed on the causes, human and economic impacts, and policy responses to massive migration. In the case of climate change, this research will have to address the conundrum of whether migration represents an example of adaptation or the failure to adapt.

References

Abadie A and Dermisi S (2011) World Trade Center tenant relocation patterns after September 11th, 2001. Paper presented at the Annual Meeting of the American Economic Association, 5 Jan 2011

Adger WN, Hughes TP, Folke C, Carpenter SR, Rockstrom J (2005) Social-ecological resilience to coastal disasters. Science 309:1036–1039

Alesch DJ, Holly JN, Mittler E, Nagy R (2001) Organizations at risk: what happens when small businesses and not-for-profits encounter natural disasters. Public Entity Risk Institute, Fairfax, VA

Alexander D (2013) Resilience and disaster risk reduction: an etymological journey. Nat Hazards Earth Syst Sci Discuss 1:1257–1284

Aminzadeh S (2007) A moral imperative: the human rights implications of climate change. Hastings Int Comp Law Rev 30(2):231–265

ARUP and Rockefeller Foundation (2014) City resilience framework. Ove ARUP & Partners International, London. https://assets.rockefellerfoundation.org/app/uploads/20150530121930/City-Resilience-Framework1.pdf

Bruneau M, Chang S, Eguchi R, Lee G, O'Rourke T, Reinhorn A, Shinozuka M, Tierney K, Wallace W, von Winterfeldt D (2003) A framework to quantitatively assess and enhance seismic resilience of communities. Earthq Spectra 19:733–752

Chang S (2009) Infrastructure resilience to disasters. The bridge: linking engineering and society. Natl Acad Eng 39(Winter):36–41

Chang S, Rose A (2012) Economic recovery to disasters. Int J Mass Emerg Dis 30(2):171–181

Chernick H (ed) (2005) Resilient city. Russell Sage Foundation, New York

Cutler H, Shields M, Tavani D, Zahran S, van de Lindt J (2015) Computable general equilibrium models and natural disasters: a note on challenges and possibilities. In: Ayyub B (ed) Economics of community disaster resilience. National Institute of Standards and Technology, Gaithersburg, MD

Cutter S (2016) The landscape of disaster resilience indicators in the USA. Nat Hazards 80:741–758

Cutter S et al (2003) Social vulnerability to environmental hazards. Soc Sci Q 84(2):241–261

Cutter S, Burton CG, Emrich CT (2010) Disaster resilience indicators for benchmarking baseline conditions. J Homel Secur Emerg Manag 7:Article 51

Dhawan R, Jeske K (2006) How resilient is the modern economy to energy price shocks? Econ Rev, Fed Reserve Bank of Atlanta 91(Third Quarter):21–32

Dormady N, Szelazek T, Rose A (2014) The potential impact of an anthrax attack on real estate prices and foreclosures in Seattle. Risk Anal 34(1):187–201

Dovers R, Handmer J (1992) Uncertainty, sustainability and change. Glob Environ Chang 2(4):262–276

Flynn S (2008) America the resilient: defying terrorism and mitigating natural disasters. Foreign Aff 87(2):2

Folke C (2006) Resilience: the emergence of a perspective for social–ecological systems analyses. Glob Environ Chang 16(3):253–267

Giesecke JA, Burns WJ, Barrett A, Bayrak E, Rose A, Slovic P, Suher M (2012) Assessment of the regional economic impacts of catastrophic events: CGE analysis of resource loss and behavioral effects of an RDD attack scenario. Risk Anal 32(4):583–600

Giesecke JA, Burns W, Rose A, Barrett T, Griffith M (2015) Regional dynamics under adverse physical and behavioral shocks: the economic consequences of a chlorine terrorist attack in the Los Angeles finance district. In: Nijkamp P, Rose A, Kourtit K (eds) Regional science matters. Springer, Heidelberg

Godschalk D (2003) Urban hazard mitigation: creating resilient cities. Nat Hazards Rev 4(3):136–143

Haimes YY (2006) On the definition of vulnerabilities in measuring risks to infrastructures. Risk Anal 26(2):293–296

Hallegatte S, Dumas P (2009) Can natural disasters have positive consequences? Investigating the role of embodied technical change. Ecol Econ 68(3):777–786

Hill R, Paton D (2005) Managing company risk and resilience through business continuity management. In: Paton D, Johnston D (eds) Disaster resilience: an integrated approach. Charles C. Thomas, Springfield, IL

Hodgkinson D, Burton T, Young L, Anderson H (2009) Copenhagen, climate change 'refugees' and the need for a global agreement. Public Policy 4(2):155–174

Holling C (1973) Resilience and stability of ecological systems. Annu Rev Ecol Syst 4:1–23

Horwich G (1995) Economic lessons of the Kobe earthquake. Econ Dev Cult Chang 48(3):521–542

Intergovernmental Panel on Climate Change (IPCC) (2014) Fifth assessment report (AR5). Switzerland, Geneva

International Federation of Red Cross and Red Crescent Societies (IFRC) (2004) World disasters report 2004: focus on community resilience. IFRC, Geneva

Kajitani Y, Tatano H (2009) Estimation of lifeline resilience factors based on empirical surveys of Japanese industries. Earthq Spectra 25(4):755–776

Kates R, Travis W, Wilbanks T (2012) Transformational adaptation when incremental adaptations to climate change are insufficient. Proc Nat Acad Sci U S A 109(19):7156–7161

Kim E, Hewings GJD, Chowoon H (2004) Application of an integrated transport network multiregional CGE model: a framework for the economic analysis of highway projects. Econ Syst Res 16(3):235–258

Levin S, Barrett S, Aniyar S, Baumol W, Bliss C, Bolin B, Sheshinski E (1998) Resilience in natural and socioeconomic systems. Environ Dev Econ 3(2):221–235

Manyena S (2006) The concept of resilience revisited. Disasters 30(4):433–450

Martin R, Sunley P (2015) On the notion of regional economic resilience: conceptualization and explanation. J Econ Geogr 15(1):1–42

Mazmanian D et al (2010) Preparing for the effects of climate change – a strategy for California. A Report of the California Adaptation Advisory Panel to the State of California. Pacific Council on International Policy

McAdam J (2011) Swimming against the tide: why a climate change displacement treaty is not the answer. Int J Refug Law 23(1):2–27

Mileti D (1999) Disasters by design: a reassessment of natural hazards in the United States. Joseph Henry Press, Washington, DC

Modica M, Reggiani A (2015) Spatial economic resilience: overview and perspectives. Netw Spat Econ 15:1–23

Norris F, Stevens S, Pfefferbaum V, Whyche K, Pfefferbaum R (2008) Community resilience as a metaphor, theory, set of capacities and strategy for disaster readiness. Am J Community Psychol 41:127–150

Okuyama Y (2015) The rise and fall of the Kobe economy from the 1995 earthquake. J Disaster Res 10(4):635–640

Okuyama Y, Hewings GJD, Sonis M (2004) Measuring economic impacts of disasters: interregional input-output analysis using sequential interindustry model. In: Okuyama Y, Chang SE (eds) Modeling spatial and economic impacts of disasters. Springer, Heidelberg

Perrings C (2006) Resilience and sustainable development. Environ Dev Econ 11(04):417–427

Piguet E, Pecoud A, de Guchteneire P (2011) Migration and climate change: an overview. Refug Surv Q 30(3):1–23

Pimm SL (1984) The complexity and stability of ecosystems. Nature 307(5949):321–326

Reggiani A, de Graaf T, Nijkamp P (2002) Resilience: an evolutionary approach to spatial economic systems. Netw Spat Econ 2(2):211–229

Rose A (2004) Defining and measuring economic resilience to disasters. Disaster Prev Manag 13:307–314

Rose A (2007) Economic resilience to natural and man-made disasters: multidisciplinary origins and contextual dimensions. Environ Hazards 7(4):383–395

Rose A (2009) Economic resilience to disasters. Community and Resilience Institute Research Report Number 8. ORNL

Rose A (2014) Economic resilience and the sustainability of cities in the face of climate change: an ecological economics framework. In: Mazmanian D, Blanco H (eds) Handbook of sustainable cities. Edward Elgar, Cheltenham, UK

Rose A (2015a) Macroeconomic consequences of terrorist attacks: estimation for the analysis of policies and rules. In: Mansfield C, Smith VK (eds) Benefit transfer for the analysis of DHS policies and rules. Edward Elgar, Cheltenham, UK

Rose A (2015b) Measuring cyber resilience. Center for Risk and Economic Analysis of Terrorism Events (CREATE), University of Southern California, Los Angeles, CA

Rose A (2016a) Construction of an economic resilience index. In: Paton D, Johnston D (eds) Disaster resilience: an integrated approach, 2nd edn. Charles C. Thomas, Springfield, IL

Rose A (2016b) Private sector co-benefits of disaster risk management. Final report to the World Bank

Rose A (2017) Defining and measuring economic resilience from a societal, environmental and security perspective. Springer, Singapore

Rose A, Krausmann E (2013) An economic framework for the development of a resilience index for business recovery. Int J Disaster Risk Reduct 5:73–83

Rose A, Kustra T (2013) Economic considerations in designing emergency management institutions and policies for transboundary disasters. Public Manag Rev 15(3):446–462

Rose A, Liao S (2005) Modeling resilience to disasters: computable general equilibrium analysis of a water service disruption. J Reg Sci 45(1):75–112

Rose A, Lim D (2002) Business interruption losses from natural hazards: conceptual and methodology issues in the case of the Northridge earthquake. Environ Hazards: Hum Soc Dimens 4:1–14

Rose A, Wei D (2013) Estimating the economic consequences of a port shutdown: the special role of resilience. Econ Syst Res 25(2):212–232

Rose A, Oladosu G, Liao S (2007a) Business interruption impacts of a terrorist attack on the electric power system of Los Angeles: customer resilience to a total blackout. Risk Anal 27:13–31

Rose A, Oladosu G, Liao S (2007b) Regional economic impacts of a terrorist attack on the water system of Los Angeles: a computable general disequilibrium analysis. In: Richardson H, Gordon P, Moore J (eds) The economic costs and consequences of a terrorist attack. Edward Elgar, Cheltenham, UK

Rose A, Oladosu G, Lee B, Beeler-Asay G (2009) The economic impacts of the 2001 terrorist attacks on the World Trade Center: a computable general equilibrium analysis. Peace Econ Peace Sci Public Policy 15(2):Article 4

Rose A, Prager F, Chen Z, Chatterjee S (2017) Economic consequence analysis tool (E-CAT). Springer, Singapore

Sachs J (2012) Climate change refugees. Sci Am 306:43

Santos JR (2012) An input-output framework for assessing disaster impacts on Nashville metropolitan region paper presented at the 20th International Input–Output Conference, Bratislava, Slovakia

Sheffi Y (2005) The resilient enterprise. MIT Press, Cambridge, MA

Sherrieb K, Norris F, Galea S (2010) Measuring capacities for community resilience. Soc Indic Res 99:227–247

Slovic P, Fischhoff B, Lichtenstein S, Roe FJC (1981) Perceived risk: psychological factors and social implications. Proc Royal Soc Lond A: Math Phys Eng Sci 376(1764):17–34

Tanner T, Surminski S, Wilkinson E, Reid R, Rentshcler J, Rajput S (2015) The triple dividend of resilience: realizing development goals through the multiple benefits of disaster risk management. Overseas Development Institute, London, UK, and World Bank, Washington, DC

Tierney K (1997) Impacts of recent disasters on businesses: the 1993 midwest floods and the 1994 Northridge Earthquake. In: Jones B (ed) Economic consequences of earthquakes: preparing for the unexpected. National Center for Earthquake Engineering Research, Buffalo, NY, pp 189–222

Tierney K (2007) Businesses and disasters: vulnerability, impacts, and recovery, Handbook of Disasters. Springer, Heidelberg

Timmerman P (1981) Vulnerability, resilience and the collapse of society: a review of models and possible climatic applications. J Climatol 1(4):396–438

United Nations Development Programme (UNDP) (2013) Absence of human security has greatly impeded Palestinian progress

Zolli A, Healy AM (2012) Resilience: why things bounce back. Free Press, New York

Adam Rose is research professor, Price School of Public Policy, and faculty affiliate, Center for Risk and Economic Analysis of Terrorism Events (CREATE), University of Southern California (USC). His primary research interests are the economic consequences of and resilience to natural disasters and terrorism, and the economics of energy and climate change policy. He is a faculty affiliate of the Center for Risk and Economic Analysis of Terrorism Events (CREATE) and of the Center for Sustainable Cities at USC, as well as a fellow of USC's Schwarzenegger Institute for State and Global Policy. He is also a faculty affiliate of the Critical Infrastructure Resilience Institute at the University of Illinois. He previously served as professor and chair of the Department of Energy and Environmental Economics, as well as professor in the Department of Geography at Penn State. Prior to that he served as professor and chair of the Department of Mineral Resource Economics and as a faculty affiliate of the Regional Research Institute at West Virginia University. Dr. Rose earned his PhD in economics at Cornell University in 1974.

Chapter 16
Disaster and Regional Research

Yasuhide Okuyama

16.1 Disaster Impact Analysis and Regional Science

Natural hazards, such as earthquakes, severe storms, flooding, wildfires, and so on, have threatened societies and economies across the world and have brought considerable damages and losses to our economy, resulting in disasters. Some recent studies indicated that the economic costs of these disasters have been rising (Adam 2013), whereas other studies showed no such trend (Okuyama and Sahin 2009). Whether or not the decisive long-run trend of increasing economic costs of disasters is confirmed can be investigated elsewhere, but what is certain is that as the world economy continues to grow and our wealth continues to rise, at risk assets and populations have also been increasing.

While the impacts of disasters sometimes become catastrophic in terms of economic damages and losses, Albala-Bertrand (2007) claimed that for most countries a severe natural hazard causes localized damages and losses on capital and activities but may not affect negatively the macro-economy in either the short-term or long-term. Except for small island nations and the least developed countries in which even a natural hazard with limited damages to their capital can still lead to the catastrophic disaster causing macroeconomic impacts, empirical analyses of disaster impacts on the national economy appear to support Albala-Bertrand's assertion. Cavallo et al. (2013) found that even very large disasters do not cause any major effect on long-run economic growth, and Felbermayr and Gröschl (2014) concluded that only temporary production losses in the year the disaster occurred are statistically significant without finding any subsequent long lasting effect on production level.

Y. Okuyama
University of Kitakyushu, Kitakyushu, Japan
e-mail: okuyama@kitakyu-u.ac.jp

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_16

At a regional level, on the other hand, the economic impacts on a disaster-struck region can become quite substantial in terms of their intensity and extent. In such cases, economic impacts result not only from the damages and losses caused by a natural hazard per se, but also from recovery and reconstruction activities after the event. These damages and losses, and the recovery and reconstruction activities are all regionalized (localized) incidents leading to quite extensive impacts and effect on the region. The sub-national (regional) analysis of disaster impacts is an important avenue for the future direction beyond national level research (Mochizuki et al. 2014), which has been conducted more often than regional empirical research on the impacts of disasters. For example, DuPont and Noy (2015) and Okuyama (2016) examined the long-run effect of the 1995 Kobe earthquake on the regional economies and found sizeable and intricate effects of short-run impacts and long-run effects from negative shocks as well as from positive demand injections of reconstruction activities. Regional analyses of disaster impacts can utilize a wide array of regional science methods and the current and future regional research agenda should include disaster impact analysis.

Meanwhile, the complexity of global production processes has made impact propagation to other regions within the country and across national borders more likely. Interregional analysis of disaster impacts has been conducted, such as the evaluation of interregional impact within a nation (Okuyama et al. 2004 and Tsuchiya et al. 2007) and across countries (Arto et al. 2015). Whereas the propagated impacts to other regions are smaller than the impacts in the region struck by a disaster, these studies find some evidence that impacts diffused interregionally, causing negative or positive impacts in other regions. Therefore, in addition to the need for regional analysis of disaster impacts discussed above, interregional analyses have also become more important in comprehensive disaster impact analyses.

In this chapter, what the regional research in disaster impact analysis will and should become in the next 50 years are presented and discussed. The following section time-travels to 50 years into the future, witnessing the launch of an integrated simulation system for disaster impact analysis, the World Disaster Impact Simulation System (WDISS). The third section describes the technical features of WDISS. The historical development of WDISS is illustrated in the last section.

16.2 Introducing the World Disaster Impact Simulation System

On January 17, 2065, an international organization, the World Disaster Consortium, introduces the World Disaster Impact Simulation System (WDISS), which enables local, regional, national government, or international organizations and their decision makers to simulate disaster process and impact with a given natural hazard for examining mitigation and recovery strategies as well as for planning preparedness and response against such events. This system utilizes a standardized platform for disaster impact analysis, constructed on a significantly extended version of the HAZUS system developed by the Federal Emergency Management Agency (FEMA) of the United States several decades ago.[1] Because earthquakes, hurricanes, typhoons, wild fires, flooding, and so on have become more intense and more frequent, due partly to global climate change and planetary geological activities, it is imperative to address how societies can deal with disaster situations. While WDISS is the state-of-the-art system for disaster impact analysis, it does not include a prediction or forecast module of natural hazard occurrences, since even now such prediction or forecast has not sufficiently progressed.[2] WDISS utilizes data generated from agencies such as geological or meteorological agencies, for hazard occurrence, intensity, projected routes, etc.

While controlling sternly and tightly the privacy and security of information used, WDISS is intended for public sector use, from the local to the international level, in order to be applicable to any natural hazard and disaster situations.[3] WDISS has an interactive interface with a superb visualization capability and includes several modules to simulate the highly complex processes of a disaster's aftermath and reconstruction on a broad range of social, economic, demographic, behavioral, and environmental dimensions. The overall structure of WDISS consists of three layers: (1) immediate damage and loss assessment; (2) comprehensive impacts estimation; and (3) analysis of recovery and reconstruction strategies and their effects.

To achieve such noble goals, WDISS has the following three major features: seamlessness; comprehensiveness; and, adaptability.

16.2.1 Seamlessness

WDISS can analyze disaster impacts seamlessly over space and time. Damages from a natural hazard are mostly localized but sometimes they reach multiple countries, for example the damages from the 2004 Indian Ocean Earthquake and Tsunami. Because higher-order effects, i.e., ripple effects, from the disaster damages can propagate across the globe via supply chains, for instance the impacts on automobile industries across the countries by the 2011 East Japan Earthquake and Tsunami, the geographical extent of simulation should not be limited to the nation hit by a natural

[1]https://www.fema.gov/hazus (accessed on January 13, 2016). The first version was released in 1997.

[2]On the other hand, the global network of multi-hazard early warning systems was greatly improved under the Sendai Framework for Disaster Risk Reduction 2015–2030 (UNISDR 2015) and had been completed during the Tokyo Framework for Disaster Risk Reduction 2030–2045, which was adopted after the 2027 Great Tokyo Earthquake.

[3]WDISS deals only with natural events, but not man-made events, such as terrorist attacks or war, because we wish that such man-made events did not exist by then.

hazard. Hence, WDISS accommodates a wide variety of spatial extents, from a local area to the entire world, depending on the distribution of damages and losses in a particular disaster situation. Additionally, a disaster simulation system should be able to model a range of activities over time, from emergency response immediately after the event occurrence to evaluation of the effects from reconstruction in the long run (Rose and Guha 2004). These phases consist of quite different activities in terms of their nature and consequences, yet their impacts overlay each other to create complex effects.

WDISS permits users to explore seamlessly impact propagations over space and time, so that decision makers and governments can decide how and to what extent they need to respond and can consider mitigating further effects. This is a major advancement over previous approaches.

16.2.2 Comprehensiveness

In earlier generations, research in disaster impact analysis was mostly limited to specific impacts, such as economic, social, environmental, or others. In WDISS, most aspects that are considered as disaster impacts are included and integrated, such as industries, institutions, capitals (physical, human, social, etc.), environment, and so forth, in order to simulate the impact propagation of a disaster comprehensively. Therefore, several aspects often omitted from impact modeling in the past can be taken into account, including financial, domestic and international aid, and environment. Financial aspects of disaster impacts, including distributional effects, have been discussed for some time (Rose 2004), but have not been integrated well into disaster impact analyses. While the financial condition of most advanced countries may not be significantly affected by even a severe natural hazard (for example, the 2005 Hurricane Katrina in the United States and the 2011 East Japan Earthquake and Tsunami), it becomes quite different for many developing countries, especially for small island nations.

In the meantime, some catastrophic natural hazards also damage the natural environment (for instance, the damaged coastal areas after the 2004 Indian Ocean Earthquake and Tsunami) and/or a drastic reconstruction strategy may alter the surrounding natural environment (for example, the extensive resettlement to higher elevation after the 2011 East Japan Earthquake and Tsunami), which would cause long term effects for the respective and surrounding ecosystems. While in the past it was considered to be inexhaustible, the environment is now considered as a limited property (capital) (Krugman 1998). Therefore, it is included in WDISS as a part of capital that can be altered during the aftermath of a disaster to evaluate the interactive and comprehensive impacts and effects from the respective event.

16.2.3 Adaptability

One of the unique and advanced attributes in WDISS is its adaptability, which comprises flexibility and resilience. Flexibility implies the flexible settings of policy variables and regional attributes. The respective local or regional government hit by a natural hazard can highlight some specific local characteristics and policy implications, such as impacts on natural resources and/or changes in inequality of income distribution, to obtain more locally oriented results. Whereas WDISS has the standardized framework for disaster impact analysis, it can emphasize a range of local attributes and change the priorities for policy variables and options. Allowing this type of analytical modification enables policymakers to define a set of alternatives and choices, while WDISS still derives a result using the standardized setting as the benchmark. Although the main features of WDISS intend toward the macro-perspective covering the affected region as a whole, there are some optional modules for micro-perspectives, such as for first responders, for evacuation strategies either in short- and long-runs, for distribution of relief goods, or for information provision to affected company's business continuity plan (BCP) operation. These modules are available in a standardized form and can be localized during ex-ante user training phases.

Resilience of a society has been a topic of discussion for the past 50 years and modeled in disaster impact analyses to some degree, for example, in the context of industrial production (Rose and Liao 2005). In addition, a broader notion of resilience was proposed as post-disaster endogenous response, in which many factors of a society cope endogenously under a disaster situation (Albala-Bertrand 2013). In WDISS, this resilience is defined as adaptability that is formalized and endogenized with exogenous policy variables, which can change the degree of adaptability under a disaster situation of a particular process. In this way, the policies to enhance adaptability of business operation in a disaster situation, for instance business continuity plan (BCP), can be evaluated for the ex-ante analysis. Furthermore, unintended but related impacts (final demand losses) caused by unfounded rumors and voluntary restraint, whose degree can be dependent more on the cultural background of a particular society, can be also evaluated as changes in social adaptability. WDISS has variables and flexible settings to model such adverse behavioral changes on the respective economy, even though they may be short-lived.

In addition to these three major features, WDISS has several methodological elements that enhance simulation capabilities and the robustness of the results. Such elements include real-time monitoring and simulation capability, stochastic simulation with a variable range of uncertainty, and disaster database and knowledge-base construction. Since these elements are more technical in nature, they are discussed below in the technical notes.

16.3 WDISS Technical Notes

WDISS overcomes several technical challenges and has achieved breakthroughs for its realization and operation. The overviews of such technical breakthroughs are briefly described below.

16.3.1 Analytical Framework and Theoretical Underpinnings

Since the data points have been limited due to the infrequent occurrence of disasters (of course, this is a good thing for society) and each natural hazard is a unique event, it is difficult to generalize how disaster impacts would develop and impact the affected region. In the past, research on regional economic impacts of disasters has been mostly empirical and case specific. This tendency influenced the framework and even the objectives of analysis, which can differ from one study to another, leading to conflicting conclusions even when the studies dealt with the same event (Hallegatte and Przyluski 2010). As such, setting the standardized framework of disaster impact in terms of the aspects, duration, and geographical space to be covered was essential to the construction of a common simulation platform for a wide variety of disaster cases. For example, in terms of duration, earthquakes happen within a few minutes but reconstruction can take several years, while flooding may last up to a few months and might require only a few more months to rehabilitate the damaged regions. Therefore, the flexibility of the definitions was a key to an effective simulation system. Seamlessness, one of the major features of WDISS, allows for suppleness of attributes so that the users can choose or explore the simulated extent and duration of a disaster.

While suppleness is necessary, robustness of simulation results across various settings is important to generate reliable policy recommendations. To make the system robust across a wide range of settings, the theoretical underpinnings of disaster processes should be placed firmly, for example how people and institutions react and respond to a disaster situation. Since the pioneering work of Dacy and Kunreuther (1969), theoretical investigations of disaster impacts and reconstruction processes had been limited for both short- and long-run effects (Okuyama 2007). Through empirical research based on comparison across disaster cases using the standardized definitions, whose findings have been accumulated in the World Disaster Database, theoretical investigations of behavioral changes under a range of disaster situations have been conducted and have developed and tested disaster theories since the late 2010s. In WDISS, some policy variables, for instance the schedule for recovery and reconstruction activities, can be controlled exogenously by users (or can be endogenous for deriving an optimal case), several behavioral variables, such as the responses and reactions of private companies and households, are endogenous to the simulation model so that the results can reflect behavioral changes and reproduce adaptive behaviors. The model also can deal with the possibility of multiple equilibria or complex system outcomes, hinted from Davis and Weinstein (2004) when they modeled the reconstruction of the Japanese industries from World War II.

16.3.2 Modeling Scheme

Because a disaster process is a multi-aspect event that impacts physical, economic, social, demographic, financial, and environmental systems, a comprehensive event simulation calls for a multi-dimensional system capable of integrating the various aspects of the impacts. Within each aspect, some factors interact while others are interdependent from each other, resulting in a multi-layer modeling structure. Combining all of these results in a large system, while persistent advancements in computational capability and data availability to this date have eliminated computer hardware or software as a binding constraint on the implementation of WDISS. Rather, the remaining issue was the modeling scheme and data structure of such a complex system. For each aspect, separate models have been available and have been employed in disaster impact analysis independently. The effective integration among these various aspects is the main and most crucial part of WDISS. In the past, multidisciplinary studies proposed a series of integrated models, such as socio-economic models, demo-economic models, and transportation (physical)-economic models (see some examples in Okuyama and Chang 2004). Comprehensive analysis of disaster impacts in WDISS required further integration of these aspects in an effective and proficient way.

To integrate multiple aspects, the data structure and the linkages among aspects have to be managed to make them consistent across the board. For instance, some physical and economic aspects use quantitative values, while social and environmental aspects often employ qualitative information. Even within quantitative data, a physical aspect is usually measured with the metric system, whereas an economic aspect is most commonly expressed in monetary scale. On top of this, to facilitate seamless integration across the aspects, micro-level data and macro-level data need to be consistent. This was actually quite difficult to do in the past. For example, information about local economies, even in advanced countries, did not add up to their equivalent data at the national level due to statistical discrepancies and measurement errors. Seamlessness over time was also problematic because of data availability, since most of the official data represented a point in time or the aggregate over a given period of time, rather than being provided in continuous form. The seamlessness of simulation does not require data in a continuous form, but the suppleness to use shorter periods, such as a week, and the consistency among various time frames had to be achieved. These issues have been addressed and mostly overcome with recent advancements in the accuracy and frequency of data collection, led by the United Nations Statistical Office (UNSO). The data used in WDISS follows mostly the UNSO systems in addition to exclusively collected data as well as on-going disaster data, such as damage and casualty data that can be collected in real-time using the dedicated satellites and monitoring systems of WDISS.

16.3.3 Methodological Elements

WDISS incorporates advanced methodological elements. One of them is the real-time monitoring and real-time simulation capability. Whether a natural hazard is occurring over a relatively short period (for example, earthquakes) or over a long period (as in some cases of flooding), timeliness of gathering and collecting damage data is essential for the relevance of simulation result. In addition, because recovery and reconstruction activities are often faster paced and more regionally concentrated than similar operations in ordinary circumstances, such as large construction projects, monitoring them in a careful and continuous manner is necessary to evaluate the progress and further effects emitting from them within the region and across other regions. Whereas such technologies for real-time monitoring systems are in an engineering domain, real-time simulation capability can be discussed in the context of various modeling frameworks, including economic, demographic, and environmental.

Another methodological element in WDISS is the capability of stochastic simulation. Stochastic processes have long been central to robust analysis of simulation forecasts or projections. Moreover, a disaster often increases uncertainty of perceptions toward the future, leading to variations in behavioral changes by the affected people and institutions. Stochastic processes are able to accommodate such fluctuations in uncertainty, while theoretical corroboration for how the alterations of behavior occur in a disaster situation determines the range of variations.

WDISS is not only a simulation system but also comprises the disaster data and knowledge base. It accumulates the simulation results as well as the actual outcomes based on the decisions made during use. These data are then stored in the system's data and knowledge base that is linked with the World Disaster Database, which includes outcomes of past disaster research, from short-run impact assessments to long-run effects analyses. The system's data and knowledge base is employed to improve the expert system through deep-learning of actual disaster processes. This also enables archiving of the disaster data and policy outcomes for ex-post analysis of a particular disaster process to examine how and to what extent the recovery and reconstruction plans alleviated the potential effects. This type of research is critical to mitigate and prepare for the future disasters.

16.4 Brief History of WDISS Development: Regional Scientists at the Helm

WDISS is a highly complex modeling system integrating various aspects from economy to environment. Its development required a broad range of expertise, from physical sciences and technology to social sciences. Therefore, WDISS became a multi-disciplinary project with the participation of numerous researchers and practitioners from different fields. In the early stage of its development, after starting the project in 2017, disaster scientists led the project. However, due to the complexity of disaster impacts and the difficulties of dealing with tradeoffs among various problem domains, the development process became disorganized. A team of regional scientists finally took over the leading role in the project around 2025, because of regional scientists' skills in integrating a diversity of empirical models. Furthermore, in terms of the seamlessness of WDISS over space and time, regional scientists had experience relevant to the analysis of disaster impacts over space (Tsuchiya et al. 2007) and time (Donaghy et al. 2007), and thus were able to extend the model to the time-space continuum in an empirical context, a challenge that had been monopolized by physicists since Einstein. Recent generations of regional scientists have broadened WDISS features to deepening our understanding of disaster processes, adding, for example, spatial analyses of disaster impacts and reconstruction effects to examine spatial externalities. Regional scientists have been particularly well suited to the leading role, continuing in the traditions of the father of regional science, Walter Isard, who declared that: "the field of Regional Science has been and is interdisciplinary" (Isard 1998).

After regional scientists took over the helm, the development process progressed very smoothly. However, the Great Tokyo Earthquake occurred in 2027, which changed our perspectives of disaster impact considerably. Not only were the numbers of casualties and damages on every aspect of the society beyond the ex-ante estimations, but so also were the impact propagations over space, time, and extent unprecedented. Every catastrophic disaster had significantly changed and expanded our understanding of disaster impacts, such as the 1995 Kobe Earthquake, the 2004 Indian Ocean Earthquake and Tsunami, the 2005 Hurricane Katrina, the 2010 Haiti Earthquake, and the 2011 East Japan Earthquake and Tsunami to name a few, and this 2027 event was not an exception.

The project team had to reconsider extensively the system structure to incorporate the lessons learnt from the 2027 event and its consequences, and finally revised its overall structure and the standardized definitions of disaster in 2032. Based on this common framework, the past disaster cases were retrospectively re-evaluated and the results were fed into the World Disaster Database for contributing to the development of new disaster theories. Scholars from a broad range of fields participated the theoretical development of the system, from disaster scientists and geophysicists to ecologists and sociologists, and even economists. The developed disaster theories in each field were harmonized for the integrated framework utilizing the technique inspired by Lisi (2007). A primary version of the unified theory of disaster was formulated in the mid 2040's, and the project team initiated programming of the simulation system based on it. The unified theory has been continuously updated employing the World Disaster Database and the system's data and knowledge base, fine-tuning the structure and parameters of the simulation system.

During the 2050's, a series of satellites and ground level equipment for the monitoring network that collect data and monitor disaster processes for the simulation system were launched and placed. At the same time, a sequence of calibrations and perfections of the system was carried out in this period using some disaster cases in this period. The final step of the system construction was to incorporate with the artificial intelligence sub-system with nano-programming (Parisi 2012) that is able to self-debug the programs and to self-update and even self-regenerate the system within the pre-determined framework. In this way, the completed WDISS system no longer has a version number, which was a common practice for software in previous generations. After the final adjustment processes and some test runs, WDISS was released in 2065.

Whereas the occurrences of natural hazards cannot be prevented, their impacts can be mitigated and reduced with appropriate countermeasures based on disaster impact analyses. The release and utilization of WDISS is expected to contribute now to the safety and security of all humankind.

References

Adam C (2013) Coping with adversity: the macroeconomic management of natural disasters. Environ Sci Policy 27(Supplement 1):S99–S111

Albala-Bertrand JM (2007) Globalization and localization: an economic approach. In: Rodriguez H, Quarantelli EL, Dynes RR (eds) Handbook of disaster research. Springer, New York, pp 147–167

Albala-Bertrand JM (2013) Disasters and the networked economy. Routledge, Oxon

Arto I, Andreoni V, Rueda Cantuche JM (2015) Global impacts of the automotive supply chain disruption following the Japanese earthquake of 2011. Econ Syst Res 27(3):306–323

Cavallo E, Galiani S, Noy I, Pantano J (2013) Catastrophic natural disasters and economic growth. Rev Econ Stat 95(5):1549–1561

Dacy DC, Kunreuther H (1969) The economics of natural disasters. The Free Press, New York

Davis DR, Weinstein DE (2004) A search for multiple equilibria in urban industrial structure. Working Paper 10252. National Bureau of Economic Research

Donaghy KP, Balta-Ozkan N, Hewings GJD (2007) Modeling unexpected events in temporary disaggregated econometric input-output models for regional economies. Econ Syst Res 19(2):125–145

DuPont W, Noy I (2015) What happened to Kobe? A reassessment of the impact of the 1995 earthquake. Econ Dev Cult Chang 63(4). doi:10.1086/681129

Felbermayr G, Gröschl J (2014) Naturally negative: the growth effects of natural disasters. J Dev Econ 111:92–106

Hallegatte S, Przyluski V (2010) The economics of natural disasters. CESifo Forum 2:14–24

Isard W (1998) The setting and introduction. In: Isard W, Azis IJ, Drennan MP, Miller RE, Saltzman S, Thorbecke E (eds) Methods of interregional and regional analysis. Ashgate, Brookfield, VT, pp 1–3

Krugman P (1998) Looking backward. In: Krugman P (ed) The accidental theorist. WW Norton & Company, New York, pp 196–204

Lisi AG (2007) An exceptionally simple theory of everything. ArXiv: 0711.0770

Mochizuki J, Mechler R, Hochrainer-Stigler S, Keating A, Williges K (2014) Revisiting the ‘disaster and development’ debate – toward a broader understanding of macroeconomic risk and resilience. Clim Risk Manag 3:39–54

Okuyama Y (2007) Economic modeling for disaster impact analysis: past, present, and future. Econ Syst Res 19(2):115–124

Okuyama Y (2016) Long-run effect of a disaster: case study on the Kobe earthquake. Singap Econ Rev 16(1):1640009-1–1640009-18

Okuyama Y, Chang SE (eds) (2004) Modeling spatial and economic impacts of disasters. Springer, Berlin

Okuyama Y, Sahin S (2009) Impact estimation of disasters: a global aggregate for 1960 to 2007. Policy Research Working Paper 4963. The World Bank

Okuyama Y, Hewings GJD, Sonis M (2004) Measuring economic impacts of disasters: interregional input-output analysis using sequential interindustry model. In: Okuyama Y, Chang SE (eds) Modeling spatial and economic impacts of disasters. Springer, Berlin, pp 77–101

Parisi L (2012) Nano architectures: the synthetic design of extensions and thoughts. In: Karatzogianni A, Kuntsman A (eds) Digital cultures and the politics of emotion: feelings, affect and technological change. Palgrave Macmillan, Basingstoke, pp 33–51

Rose A (2004) Economic principles, issues, and research priorities in hazard loss estimation. In: Okuyama Y, Chang SE (eds) Modeling spatial and economic impacts of disasters. Springer, Berlin, pp 13–36

Rose A, Guha GS (2004) Computable general equilibrium modeling of electric utility lifeline losses from earthquakes. In: Okuyama Y, Chang SE (eds) Modeling spatial and economic impacts of disasters. Springer, Berlin, pp 119–141

Rose A, Liao SY (2005) Modeling regional economic resilience to disasters: a computable general equilibrium analysis of water service disruptions. J Reg Sci 45(1):75–112

Tsuchiya S, Tatano H, Okada N (2007) Economic loss assessment due to railroad and highway disruptions. Econ Syst Res 19(2):147–162

UNISDR (2015) Sendai framework for disaster risk reduction: 2015–2030. UNISDR, Geneva

Yasuhide Okuyama is professor, University of Kitakyushu, Japan. His research interests center on economic impact of disasters; regional science; input-output analysis; and urban and regional planning. Previous faculty positions were at the International University of Japan, West Virginia University, and the University at Buffalo, the State University of New York. Dr. Okuyama earned his doctoral degree in regional planning from the University of Illinois at Urbana-Champaign in 1999.

Chapter 17
Regional Sustainability and Resilience: Recent Progress and Future Directions

Elena G. Irwin, Tim Jaquet, and Alessandra Faggian

17.1 Introduction

Heightened awareness of the interdependencies among the economy, environment and society has spurred interest in aspects of community and regional well-being that go beyond a sole focus on economic growth. Sustainability—concerned with societal well-being and the maintenance of natural, manufactured and human capital and other community assets over the long run—and resilience—focused on short-run recovery and adaptation to negative external shocks—have emerged as two additional criteria to guide policy decisions. Policymakers and stakeholders are increasingly concerned about the vulnerabilities of cities and regions to multiple stressors and shocks, including environmental (e.g., climate change, water scarcity), economic (e.g., recessions), and social (e.g., police violence, racial unrest). Creating more sustainable and resilient places requires not only an understanding of these complex processes and their potential costs to society, but also the responsiveness of individuals and communities to these events in ways that may either reinforce or offset these costs.

To illustrate, consider the potential impacts of climate change. Increasing average temperatures and precipitation, rising sea levels and more frequent and extreme weather events, such as droughts and floods, are projected under a range of current and alternative future growth scenarios (IPCC 2014) and are expected to impact regional and national economic productivity (e.g., Zivin and Neidell 2014; Schlenker and Roberts 2009). Studies reveal not only the sensitivity of economic productivity to temperature increases, but also the highly non-linear responses of economic production to such increases (Burke et al. 2015). Individual and

E.G. Irwin (✉) • T. Jaquet • A. Faggian
Department of Agricultural, Environmental and Development Economics, The Ohio State University, 2120 Fyffe Road, Columbus, OH, 43210, USA
e-mail: irwin.78@osu.edu; jaquet.5@osu.edu; faggian.1@osu.edu

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_17

community adaptations to climate change will offset potential future losses. If climate change causes future households to face increased summer temperatures, for example, then this will induce firms responding to large future profits to develop new efficient air conditioning products and housing (Kahn 2016).

As Martin and Sunley (2015) note with respect to resilience, local policy developments have outpaced our scientific understanding of how concepts of sustainability and resilience apply to a region or community. For example, most cities in the U.S. have an office that is charged with developing and implementing sustainability goals, and more than half the states have completed, or are in the process of developing, a comprehensive climate action plan. Given the public concerns over climate change and other environmental, economic and social shocks and stressors that challenge the wellbeing of individuals and communities, questions of how cities and regions can develop in more sustainable and resilient ways constitute an important research agenda for regional scientists.

The purpose of this chapter is to highlight current research approaches to regional sustainability and resilience and discuss future directions that will shape this research agenda. Before proceeding, we pause to ask an important clarifying question: are the factors that affect the sustainability or resilience of a region any different than those that affect regional growth? In other words, are new concepts and methods really justified? It is clear that sustainability is a different concept than economic growth. While both seek to assess well-being, they are concerned with different aspects of social welfare. Economic growth reflects technological innovations and improvements in efficiency. It is typically measured in terms of the gross domestic product (GDP) of a region or country—defined as the value of all goods and services produced in the region in a given time period—or in other words, the value of all the changes in manufactured capital over time. Sustainability is concerned with changes in a broader set of assets—including manufactured, human, natural, social, and institutional capital—and is grounded in an ethical view that seeks an equitable distribution of these assets over time (intergenerational equity) and across people and regions at a given point in time (intragenerational equity). Alternatively, resilience is focused on shocks or disturbances that cause a deviation, either temporarily or permanently, from long run trajectories. Resilience places a particular focus on negative shocks and periods of stress; rather than factors that drive growth during normal circumstances, the focus is on the processes that allow a local area to endure adversity and "bounce back" after a shock. Uncertainty and the potential for crossing thresholds or tipping points associated with some stressors increase the importance of understanding questions about sustainability and resilience. These considerations set questions of sustainability and resilience apart from much of the literature on economic growth.

The rest of the paper is organized as follows. First, we discuss the different ideas that have been used to conceptualize resilience and sustainability in general, and how they have been specifically operationalized at a regional scale. We then consider factors that will shape future policy needs and research opportunities for scholarship on regional sustainability and resilience topics, focusing in particular on the challenges of modeling the openness and spatial interdependencies of regions;

key interactions among environmental and economic systems; and the welfare impacts and trade-offs that are imposed by sustainability and resilience policies. We conclude with a brief discussion of why these are topics of central importance to regional science.

17.2 Defining and Measuring Regional Sustainability and Resilience

Local economies are dynamic and consist of many interrelated components. When studying the long-term trajectory of a region and its response to an external shock, these components and their interactions create a dimensionality that makes analysis very complex (Anderies et al. 2013). A challenge in evaluating the sustainability or resilience of a region is to necessarily abstract from much of this complexity, but retain enough detail so that key trade-offs that result from these interactions can be considered. Doing so depends on how these concepts are defined and operationalized. Here we highlight current approaches to defining and measuring sustainability and resilience respectively.

17.2.1 Sustainability

Sustainability is fundamentally concerned with intergenerational equity and the importance of maintaining some type of balance between the economy and the environment, so that future generations have sufficient means to meet their own needs.[1] Weak sustainability is defined as non-declining social welfare over time, and focuses on the aggregation of different types of capital that form what Dasgupta (2001) calls "society's productive base," and maintaining the aggregate value of these capital stocks—called comprehensive or inclusive wealth—over time. Definitions of the various types of capital can vary, but the accounting should encompass not only manufactured capital stocks, but also natural capital stocks that include earth minerals and metals, land, and stocks of fish, trees and water; human capital that includes education, skills, and knowledge generated from R&D; institutional capital that encompasses all forms of good governance, including strong and transparent institutions and a lack of corruption; and social capital that

[1]The starting point for most discussions of sustainability is the famous Brundtland Commission report (1987) in which sustainable development is defined as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs ... At a minimum, sustainable development must not endanger the natural systems that support life on earth." Sustainability also encompasses intra-generational equity and the distribution of resources and consumption across developed and developing regions as well as at more localized scales, e.g., within a city.

includes the networks of relationships among individuals within a community. Weak sustainability assumes that it is possible to substitute among different types of capital, so that it is possible to offset a decline in a capital stock through investments in other capital stocks. For example, extraction of oil or gas from a region diminishes its natural capital, but severance taxes that generate revenues that are then invested in new roads, schools or other public services in the region can compensate for this loss by investing in these other community assets. If the community benefits from these investments are sufficient to offset the social costs of extraction, then the change is said to be weakly sustainable.[2] Evaluating these tradeoffs relies on being able to value these capital stocks according to their shadow prices, including the many types of natural, social and human capital that are public goods and whose marginal social value is not fully reflected in the market.

On the other hand, strong sustainability emphasizes physical limits to the consumption of resources and defines sustainability as meeting specific environmental limits, e.g., in terms of resource extraction and pollution impacts. A recent example is the notion of planetary boundaries (Rockström et al. 2009; Steffen et al. 2015) defined as scientifically based levels of human perturbation of the Earth system beyond which the functioning of basic biophysical processes—e.g., climate change, biosphere integrity, land-system change, freshwater use, ocean acidification—may be substantially altered. The implication is that transgressing one or more of these boundaries creates substantial risk of destabilizing the current state of the Earth system, due to non-linear changes, and that these changes may generate enormous societal impacts. However, this approach does not attempt to value these impacts and therefore does not rely on quantifying trade-offs to determine these boundaries.

Given its all-encompassing nature, a central challenge is to operationalize definitions of sustainability at local or regional scales. Strong sustainability measures, such as the ecological footprint (e.g., Borucke et al. 2013) and material flows analysis (e.g., Bringezu et al. 2003), penalize regions for importing goods and services from other countries or regions and therefore seem to imply that self-reliance increases sustainability. This is nonsensical from an economics viewpoint since trade is welfare enhancing, and autarky would imply highly inefficient regions with low levels of social welfare. On the other hand, measures of comprehensive wealth have mostly been applied at the country level (Arrow et al. 2012) given the added difficulty of measuring sustainability in open economies.

In light of these challenges, by far the most common approach to operationalizing sustainability at local and regional scales is with indicators that track key environmental, economic and social variables. This approach may rely on stakeholder input

[2]While there is a strong correspondence between intertemporal benefit cost analysis and weak sustainability, efficiency and sustainability are not the same thing. An efficient outcome is one in which the present value of net benefits are maximized; a sustainable outcome is one in which these net benefits are non-declining over time. Discounting plays a critical role in determining whether an outcome is either efficient or sustainable. For more discussion, see Irwin et al. (2016a).

or best practices defined by a third party[3] to determine the set of indicator variables. In addition to tracking trends over time, specific targets may be determined, e.g., 30% of all land should be open space or 20% increase in energy efficiency by a target date, which then provides a set of strong sustainability rules to guide decision making (Mori and Christodoulou 2012; Rennings and Wiggering 1997). Aggregation methods can be used to construct an overall index of sustainability using stakeholder inputs or expert opinion to assign relative weights (e.g., Dovern et al. 2014; Sullivan 2002). While indicators usefully track trends over time and provide a means of determining strong sustainability based on specific targets or limits, they do not provide an understanding of the underlying processes that determine changes in indicators or of the tradeoffs among indictors. Thus, they are limited in terms of guiding policies that seek to manage these tradeoffs and, unless they are translated into measures of costs and benefits, cannot be used to evaluate weak sustainability.

17.2.2 Resilience

Three broad definitions of resilience have emerged in the literature (Martin and Sunley 2015; Pendall et al. 2010) and each relates to a slightly different key concept (Table 17.1).[4]

The idea of recovery, coined "engineering resilience" by Holling in 1973, is borrowed from the physical sciences and refers to the ability of a material to recover its original position after an external shock. Because of its clarity and simplicity, this definition of resilience has become very popular in the study of regional resilience. Moreover, it is easy to measure this type of resilience by looking at the trend over time of indicators, such as employment or GDP, which are observable and do not require any assumption about the nature of the underlying processes. However, looking at only one outcome variable does not give a full picture of the complexity of how local systems react to shocks. Differences across models and uncertainty about how these factors interact make it difficult to compare or synthesize the results of different resilience studies. Even when multiple indicators are chosen, there is a similar subjectivity in how the researcher decides to weigh the different elements of the index (Davoudi and Porter 2012). The time frame for recovery also strongly shapes the results. Looking at job recovery 1 year versus 3 years after a recession, for example, can paint a very different picture about which factors are important.

[3]For example, Sustainability Tools for Assessing and Rating (STAR) Communities is a nonprofit organization that provides sustainability certification of cities and communities (http://www.starcommunities.org/)

[4]It is worth noting that there is nothing concrete about the titles given to the definitions below. These words show up frequently in the literature, often used in different contexts depending on the discipline. Because there are no established universal titles for these concepts, they were chosen here simply to facilitate reference.

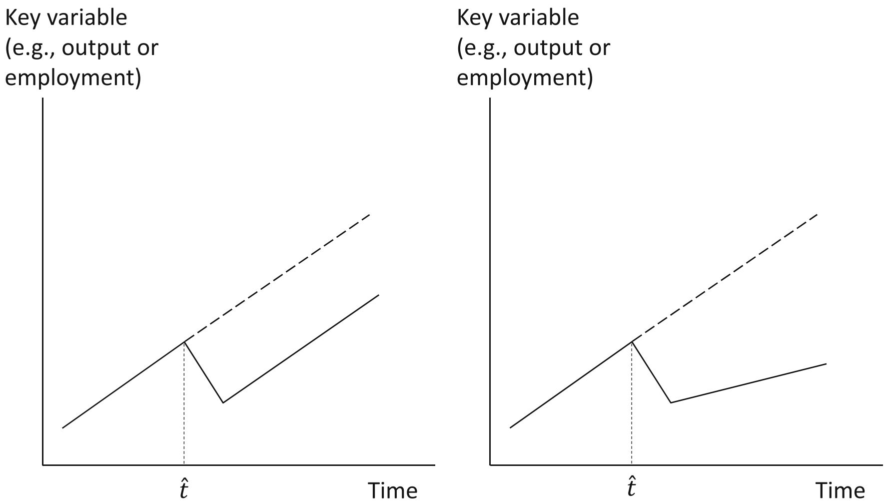
Table 17.1 Concepts of resilience

Definition/Idea	Source	Interpretation	Economic parallel
Recovery Ability to bounce back from a shock	Engineering and Physical Sciences	The speed and degree to which something returns to its pre-shock state	The belief that an economy has a global equilibrium state to which it will eventually return. Any deviations are temporary disturbances. Primarily focused on speed of recovery or level or recovery after a certain period of time
Resistance Ability to remain unaffected by a shock	Ecology, Social Ecology, and Environmental Sciences	The amount of force a system can sustain without being shifted from its current state to a new equilibrium	The size of the deviation in a global equilibrium framework, or the relative distance from a threshold or likelihood of a regime shift in models with multiple equilibria
Robustness Ability to adapt or evolve in response to change	Psychology and Organizational Theory	Ability of an individual, system, or organization to maintain its core functions in response to negative shocks through adaptation	Closest to work in evolutionary economics; focuses on industry composition and diversity, related and unrelated industries, and the role of human capital in the development and adoption of new technology

Holling (1973) also puts forth a second definition that he calls “ecological resilience” that relates to the concept of resistance to shocks. Defined by ecologists as the size of force that is necessary to move a system to a new equilibrium, in economics it is usually measured by the size of the disturbance (e.g. Balland et al. 2015). While clearly different than the notion of recovery, these concepts are related: the time it takes to return to equilibrium may often be proportional to the magnitude of disturbance (Simmie and Martin 2010). However, this second definition takes on additional value in systems subject to tipping points that may shift the economy to a qualitatively different state—a so-called regime shift. If these shifts are irreversible, this renders a community’s capacity to bounce back to the original state irrelevant.

A third definition of resilience focuses on the idea of robustness. Robustness has been described as “the ability of the system to withstand either market or environmental shocks without losing its ability to allocate resources efficiently” (Perrings 2006, p. 418) or “the ability of organizations to maintain their core functions in the face of disturbance by anticipating key events from emerging trends and constantly adapting” (Marcos and Macaulay 2008, p. 1). The concept of robustness is a more dynamic interpretation of resilience, and adaptation plays a key role (Folke et al. 2010; Boschma 2015). Robust communities are typified not only by the ability to mitigate shocks, but to adapt, reorganize and capitalize through increased “functionality” in accordance with the changes experienced (Anderies et al. 2013). While conceptually appealing, this approach to defining resilience is even more difficult to measure than recovery or resistance. Aggregate measures, such as total employment or per-capita income, may give some idea of whether a region has recovered or not, but do not give any insight into any possible restructuring that may have happened or the overall impact on social welfare. Aggregate employment measures, for instance, do not capture changes in the sectoral mix of employment.

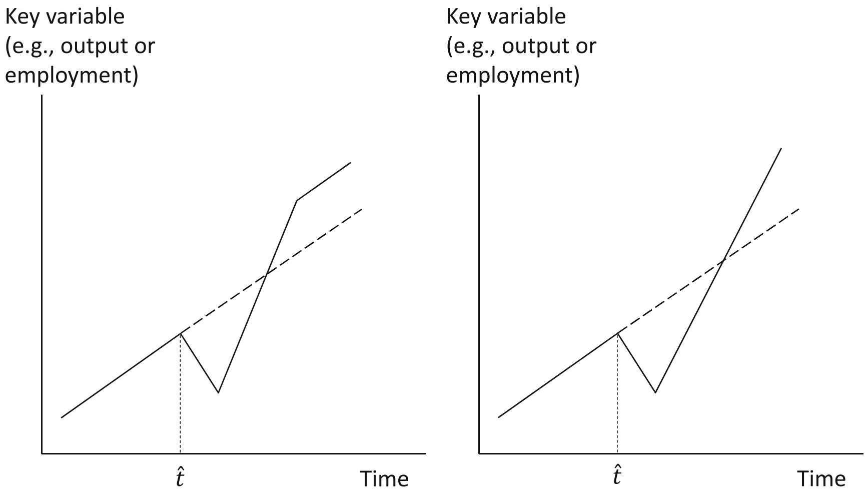
Relating all three definitions is difficult and in some cases these definitions can be at odds with each other. For example, mechanisms that may help a community to be more resistant to a negative shock and thus prevent a shift to an inferior state may also make governance systems less flexible, thereby reducing the community's ability to adapt (Anderies 2015). The concept of hysteresis offers a potentially unifying framework. Hysteresis suggests that a one-time impact can shift the long-term trajectory of an area (Romer 2001). Figures 17.1 and 17.2, borrowed from Martin (2012), illustrate this concept with possible responses of a key variable, e.g., employment or output, to a negative shock. The graphs show that the impact can result in a linear shift of the growth path (Figs. 17.1a and 17.2a) or a lasting impact on the growth rate itself (Figs. 17.1b and 17.2b). Recovery is indicated by the slope of the line after the shock and how quickly the key variable returns to the dotted line; resistance is indicated by the distance below the dotted line the variable initially falls; and robustness is represented by how far above the dotted line the variable rises before leveling off (Fig. 17.2a) or the difference in slope between the dotted line and the new growth rate if the trajectory persists (Figs. 17.1b or 17.2b). Thus resilience is



(a): A linear shift downward in growth following a shock in time $\hat{t}$

(b): A shift downward and persistent negative effect on the growth rate following a shock in time $\hat{t}$

Fig. 17.1 Negative hysteresis. (**a**) A linear shift downward in growth following a shock in time $\hat{t}$, (**b**) a shift downward and persistent negative effect on the growth rate following a shock in time $\hat{t}$



(a): A temporary shift downward in growth following a shock in time $\hat{t}$ followed by adaptation leading to an improved state and return to pre-shock growth rate

(b): A temporary shift downward in growth following a shock in time $\hat{t}$ followed by adaptation leading to an improved state with a higher growth rate

Fig. 17.2 Positive hysteresis. (**a**) A temporary shift downward in growth following a shock in time $\hat{t}$ followed by adaptation leading to an improved state and return to pre-shock growth rate, (**b**) a temporary shift downward in growth following a shock in time $\hat{t}$ followed by adaptation leading to an improved state with a higher growth rate

a multi-dimensional concept and, rather than providing competing definitions, these definitions describe different aspects of a larger dynamic system (Martin 2012).

Beyond the challenges of defining resilience, operationalizing these definitions through empirical measurement is equally daunting. As Carpenter et al. (2001) points out, the initial task of implementing any type of empirical analysis is to answer questions regarding the resilience "of what" and "to what." Regarding the "to what" question, much of the research has dealt with exogenous global recessions (Martin 2012; Davies 2011; Cerra and Saxena 2008), e.g., the oil crisis in the early 1980s or the global recession of 2008–2009. These studies leverage the exogeneity of the shock, known shock characteristics, and broad geographical impact to empirically implement the analysis, albeit degree of impact on individual communities can be difficult to quantify. Other studies simply look at growth paths and look for periods of negative growth. While the specific shocks may be unknown, the analysis elicits traits that appear to be correlated with more frequent, more severe, or longer lasting recessionary periods (Balland et al. 2015; Augustine et al. 2013; Ormerod 2010). These and other studies vary in geographic scope, e.g., from metropolitan (Augustine et al. 2013; Balland et al. 2015) to larger regions (Lagravinese 2015 (Italy); Cellini and Torrisi Cellini and Torrisi 2014 (Italy);

Fingleton et al. 2012 (UK), or country level analysis (Ormerod 2010; Cerra and Saxena 2008). Alternatively, there have also been more in-depth, qualitative case studies of a specific area's traits or the role of a certain trait's role in the area's economy (Evans and Karecha 2014; Treado and Giarratani 2008; Foster 2007).

The "of what" question is more complicated because the answer partially depends on the chosen definition of resilience. Most studies focus on economic indicators such as employment levels (Lagravinese 2015; Fingleton et al. 2012), employment growth rate (Augustine et al. 2013), per-capita GDP (Cellini and Torrisi 2014) or national GDP (Cerra and Saxena 2008). Balland et al. (2015) take a different approach, attempting to look at economic drivers by analyzing patent generation as a proxy for local innovation capability. While using a single variable as a proxy for regional resilience has its advantages—primarily simplicity and ease of interpretation—we argue that a better approach is to look at more aggregate ways of evaluating regional resilience, which account for the system complexity of local areas and assess tradeoffs using a measure of aggregate social welfare. We develop these points further in the next section.

17.3 Future Research Directions, Opportunities and Challenges

Where will regional research on sustainability and resilience take us in the future? What societal issues will be the most pressing and what are the likely research challenges? Undoubtedly technological innovations and human adaptations will continue to spur economic growth, globalization and the increasing wellbeing of many. On this basis, some may argue that concerns over regional sustainability and resilience will wane in the future—a belief that we would argue is far too optimistic. Growing demand for global resources coupled with increased regional interdependence, resource scarcities, environmental degradation, and uneven distribution of societal gains and losses will push sustainability and resilience questions even further to the forefront of research and policy agendas. Taken together, these trends suggest that future economic growth will be conditional on achieving and maintaining sustainable and resilient systems—e.g., low-carbon energy resources, drought-resistant agriculture, renewable industrial feedstocks, and more equitable communities. In a future of ever greater economic interdependence, resource scarcity and social connectivity, we anticipate that advancing sustainability, resilience and equity goals will become the key policy objectives of communities.

Future research on sustainability and resilience topics will progress beyond questions of definition and measurement to a focus on modeling and welfare assessment. These efforts will benefit from increasing interdisciplinary collaborations among social and natural scientists and the "big data" revolution that is already underway (Miller 2015). Advancements in data collection—including real time sensor technologies, geographic information systems, and increased access to administrative data sources—are enabling data to be created at spatial and temporal resolutions that were previously impossible. However, before we can capitalize on these new resources, significant investment must be made into hardware, human capital, and developing new data analysis techniques to manage data sets of this magnitude (Einav and Levin 2014). Creative insight that identifies what new questions we can answer, and discerning which parts of this massive data are relevant, will become increasingly valuable skills in this new landscape.

Future policy needs and data-rich research opportunities underscore the importance and feasibility of developing spatially and temporally explicit models that can be usefully applied to assess the sustainability and resilience of regions. In the remainder of this section we focus on key research areas that we view as particularly germane to addressing questions of regional sustainability and resilience: (1) open economies and spatial interdependence, (2) dynamic aspects of regional economies and interdependence with environmental systems, (3) welfare-based assessment, and (4) trade-offs between sustainability and resilience goals.

17.3.1 Open Economies and Spatial Interdependence

Future innovations in communications and transportation imply that economic, social and political interdependencies at regional to global scales will continue to be defining features of regions. However, many approaches to assessing the sustainability or resilience of a region ignore the fundamental openness of regions. For example, an indicator approach tracks many of the variables that change over time due to this interdependence—e.g., population, income, sectoral growth—but does not seek to account for their dependence on other regions. Other approaches seek to make trading relationships explicit. For example, so-called consumption-based accounting (Wiedmann et al. 2015) attempts to trace the embedded emissions or resource extraction associated with consumption of final goods based on global trade flows among countries. Thus, while China is the largest producer of CO_2, an accounting of consumption-based CO_2 emissions reveals that many developed countries have much higher levels of embodied emissions based on their demands for products produced in China and elsewhere (Davis and Caldeira 2010). Such analyses are interesting, and provide another view of trading relationships, but cannot be directly interpreted in terms of sustainability. A weak sustainability approach would require accounting for the externalities generated, but this accounting is done on the basis of the welfare of residents living in the country or region of interest. So, for example, a sustainability analysis of the U.S. would account for the externalities associated with production in the U.S. and elsewhere that impacts the welfare of U.S. residents. This would include all domestic externalities associated with production in the U.S., as well as well as externalities from foreign production that affect U.S. residents (e.g., see Arrow et al. 2012). It would not include the regional externalities associated with the production of goods in China, despite the fact that some portion of these goods are produced for the U.S. market, assuming that these regional externalities only impact the welfare of Chinese residents.

Openness introduces other challenges into the analysis of the sustainability or resilience of a region. In assessing weak sustainability, changes in household or firm location cannot be treated as exogenous; therefore, explicitly modeling of population and employment changes, and how this impacts the value of the region's aggregate capital stocks, is difficult, yet essential. This interdependence has an important spatial dimension, since regions that are closer together tend to be more dependent.

The role of spatial effects is also important in assessing resilience. From the national perspective, resilience may be maximized by allocating resources to highly specialized regions if the benefits from this specialization spillover to neighboring regions or extend nationally. However, scarcity implies that there may often be a competitive dynamic. For example, one of the factors that made Munich more resilient after World War II was the relocation of major industry from Berlin (Evans and Karecha 2014). If resilience becomes simply a measure of displacing hardship and not actually mitigating negative shocks, this will have implications for aggregate social welfare, sustainability, and resilience of the larger region.

17.3.2 Integrated Modeling of Economic and Environmental Systems

While structural models of regional economies have a long history in regional science, current empirical approaches favor reduced-form econometric analysis that can isolate the causal effect of a policy or key variable. Both approaches are critical in assessing regional sustainability and resilience, but much more work is needed to advance the state of the art of integrated models of regional economies and environmental processes. As Irwin et al. (2016a) describe in more detail, integrated economic-ecological models are needed to generate what Arrow et al. (2003) and others (Dasgupta 2001, 2009) refer to as an economic "program" or forecast that describes the evolution of capital stocks and resource consumption over time. This is necessary to calculate shadow prices, which are essential for weak sustainability and intertemporal cost benefit analyses. Estimating the shadow price of resilience is an important and emerging research area with many challenges and opportunities for future contributions (Barbier 2016; Irwin et al. 2016a). Finally, dynamic models that describe the state transitions of key capital stocks are important for representing the underlying dynamics of a system, analyzing regime shifts from one domain of attraction to another, and assessing trade-offs associated with a more or less resilient state. For example, Chen et al. (2009) develop a dynamic model of a regional economy with endogenous environmental amenities and examine the trade-offs between the resilience of a "balanced growth-pollution" steady state and the efficient outcome that maximizes net benefits.

Integrated assessment modeling at the global scale is well established (e.g., Manne et al. 1995; Parson and Fisher-Vanden 1997; Sokolov et al. 2005). This approach seeks to understand the environmental and economic impacts of climate change under baseline and alternative scenarios, e.g., by combining environmental process models of the climate system with global economic and energy models. Integrated modeling at regional scales is an emerging area of study. For example, Platform for Regional Integrated Modeling and Analysis (PRIMA)[5] is an integration of regional climate, hydrology, land use, economics and energy systems developed to study interactions of climate, energy, water, and land at local and regional scales (Kraucunas et al. 2015). The model allows for substantial spatial and agent heterogeneity, but lacks an equilibrium model of the regional economy. Rose and Liao (2005) develop a computable general equilibrium model that links production function parameters to hypothetical producer adaptations to a disaster by developing algorithms for recalibrating production functions to empirical or simulation data.

A primary question in developing integrated models is the level of detail that is needed to account for the key sources of spatial and agent heterogeneity while also accounting for equilibrium-based economic flows across regions. For example, capturing the market feedbacks between individual decisions and market prices in a way that is consistent with theory is more difficult in models that incorporate multiple sources of spatial and agent heterogeneity (Chen et al. 2011). While standard economic models readily represent market feedbacks through the assumption of aggregate demand and supply functions that equilibrate via market prices, this top–down equilibrium approach is often at odds with agent-based models that use a bottom–up modeling approach (National Resource Council 2014). Chen et al. (2016) develop a model of optimal bidding that accounts for market competition among agents with limited heterogeneity among agents and across space, but much more work is needed to bridge the gap between equilibrium-based economic models and non-equilibrium agent-based models (Irwin 2010; Irwin and Wrenn 2014). In addition, accounting for behavioral heterogeneity, e.g., differences in motives and perceptions among agents, is an important step towards greater realism (Irwin et al. 2016b).

Innovations in data collection will greatly enhance the feasibility of integrated modeling and the development of data-driven models that can better represent real-world systems using real-time, detailed data in understandable and interactive ways (Miller 2015). While this type of empirically-based simulation modeling is not new, the addition of drastically improved data sources greatly increases both the accuracy, and the possible scope of these models (Townsend 2015). Coupling these models of socio-economic systems with other models, such as climate change, allow us to test the sensitivity of future projections to different parameters. Such systems can also be used to run repeated simulations of a wide variety of negative shocks and increase our understanding of how different parts of the system respond and why.

[5]For more information on PRIMA see: http://prima.pnnl.gov/prima-platform-regional-integrated-modeling-and-analysis

Additionally, such data-rich environments will improve reduced form model estimation, which will continue to play a critical role in sustainability and resilience modeling. For example, integrating data on market outcomes, such as housing prices and wages, with disaggregate data on households, firms, and spatial attributes is a critical need for applied welfare analysis (Bieri et al. 2014).

Despite these advantages, there are many challenges associated with this new data. Different rates of adoption or variability in the quality or collection techniques of local datasets may leave geographic holes in certain data sets. While these data sets have the ability to be very specific, lack of integration across data sets may limit the ability to track individuals, business, or properties, for example, and linking the necessary information to answer broader questions. In addition, along with newfound abilities to capture massive volumes of information arise questions of what information is relevant. "Big data" requires significant investment just to process and maintain. New methods, models and techniques are needed for making these data useful. It will take substantial interdisciplinary collaboration to properly manage and operationalize these data.

17.3.3 Welfare-Based Assessment

The ultimate goal of any sustainability or resilience assessment is to evaluate the welfare impacts of a change in some aspect of the economy, e.g., a negative shock in the case of resilience or a change in a policy, technology or other key variable in the case of sustainability. This underscores the importance of establishing a welfare theoretic basis for evaluating sustainability and resilience at regional scales. Weak sustainability derives from the same underlying welfare foundation as cost benefit analysis and therefore, provides a welfare basis for evaluating sustainability (Irwin et al. 2016a).

Evaluating welfare changes at regional scales must account for the mobility of people: households and firms make utility and profit maximizing location decisions, which in spatial equilibrium tends to equalize utility across regions, making the marginal person or firm indifferent to their location. Given a spatial equilibrium, critical differences in social welfare across regions are reflected in housing prices and wages that capitalize differences in a myriad of private and public goods. This includes the positive externalities associated with a well-educated society, for example, and the many so-called provisioning and regulating ecosystem services—for example, natural vegetation that provides open space amenities and absorbs nutrient run-off that improves water quality and human health. Disentangling these many capitalized values is difficult, if not impossible in many cases, but critical for assessing the multiple ecosystem services and trade-offs that are generated by different land uses and investments in public goods. Thus, future research should continue to estimate amenity expenditures (e.g., Albouy 2015; Bieri et al. 2014) and seek to estimate the welfare effects of these many spatially differentiated nonmarket goods and services. The increasing availability of spatially explicit data at highly disaggregated scales is crucial for advancing this research.

A welfare-based approach provides a means of aggregating the numerous detailed interactions among people, firms, policymakers, and other agents that influence the flows of capital stocks over space and time that ultimately determine the welfare of residents within a region. As Arrow et al. (2003), Dasgupta (2009), and others have shown, it is theoretically possible to derive a correspondence between social welfare and a linear index of comprehensive wealth.[6] The analysis starts with the standard Ramsey-Koopmans form of intergenerational social welfare, $V(t) = \int_t^\infty \left[U\left(C\left(\tau\right)\right)e^{-\rho(\tau-t)}\right]d\tau$, where $C(t)$ is the rate of aggregate consumption, U is per period social utility and ρ is the social discount rate. Given that production depends on capital stocks, and assuming that the underlying institutional, economic and social processes that determine the allocation of stocks over time can be represented, a linear index of well-being for a given time period t, $W(t)$ can be derived from $V(t)$. This equates to the aggregate value of all productive capital stocks in the economy evaluated at constant shadow prices: $W(t) = \sum_i p_i(t) S_i(t)$, where $p_i(t)$ is the shadow price of capital stock $S_i(t)$ in time period t. Evaluated at a point in time, this provides a local test of weak sustainability: the economy is sustainable in time t if $\Delta W(t) > 0$. In practice, this capital-theoretic approach relies on a full accounting of capital stocks and shadow prices, something that is impossible. Nonetheless, even a partial accounting of capital stocks with approximate shadow prices will generate a much more holistic assessment of sustainability that goes beyond incomplete measures, such as GDP, or ad hoc measures, such as the Genuine Progress Indicator (GPI) or Human Development Index (HDI).

A welfare-based approach provides a useful endpoint for assessing resilience as well, given that there are many trade-offs inherent in the different types of resilience. Is resilience simply a trait that helps across all aspects, or does boosting resistance necessarily reduce the ability to recover? Are there systems that speed up recovery but limit the ultimate level of growth? Incorporating the value of resilience into a weak sustainability analysis is conceptually possible if resilience is viewed as a capital stock, an approach that has been developed by Mäler (2008) and Mäler and Li (2010). Walker et al. (2010) model the resilience of the Goulburn-Broken Catchment (GBC) in South East Australia to a salinity threshold. They measure resilience of a non-saline state by the physical distance from the water table to the threshold using historical data. They estimate the probability that the system will shift from a non-saline to the saline state as a function of this distance and use market prices to translate this probability into value terms.

Currently, most of the work on sustainability and resilience is not grounded in a welfare-based approach and, instead, focuses on the associations or trade-offs among specific key variables. This kind of analysis will continue to be important, both as a means of improving our understanding of the many trade-offs inherent in regional economies and for informing policies that target specific sustainable

[6]For a more details regarding this approach, see Irwin et al. (2016a).

development goals. For example, analysis that reveals the correlations between human activities, e.g., energy usage, transportation modes, urban density, and carbon emissions (e.g., Glaeser and Kahn 2010) are critical for guiding policies that would seek to reduce these emissions.

17.3.4 Synergies and Trade-Offs Between Sustainability and Resilience

Both sustainability and resilience are normative concepts that offer a means for assessing societal wellbeing. These concepts are often used in tandem, much like we do here, but they are distinct. Understanding the conditions under which investments to make communities more resilient lead to more versus less sustainable consumption is critical for guiding policy. While some actions may enhance both sustainability and resilience, e.g., collecting rainwater can make a drought-prone community both more sustainable and resilient, there often are trade-offs between these two goals and one is not a sufficient condition for the other (Derissen et al. 2011). For example, investing in built infrastructure contributes to regional resilience. However, Australia has invested billions in desalination plants that are currently unused (Radcliffe 2015). What is the trade-off between improved capacity for resilience to drought and inefficient investment that may not be sustainable in the long run? The time period over which sustainability and resilience are assessed also matters. An intense period of spending in a crisis may prevent a regime shift to a less desirable equilibrium—e.g., as many argued was necessary for avoiding an even worse outcome than the 2007–2008 Great Recession. This heightened level of consumption may create a trade-off by improving resilience and reducing sustainability in the short run, but may improve both in the longer run. How should such a large, one-time investment in preventative measures or technology be weighed against uncertain future benefits and over what time period? Answering such questions not only requires prioritization of societal goals by stakeholders, but it also necessitates a deeper understanding of the systems themselves to accurately model the probabilities surrounding such shocks and estimate the opportunity costs and realistic benefits of alternative courses of action.

17.4 Conclusions

Our physical, social, and economic systems are characterized by growing interdependencies whose complexities we are far from measuring or modeling. Nonetheless, the development and application of concepts like sustainability and resilience require that we increase our understanding of the relationships among these systems and their contributions to social welfare. Doing so will require innovative models, vastly more data, and new methods to integrate these models and data across spatial, temporal and organizational scales. This will also require collaborations among scientists from a range of social and natural sciences to advance integrated models of economic, social and environmental processes. Given the long history of collaborative work across disciplines in regional science, and a central focus on the spatial and dynamic aspects of systems, regional science is well-positioned to take on these challenges. Indeed, and as we have highlighted here, regional scientists have already made important contributions. Trends imply that an understanding of regional sustainability and resilience will become even more pressing for the future development of cities and regions and thus we expect that research on these topics will be at the forefront of regional science scholarship for decades to come.

Acknowledgments We gratefully acknowledge support from National Science Foundation Grants CNH 1114934, WSC 1058059, DEB 0410336.

References

Albouy D (2015) What are cities worth? Land rents, local productivity, and the total value of amenities. Rev Econ Stat. doi:10.1162/REST_a_00550
Anderies JM (2015) Managing variance: key policy challenges for the Anthropocene. Proc Natl Acad Sci USA 112(47)
Anderies JM, Folke C, Walker B, Ostrom E (2013) Aligning key concepts for global change policy: robustness, resilience, and sustainability. Ecol Soc 18(2):8–24
Arrow KJ, Dasgupta P, Mäler KG (2003) Evaluating projects and assessing sustainable development in imperfect economies. Environ Resour Econ 26(4):647–685
Arrow KJ, Dasgupta P, Goulder LH, Mumford KJ, Oleson K (2012) Sustainability and the measurement of wealth. Environ Dev Econ 17(3):317–353
Augustine N, Wolman H, Wial H et al (2013) Regional economic capacity, economic shocks and economic resilience. MacArthur Foundation Network on Building Resilient Regions. Working Paper, May
Balland PA, Rigby D, Boschma R (2015) The technological resilience of US cities. Camb J Reg Econ Soc 8(2):167–184
Barbier (2016) Sustainability and development. Ann Rev Resour Econ (in press)
Bieri DS, Kuminoff NV, Pope JC (2014) National expenditures on local amenities. Department of Economics, Arizona State University, Manuscript
Borucke M et al (2013) Accounting for demand and supply of the Biosphere's regenerative capacity: the National Footprint Accounts' underlying methodology and framework. Ecol Indic 24:518–533
Boschma R (2015) Towards an evolutionary perspective on regional resilience. Reg Stud 49:733–751
Bringezu S, Schütz H, Moll S (2003) Rationale for and interpretation of economy-wide materials flow analysis and derived indicators. J Ind Ecol 7(2):43–64
Brundtland GH (1987) Report of the World Commission on environment and development: "our common future". United Nations, New York. www.un-documents.net/our-common-future.pdf. Accessed 20 July 2016
Burke M, Hsiang AM, Miguel E (2015) Global non-linear effect of temperature on economic production. Nature 527(7577):235–239

Carpenter S, Walker B, Anderies JM et al (2001) From metaphor to measurement: resilience of what to what? Ecosystems 4(8):765–781
Cellini R, Torrisi G (2014) Regional resilience In Italy: a very long-run analysis. Reg Stud 48(11):1779–1796
Cerra V, Saxena SC (2008) Growth dynamics: the myth of economic recovery. Am Econ Rev 98(1):439–457
Chen Y, Irwin EG, Jayaprakash C (2009) Dynamic modeling of environmental amenity-driven migration with ecological feedbacks. Ecol Econ 68:2498–2510
Chen Y, Irwin EG, Jayaprakash C (2011) Incorporating spatial complexity into economic models of land markets and land use change. Agric Resour Econ Rev 40(3):321–340
Chen Y, Irwin EG, Jayaprakash C, Park K (2016) An agent based model of land development in an urbanizing region. Comput Environ Urban Syst (forthcoming)
Dasgupta P (2001) Human well-being and the natural environment. Oxford University Press, London
Dasgupta P (2009) The welfare economic theory of green national accounts. Environ Resour Econ 42(1):3–38
Davies S (2011) Regional resilience in the 2008–2010 downturn: comparative evidence from European countries. Camb J Reg Econ Society 4(3):369–382
Davis SJ, Caldeira K (2010) Consumption-based accounting of CO_2 emissions. Proc Natl Acad Sci 107(12):5687–5692
Davoudi S, Porter L (2012) Resilience: a bridging concept or a dead end? Plann Theory Pract 13(2):299–307
Derissen S, Quaas MF, Baumgärtner S (2011) The relationship between resilience and sustainability of ecological-economic systems. Ecol Econ 70(6):1121–1128
Dovern J, Quaas MF, Rickels W (2014) A comprehensive wealth index for cities in Germany. Ecol Indic 41:79–86
Einav L, Levin J (2014) The data revolution and economic analysis. Innov Policy Econ 14:1–24
Evans R, Karecha J (2014) Staying on top: Why is Munich so resilient and successful? Eur Plan Stud 22(6):1259–1279
Fingleton B, Garretsen H, Martin R (2012) Recessionary shocks and regional employment: evidence on the resilience of U.K. regions. J Reg Sci 52(1):109–133
Folke C, Carpenter S, Walker B, Scheffer M, Chapin T, Rockström J (2010) Resilience thinking: integrating resilience, adaptability, and transformability. Ecol Soc 15(4):20–29
Foster KA (2007) A case study approach to understanding regional resilience. University of California, Institute of Urban and Regional Development, Berkeley, CA. http://hdl.handle.net/10419/59413
Glaeser E, Kahn ME (2010) The greenness of cities: carbon dioxide emissions and urban development. J Urban Econ 67(3):404–418
Holling CS (1973) Resilience and stability of ecological systems. Annu Rev Ecol Syst 4:1–23
IPCC (2014) Climate change 2014: synthesis report. In: Core Writing Team, Pachauri RK, Meyer LA (eds) Contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change. IPCC, Geneva, 151 pp
Irwin EG (2010) New directions for urban economic models of land use change: incorporating spatial dynamics and heterogeneity. J Reg Sci 50(1):65–91
Irwin EG, Wrenn D (2014) An assessment of empirical methods for modeling land use. In: Duke JM, Wu J (eds) The Oxford handbook of land economics. Oxford University Press, London
Irwin EG, Gopalakrishnan S, Randall A (2016a) Welfare, wealth and sustainability. Ann Rev Resour Econ 8:19.1–19.22. doi:10.1146/annurev-resource-100815-095351
Irwin EG, Campbell J, Wilson R, Faggian A, Moore R, Irwin N (2016b) Human adaptations in food, energy and water systems. J Environ Stud Sci 6(1):127–139. doi:10.1007/s13412-016-0375-8
Kahn ME (2016) The climate change adaptation literature. Rev Environ Econ Policy. doi:10.1093/reep/rev023

Kraucunas I et al (2015) Investigating the nexus of climate, energy, water, and land at decision-relevant scales: the Platform for Regional Integrated Modeling and Analysis (PRIMA). Clim Chang 129(3–4):573–588
Lagravinese R (2015) Economic crisis and rising gaps North-South: evidence from the Italian egions. Camb J Reg Econ Soc 8:331–342
Mäler KG (2008) Sustainable development and resilience in ecosystems. Environ Resour Econ 39(1):17–24
Mäler KG, Li CZ (2010) Measuring sustainability under regime shift uncertainty: a resilience pricing approach. Environ Dev Econ 15(6):707–719
Manne A, Mendelsohn R, Richels R (1995) MERGE: a model for evaluating regional and global effects of GHG reduction policies. Energy Policy 23(1):17–34
Marcos J, Macaulay S (2008) Organisational resilience: the key to anticipation, adaptation and recovery. Cranfield School of Management, Cranfied University, Manuscript
Martin R (2012) Regional economic resilience, hysteresis and recessionary shocks. J Econ Geogr 12(1):1–32
Martin R, Sunley P (2015) On the notion of regional economic resilience: conceptualization and explanation. J Econ Geogr 15(1):1–42
Miller H (2015) Data-driven geography. GeoJournal 80(4):449–461
Mori K, Christodoulou A (2012) Review of sustainability indices and indicators: towards a new City Sustainability Index (CSI). Environ Impact Assess Rev 32(1):94–106
National Research Council Geographical Sciences Committee (2014) Advancing land change modeling: opportunities and research requirements. National Academies Press, Washington, DC
Ormerod P (2010) Risk, recessions, and the resilience of the capitalist economies. Risk Manage 12(1):83–99
Parson EA, Fisher-Vanden K (1997) Integrated assessment models of global climate change. Annu Rev Energy Environ 22(1):589–628
Pendall R, Cowell M, Foster KA (2010) Resilience and regions: building understanding of the metaphor. Camb J Reg Econ Soc 3(1):71–84
Perrings C (2006) Resilience and sustainable development. Environ Dev Econ 11:417–427
Radcliffe J (2015) Water recycling in Australia—during and after the drought. Environ Sci: Water Res Technol 1(5):554–562
Rennings K, Wiggering H (1997) Steps towards indicators of sustainable development: linking economic and ecological concepts. Ecol Econ 20(1):25–36
Rockström J, Steffen W, Noone K, Persson Å, Chapin FS, Lambin EF, Lenton TM, Scheffer M, Folke C, Schellnhuber H, Nykvist B (2009) A safe operating space for humanity. Nature 461(7263):472–475
Romer R (2001) Advanced macroeconomics. McGraw Hill, New York
Rose A, Liao SY (2005) Modeling regional economic resilience to disasters: a computable general equilibrium analysis of water service disruptions. J Reg Sci 45(1):75–112
Schlenker W, Roberts MJ (2009) Nonlinear temperature effects indicate severe damages to US crop yields under climate change. Proc Natl Acad Sci 106(37):15594–15598
Simmie J, Martin R (2010) The economic resilience of regions: towards an evolutionary approach. Camb J Reg Econ Soc 3:27–43
Sokolov AP, et al (2005) MIT Integrated Global System Model (IGSM) version 2: model description and baseline evaluation. Technical report, MIT Joint Program on the science and policy of global change
Steffen W, Richardson K, Rockström J, Cornell SE, Fetzer I, Bennett EM, Biggs R, Carpenter SR, de Vries W, de Wit CA, Folke C (2015) Planetary boundaries: guiding human development on a changing planet. Science 347(6223):1259855.
Sullivan C (2002) Calculating a water poverty index. World Dev 30(7):1195–1210
Townsend A (2015) Cities of data: examining the new urban science. Public Cult 27:201–212
Treado CD, Giarratani F (2008) Intermediate steel-industry suppliers in the Pittsburgh region: a Cluster-based analysis of regional economic resilience. Econ Dev Q 22(1):63–75

Walker B, Pearson L, Harris M, Maler KG, Li CZ, Biggs R, Baynes T (2010) Incorporating resilience in the assessment of inclusive wealth: an example from South East Australia. Environ Resour Econ 45(2):183–202

Wiedmann TO et al (2015) The material footprint of nations. Proc Natl Acad Sci 112(20):6271–6276

Zivin JG, Neidell M (2014) Temperature and the allocation of time: implications for climate change. J Labor Econ 32(1): 1–26.

Elena G. Irwin is professor, Department of Agricultural, Environmental, and Development Economics, Ohio State University. Her research focuses on the economics of land use change; integrated models of land use and ecosystem services in urban, urbanizing, and agricultural regions; and sustainability measurement and assessment. She is currently a member of the US EPA Advisory Committee for the Sustainable and Healthy Communities Program, an elected board member of the Association of Environmental and Resource Economists and the faculty director the Sustainable and Resilient Economy program at Ohio State University. Dr. Irwin earned her Ph.D. in agricultural and resource economics from the University of Maryland in 1998.

Tim Jaquet is a PhD student in Agricultural, Environmental, and Development Economics (AEDE) at Ohio State University. His primary research interests are regional economics and development, with a focus on regional resilience, housing, and migration. He holds a Bachelor of Arts in Business Administration from Eastern Mennonite University, and has experience in community development in both domestic and international settings. In addition to working on his doctoral dissertation research, he currently teaches a course in environmental economics.

Alessandra Faggian is Professor and Graduate Studies Chair at the Ohio State University, AED Economics Department. Her research interests lie in the field of Regional and Urban Economics, Demography, Labour Economics and Economics of Education. Her publications cover a wide range of topics including migration, human capital, labour markets, creativity and local innovation and growth. She is co-editor of Papers in Regional Science and is on the editorial board of the Journal of Regional Science and Regional Science, Regional Studies. In 2015 she was awarded the Geoffrey Hewings Award by the North American Research Council.

Part IV
Migration, Demography, and Human Capital

Chapter 18
Directions in Migration Research

Peter V. Schaeffer

18.1 Introduction

Migration is a major driver of demographic change. As birth rates declined, migration, domestic and international, became the major force for demographic change in advanced industrialized countries. The importance of immigration is particularly apparent, for example, in Germany where, in 2014, there were 868,373 deaths versus 714,927 live births (net change: −153,446) (Federal Statistical Office of Germany 2015a). In the same year, net migration added 550,483, mostly younger, residents whose fertility then also contributed to the number of births. However, while in 1995, births to immigrant parents accounted for over 13% of all births, by 2010 that number had dropped to approximately 5% (Federal Statistical Office of Germany 2012). This development demonstrates that, unless immigration accelerates over time, it can slow but not permanently halt the decline of an aging population (e.g., Federal Statistical Office of Germany 2015b).

In the developing world, falling birth rates and large internal rural-to-urban migrations are changing the population distribution and shaping the demographic structures in urban and rural regions, respectively. While urbanization in the United States, Western Europe, and Japan is no longer a major contributor to demographic and economic change, it is not yet finished and, on a world scale, has only recently reached 50% of the world's population. In addition to domestic migration, international migration has also changed. Today, more countries are experiencing large outflows or inflows and some countries are immigration destinations, while simultaneously sending many if its citizens to other countries. Although debates about brain drain are not new, international competition for highly skilled workers

P.V. Schaeffer
Division of Resource Economics and Management, Faculty Research Associate, Regional Research Institute, West Virginia University, Morgantown, WV, 26508-6108, USA
e-mail: peter.schaeffer@mail.wvu.edu

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_18

is a relatively new phenomenon that has given rise to renewed research on the brain drain. The H-1B visa program in the United States (Schaeffer and Kahsai 2009), which dates to 1990 (though there were a few small earlier programs), is a policy aimed at highly skilled immigrants. Several other countries, for example Australia, Canada, and Germany, have similar programs.

The immense political changes that occurred during the twentieth century are less regularly citied as a factor fueling migrations on a global scale, but a comparison of a map of the world in 1900 and today shows how dramatically things have changed. In 1900, European powers and the ascendant United States controlled large areas in Asia and Africa, but by the end of the century almost all European and U.S. colonies had gained their independence and Europe had lost its former global political and economic leadership to the United States. In addition, China was reemerging as a major power and other Asian nations had also become important economic players. In the aftermath of the end of European colonialism, the world has experienced internal conflicts, wars over disputed territories between former colonies, as well as the splitting up of colonial entities, most famously that of India, and Pakistan. In Europe, the fall of communism and disappearance of the Iron Curtain eventually saw the peaceful breakup of Czechoslovakia into the Czech Republic and Slovakia, which became effective on January 1, 1993. The former Socialist Federal Republic of Yugoslavia broke up into Croatia, Slovenia, Serbia (which is now further subdivided along ethnic lines into the Autonomous Province of Kosovo and the Autonomous Province of Vojvodina), Bosnia and Herzegovina, and the Former Yugoslav Republic of Macedonia. This sometimes violent breakup generated a large refugee stream, a type of migration that has been almost completely ignored in regional science to date, in spite of the large number of refugees world-wide. The conflict over the use of the name Macedonia between Greece and the Former Yugoslav Republic of Macedonia, though it has not been violent, serves as an illustration of the great potential for conflict within and between countries in the aftermath of large political shifts. In Africa, Eritrea seceded from Ethiopia after a war that began in 1961 and ended in 1991 and South Sudan seceded from the Republic of Sudan in 2011. The conflict over the 2014 annexation of the Crimean Peninsula by Russia is only one of the currently latest, but not the final such event in our rapidly changing world. As we extend regional science research to countries that deal with border disputes or internal conflicts, more attention needs to be paid to the political context of population movements.

Therefore, because of the importance of migration, in this chapter we discuss anticipated directions in migration research in regional science over the next several decades. We are not interested in small-scale migrations; they usually pose few serious challenges. Instead, this chapter focusses on large migration movements, which, because of their size, often disrupt and change the inherited order of things.

To systematically discuss different aspects of and issues in migration, in Sect. 18.2 we first develop a classification system. Utilizing this classification system, Sect. 18.3 identifies important issues that, we believe, will require more attention in the future than they have received in the past. Thus, they represent new directions in migration research in regional science. A summary and conclusion in Sect. 18.4 completes this chapter.

18.2 Classifying Migrations

To organize our discussion of future directions in migration research, we propose a system for classifying different migrations. We find the first two criteria by considering individuals' reasons for migrating. If an individual is currently in an optimal location, subject to constraints and available information about alternative locations, then the only reason for moving to a different location is if (a) something related (internal) to the individual changes, or, (b) conditions external to the individual change. External changes can be economic, political, or environmental, or consist of new information about alternative locations.

Changes internal to a person lead to what we call natural migrations. Many of the reasons for natural migrations are tied to a person's life cycle, such as graduation from high school or college, marriage, job promotions, employment changes, and retirement. We further refine natural migrations by distinguishing between migrants who are constrained by labor market considerations and those who are not. We make this distinction because labor market participants usually face locational constraints related to the nature of their employment. However, not all labor market participants experience such constraints as new communication technologies have rendered some employment, and self-employment, more footloose.

In contrast to natural migrations, structural migrations are caused by a change or changes that are external to individuals, for example, the decline of a once dominant industry or the emergence of a vibrant new industry elsewhere. In addition to natural and structural migration, we distinguish between domestic and international migration because of the different rules and constraints that apply to them. Because we focus particular attention on structural migrations, we provide additional criteria for a finer separation between different structural migrations, but not for natural migrations. The remainder of this chapter will deal almost exclusively with the former.

The additional criteria for classifying structural migration movements are, first, how quickly they emerge (suddenly vs. gradually), second, the time span over which they occur (short versus long) and, third, whether they are voluntary or forced.[1] This scheme results in 16 different types of structural migrations (Table 18.1).

Some ambiguity is inherent in some of the criteria used in this classification scheme. For example, the difference between domestic and international migration is not clear-cut in the case of an EU citizen moving from one EU member country to another, since this migrant has a right to a work and residence permit in the destination country and is even eligible to run for local public office. This is not the

[1] As an aside, note that by definition all natural migrations are voluntary.

Table 18.1 General migration classification, without causes

		Structural migration (external to individual)		
	Natural migration (internal to individual)	Speed of emergence	Duration of movement	Voluntary vs. involuntary
Domestic migration	1. Faces labor market constraint(s)	Gradual	Short	1. Voluntary
				2. Involuntary
			Long	3. Voluntary
				4. Involuntary
	2. No labor market constraints	Sudden	Short	5. Voluntary
				6. Involuntary
			Long	7. Voluntary
				8. Involuntary
International migration	3. Faces labor market constraint(s)	Gradual	Short	9. Voluntary
				10. Involuntary
			Long	11. Voluntary
				12. Involuntary
	4. No labor market constraints	Sudden	Short	13. Voluntary
				14. Involuntary
			Long	15. Voluntary
				16. Involuntary

way we have traditionally thought about "international migration." However, at this time this is still the exception rather than the norm.

The terms "gradual" and "sudden" are also ambiguous and need to be specified empirically to be useful, since there is a continuum from sudden to gradual without a clear dividing line between the two. In reality, it is also not always easy to make a clear distinction between involuntary (forced) and voluntary actions, as illustrated by the struggles to distinguish between "real" and "bogus" refugees (Schaeffer 2010).

The typology in Table 18.1 can be further refined based on the cause of the external change that triggers structural migrations. The main causes we consider are economic, environmental, and political (persecution, civil unrest, civil war, or war between countries). The refinements are presented in Table 18.2 with natural migrations omitted to save space. Of course, it is possible that more than one cause is at work at the same time (e.g., Borger 2008; see also Schleussner et al. 2016, for the relationship between climate-related disaster and conflict). Ignoring this possibility, Tables 18.1 and 18.2 distinguish between 52 different migration types based on their nature (internal-external), speed and duration, voluntary or involuntary, causes (economic, environmental, political), and domestic[2] or international. Note that in Table 18.2, structural migrants who move because of economic changes will usually be, or seek to become, labor market participants. While structural migrants who move only because of political or environmental changes may also wish to work, economic reasons did not trigger their decision to move.

18.3 New Migrations, New Questions

18.3.1 Natural vs. Structural Migrations

There is in regional science and related disciplines a large body of theoretical and empirical research on natural migrations. Natural migrations are moves that occur regularly time after time and they usually change gradually over time, making adjustments easy. They are linked to important life-cycle events such as graduation from high school or college, marriage, discharge from military service, retirement, and career-related mobility. Although most natural migrations occur within a country, because of increasing global economic integration, international natural migrations are on the rise. This is particularly obvious within the European Union, but is also occurring elsewhere.[3] The theoretical foundation for these migrations

[2]Domestic migration is often referred to as internal migration, as well. Because we refer to the causes of natural migrations as being internal to the migrant, in this chapter we will always use domestic migration to avoid the possibility of confusion.

[3]International natural migrations are not a new phenomenon, but their magnitude has increased. The tradition of the journeyman who travels from employment to employment for some time after

Table 18.2 Structural migration classification, with causes

Structural migration									
	Economic causes			Environmental causes			Political causes		
	Speed of emergence	Duration	Voluntary vs. forced	Speed of emergence	Duration	Voluntary vs. forced	Speed of emergence	Duration	Voluntary vs. forced
Domestic	Gradual	Long	Involuntary	Gradual	Long	Involuntary	Gradual	Long	Involuntary
			Voluntary			Voluntary			Voluntary
		Short	Involuntary		Short	Involuntary		Short	Involuntary
			Voluntary			Voluntary			Voluntary
	Sudden	Long	Involuntary	Sudden	Long	Involuntary	Sudden	Long	Involuntary
			Voluntary			Voluntary			Voluntary
		Short	Involuntary		Short	Involuntary		Short	Involuntary
			Voluntary			Voluntary			Voluntary
International	Gradual	Long	Involuntary	Gradual	Long	Involuntary	Gradual	Long	Involuntary
			Voluntary			Voluntary			Voluntary
		Short	Involuntary		Short	Involuntary		Short	Involuntary
			Voluntary			Voluntary			Voluntary
	Sudden	Long	Involuntary	Sudden	Long	Involuntary	Sudden	Long	Involuntary
			Voluntary			Voluntary			Voluntary
		Short	Involuntary		Short	Involuntary		Short	Involuntary
			Voluntary			Voluntary			Voluntary

is well-established and includes the contributions of Sjaastad (1962) on migration as human capital investment, Mincer (1978) on family migration, and Schaeffer (1985) on repeat migration linked to changes in marketable human capital. Natural migrations are compatible with dynamic equilibrium because they are part of the system and not a response to external shocks.

In contrast to natural migrations, structural migrations are responses to significant social, technological, economic, political, or environmental changes. Some changes are sudden and unexpected. This is particularly true of environmental shocks. Examples include the 2004 earthquake in the Indian Ocean that triggered a massive tsunami that hit coastal communities in several Asian nations and resulted in the loss of more than 200,000 lives. Hurricane Katrina in 2005 devastated New Orleans so that tens of thousands of individuals had to be relocated to other parts of the United States, and the city was permanently changed physically and socially. A strong earthquake in 2011 caused a tsunami that destroyed coastal towns on the northeastern shores of Japan's Honshū Island, cost almost 16,000 lives, and triggered a nuclear disaster at Fukushima. In the aftermath, a large region around the nuclear plant had to be evacuated for years to come. But the most famous modern man-made environmental disaster probably is the 1986 meltdown at the Chernobyl nuclear power plant in the Ukraine. This accident contaminated a large area and about 900 km^2 (about 350 square miles) had to be declared an "exclusion zone." As in Fukushima, tens of thousands of inhabitants had to be relocated. Most recently, in May 2016, close to 100,000 residents in Alberta, Canada, had to be evacuated from a large wildfire, and Davenport and Robertson (2016) published an article in the New York Times "Resettling the first American 'climate refugees'." This last example is of an event that has been developing over a long time and did, therefore, not come as a surprise. The other event occurred suddenly and with little or no warning. The examples demonstrate the potential of structural migrations to disrupt social, political, and economic conditions, both in the migrants' region of origin as well as destination region.

At the present, the geographic focus of most regional science research is on member countries of the Organisation for Economic Co-operation and Development (OECD), that is, the world's economically most successful countries. This reflects the membership of the largest regional science organizations,[4] which were first established in North America and Europe. This geographical focus will likely become broader, as has already happened with the formation of the Regional Science Association of the Americas, the growing number of national regional science organizations in Latin America, and the inclusion of the Moroccan Regional

completion of the apprenticeship has its roots in the Middle Ages and survived into the twentieth century among German carpenters.

[4]The Regional Science Association International (RSAI), which was first established in the United States under the leadership of Walter Isard, the North American Regional Science Council (NARSC) and its affiliated member organizations (Canadian, Western, Mid-continent, and Southern Regional Science Associations, respectively), and European Regional Science Association (ERSA) and its 16 affiliated associations.

Science Association in the European Regional Science Association. The inclusion of more developing and emerging economies into national, regional, and international regional science organizations will result in more attention being paid to issues less commonly encountered in OECD countries.

In addition to broadening the geographical foci of regional science, a greater diversity of disciplines in migration research, would contribute to a broader perspective on many research questions. For example, the North American Regional Science Council's web page lists seven interest groups: GeoComputation, Regional Development, Industry Studies, Location Analysis, Rural Development, Transportation, and Urban Economics. Most of these focus on economic or technical issues. Demography, sociology, or psychology are not well represented. The council members of the Regional Science Association International similarly show a predominance of economics and geography.[5]

The absence of demographic issues as a major focus should worry regional scientists, because we are in the midst of an unprecedented demographic shift. While large migrations occurred at many stages in human history, the increase in longevity, decline of fertility rates, and aging of populations have no counterpart. Although much longer in the making—modern urbanization began in the late eighteenth century in England—the degree of global urbanization, which passed the 50% mark only around 2010, is also without historical precedent and has not yet run its course.

In summary, natural migrations are important and will continue to be an integral part of our social, economic, and cultural fabric, but institutions that deal with them, for example in job and housing markets, have existed and been tested for a long time. Most changes associated with natural migration are gradual and, therefore, much easier to anticipate and deal with than the often sudden changes that trigger structural migrations. Thus, the need for new research concerning natural migrations, particularly in the economically most advanced countries, does not appear to be pressing. The one exception is with respect to family migration, that is, the joint migration of more than one decision maker. While the theory of family migration is well established, data limitations have hindered progress in empirical research of this type of migration. New sources of data, such as data from mobile phones, hold some promise that this could change in the future. It is one of many potential changes in empirical regional science research tied to "big data."

While generation after generation repeats natural migrations, structural migrations, though they can last for a long time, are not similarly repetitive. Because of this, they are more challenging to deal with, particularly when they occur

[5]Of 19 council members on whom we could find information on the Regional Science Association International's (RSAI) homepage, we determined that ten were economists, six geographers, two planners, and one an engineer. The quantitative orientation of the council members was even more lopsided. http://www.regionalscience.org/index.php?option=com_k2&view=item&layout=item&id=378&Itemid=590, downloaded on May 6, 2016. The leadership of the North American Regional Science Council is also dominated by economists who hold ten of 19 leadership positions. http://www.narsc.org/newsite/background-history/narsc-council-directory/, downloaded on May 9, 2016.

unexpectedly. Nevertheless, the study of such events may not only yield lessons concerning a specific structural migration, but generalizable lessons that can also be applied to new structural migrations in the future.

18.3.2 Domestic Structural Migrations

18.3.2.1 Environmental Disasters

Most individuals who have been displaced by an environmental disaster will relocate, or be relocated, to another region in the same country, though migrations beyond national borders also occur. Of the 65.3 million displaced persons in 2015, over 40 million were internally displaced (UNHCR 2016). If displaced individuals can return home fairly quickly, then the most significant challenges will be related to logistics. However, if a quick return is infeasible, then the integration of displaced populations in new locations is a significant additional challenge. Countries where the vast majority of citizens share a common language, for example the United States, France, and Germany, do not have to contend with one of the most formidable barriers to the economic and social success of resettled individuals. The migration literature is unanimous in finding that language skills are one of the best indicators of assimilation and integration into the host society, as well as of economic success in a new location, in general.

Among the many aspects of potential interest to regional scientists in large population displacements are impacts on the demographic structure of origin and destination regions, and the long-term effects of a sudden population loss or gain, respectively. Past research taught us that migrants rarely if ever are a representative sample of the origin region's population, but biased in favor of the young and educated. Therefore, a large permanent population loss or gain, particularly if it occurs over a short period, is likely to have important social and economic implications. Even if a disaster indiscriminately pushes out everyone in the affected region, the young and educated tend to have more options to choose from and the impacts will, therefore, still not be uniform across receiving regions. Immigrants also tend to be biased towards the young and not representative of the demographic structure of the destination region, which they will, therefore, change if they arrive in large numbers.

Finally, in countries with regionally distinct populations, a large internal population shift may challenge political stability. There are many ethnically, culturally, and linguistically diverse countries and this is, therefore, a concern shared by many national and subnational governments. It can even turn into a foreign policy challenge, particularly when a subnational region is made up of ethnic groups related to the population of a neighboring country, although we expect that this is more likely in the case of displacement because of an internal conflict than because of an environmental disaster.

18.3.2.2 Oppression, Civil Unrest, and War

Civil unrest or wars add political dimensions and violence not usually present in internal structural migrations that have other causes. In addition, government institutions are often damaged by unrest and war and important actors in government and private organizations may have been drawn into the war effort and their resources and expertise may, therefore, not be available to assist the displaced population. International organizations may strive to fill the gap, but coordinating the efforts of different aid organizations from a variety of countries is no small task and often also has political undertones. The aftermath of the disastrous 2010 Haiti earthquake demonstrated the difficulties that arise when a national government has become incapacitated to assume its expected leadership role, and the war in Syria provides a current example.

Oppression can also trigger migrations. Thus, the Great Migration of African Americans out of the southern U.S. can be interpreted not only as a response to better economic opportunities elsewhere, but also as a political migration. Empirically, it is not always easy to distinguish economic from political motives, and it could be in the interest of migrants who move for political reasons to publicly claim that the move was economically motivated. For example, in the post-World War II era, a Spaniard who was opposed to the Franco regime could leave as a guest worker for France, Germany, or Switzerland, without compromising his ability to return for visits or the fear of repercussions for relatives or close friends left behind (see Schaeffer 2010 for a theory of political migration).

Finally, we call attention to the potential of turning an internal into an international conflict if one of the ethnic groups involved in the struggle has co-ethnics in other countries. This potential is particularly vividly demonstrated by the 2014 annexation of the Crimean Peninsula by Russia, as well as the tensions between Russia and the western alliance that intervened in the Kosovo Conflict in 1998–1999. The possibility of an internal conflict turning into an international one is a complicating factor in designing policies to address internal strife,[6] unrest, or civil war. More detailed statistical information on internally displaced populations can be found in publications of the Internal Displacement Monitoring Centre (2015, 2016).

18.3.2.3 Domestic Structural Migrations in Response to Economic Changes

Even positive economic change can be disruptive. Industrialization created many new opportunities and improved the economic well-being of almost everyone, but it also displaced many traditional businesses, changed industries, and the ways in

[6] Internal strife can also be due to criminal activities. For example, violence related to drug cartels is responsible for internal displacement in Mexico and gang violence is at least partially responsible for the flight of youths to the United States from Central America, particularly Honduras.

which we work. The concentration of industries in urban areas was responsible for the initial urbanization process. In recent decades, the decline of traditional manufacturing industries in the northeastern United States triggered an outmigration to the south and the west that changed the economic and political landscape of the United States. The decline of upper Midwest cities like Detroit has similarly rearranged the ranking of U.S. cities. The earlier (in the United States and Europe) urbanization process shifted populations from rural to urban places, and many formerly rural places became urbanized. Politicians in rural regions understood that urbanization reduced their political power at the national level. In the highly urbanized United States a recent, unsuccessful, court challenge to the allocation of representatives on the basis of population size rather than the number of citizens (CBS 2016), demonstrates the continuing political conflict potential of urbanization. The nature of cities also changed with immigration from rural places, a process that was accelerated by changes in transportation infrastructure. Large migrations over a short time period are particularly disruptive, as they allow no time for gradual adjustment. Because of such concerns, some countries, including the PR China, imposed restrictions on domestic migrations, which therefore became more similar to international migrations, where legal migration is possible only with the permission of the receiving country and, in some cases, the sending country, as well.

Industrialization has also been supported by massive infrastructure projects, for example, dams for irrigation and electricity generation (e.g., Aswan High Dam in Egypt—Nile, Atatürk Dam in Turkey—Euphrates, and the Three Gorges Dam in China—Yangtze). Such projects frequently permanently displace many people and households. While they help create new opportunities, among the displaced are farmers who lost their land and other small self-employed individuals. Although large infrastructure projects usually enhance the efficiency of national economies, we should not ignore equity issues associated with displacement.

In summary, massive demographic shifts between regions can cause conflict. Even if there is no open conflict, behind the scenes, changing political power usually leads to policy changes, many of which have regional impacts. Some policies reinforce trends, while others strain to oppose them, and yet others, to assist the losers in the process of change. All of these are questions of high relevance to regional science research, but the dearth of consistent long-term data has made it difficult to conduct empirical studies and learn systematic lessons on how to deal with ongoing changes from the same, or new disruptive processes.

18.3.3 International Migration, in General

18.3.3.1 Sending Regions

Large population losses over just a few years frequently have negative impacts on the sending regions. The most widely researched of these impacts is the brain drain. Stark et al. (1997) present a theoretical argument that brain drain may actually increase the native human capital beyond what would have occurred without it, but empirical evidence in support of this contention is lacking. The original view that brain drain has negative impacts persists.

The negative impact of the brain drain can be severe, as in sub-Saharan Africa, which has lost so many health professionals to emigration that the integrity of even basic health care services is threatened in some regions. The guest workers of the 1950s–1970s from southern Italy and Spain to northern Europe, later followed by workers from the former Yugoslav Socialist Republic and Turkey, or the Mexicans and Central Americans entering into the United States, were not particularly well educated individuals, but their loss in large numbers over a period of sometimes only two or three decades, left the remaining population with an imbalanced demographic structure, since migrants are not representative of their native regions' populations. In addition, it takes initiative and an ability to cope with risk to migrate to a foreign country. For individuals with relatively little education, to move to a country (or region within a country) with different laws, language, customs, and food is intimidating. Those who move show courage, which could otherwise be used at home in a variety of ways, including entrepreneurial initiatives or civic and political leadership. Emigration from shrinking communities may rob them of future leaders at a time when they need them particularly badly. Since immigrants are also drawn from among the young and may additionally be gender-imbalanced, long-term effects are likely to include negative aspects, but systematic empirical evidence, positive or negative, of entire sending regions is lacking. With large migration movements spreading to ever more countries, these are issues in need of attention.

Because of the dominance of economists and economic geographers in our discipline, and particularly in leadership positions, the regional science literature pays more attention to economic than, for example, social impacts. Sociologists are more likely to address demographic issues and study populations "left behind," and psychologists to psychological costs of tied movers (e.g., McCollum 1990; in the international context, see Groysberg et al. 2011 for a more recent contribution). That is why a concerted effort to invite and include a more diverse mixture of scholars is important for the future viability and innovative capacity of regional science migration research.

There also seem to be cycles in large migration movements. First, not all migrants stay. In the great immigration to the United States that began in the nineteenth century and lasted until the start of World War I, about one-third of immigrants eventually returned home. Similarly, of the more than 500,000 Italian immigrants to Switzerland in the post-war period, about a third left to go "home." Most recently, Gonzales-Barrera (2015) reports that net immigration of Mexicans to the United States has turned negative, with some 140,000 fewer immigrants than return migrants over the period from 2009 to 2014. The major reason cited by return migrants was family reunification.

18.3.3.2 Receiving Regions

Large inflows of foreign nationals can create serious challenges, even when there is a close cultural affinity between natives and newcomers. This was demonstrated by the inflow of highly skilled German professionals into Switzerland in the first decade of this century. Their addition to the labor force was a gain for the economy, but in tight housing markets with vacancy rates below 1%, particularly in the Zurich metropolitan region, the newcomers, whose earnings gave them significant purchasing power, "crowded out" established renters from many housing markets. The resulting discontent fed xenophobic sentiments that were also expressed at the ballot box. In general, a large population inflow into a region—for whatever reason—increases competition for scarce resources that will last until regional markets and public institutions have caught up with the added demand. In the interim, crowding out effects could turn natives against the newcomers, even if initial attitudes were welcoming. A recent example of such a reversal was witnessed in 2015–2016 in Germany, where the initial welcome extended to Syrian refugees has cooled and authorities had to worry about sporadic violence and the firebombing of asylum centers. The presence of large numbers of immigrants in Zambia has also triggered violent reactions (BBC 2016). South Africa similarly has experienced sometimes violent opposition to the arrival of large numbers of often illegal immigrants. In Europe, the former head of Britain secret service MI6, Richard Dearlove, recently warned of a popular uprising in the European Union in response to the Syrian refugee immigration crisis.

An additional risk in receiving regions relates not to immigrants, but to their offspring, frequently referred to as the second generation. Children of immigrants who obtain their education in the schools of their parents' host country, and who speak the country's language without accent, have different expectations than their parents. Many of them consider the country of their birth their home country[7] and react with disappointment and even anger to any real or perceived rejection (e.g., Portes and Stepick 1994).[8] Possible consequences of the failure to integrate the "second generation" were demonstrated in the 2005 riots in France that followed the accidental deaths of two youths who had been hiding from the police. The role played by Belgium citizens in the terror attacks in Paris in 2015 and Brussels in 2016 are even more urgent reminders of the importance of learning more about the integration and assimilation process (see Schaeffer 2006, for an outline of a theory of assimilation).

---

[7]In those countries where citizenship is awarded to all children born there, they are in fact citizens with all rights and obligations. But even in countries where this is not the case, such as in Switzerland or Israel, the second generation may feel more at home in the country of their birth than the one of their citizenship.

[8]An indication of differences of identification with the country of birth vs. the parents' country is the case of two soccer stars, born in Switzerland to immigrant parents from Albania, one who is playing for the Albanian national team and the other for the Swiss national team.

18.3.4 Political International Migrations: Refugees

Political migrations pose particular problems. International migrants are people who seek to play a role in the host economy, even those with relatively little education. By contrast, refugees are a mixed group that often includes individuals who are not—or not immediately—ready to become part of the host economy. They are also more likely to include dependents—young and old—than recent labor immigrants. In fact, a relatively new phenomenon has been the immigration to the United States of Honduran minors fleeing violence (Robles and Shear 2014) and of child refugees from Syria to Europe. These young refugees, in particular, are dependent on the host government's support for a period that can last relatively long. Unlike labor migrants who can often rely on co-nationals and co-ethnics who had arrived earlier, refugees rarely have access to similarly well-established networks to ease the challenge of making a new beginning in a foreign land. At the same time, refugee children may make demands on school systems that are expensive to meet. For example, children who witnessed the violence of war may initially be too traumatized to focus on conventional learning but instead need counseling to cope with their experiences. Thus, refugees to Switzerland from the Kosovo conflict at the end of the twentieth century included many children who had to adapt to a new environment and language, make up for weeks, months or even longer periods of schooling missed because of the war, and deal with traumatic experiences, to boot. Serving such children is an expensive long-term task.

There is also the question of the spatial distribution of refugees, if their numbers are very large. In terms of the ability to accommodate them, distributing refugees among a large number of communities has many advantages, but it robs them of the ability to establish dense ethnic networks and create markets big enough for ethnic businesses to flourish. Such businesses are an opportunity for entrepreneurial members of the refugee community to establish an existence in the host country and a scarcity of such opportunities might slow down their overall economic integration. This has to be balanced, however, against the risk of establishing large inward-looking refugee communities that could slow down the social integration into the host community. The downside of establishing large refugee camps are demonstrated by the experience with Palestinian refugee camps in Lebanon, the first two of which date to 1948 and still exist today. Although there are complex reasons for this state of affairs in this particular instance, but clearly, the human and political costs, domestic and international, are high.

At present, the vivid images on television and in the press of Syrian refugees stranded in Greece and the Balkans have made us only too aware of the challenges of political migrations. An initially welcoming mood in Germany has turned as more and more refugees arrived. Unfortunately, refugee crises have been a recurring event. The expectation at the time when the United Nations High Commission for Refugees (UNHCR) was founded in 1950 to deal with European refugees after World War II was that the need for its services would eventually decline significantly; this did not materialize. The UNHCR reports that the number of refugees has been increasing from 10,489,000 people in 2012 to 14,385,300 in 2014 (+37%). While the number of refugees increased, the number of those who were able to return home fell to 126,800, the lowest since 1983 (*Statistical Yearbook,* UNHCR 2015a: Table 5.2, p. 66).[9] This recent increase in refugee numbers comes after a period of decline, an indication of the present unpredictability of refugee numbers very far into the future. For statistics on refugees and asylum seekers in the United States see Mossaad (2016). Although Syrian refugees and asylum seekers in Europe are currently in the headlines, Africa has the largest number of refugees (UNHCR 2015b).

The political stress that can result from large refugee streams is also demonstrated by the Syrian refugee crisis, which is threatening the cohesion of the E.U. (e.g., Spiegel Online 2016). Beyond the acceptance or non-acceptance of refugees, refugees themselves can become a political problem, even a foreign policy problem. For example, political and military activities among Palestinian refugees in Lebanon were at the heart of the 1982 Israel-Lebanon war. The large number of Kurdish immigrants in Germany similarly has brought conflicts in Turkey, Iraq, and Syria to German cities, for example, in large demonstrations. Some activities by refugees have created diplomatic problems for the host country.

18.3.5 Environmental International Migrations

The last two decades have seen a significant growth of research on resilience (e.g. World Conference on Disaster Reduction 2005 for the Hyogo Framework for Action). Most of this research deals with the affected communities and regions. However, the examples of Chernobyl and Fukushima show that in the case of dealing with displaced populations, including international refugees, we also need to identify likely receiving regions and examine their resilience to assure that they are at least sufficiently prepared to deal with the immediate aftermath of a large inflow of people (e.g., Renaud and Perez 2010).

Environmental international migrations are a relatively new phenomenon and laws and rules, and even the definition of what constitutes an environmental refugee, are not yet settled, though the United Nations has published guidelines (McAdam 2011). A summary of research and definitions until about 2000 can be found in Black (2001). The lack of agreement makes the politics of environmental refugee migrations particularly difficult. This is an emerging topic of research and is relevant in many regions. Although there is not much research on this topic in regional science, it is not new. The 1930s Dust Bowl, which triggered the emigration of many

[9]Not all displaced individuals are recognized as refugees. In addition, they also include internally displaced persons who are not "technically" regarded as refugees. This is why the number of displaced people is much larger than the number of refugees. The number of displaced persons increased from 59.5 million in 2014 to 65.3 million in 2015; that is one of every 113 people currently alive (UNHCR 2016). About a third are officially recognized refugees.

from Oklahoma to California, has become part of the collective national memory of the United States, particularly through John Steinbeck's (1939) novel *The Grapes of Wrath*.

International Migration, the official journal of the United Nations' International Organization for Migration, in 2011 dedicated a special issue to environmentally induced migrations. One of the pioneers of research on this topic, Fabrice Renaud, coauthored a noteworthy article that lays out a framework and agenda for research (Renaud et al. 2011). More research is urgently needed if some of the most pessimistic predictions concerning the number of environmental refugees come true. A recent special report by a large Swiss newspaper writes of about 200 million environmental refuges, though the time frame over which this is expected to occur is not mentioned (Jeska 2016). Borger (2008) provides complementary information, including a map that shows where environmentally displaced people might come from, but stresses the connection between climate change, conflict, and forced displacement. In other words, there is an overlap of conflict and environmentally induced displacement.

18.4 Conclusions

In this chapter we argue for prioritizing research in new topics in migration research, particularly those related to structural migrations. Because large scale economic, political, environmental and demographic changes trigger migration responses that are unique, we need to find currently missing information and develop theories to address these changes. Many changes are already under way with some of them continuing for some time, while others have not yet fully started. An example of the latter are environmentally caused displacements, but their growing importance is very likely. Although natural migrations also continue to be an important social and demographic phenomenon, our understanding of them is advanced and sufficient for policy design and recommendations, which is not the case for many of the structural migrations.

The successful research into ongoing and emerging large scale structural migrations, both domestic and international, requires a broader disciplinary range than has been the case over the last few decades, when economics and geography have been dominant in shaping the development of regional science methods and foci.

In addition, economic growth has transformed the world and more scholars from developing and emerging economies are joining the regional science community. They will introduce new questions that reflect the needs and interests of their countries. We hope that regional science will welcome these newcomers and see in their joining our efforts an opportunity to further advance our discipline not only along established lines of research, but also by expanding its scope. We believe that failure to do so could stifle the future growth and development of regional science and over time diminish its relevance.

References

BBC (2016) Zambia xenophobic riots: two burned alive in lusaka, April 20. http://www.bbc.com/news/world-africa-36092917. Downloaded 16 May 2016

Black R (2001) Environmental refugees: myth or reality? New issues in refugee research. Working Paper No. 34. http://www.unhcr.org/3ae6a0d00.html. Downloaded 16 May 2016

Borger J (2008) Conflicts fuelled by climate change causing new refugee crisis, warns UN. The Guardian, June 16. http://www.theguardian.com/environment/2008/jun/17/climatechange.food. Downloaded 16 May 2016

CBS (2016) Supreme court rejects challenge to "One person one vote" rule. CBS News, April 4. http://www.cbsnews.com/news/supreme-court-rejects-challenge-to-one-person-one-vote-rule/. Downloaded 9 May 2016

Davenport C, Robertson C (2016). Resettling the first American 'Climate refugees'. New York Times, May 2. http://www.nytimes.com/2016/05/03/us/resettling-the-first-american-climate-refugees.html Downloaded 12 May 2016

Federal Statistical Office of Germany (2012) Geburten in Deutschland. Ausgabe 2012. Wiesbaden, Germany, available in German only. https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Bevoelkerungsbewegung/BroschuereGeburtenDeutschland0120007129004.pdf?__blob=publicationFile. Downloaded 15 Dec 2015

Federal Statistical Office of Germany (2015a) Population. https://www.destatis.de/EN/FactsFigures/SocietyState/Population/Population.html;jsessionid=1D5BBC7FEFFD9C9716341494D4F047E6.cae2. Downloaded 15 Dec 2015

Federal Statistical Office of Germany (2015b) Germany's population by 2060. Wiesbaden, Germany. https://www.destatis.de/EN/Publications/Specialized/Population/GermanyPopulation2060_5124206159004.pdf?__blob=publicationFile. Downloaded 15 Dec 2015

Gonzales-Barrera B (2015) More Mexicans leaving than coming to the United States. Pew Research Center. http://www.pewhispanic.org/2015/11/19/more-mexicans-leaving-than-coming-to-the-u-s/. Downloaded 16 May 2016

Groysberg B, Nohria N, Herman K (2011) The expat dilemma. Harv Bus Rev 89(11):150–153

Internal Displacement Monitoring Centre (2015) Global overview 2105: people displaced by internal conflict and violence. Geneva, Switzerland. Available via http://www.internal-displacement.org

Internal Displacement Monitoring Centre (2016) GRID 2016: global report on internal displacement. Geneva, Switzerland. http://www.internal-displacement.org/globalreport2016/. Accessed 23 May 2016

Jeska A (2016) Natur gegen Mensch. Special online report by Tages-Anzeiger. http://webspecial.tagesanzeiger.ch/longform/klimaflucht/fluechtlingsstatus/. Viewed 16 May 2016

McAdam J (2011) Climate change displacement and international law: complementary protection standards, Legal and protection policy research series. United Nations High Commissioner for Refugees, Geneva

McCollum AT (1990) The trauma of moving: psychological issues for women, Sage library of social research 182. Sage, Newbury Park, CA

Mincer J (1978) Family migration decisions. J Polit Econ 86(5):749–773

Mossaad N (2016) Refugees and asylees: 2014. Annual flow report, April. U.S. Department of Homeland Security, Office of Immigration Statistics

Portes A, Stepick A (1994) City on the edge: the transformation of miami. University of California Press, Berkeley, CA

Renaud F, Perez R (2010) Climate change vulnerability and adaptation assessments. Sustain Sci 5:155–157

Renaud F, Dun O, Warner K, Bogardi J (2011) A decision framework for environmentally induced migration. Int Migr 49(S1), e30–e49. http://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2010.00634.x/epdf. Downloaded 16 May 2016

Robles F, Shear MD (2014) U.S. Considering refugee status for hondurans. New York Times, July 24. http://www.nytimes.com/2014/07/25/world/americas/administration-weighs-plan-to-move-processing-of-youths-seeking-entry-to-honduras-.html. Downloaded 12 May 2016
Schaeffer P (1985) Human capital accumulation and job mobility. J Reg Sci 25(1):103–114
Schaeffer P (2006) Outline of an economic theory of assimilation. J Reg Anal Policy 36(2):153–160
Schaeffer P (2010) Refugees: on the economics of political migration. Int Mig 48(1):1–22
Schaeffer P, Kahsai MS (2009) Evaluating the H-1B visa program: theoretical issues. China-USA Bus Rev 8(11):44–54
Schleussner C-F, Donges JF, Donner RV, Schellnhuber HJ (2016) Armed conflict risks enhanced by climate-related disasters in ethnically fractionalized countries. Proc Nat Acad Sci U S A. http://www.pnas.org/content/early/2016/07/20/1601611113. Downloaded 28 July 2016
Sjaastad LA (1962) The Costs and Returns of Human Migration. J Polit Econ 70(5 Supplement):80–93
Spiegel Online (2016) Refugee crises pushes Europe to the brink, March 4. http://www.spiegel.de/international/europe/border-closures-spell-refugee-back-up-in-greece-a-1080643-druck.html. Downloaded 12 May 2016
Stark O, Helmenstein C, Prskawetz A (1997) A brain gain with a brain drain. Econ Lett 55(2):227–234
Steinbeck J (1939) The grapes of wrath. Viking, New York
UNHCR (2015a). UNHCR statistical yearbook 2014. United Nations High Commissioner of Refugees, Geneva. http://www.unhcr.org/statistics/country/566584fc9/unhcr-statistical-yearbook-2014-14th-edition.html. Downloaded 23 May 2016
UNHCR (2015b) UNHCR midyear trends 2015. United Nations High Commissioner of Refugees, Geneva. http://www.unhcr.org/statistics/mid2015stats.zip. Downloaded 23 May 2016
UNHCR (2016) Global trends: forced displacement in 2015. United Nations High Commissioner of Refugees, Geneva. https://s3.amazonaws.com/unhcrsharedmedia/2016/2016-06-20-global-trends/2016-06-14-Global-Trends-2015.pdf. Downloaded 21 June 2016
World Conference on Disaster Reduction (2005) Hyogo framework for action 2005–2015: building the resilience of nations and communities to disasters. http://www.unisdr.org/2005/wcdr/intergover/official-doc/L-docs/Hyogo-framework-for-action-english.pdf. Downloaded 16 May 2016

Peter V. Schaeffer is professor, Division of Resource Economics and Management, West Virginia University (WVU). His primary research interests are regional economic policy; international labor migration; job mobility; natural resource management; and historic preservation. He is a faculty research associate in WVU's Regional Research Institute and adjunct professor in the Department of Economics. Previous faculty positions were at the Universities of Colorado–Denver, Illinois at Urbana–Champaign, and one year as visiting professor at the Swiss Federal Institute of Technology–Zurich. Dr. Schaeffer earned the Ph.D. in economics from the University of Southern California in 1981.

Chapter 19
Human Capital Research in an Era of Big Data: Linking People with Firms, Cities and Regions

K. Bruce Newbold and W. Mark Brown

19.1 Introduction

For regions that are seeking to attract and retain skilled workers, human capital, or the stock of knowledge inclusive of creative skills embodied within a population that is capable of producing economic value, has been identified as vital to economic growth and opportunity for communities. Human capital is an important source of growth for cities and a driver of wage levels, with a positive association between initial levels of human capital and long-run employment growth (Behrens et al. 2014; Combes et al. 2012; Glaeser et al. 1995; Glaeser 1999; Glaeser and Maré 2001; Glaeser and Saiz 2003; Shapiro 2005). Partially reflecting selective migration flows, differences in human capital are also a major determinant of wage differences across cities (Beckstead et al. 2009; Combes et al. 2008; De la Roca and Puga 2015; Rosenthal and Strange 2008; Yankow 2006) and have been associated with both private and public benefits to society, including increased productivity (Moretti 2004).

Human capital tends to concentrate in large cities, reflecting a mix of immigration (with immigrants bringing high levels of human capital), internal migration, and greater investment in human capital owing to greater educational opportunities, greater employment opportunities, and greater returns to that investment in education through higher wages (Brown et al. 2010; Brown and Scott 2012; De la Roca and Puga 2015). Conversely, levels of human capital are typically less in smaller urban or rural areas, reflecting fewer opportunities to build and grow human capital,

K.B. Newbold (✉)
School of Geography & Earth Sciences, McMaster University, Hamilton, ON, L8S 4K1, Canada
e-mail: newbold@mcmaster.ca

W.M. Brown
Economic Analysis Division Statistics Canada, Ottawa, ON, K1A 0T6, Canada
e-mail: mark.brown@canada.ca

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_19

as well as fewer opportunities within the labor market to utilize it (Combes et al. 2012; De la Roca 2015).

While there is a strong consensus within the literature with respect to the impact and benefits of human capital on regional economic growth, the literature continues to explore why human capital is unevenly distributed and what prompts individuals to invest (or continue to invest in) their own human capital, amongst other questions. Following the evaluation of the differential growth of human capital across the urban hierarchy, Brown et al. (2010: 1584) concluded that

> The implication is that human capital may be endogenous to cities and, to the extent that human capital drives economic growth and development, the role of cities as drivers of economic growth looms ever larger.

If this is indeed the case—that human capital is endogenous to cities—then it perhaps means that we have reached the limit of what current methodologies and data can uncover. Moreover, issues of endogeneity are seen elsewhere in the regional science and human capital literature. Newbold and Brown (2015) noted the potential for endogeneity between city size and university attendance, including parental education levels and university access as determinants of university participation amongst youth. Likewise, in exploring the location decisions of workers, Breinlich et al. (2013: 4–5) write:

> Gennaioli et al. (2013) emphasize that regional output per capita is strongly correlated with average human capital. In a regression of output per capita on average years of education and country dummies, using 1500 regions across 105 countries, they find that education explains 38% of the variation in output per capita within countries. This is striking, but nobody would propose that regions within a country are each endowed with a fixed stock of skilled workers. Investigating the association requires models in which regional factor supplies are endogenous to the location decisions of workers and firms.

This therefore raises the question of whether there are other opportunities for continuing to explore the importance of human capital through other means. While new and more specialized econometric methods will likely aid our quest, the use of so-called big data may also provide data-driven opportunities to engage research, with big data offering new data sources, greater amounts of data in real time, and greater granularity and scale of data, enabling measurement of aspects of labor market, job matching, skills acquisition and training that have not been possible before.

If we are to utilize big data to explore regional science questions, and particularly human capital questions, what is that we might want to understand? A central question may be whether we can determine or better identify and unpack the endogeneity inherent between human capital and cities. However, to answer this, we likely need to consider two inter-related questions: (1) What prompts an individual to relocate or to invest in human capital to provide a more nuanced, micro perspective of the process?, and (2) are there other non-economic effects (Finnie 2012) that influence decisions related to human capital acquisition? This chapter, therefore, explores the opportunities and pitfalls of using big data to answer questions related to human capital formation.

19.2 Introducing Big Data to Regional Science: Potential, Sources, Problems and Opportunities

19.2.1 Big Data: Big Potential?

Big data is increasingly used in human resources and labor force analytics. In particular, private industry has seen an explosive growth of big data in almost every industry and business area (McKinsey Global Institute 2011). While the use and analysis of big data is still more common within industry as compared to academia (Jin et al. 2015), academics are also turning to big data, with an increasing number of examples of its use to explore questions related to urban planning, education, wages, and health care. The rise of the 'smart city', for example, and an increasing number of data streams from diverse areas including merchandising and sales, GPS and transportation networks and related mobile technologies means that there are multiple research applications, some of which could be defined as regional science. Similarly, the use of data embedded in social media content, including Twitter and Facebook, could enable understanding the role of social connections in education, career, or employment choices and searches.

To date, much of the existing work on human capital relies on traditional data surveys such as the census and labor force surveys. Each has its benefits and costs, but overall, they tend to be expensive, sometimes infrequently collected (as in the case of the census, for example), limited in terms of the scope of questions asked, and are often cross-sectional only. Consequently, it is frequently difficult to understand the temporal and causal processes and details of human capital generation. Further, the census and related files have been seen by critics as being intrusive and out-of-date (Shearmur 2010), notions that led, for instance, to the cancellation of the US long-form census in 2000. Likewise, the UK government suggested that their census could be dropped, with a mix of administrative and survey data used in its place. Such examples of governments scaling back large scale data collection have led to the consideration of other data sources. Indeed, governments in the UK and elsewhere have suggested that big data is capable of describing a population. However, given the limitation of big data noted below, this remains an open question.

The turn toward administrative files by governments could be considered a form of big data. While there is no universal definition and acceptance of what big data is, it is generally defined in terms of the "4 V's": *volume*, *velocity*, *value*, and *variety*. A fifth 'V', *veracity*, references uncertainties in the data (Jin et al. 2015), issues that will be explored later in this chapter. With billions of data points captured on millions of users each day through various sources, inclusive of some form of location measures, big data could potentially augment sparse, traditional data sources, helping to answer old questions and raise new ones. Big data may also enable greater interactive or dynamic analysis, responding to events and providing real-time analytical capabilities.

The emergence of big data offers a number of advantages for researchers, including altering the scale and granularity of research questions and analysis, such as individual details versus much coarser-grained scales such as census tracts typically found in the census or related data sets. Additionally, big data offers improved precision in both space and time (i.e., real-time measures and accurate geolocation), the potential to select constructed samples that meet special requirements, and increased robustness given the ability to look at different subsamples by geography, demography, etc. (Horton and Tambe 2015). Big data also allows exploration of outcomes from natural experiments, whereby the researcher is able to manipulate the data to test responses (Einav and Levin 2014; Horton and Tambe 2015). Examples include analyses of future earnings amongst students and their random assignments to classrooms (Chetty et al. 2011) and estimating the impact of Medicaid on health and financial measures (Taubman et al. 2014).

19.2.2 Big Data: Big Sources?

Despite the previous definition of what constitutes big data, it potentially includes administrative data that are increasingly available (Einav and Levin 2014; McKinsey Global Institute 2011), given the sheer size (volume) of data and the ability to link across different data files (variety). Governments routinely collect detailed data on individuals and corporations, including health, education, taxation, and spending. Such data files typically have high data quality and a long-term panel structure. In the US, sources of administrative data include Social Security, the Internal Revenue Service, and Medicare and Medicaid, all of which have a panel structure. Statistics Canada's new Canadian Employer-Employee Dynamics Database (CEEDD) links tax files, immigration records, and Record of Employment information to create a longitudinal (2001–2010) employer–employee dataset covering all workers, including the location of workers and their employers. Similarly, data files such as Statistics Canada's Longitudinal Administrative Database (LAD), which covers all Canadian tax filers and their yearly tax return, could be considered as 'big' data. In the UK, the Office for National Statistics (ONS) uses administrative data, including National Health Service Central Register data and Patient Registers to estimate yearly migration flows (Thomas et al. 2012). Administrative data are already widely used, but additional opportunities remain, particularly when databases are linked, thereby increasing the power of the data, a point reinforced by Einav and Levin (2014).

While administrative data files may be considered as big data, the term is more commonly associated with data generated through the internet and other sensor platforms. Potential sources include:

- '*Passive*' *collectors of data*: Passive collectors of data include cell phones and GPS transponders that collect data including mobility and other user information with high spatial and temporal resolution. The large number of platforms [estimated at over 7 billion worldwide (Rango 2015)] suggests the potential depth and breadth of this data.

- *Private sector:* The private sector routinely collects and aggregates data through such means as loyalty programs, with information including customer purchasing behaviors, basic demographic information of the consumer, and related financial transactions. Similarly, banks and credit companies collect detailed data on household and business financial interactions.
- *Social media:* Social media websites such as Facebook, Twitter and Instagram offer a wealth of social network information. Social media data offer potential insight(s) into the role of social networks [data that are often hard to gather or quantify (Einav and Levin 2014)], as well as employment by recording labor market searches and employment history as individuals post or tweet changes to employment status. Additionally, user activity is geolocated, allowing estimation of mobility patterns.
- *Labor intermediaries*: Labor intermediaries include job websites, collaborative platforms, and on-line education platforms, all of which offer the potential and opportunity to explore concepts related to human capital development. On-line education platforms such as Coursera or Udacity collect information on how individuals acquire new skills. Such data are potentially useful for learning about who selects into education, along with how and what they learn—that is, what skills are being upgraded. Job websites such as Monster, CareerBuilder, and LinkedIn capture information including employment networks, employment histories, employment opportunities, and the matching between workers and employers, while providing details about dates and timing of job entry/exit, employment search activities, type of job, skills, education, and location (Horton and Tambe 2015). Such information potentially highlights the role of networks in job searches, including the relative role proximate and distal networks offer.
- *Collaborative platforms*: Platforms such as GitHub, Slack, and Sourceforge often contain human capital or career information about individuals, enabling analysis of how contributions to these sites impact short-run labor outcomes or career trajectories of individuals.

Several examples demonstrate the potential use of big data in considering regional science questions. Einav and Levin (2014) highlight a number of examples of the use of administrative data, including analyses of economic mobility and health care spending, amongst other examples. Other examples include the use of LinkedIn to examine labor mobility (Ge et al. 2014), Sobolevsky et al.'s (2014) use of bank card transactions to identify mobility patterns in Spain, Wu et al.'s (2016) use of Weibo (Chinese Facebook) to look at mobility patterns, and Marinescu and Rathelot's (2013) use of CareerBuilder data to explore the spatial mismatch between the residential location of job seekers and employment opportunities. Marinescu and Wolthoff (2015) also used CareerBuilder data to explore how language appearing in job advertisements affected the labor matching process, finding that the language used in the job title could explain as much as 80% of the variation in the education and experience of applicants. Search engines, including Google, have also been used for short-run forecasts of unemployment, consumer confidence, and retail sales by aggregating search terms (Choi and Varian 2012; Varian and Scott 2014).

Social media websites have also been mined for data, with Twitter used to measure labor market flows (Antenucci et al. 2014), while various new tools enable researchers to convert data into information (see http://mashable.com/2009/05/03/twitter-research-tools/#ak5HSOUWE8qC for examples of such tools) and an on-line search revealed multiple requests for Twitter data. Cell phones have also emerged as a generator of big data. While most existing examples have relied on small sub-samples of phone records, Rango (2015) suggested that cell phone call logs could be used to monitor mobility, particularly international mobility and refugee movements where other administrative data are limited. Toole (2015) has used cell phone logs to examine a series of travel and mobility related phenomena, including travel demand, mobility and social behavior in cities, with results revealing (amongst other findings) declines in social behaviour and mobility following a layoff event. Papinski and Scott (2013) and Papinski et al. (2009) used GPS data from smartphones to explore commute and trip behaviour, with the GPS data augmented through more traditional survey data to provide the context of commute trips.

19.2.3 Big Data: Big Problems?

Like all data, users of big data must recognize its problems and shortcomings, an issue that was highlighted by Finance Canada's use of labor intermediary data in Budget 2014 to estimate job vacancies in Canada. The job vacancy rate was calculated as the number of online job postings divided by labour demand, which is total job postings plus occupied positions. Job vacancy rates were based on job postings 'scraped' from Kijiji and related job websites. It was later realized, however, that Kijiji posted the same job vacancy multiple times, potentially inflating the vacancy rate (see Curry 2014). Since then, Statistics Canada's Job Vacancy and Wage Survey program was implemented to determine the level of unoccupied positions.

To date, big data remains a *possible* source of data for regional scientists, but it has not yet realized its full potential. Instead, it is constrained by multiple issues. For example, access to administrative data is often limited owing to confidentiality and privacy concerns. Data drawn from social media sources are dependent upon personal reporting, meaning that updates may be infrequent, incomplete, or incorrect. Likewise, access to data generated through other means, including privately collected data, the internet, mobile computing devices, or other sensors may be corporate property, with dissemination to users limited.

Computationally, big data implies the need to process large quantities of data. Given system complexity, computational complexity, and data complexity owing to its volume and/or velocity (Jin et al. 2015), computing and software environments, including high speed computing capable of working with big data is an issue

(Lee and Kang 2015). This may include knowledge and abilities to clean the data, including imputing missing values, removing extraneous information, and organizing large-scale, unstructured data (Einav and Levin 2014; see also Thomas et al. 2012 for an example of the cleaning of a large commercial data set in the UK context). Consequently, given the sheer size and volume of data, data processing and analysis may require users to learn new tools, programming techniques, and programing languages.

Beyond analytical and computational issues, a series of other issues are associated with big data. Despite the fact that proponents of big data have positioned it as a complete and exhaustive data source, which may therefore more closely resemble the universe as opposed to a sample, it is still subject to sample bias owing to such things as the regulatory environment and the technology and platform used, affecting its veracity (Kitchen 2014). In addition, big data is affected by the consistency and continuity of what is (and is not) collected over time, an issue that may be particularly relevant given privacy concerns. The lack of a clearly defined sampling frame may be the largest issue faced by users of big data, given that the sampling frames of most existing big data sources often vary across platforms and source(s). The lack of a defined sample frame is in stark contrast to the defined sampling frames associated with the census or other similar publically available data files. For example, the reasons that users participate in various online tools such as job websites have implications for who appears in the sample, why and when they choose to participate, and the accuracy of the information that is provided by the user.

In other cases where data are collected passively, there is likely to be little additional contextual information associated with it. Consider GPS tracking: regardless of whether we are tracking an individual and their commute behaviors or a transport truck on its delivery route, there is little other contextual information associated with the sample. Was the stop for pickup or delivery? What was delivered or picked up and what was its value? Was the time lag on the stop indicative of the actual time to transfer goods? Who was in the vehicle? Only an additional survey would elucidate these questions by providing knowledge of the purpose of the stop and goods moved. Likewise, big data may not incorporate many of the traditional baseline and contextual socio-economic or socio-demographic data, including measures of sex, education, family type, marital status, etc., that regional scientists typically include in analyses of human behavior. Big data may only be able to get around this shortfall if the data are longitudinal such that effects are fixed over time, a complementary survey is compiled (i.e., Papinski and Scott 2013; Papinski et al. 2009), and/or data sources are linked (such as the CEEDD file). In the latter cases, linking to other survey tools or data files provides contextual effects including socio-demographic and socio-economic indicators that regional scientists typically rely upon, and therefore maximizes the value of big data sources. Doing so also ensures confidence that the data reflect real characteristics and behaviour.

Big data also requires that it is benchmarked relative to other data sources so as to "... have confidence that the data reflect real characteristics and behaviour" (Thomas et al. 2012: 20). Such benchmarking is likely to be made to other existing sources. In the case of Thomas et al. (2012), who evaluated the usefulness of a large commercial survey, benchmarking was conducted relative to census and other survey data, with the authors benchmarking at the aggregate (areal) level, at the micro level (i.e., age, ethnic, and gender distributions), and spatially (moves or linkages across regions). Of course, such benchmarking only works when there are comparable and reliable data sources to benchmark against.

Therefore, any inferences drawn from the data must recognize the potential issues around the sample frame. For regional scientists, what does this mean for the analysis of human capital and related questions? Shearmur (2015) argues that big data can reveal different things and ideas than a census (or related data) can, but that it can't replace a census. Despite the opportunities inherent in big data, it is not about a society but about its users and markets, meaning that individuals who fall out of the market do not appear in the data. Likewise, big data can imply relationships, but become problematic owing to sampling bias, self-selection bias, and ecological fallacy.

19.2.4 Big Data: Big Opportunities?

So let us now return to the questions of the need to better identify the economic and non-economic city-based factors that influence individual decisions to invest in human capital, to accumulate human capital more rapidly, and to favour relocating to larger urban centres. Much of the work to date has implied the influence of cities on these decisions by making a series of related arguments: Cities attract workers with more skills because they have thicker labour markets; Cities induce their citizens to invest in human capital because returns to investment are higher in cities (the signal strength is greater); Cities accelerate income growth for highly skilled workers because they provide greater opportunity to learn from talented peers and to move from organization to organization with different forms of intangible capital. In all of this work, cities play the role of overcoming space. As noted by Glaeser (2011: 6)

> Cities are the absence of physical space between people and companies. They are proximity, density, closeness. They enable us to work and play together, and their success depends on the demand for physical connection.

While interaction across space is not free, cities are expected to ease and make more effective economic and social interactions. To date, the role of cities has been identified indirectly through outcomes (e.g., location choices of degree holders) via census, survey and administrative data sources. The challenge is assessing more directly the mechanisms thought to facilitate the accumulation of human capital in cities, a question where big data may begin to play a role.

If the underlying mechanisms thought to drive and accelerate the accumulation of human capital in cities are driven by work-based and social networks, then many of the big data sources (e.g., various social median platforms) come to the fore. Do we observe in social media and services like Linkedin thicker networks in cities that workers can learn from? Following from this, are cities and social networks complements, with social networks making it easier to find people with complementary skills that workers can emulate and learn from? How do social networks influence the relocation of workers? Do networks follow or lead the decision to move? The identification of various networks is a powerful tool, but leveraging this power ultimately means connecting these with traditional sources of data. Being connected to a professional network via Linkedin may facilitate connections, but it is only if these connections are associated with outcomes (e.g., career progress or wage gains) that its strength is revealed. In a nutshell, the power of these data stems from using them to leverage more traditional data sources. This has the additional benefit of providing the proper weights and basic social demographic information that otherwise might limit the usefulness of big data. It has the challenge of finding ways to link big data harvested from social media to more traditional administrative and survey files held by statistical agencies. Identifiers have to be found to make the link and mechanisms have to be put in place that assure the confidentiality of the individual microdata.

19.3 Conclusion: Where Does Big Data Fit in the Regional Science Agenda?

Overall, there is likely tremendous value associated with the use of big data, with Einav and Levin (2014: 715) commenting that "big data is already allowing for better measurement of economic effects and outcomes and is enabling novel research designs across a range of topics". The benefits of its use include real time analysis, scale/scope of data, and the potential to identify relationships and correlations that might not have been observed using classical methods and data sources. Indeed, a report by the McKinsey Global Institute (2011) concluded that the greatest benefits of big data would be to generate new value or increase productivity, such as through firms who are able to utilize the data for greater market penetration, and/or growth opportunities for firms that analyze data. Over time, big data will likely affect the type and diversity of questions regional scientists can ask. For regional scientists interested in human capital issues, the use of big data also may provide insight into the current endogeneity issues.

But big data will not provide all of the answers on its own. Pitfalls, including data availability, representativeness, computing and software issues, and contextual content remain. Existing big data, including that from labor intermediaries and passive sources are scattered and heterogeneous. If the data is to be used effectively, the scale and complexity of the data pose practical, technical, and methodological challenges. Nevertheless, potential regional science questions can be easily identified. Movements to liberate data including 'open data' and 'smart cities' may increase data availability, but other proprietary data are less likely to be widely available to academic researchers. The greatest ability of big data may be found when it can be leveraged with other data, including survey data, thus helping to minimize some of these issues, or when it can be used to explore natural experiments (Glaeser et al. 2015).

As interest in big data has grown, an epistemological debate has emerged within the literature: Does the analysis of big data require theory for understanding? Has the emergence of big data created a new research paradigm of data-intensive research? There is certainly a thriving discussion within the current literature (i.e., Chang et al. 2014; Guszcza and Richardson 2014; Kitchen 2014). Chang et al. (2014), for example, discuss notions of a new 'computational science', while Kitchen (2014) suggests that a paradigmatic shift is occurring as analytical social science moves from a knowledge-driven science to a data-driven science, meaning that big data has prompted some to suggest that analysis and interpretation is now a-theoretical. In a related way, the role of inference changes (and perhaps disappears) as a sample approaches the universe, such that inference approaches pure description. That is, big data (largely given its volume and velocity) is reconfiguring how science is done, with the data able to speak for itself, removing the need for theory, hypotheses, and models.

However, it is one thing to identify patterns, which the new big data science argues is better, and another to understand why such patterns arise. Such reasoning is reflective of Popper's view that theory is embedded in data and that theory enables understanding of the phenomena at hand. For example, big data might tell us that someone has moved, but it would not provide reasons for the move and what the individual was doing before or after the move. Instead, theory and inference must be used to decipher the motivations, a process that is guided by theory. As Shearmur (2015: 2) argues "*there are many understandings of social phenomena, not a single understanding, and these differing perspectives are irreconcilable with the determinism implicit behind Big Data rhetoric*." Further (p. 3), Shearmur notes

> As such, for the purposes of social science (and urban geography in particular), they [big data] can serve to generate hypotheses inductively: but however useful inductive reasoning may be, theory and imagination are necessary to understand and interpret observations.

Consequently, theory must remain embedded in the analysis of big data.

References

Antenucci D, Cafarella M, Levenstein M, Re C, Shapiro M (2014) Using social media to measure labor market flows. National Bureau of Economic Research (NBER) working paper no. 20010

Beckstead D, Brown WM, Gou Y, Newbold KB (2009) Cities and growth: human capital and urban–rural income disparities in Canada. Draft working paper, Micro-economic Analysis Division, Statistics Canada

Behrens K, Duranton G, Robert-Nicoud F (2014) Sorting, selection, and agglomeration. J Polit Econ 122(3):507–553

Breinlich H, Ottaviano GIP, Temple JRW (2013) Regional growth and regional decline. CEP discussion paper no 1232

Brown WM, Newbold KB, Beckstead D (2010) Growth and change in human capital across the Canadian urban hierarchy, 1996–2001. Urban Stud 14(7):1571–1586

Brown WM, Scott DM (2012) Human capital location choice: accounting for amenities and thick labor markets. J Reg Sci 52(5):787–808
Chang RM, Kauffman RJ, Kwon Y (2014) Understanding the paradigm shift to computational social science in the presence of big data. Decis Support Syst 63:67–80
Chetty R, Friedman JN, Hilger N, Saez E, Schanzenbach DW, Yagan D (2011) How does your kindergarten classroom affect your earnings? Evidence from project star. Q J Econ 126:1593–1660
Choi H, Varian H (2012) Predicting the present with google trends. Ekon Istraz 88:2–9
Combes PP, Duranton G, Gobillion L (2008) Spatial wage disparities: sorting matters! J Econ 62(2):723–742
Combes PP, Duranton G, Gobillon L, Roux S (2012) Sorting and local wage and skill distributions in France. Reg Sci Urban Econ 42:913–930
Curry B (2014) Job-vacancy rate plunges as Tories drop Kijiji data. The Globe and Mail, 5 May 2014. http://www.theglobeandmail.com/report-on-business/economy/jobs/ottawa-adjust-labour-data-raising-questions-about-national-skills-shortage/article18457198/. Accessed 15 Nov 2015
De la Roca J (2015) Selection in initial and return migration: evidence from moves across Spanish cities. Discussion paper
De la Roca J, Puga D (2015) Learning by working in big cities. Discussion paper 9243, Centre for Economic Policy Research
Einav L, Levin J (2014) Economics in the age of big data. Science 346(6210):715–721
Finnie R (2012) Access to post-secondary education: the importance of family culture. Child Youth Serv Rev 34(6):1161–1170
Ge C, Huang KW, Png IP (2014) Engineer/scientist careers: patents, online profiles, and misclassification bias. Online profiles misclassification bias
Gennaioli N, La Porta R, Lopez-de-Silanes F, Shleifer A (2013) Human capital and regional development. Q J Econ 128(1):105–164
Glaeser EL (1999) Learning in cities. J Urban Econ 46:254–277
Glaeser EL (2011) Triumph of the city: how our greatest invention makes US richer, smarter, greener, healthier and happier. Penguin, New York
Glaeser EL, Maré DC (2001) Cities and skills. J Labor Econ 19(2):316–342
Glaeser EL, Saiz A (2003) The rise of the skilled city. Working paper no 10191, National Bureau of Economic Research, Cambridge, MA
Glaeser EL, Scheinkman JA, Shleifer A (1995) Economic growth in a cross section of cities. J Monet Econ 36(1):117–143
Glaeser EL, Kominers SD, Luca M, Naik N (2015) Big data and big cities: the promises and limitations of improved measures of urban life. Working Paper 16-065. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2694723. Accessed 6 June 2016
Guszcza J, Richardson B (2014) Two dogmas of big data: understanding the power of analytics for predicting human behavior. Deloitte Review, 15
Horton JJ, Tambe P (2015) Labor economists get their microscope: big data and labor market analysis. Big Data 3(3):130–137. doi:10.1089/big.2015.0017. Accessed 6 June 2016
Jin X, Wah BW, Cheng X, Wang Y (2015) Significance and challenges of big data research. Big Data Res 2:59–64. doi:10.1016/j.bdr.2015.01.006. Accessed 6 June 2016
Kitchen R (2014) Big data, new epistemologies and paradigm shifts. Big Data Soc:1–12. doi:10.1177/2053951714528481
Lee JG, Kang M (2015) Geospatial big data: challenges and opportunities. Big Data Res 2:75–81
Marinescu IE, Rathelot R (2013) Mismatch unemployment and the geography of job search. Available via http://www.sole-jole.org/15260.pdf. Accessed 24 May 2016
Marinescu IE, Wolthoff R (2015) Opening the black box of the matching function: the power of words. Available via http://ftp.iza.org/dp9071.pdf. Accessed 2 Nov 2015
McKinsey Global Institute (2011) Big data: the next frontier for innovation, competition, and productivity. Available via http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation. Accessed 24 May 2016

Moretti E (2004) Human capital externalities in cities. In: Henderson JV, Thisse J-F (eds) Handbook of regional and urban economics, vol 4. Elsevier, Amsterdam, pp 2243–2291
Newbold KB, Brown WM (2015) The urban-rural gap in university attendance: determinants of university participation amongst Canadian youth. J Reg Sci 55(4):585–608
Papinski D, Scott DM (2013) Route choice efficiency: an investigation of home-to-work trips using GPS data. Environ Plan A 45:263–275
Papinski D, Scott DM, Doherty ST (2009) Exploring the route choice decision-making process: a comparison of planned and observed routes obtained using person-based GPS. Transp Res F 12:347–358
Rango M (2015) How big data can help migrants. Available via https://agenda.weforum.org/2015/10/how-big-data-can-help-migrants/. Accessed 2 Nov 2015
Rosenthal SS, Strange WC (2008) The attenuation of human capital spillovers. J Urban Econ 64:373–389
Shapiro JM (2005) Smart cities: quality of life, productivity, and the growth effect of human capital. Working paper no 11615, National Bureau of Economic Research, Cambridge, MA
Shearmur R (2010) Editorial—a world without data? The unintended consequences of fashion in geography. Urban Geogr 31(8):1009–1017. doi:10.2747/0272-3638.31.8.1009
Shearmur R (2015) Dazzled by data: big data, the census and urban geography. Urban Geogr 36(7):965–968
Sobolevsky S, Izabela S, Des Combes RT, Hawelka B, Arias JM, Ratti C (2014) Money on the move: big data of bank card transactions as the new proxy for human mobility patterns and regional delineation. The case of residents and foreign visitors in Spain. Paper presented at the 2014 IEEE international congress on big data. doi:10.1109/BigData.Congress.2014.28
Taubman SL, Allen HL, Wright BJ, Baicker K, Finkelstein AN (2014) Medicaid increases emergency-department use: evidence from Oregon's health insurance experiment. Science 343:263–268
Thomas M, Gould M, Stillwell J (2012) Exploring the potential of microdata from a large commercial survey for the analysis of demographic and lifestyle characteristics of internal migration in Great Britain. Department of Geography, University of Leeds, Working paper 12/03
Toole JT (2015) Putting big data in its place: understanding cities and human mobility with new data sources. Dissertation, Massachusetts Institute of Technology
Varian H, Scott S (2014) Predicting the present with Bayesian structural time series. Int J Math Modell Numer Optim 5:4–23
Wu W, Wang J, Dai T (2016) The geography of cultural ties and human mobility: big data in urban contexts. Ann Assoc Am Geogr 106(3):612–630
Yankow JJ (2006) Why do cities pay more? An empirical examination of some competing theories of the urban wage premium. J Urban Econ 60(2):139–161

K. Bruce Newbold is professor, Geography & Earth Sciences (SGES), McMaster University. He is the current Director of SGES, and the former Director of the McMaster Institute of Environment and Health. His primary research interests are related to population geography, including immigration; health; commuting behavior; mobility; and human capital development. Previous faculty positions were at the University of Illinois Urbana-Champaign, and he has been a visiting professor at the University of Glasgow and University of California San Diego. Dr. Newbold earned his Ph.D. in geography from McMaster University in 1994.

W. Mark Brown is the chief of Regional and Urban Economic Analysis, Economic Analysis Division, Statistics Canada. His primary research interests are international and interregional trade; urban and regional productivity; and human capital migration and growth. He was a faculty member in the departments of Geography and Agricultural Economics at McGill University before taking his current position at Statistics Canada. Dr. Brown earned his Ph.D. in Geography from McMaster University in 1999.

Chapter 20
The View from Over the Hill: Regional Research in a Post-Demographic Transition World

Rachel S. Franklin and David A. Plane

20.1 Introduction: Demography as a Driver of Regional Economies, Policy and Research

People age. So do disciplines, theories, and methods. Just as people may pass their prime and become 'over the hill,' so too may ideas or received wisdom. In this chapter we address aging *and* regional science and attempt to 'look over the hill' at some future directions for regional research, as suggested by portending demographic change. It is our contention that population trends are both an important topic for regional scientists to engage with, going forward, and that they are key drivers of regional economies and the full spectrum of regional policy issues for which the analytical methods of regional science are ideally suited.

The multidisciplinary field of regional science was born and nurtured to maturity in the immediate post-World War II era. It was a period when rapid advances were being made in many fields of science, and new electronic and computer technologies were being invented. Large-scale modeling and normative approaches to urban and regional planning were coming in vogue.

Looking back down the hill of the past 60 years, regional science has come to be most closely associated with its methods; many believe that developing a toolkit of useful methods is what constitutes the field's greatest success. To look up ahead, towards the future frontiers of the field, we take as a given that there will continue to be an important role for quantitative approaches to regional problem-solving and

R.S. Franklin (✉)
Brown University, Providence, RI, USA
e-mail: Rachel_Franklin@brown.edu

D.A. Plane
University of Arizona, Tucson, AZ, USA
e-mail: plane@email.arizona.edu

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 1*,
Advances in Spatial Science, DOI 10.1007/978-3-319-50547-3_20

policy formulation. We presume, furthermore, that the critical problems and policy needs for future regional science research will arise in a world very different from the one that shaped the field at its founding. To scope out key aspects of that future world, we take as our lens the world's evolving demography.

During the more than half-century that has passed since the 1950s, the world's countries have been undergoing profound demographic transformations. Slower rates of population growth or decline are becoming the new norm, and populations most everywhere have begun to age. Since regional science's salad years, social norms in many areas of the world have drastically transformed as wage labor forces have expanded to encompass most all adults: women as well as men. Concomitantly, the composition of households has been radically altered as prospective parents face a new calculus about when and how many children to bring into the world; norms regarding desirable household size have been downsized. Linkages between demography and economics are strongly forged because of the co-dependence of the labor force and child-rearing decisions made by individuals and households. And the size of successive generations affects individuals' economic prospects and decisions (Easterlin 1980).

Demography is a particularly powerful lens through which to scope out the future because of two simple, obvious, yet highly salient facts about individual lives (Plane 2012). The first is the inevitability of aging: it is an inconvenient truth that, year-by-year, most everyone gets a year older. And the second is that, at different points in their life courses, people tend to be doing different things. As a result of these two inescapable regularities, future human populations can be projected at broad geographic scales with relatively high confidence. So, too, the future age, sex, race and ethnic compositions of those populations are fairly knowable. Projections are, of course, always subject to a few caveats: calamitous events could lead to unanticipated, significant spikes in mortality; radical breakthroughs in extending maximum human life spans, currently thought unlikely, might occur; and abrupt changes in individual preferences and cultural norms could alter the relatively steady, long-term trends with respect to desired family size and to the number and timing of births.

Less knowable than future mortality and fertility levels is how people will be redistributing themselves in the future. Migratory flows wax and wane in response to complex systems of pushes and pulls. For finer-scale geographic units, including the various types of regions where regional scientists focus their efforts, migration is often the largest component of population change. The methods of regional science, however, are all about understanding how complex, spatially interacting systems function. So it is perhaps unsurprising that it is migration—and, more specifically, how migration affects regional labor supply and demand—that has occupied most of the attention of regional scientists working in the field's interface zone with demography and population geography (Plane and Bitter 1997; Plane 2000).

Largely a First World scholarly movement in its formative years, regional science's methods of analysis were developed in response to the regional issues and policy needs of the 1950s, 1960s, and 1970s in the countries of North America, Northern and Western Europe, Japan and Korea, and Australasia. In these countries, where the field's initial adherents were proselytized, the subject matter chosen for regional analysis reflected some common contexts, problems, and policy needs. During the first three decades of regional science, most of these countries had rapidly expanding populations due to post-World War II 'baby booms,' and their industrial-driven economies were expanding at record paces, as science and technological innovation blossomed.

Large-scale changes in both demographic and economic trends drove patterns of regional population growth and redistribution. Suburbanization and urban sprawl entered the lexicon and became the primary focus of regional planners and a land development industry. The baby booms fueled massive investments in new housing stock to accommodate newly formed, multi-child families. At the same time, strong labor demand drew workers to rapidly growing and expanding metropolitan areas. For decades, labor-force entrants had been leaving rural areas, where agricultural production increasingly substituted machines for hand work, and moving to the cities. It was in the expanding urban areas where manufacturing and service industries could reap seemingly inexhaustible economies of agglomeration to provide new goods and services for expanding and increasingly prosperous consumer markets.

But as regional science continued to develop from its initial set of approaches appropriate for responding to these post-war trends, newly evolving demographic and economic realities in the most developed countries where regional science was pioneered began to fundamentally alter policy and research needs. Those realities reflected the repercussions of all these countries completing their long-term fertility transitions.

In the United States, at the height of the baby boom in the late 1950s, total fertility peaked at close to 3.7 births per woman; by the 1970s the U.S. Total Fertility Rate (TFR) had dropped below 2.0 births per woman. Ever since, U.S. fertility has stayed plateaued at—or just slightly below—the long-term replacement level of 2.1. As of 2014, the U.S. TFR stood at 1.9 (Population Reference Bureau 2014). In all the other Most Developed Countries (MDCs), fertility has dropped below replacement level, in many cases well below.

Looking 'over the hill,' the prognosis for MDCs is for very little future population growth or, in many cases, population declines. With working and child-bearing age cohorts no longer replacing themselves, developed countries' populations, absent significant immigration, will rapidly age, with reduced youth dependency but increased elderly dependency.

Along with the new demographics, recent decades have witnessed the emergence of a globalizing economic world. The people and regions of the MDCs have been grappling with the repercussions of the 'hollowing phenomenon' and post-industrialization, out-sourcing and off-shoring, the emergence of significant new consumer markets within parts of the still developing world, and the whole interrelated bundle of issues wrapped up in the new webs of international economic relationships.

We think one of the first harbingers of how regional science needs to be prepared to cope with new demographic and economic realities took place in the realm of population redistribution. It was a phenomenon that came to attention back in the 1970s and early 1980s, when many of the trends just described were only slowly coming into focus. The unanticipated population phenomenon was a set of internal migration 'turnarounds' that took place across the MDCs, as long-standing urbanization flows halted and what came to be called 'counter-urbanization' flows emerged. For a time, non-metropolitan growth outstripped metropolitan, and peripheral regions were challenging the hegemony of these traditional, manufacturing-dominant cores (Vining and Kontuly 1978; Brown and Wardwell 1980). Regional scientists, with their toolkits of analytic and data-driven methods, were perfectly positioned to leap to the fore in studying and seeking to explain this empirically observed phenomenon. Not coincidentally, it was found that these trends occurred as the baby-boomers were coming of age and flooding regional labor markets (Plane and Rogerson 1991a).

As the decades have unfolded since, the turnaround was not sustained, coming and going rather like an El Niño weather pattern (Champion 1988; Long and Nucci 1997). Regional scientists have been presented with many challenging new trends to analyze and to contribute to policy-making. For instance, in the U.S. beginning around 2000, when the huge Millennial or baby-boom-echo generation was coming of age and again pressuring labor markets, young professional gentrifiers and the real estate industry have transformed and revived the urban cores of metropolitan areas.

In the later sections of this chapter we will flesh out some ideas about where, more specifically, demographic trends will create needs and opportunities for new types of regional analysis. But all such considerations also need to take into account an overarching economic challenge for the MDCs as they move into the future. As the historical 'demographic divide' and economic dichotomy of 'have' and 'have not' countries dissipate, what role will these countries' firms and households play in the world?

Of late, the regional science movement has been rapidly diffusing throughout the developing world. As the differences between MDC and LDC (Lesser Developed Country) economies lessen, some of the issues for future regional science research are similar. Yet in the next 50 years or so, the LDCs will be grappling with some of the same types of challenges the MDCs faced in the past half century.

The still-developing countries, with the notable exception of some in sub-Saharan Africa, have now moved out of the early-expanding stage of their demographic transitions—in which fertility remains high and population growth explodes while mortality is reduced—into the late-expanding stage, during which total fertility rates rapidly drop towards or even, in a few cases thus far, below replacement level. The former Second and Third World countries' fertility declines have taken place much more abruptly than in the First World.

At a regional and local scale, as previously noted, settlement patterns in much of the developed world have been characterized by urban sprawl and suburbanization, as well as periodic counter-urbanization starting in the 1970s. Meanwhile, in the developing world, rapid urbanization has become the norm, with the growth of many new mega-cities brought about by massive rural-urban migration. The still developing countries are, or in most cases soon will be, passing through the final stages of demographic transition and grappling with the attendant economic opportunity presented by the so-called 'demographic dividend': that sweet spot in time when there is lessened youth dependency plus the bulk of the adult population is still of prime working age and more women enter the wage labor force.

In the next nine short sections, we attempt to sketch out some of the changing needs and issues for regional research as global population growth slows, as labor forces age in both more developed and rapidly developing countries, as advances in gender equality proceed apace, and when in many regions little population change, or even population losses, become as likely as continued population growth. Similarly, we discuss the implications for regional research of trends in demographic composition, such as new, small-family norms with more people living alone or in nontraditional household arrangements.

We don't pretend that our nine areas constitute the full over-the-hill landscape of such needs and issues. We try to be a bit controversial in order to provoke thought. We recognize that, in many cases, alternative arguments can be made about the repercussions of projected demographic trends. But then that's what the mission of regional science research is all about: to gain understanding of the competing forces at work in highly complex, spatially interacting, social and economic systems, as well as to contribute to formulating regional policies that help solve important, real-world problems.

20.2 Residential Location and Households

Within countries that have completed the demographic transition, aging populations are now living in very different household arrangements than was typical during the 1950s and 1960s baby-boom period, when the nuclear-family with two parents and multiple children at home was the archetype. For the U.S., the rise in non-family households is dramatically illustrated by Fig. 20.1, which shows the trends since 1940 in percentages of households of various types. The percentage of single-person households has more than doubled since 1960, increasing from 13.1 to 28.0% of all households. Household composition shapes the demand for housing. The U.S. has long had exceptionally high rates of homeownership by world standards. But a long-term upward trend in people buying homes peaked in 2005, when 69.3% of householders were owners. By 2014, the rate had fallen to 64.6% of households living in owned quarters.

After the bursting of the housing bubble in 2007, and during the subsequent Great Recession that it brought on, the U.S. housing construction industry discovered a 'new' market: the old downtowns of large cities. Many of the new, smaller households' preferences shifted away from choosing more space at the price of longer commutes, towards shorter commutes and less space. These demographically fueled market trends should come as no surprise to regional scientists reared on monocentric bid-rent theory. That downtowns became hotspots for housing

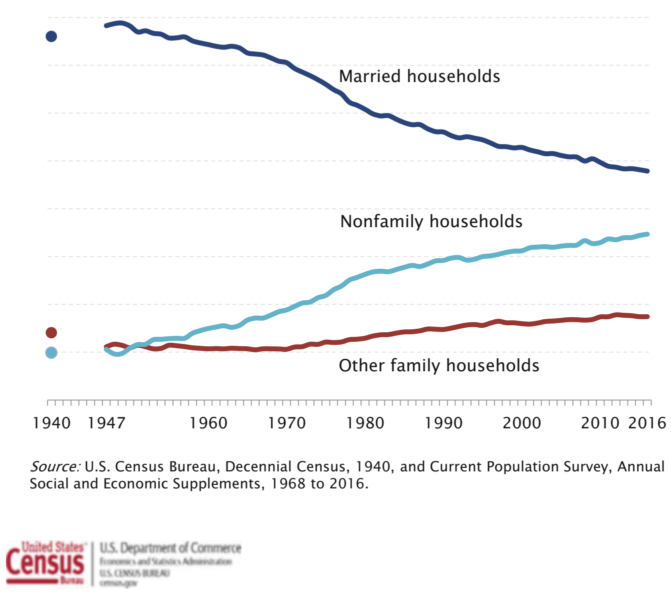


Fig. 20.1 Trends in prevalence of U.S. Households by type (Source: https://www.census.gov/hhes/families/, last accessed January 27, 2017)

also reflected the success of major cities' CBDs recapturing their role as their metropolitan area's entertainment and social hub. But it was not coincidental that the Millennial generation—now the U.S.'s largest, exceeding even that of the baby boomers—were just then reaching the emerging-adult stage of life. So too, a portion of the baby-boomers who had recently become empty-nesters found an inward move from the suburbs appealing at their stage of life, when downsizing housing makes sense.

For the future of regional science research, we see as a first critical issue the 'right-sizing' and 'right-matching' of housing stock with households rapidly shifting in terms of both composition and preferences for location. This involves, as well, a shift in the composition of housing types: apartments, condos, and townhouses becoming more attractive options; large, single-family, detached homes becoming less popular. Affordability of housing in newly prized more central locales, and a likely glut of now-aging suburban homes, presents many challenges and opportunities. The spatial repercussions of changing household composition and housing needs and preferences also, of course, impinge on a whole host of other, prime regional science research areas, such as job location, transportation, infrastructure provision, public finance, and so forth—some of which we shall detail in subsequent sections of this chapter.

In addition to the intraurban dimension, the aging of MDC populations and the shifting proportions of people passing through various life stages also have highly salient repercussions for residential redistribution at the scale of metropolitan areas and broad regions (Bitter and Plane 2012). Plane and Heins (2003) show how the age-profiles of migration vary between pairs of U.S. metropolitan regions. The different age structures reflect a whole host of factors, including the functional specialization of various cities' economies, complementarities of climate, cost of living and a whole host of other differentials that regional scientists have long taken an interest in.

Another of the salient factors is one for which intriguing new empirical regularities are being found: metropolitan area size (see, e.g., Plane et al. 2005; Plane and Jurjevich 2009). We think extensions of regional scientists' traditional central place theories and economic base models that take into account the newly emerging realities of post-industrial, globalized economies, together with the factors that motivate populations at different life course stages, could aid in understanding of spatial shifts in many key aspects of the space-economy: those of industries, employment, income and wealth, among others.

This discussion of residential location and households has focused, thus far, largely on MDC patterns. There is an equally key need for regional science research on this topic in LDC contexts. As most LDCs have now begun to go through the final stages of their extremely rapid demographic transitions, it is unlikely that patterns of suburbanization and exurbanization will play out in exactly the same way as they did during earlier decades in the First World. What is more, the very different contexts and cultures found across the former Third World have resulted in a variety of historic urban forms, and there is no reason to presuppose that cities will develop in lock-step ways.

With globalization, we have also entered a new age of heightened international migration. United Nations (2015) estimates suggest that the number of people migrating between countries surged by 41% between 2000 and 2015. Debates about border security, new political structures such as the European Union, and refugee movements overlay an additional level of complexity on migration research when moving up to the international scale from studies of intraurban residential relocation and internal or domestic migration.

20.3 Transportation and Accessibility

In discussing residential location, we have already hinted at some needs for future studies in the realm of transportation and accessibility. Demographic change, coupled with technological change, is reshaping the terrain for transportation studies.

In MDCs, as higher fractions of the population enter the post-working ages, what will be the implications for cities that have developed along patterns reflective of extreme auto dependence? Early-career young people, of late, have been exhibiting a reluctance to being auto dependent. With new modes of shopping through e-commerce and a strong trend towards interacting within the virtual world of the web and cellphones rather than through physical movement (e.g., texting and emailing, rather than making trips to the post office to mail letters; spending time on social media, rather than attending meetings of social organizations), the accessibility considerations of people at various stages of life are undergoing interesting transformations, with attendant implications for business location and transportation and service-provision planning and policy. The lives of even the least tech-savvy oldest cohorts are being impacted, and now, the world's first computer generation, the baby boomers, are entering the elderly ranks.

Another topic for further exploration connected to demographic aging is to reorient thinking about inter-generational spatial accessibility. We think the co-location of the elderly and their adult children will assume a more central role. With longer life spans and greater years of good health, thinking needs to move away from a three-generation (youth, adult, elderly) to a four-generation model. Healthier younger elderly and late-in-life dependent elderly have very different daily activity and yearly activity patterns and spatial orientations. In the future, regional science studies on issues pertinent to both these groups need to become a more significant part of the field.

In the developing country context, rapid rural-to-urban migration has led to the growth of megacities lacking much-needed basic infrastructure, and lack of transportation and accessibility are of critical concern. While some of the experience of MDCs can be brought to bear, the current situation in LDCs differs. The current boom period of urban growth is taking place at the beginning of the twenty-first century, when automobiles rule the roost. During the nineteenth century, when immigrants were pouring into the East Coast cities of the U.S. leading to unprecedented growth rates and densities, new forms of public mass transportation were being critically demanded to expand the distances that could be feasibly reached by people traveling on foot. The problems then were of excessive density in the constrained radius where people and jobs had to locate. LDCs now and in the future must grapple with extreme congestion on roadways and with air pollution.

20.4 Amenities and Quality of Life

Inherent to much regional research is the assumption that some places are more desirable than others and that the various dimensions of desirability are objective and measurable. Research on migration and economic development, in particular, has emphasized the role of 'quality of life' variables in explaining regional differences in in-migration, firm location, or innovation, for example (see e.g., Mulligan and Carruthers 2011; Partridge et al. 2010; Whisler et al. 2008). Data products such as the U.S. Department of Agriculture's natural amenities scale attach values to different environmental characteristics of counties and assign them to a range of categories from high to low amenities. However, while some aspects of quality of life are likely to be universal—low crime, for example—others are surely more subjective and depend on the demographic group under consideration (Ruth and Franklin 2014). While younger age cohorts may evince a preference for the great outdoors or coastal proximity, perhaps older cohorts prefer locations with less climatic variability. Moreover, in a world of limited resources, even universal quality of life variables may rise or sink in relative priority, depending on the group affected. Older cohorts may share a preference for coastal locations with the rest of the population, but might rank outdoor environment lower than proximity to healthcare.

Another example comes from school funding. In general, good public schools are an indication of high quality of life, especially in the U.S., where property values are tightly linked to perceived public school quality. Older populations, however, may be not only less likely to care about schools, but may also shy away from diverting funding from healthcare to schools when forced to choose.

It appears to us there are at least two ways that aging populations may bring about a change in research employing quality of life (QoL) indices or variables. First, regional researchers may need to adjust the actual measures used to judge quality of life. These may be amenity variables, but might also be the weighting used in developing composite measures. As discussed above regarding transportation, the aging of the population may shift good public transportation from a 'desirable' to a 'must have' in order for quality of life to be preserved. Further research will certainly be necessary to understand how different groups weight various amenities, whether public transportation, climate, healthcare, or entertainment options.

Second, and related, regional variation in population composition may lead to increased heterogeneity in responses to typical amenities, however they are measured. While we may learn that past measures of QoL no longer hold for older groups, it may also be that differences in rates of aging and stocks of older people in different locations mean that there is less agreement about how desirable any one place might be—and evaluations of quality of life may evolve rapidly, as the characteristics of the people living in an area change. Perhaps for smaller areas, future quality of life measures will need to be more demographic specific.

20.5 Economic Development

When the economic effects of aging are talked about, they are generally framed as national-scale concerns about labor supply and elderly dependency. Contemplating projected reductions in proportions of the population in labor force ages, questions are raised about who the workers of the future will be. Concomitantly, the question gets asked: Who is going to fund the medical and other care needs of more old folks? Our focus here, however, goes beyond such national scale debates. For visualizing the frontiers of regional science research, we need to consider the effects of slow or negative overall population growth and older age structures at the scales of regions, states, metropolitan areas, and neighborhoods.

A first observation: Population is likely to continue to redistribute in the absence of overall growth. Age composition shifts, themselves, will be a driver of such movements because of the different preferences and incentives at the various stages of life. In the U.S., with its high rates of home-ownership, young people early in their careers pay for housing, whereas for many elderly persons the home constitutes the majority of their equity. 'Equity migration' is likely to become even more important. We would note that regional scientists over the last several decades have done interesting and important studies on elderly migration, natural-amenity driven economies, and the significant role of transfer payments in rural economic development.

Taking no or negative national population growth down to regional scales should finally hasten the laying to rest of the old 'growth' versus 'development' debate—because growth may not be a realistic goal for those in many regions to pursue! Although in theory we know growth and development are not synonymous, growth-promotion is still the mode within which many real-world development agencies operate, and it's the number of new firms, jobs, and so forth brought to town by which their efforts get evaluated. Both public and private sector players have a stake in attracting more jobs and people to town. Numerous non-basic sectors benefit from expanding markets.

A big picture question, though, is to what extent do recruitment efforts distort the otherwise optimal locations predicted by our theories of industrial and retail location? And how much is being given up in the incentives and tax breaks that get offered when communities are played off against one another by firms adding or moving jobs—jobs that will surely locate somewhere? Development may not be a zero-sum game. It may be the case, as succinctly summed up by former Evansville, Indiana mayor, Benjamin Bosse, "When everybody boosts, everybody wins."

With no longer a constantly increasing national population to draw upon and with a shrinking number of young labor entrants—those prospective employees the most willing to relocate—we suspect many fewer communities will be focused on attracting and dealing with the repercussions of growth, such as the need for new infrastructure. Development efforts will increasingly turn to strategies for increasing the quality of the jobs held by the local labor force and enhancing the quality of life for a stable or declining pool of citizens.

During the expanding stages of the demographic transition, huge societal investments are required to provide for children. As the bases of population pyramids shrink away, and age structures begin to look more like rockets or Japanese lanterns, fewer absolute resources will need to be devoted to youth dependents. Education, we would hope, would not drop from a prominent place in public discourse. Rather, education needs to become perceived as ever and ever more important to provide new entrants and retrain existing employees for the smaller, smarter, tech-savvy workforces of the future. Focus can shift, however, from building new school facilities, to other types of infrastructure, which often has been in short supply in high-growth environments.

For adults, concepts of work-life balance and retirement need to be rethought. New employment options are needed for balancing the competing demands on time of career-advancement and child rearing in two-worker or single-parent households. Elderly subpopulations, unlike youth, need not be viewed as 'dependents.' From an economic development standpoint, retirees are critical for supporting the communities' cultural and educational institutions. And with the span of healthy years of life dramatically increased from the norms of the past, people past traditional retirement ages are a still largely under-tapped source for part-time labor and volunteer positions. Some creative competition between communities could be useful in experimenting with how private and institutional employers can make best use of the diversity of age structures likely to be found at the local scale. Regional science models need to build in more concepts of the life course to be relevant to such efforts.

20.6 Migration

If we were to issue the Ten Commandments for future migration research, then one of them would cite respect for the connection between life course stage (or age) and human movement (cf. Plane and Rogerson 1991b). The life course, in fact, matters for everything; note how often it has been invoked in this chapter. Where migration is concerned, age impacts whether people move, why they move, how far they move, and where they go. And regional scientists have not only devoted attention to the study of migration behavior itself, but also the role of migration in regional development, broadly speaking. This suggests that changes in age structure are likely to directly affect migratory behavior but will also indirectly affect regional systems as a whole.

At any given age, most individuals are stayers, not movers. However, the propensity to, or likelihood of, moving varies considerably by age. As we age, children, employment, and homeownership tie us to place, decreasing the probability of moving. In most countries, migration peaks in early adulthood, before and during the early years of household formation. In the U.S. and some other countries, a retirement bump is discernible around retirement age. Figure 20.2 takes the specific case of recent migration in the U.S., as captured by the Current Population Survey.

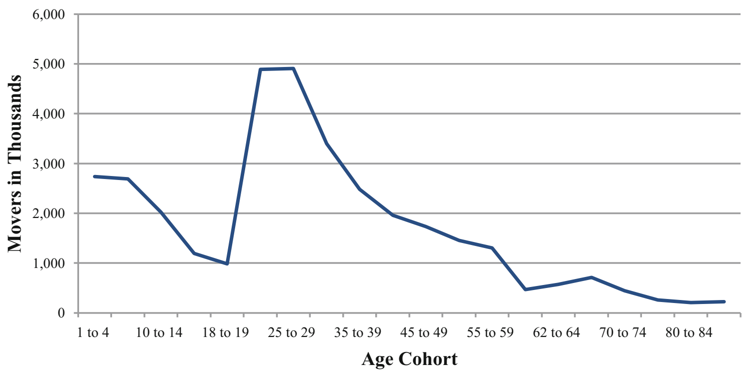


Fig. 20.2 Mobility by age cohort, United States, 2014–2015, current population survey

Here, the retirement migration bulge is quite apparent, as is—interestingly—a miniature pre-retirement bump. This latter likely reflects empty-nester moves in advance of retirement.

Depending on the overall age structure of a place, then, we can expect fewer movers, but then also potentially a great wave of retirement moves as the baby boomer generation continues moving through this stage of the life course.

Population aging is also likely to bring changes to labor markets, though, shifting or extending the transition from employed to retired. How aging might affect the relationship of migration to the later stages of the life course is complicated. Later retirement might push the retirement bump out to later years, but not change the basic shape of the curve. If individuals increasingly choose to migrate first, work longer, and then retire later, however, the bump might shift to earlier ages. A third option, that the transition stretches out (both at the individual and group levels) might lead to more of a "speedbump" in the life course, as opposed to an actual bump.

What about migration distance and the older migrant? The archetypal aged migrant moves a relatively long distance, in the U.S. from, say, an inhospitable Midwest suburban location to the sunny South. Migration could also occur over short distances, though, as individuals keep their local environment but simply downsize their housing unit. In the future, as the population as a whole ages, this means we may see wholesale neighborhood transitions, with residents either moving nearby or to farther off climes. It's possible, though, that we may also observe cohort changes in migration behavior, if not in terms of distance then at least in terms of motivation for covering that distance. For example, as suggested above, maybe future generations of retirees will prioritize locations closer to children and grandchildren over shuffleboard in Florida.

As intimated in Sect. 20.4 above, older people may also prefer different sorts of locations than their younger selves or younger peers. This is already well documented for metropolitan area size (Plane et al. 2005), so the real question is how much established patterns are likely to change in the future. An aging population may lead to changes in the observed spatial patterns of migration. This could mean increased flows to traditional retirement destinations (e.g., Florida or college towns) but could also reinforce more recently observed trends: perhaps middle class retirees will throng central cities. Certainly *more* older people suggests larger flows to these destinations, potentially increasing the impact this demographic has on these places, especially if smaller and less accustomed to older people.

20.7 Resource Consumption

Anecdotally, at least, age is associated not only with different rates of resource consumption but also with *attitudes* towards resource consumption. Members of the Great Depression generation in Western countries are often stereotyped as conservative towards wasting of personal resources, but more sanguine about environmental resources. Discussions about environmental activism often assume that younger age cohorts are more involved, whether or not the evidence exists to substantiate this assumption. In any case, care for the world around us is assumed to vary with age. Without knowing exactly what to expect, this suggests that population aging will lead to attitudinal shifts towards the natural world, whether resource exhaustion, climate change, or environmental pollution.

Where actual natural resource consumption is concerned, stage of life course is likely important but also interacts with stage of economic development. One standard perspective on the link between economic development and resource consumption or environmental degradation is the environmental Kuznets curve, or EKC. This hypothesis argues that pollution and economic degradation will increase as a country develops economically and that the priority for the government and society will be economic development. At a certain stage of development, though, attention will shift towards conservation and environmental improvement. The literature supporting the existence of the EKC is mixed, but as it directly connects stages of development to resource consumption and degradation it remains a helpful construct.

Demographic change merits a place within this construct as well (Franklin and Ruth 2012). As countries develop, exploiting the natural world around them, they are also likely going through substantial demographic change; fertility, mortality, female labor force participation and—by extension—age structure are all shifting. As areas develop, their age structure moves away from a traditional age pyramid to more of an onion shape and, potentially, eventually something resembling an inverted pyramid. Extending the basic EKC relationship, both stage of development and age structure are likely to impact resource consumption, measured in myriad ways.

Aging populations in developed countries may interact with resource consumption in a variety of ways. Shifts to smaller housing units, especially if in denser urban environments, may reduce energy consumption, as may shifts from automobile dependence to public transportation. On the other hand, once freed of day-to-day life responsibilities, older individuals may decide to travel more, drive larger cars, or simply just buy more stuff.

In developing countries, in contrast, economic development and aging will almost certainly lead to increased resource consumption, simply because the starting base was lower. Increased ability to consume resources will occur in a much different context than previous generations in different locations. Future older cohorts in developing countries will perhaps benefit from enhanced awareness of the value of the environment. Just as we speak of technological leapfrogging in terms of economic development in these areas, perhaps the environment and resource consumption will benefit from leapfrogging over the high-consumption period that has so affected developed countries and thereby the entire world.

20.8 Human Capital

Human capital, either as an individual attribute or a characteristic of regions, is the fuel that drives many migration and regional development models. Generally measured in terms of educational attainment, human capital is an individual quality that increases mobility and, holding other factors constant, also improves individual well-being. For regions, larger human capital stocks may permit growth or development by attracting workers and jobs. Innovation, entrepreneurship, and economic growth—to take just a few popular regional science topics—all depend to a considerable extent on human capital as an input.

At the present state of the world, educational attainment tends to vary by age cohort. In general, younger age groups are more educated (in terms of degrees earned, at any rate) than older groups, reflecting recent changes in economic structures and policy priorities. Table 20.1 shows educational attainment for a rough set of age cohorts for the U.S. as a whole. The increases in educational attainment over time are clear, as is a slight dip for the 35–44 cohort (what is really notable here is the relatively high educational attainment of those benefiting from the U.S. post-World War II GI Bill). The table suggests that age is associated with lower levels of human capital. In terms of *aging*, though, the table suggests something else. As the population ages, it should become more educated; overall human capital stocks should increase but the aged should also become more educated on the whole.

From a regional research perspective, this is of course fascinating, because not all areas are equally aged (see the section on population composition below) and not all will benefit from an equally educated aged population. As pressure increases to expand the length of one's working life (i.e., to retire later), this has all sorts of implications for the spatial distribution of human capital stocks and also for how we define employment. Places with more educated older people may benefit, as will regions that can attract these individuals. More educated individuals may also seek to adapt the contours of 'work,' opting for seasonal or part-time employment where they have the choice. Because of their sheer numbers, it will be interesting to see how shifts in age structure and human capital impact employment.

Table 20.1 Educational attainment in the United States by age, 2000

Age cohort	Percent population with at least a bachelor's degree
25–34	27.54
35–44	25.88
45–64	26.39
65+	15.39

Source: U.S. Census Bureau, Decennial Census

20.9 Population Composition

Aging is itself an element of population composition. In the United States it's generally assessed alongside race and ethnicity, as well as sex, nativity, and socio-economic status. Other countries may categorize their inhabitants along other lines (e.g., social class or religion), but the desire to describe groups using a prescribed set of characteristics is the norm. From a regional perspective, the subject of population composition is intriguing because it varies considerably across space and depends on the spatial unit employed. In the U.S., for example, a frequent topic of discussion is the impending shift to a minority-majority country, as national population composition moves away from its historic, majority white, non-Hispanic nature (see, for instance, Johnson and Lichter 2008 or much of the recent writing of William Frey). At the state, region, or county levels, however, the shift to minorities being in the majority has either already occurred or is seemingly decades away from possibility (Franklin 2014). When the demographic character of places varies significantly with spatial scale, it is easy to see how national-level discourse on race relations, politics, or economic distress might be interpreted differently.

The aging of the population both affects the other population composition characteristics and is affected by them, in turn. Different demographic subgroups have different life expectancies and fertility rates—some live longer, and some produce more children than others. Both of these factors will affect how old the population is, but also the pace at which it will age, if at all. In most countries, because women usually outlive men, old places also tend to be more female. Life expectancy is also associated with socio-economic status and race and ethnicity, so that the mix of groups within an area will likely affect both how old its overall population is at the moment, as well as how quickly it will get older.

What this means for the aging process of areas is that, although countries as a whole may get older, the process at smaller geographic scales is likely to be quite heterogeneous and, once set in motion, may proceed at quite different paces in different areas. In addition, the path dependency associated with aging populations—fewer young people producing children, leading to an even older population—may in turn affect population composition. Or, alternatively, the differential aging of various sub-groups will lead to increased variability over time in the demographic composition of areas.

All of these considerations about age composition and aging, in turn, play into many of the other topics we have discussed above. As the demographic character of a local area changes, it becomes more or less attractive to firms and migrants. It also shifts preferences for amenities and transportation. And, of course, individuals of all ages are free to move, which affects population composition in origins and destinations.

20.10 The Importance of Spatial Scale

One final future regional science research agenda topic that we would like to call to attention is the need to focus creative thought on questions about spatial scale. This is a general concern of ours for the future of the field, as well as one highlighted by our considerations of the impacts of slowing or negative population growth, aging, and age-composition change. Trends at the national level and those at the local may increasingly diverge.

People can no longer be considered fixed geographic points, living within sets of well-bounded, mutually exclusive regions. The very essence of regional science—the region—needs to be rethought in relation to the realities of a globalized world. Many activities that were once done in person by traveling about within fairly narrowly constrained time-space prisms (as so aptly set forth by Hagerstrand 1970), now take place in virtually frictionless cyberspace. Interestingly, as the world becomes ever more globalized, the local realm seems to be being redefined and gaining new importance (Plane 2016).

At different life stages, people are now making spatial choices very unlike those that were possible, or thought desirable, during the initial years of regional science. Data from the last two U.S. censuses show young adults flocking to urban cores, as are some empty-nesters (Wilson et al. 2012). Even the notion of a single, usual place of residence is becoming problematic, as increasing numbers may choose or need to live in multiple locations, whether in seasonal, weekly, or other patterns throughout the year. U.S. presidential candidate, John McCain, when quizzed by a reporter during the 2008 election campaign about which of his several residences he considered home, replied that he would have to check with his staff!

20.11 Conclusions

Times change. Fashions change. Often, for researchers with a demographic bent, it can be difficult to ascertain whether observed changes are due to age, period, or cohort effects. In the case of migration and aging, for example, is it the case that we observe a bump in retiree migration because that's just what *older* people do? Or is it because of the way economic and social life are *currently* organized? Or, another possibility, are observed patterns a function of *these* older people: these folks who had the benefit of being born in the halcyon years immediately following the second world war? The answer is, of course, that all three effects are important, but the relative importance of each makes predicting future patterns tricky. As usual, we are only likely to understand what is happening after it has already occurred.

In this chapter, we have suggested areas where we believe regional science, and regional research in general, will be affected by aging and an older population. We have done our best to make some prognostications and have also, we believe, suggested some fruitful avenues of research for those interested in exploring how aging already matters. For regional science to remain in touch with real world behavior, and for research findings to have any resemblance to real world behavior and policy, our theories, models, and measures all require a closer look for assumptions we make about the human actors in our models. It is, after all, important that, as each of us individuals inevitably climbs the hill towards obsolescence (one year at a time), our field is ever renewing and rejuvenating itself, preparing for what lies over the hill.

References

Bitter C, Plane DA (2012) Housing markets, the life cycle, and migration up and down the urban hierarchy. In: Clapham DF, Clark WAV, Gibb K (eds) Handbook of housing studies. Sage, Los Angeles, pp 295–311

Brown DL, Wardwell JM (1980) New directions in urban-rural migration: the population turnaround in rural America. Elsevier, Amsterdam

Champion AG (1988) The reversal of migration turnaround: resumption of traditional trends? Int Reg Sci Rev 11:253–260

Easterlin RA (1980) Birth and fortune: the impact of numbers on personal welfare. University of Chicago Press, Chicago

Franklin RS (2014) An examination of the geography of population composition and change in the United States, 2000–2010: insights from geographical indices and a shift–share analysis. Popul Space Place 20:18–36

Franklin RS, Ruth M (2012) Growing up and cleaning up: the environmental Kuznets curve redux. Appl Geogr 32:29–39

Hagerstrand T (1970) What about people in regional science? Pap Reg Sci Assoc 24:7–21

Johnson KM, Lichter DT (2008) Natural increase: a new source of population growth in emerging Hispanic destinations in the United States. Popul Dev Rev 34:327–346

Long L, Nucci A (1997) The 'clean break' revisited: is US population again deconcentrating? Environ Plann A 29:355–366

Mulligan G, Carruthers J (2011) Amenities, quality of life, and regional development. In: Marans RW, Stimson RJ (eds) Investigating quality of urban life. Springer, Dordrecht, pp 107–133

Partridge MD, Ali K, Olfert MR (2010) Rural-to-urban commuting: three degrees of integration. Growth Chang 41:303–335

Plane DA (2000) The role of regional science in migration research. Rev Reg Stud 29:79–83

Plane DA (2012) What about aging in regional science? Ann Reg Sci 48:469–483

Plane DA (2016) Regional science in a more/less world. Presidential address delivered at the annual meeting of the Western Regional Science Association, Waikoloa, Hawaii, Feb 2016

Plane DA, Bitter C (1997) The role of migration research in regional science. Pap Reg Sci 76:133–153

Plane DA, Heins F (2003) Age articulation of U.S. Inter-metropolitan migration. Ann Reg Sci 37:107–130

Plane DA, Jurjevich JR (2009) The patterns and repercussions of age-articulated migration. Prof Geogr 61:4–20

Plane DA, Rogerson PA (1991a) Tracking the baby boom, the baby bust, and the echo generations: how age composition regulates US migration. Prof Geogr 43:416–430

Plane DA, Rogerson PA (1991b) The ten commandments of migration research. In: Boyce DE, Nijkamp P, Shefer D (eds) Regional science: retrospect and prospect. Springer, Berlin, pp 15–42

Plane DA, Henrie CJ, Perry MJ (2005) Migration up and down the urban hierarchy and across the life course. Proc Nat Acad Sci U S A 102:15313–15318

Population Reference Bureau (2014) 2014 world population data sheet. PRB, Washington DC

Ruth M, Franklin RS (2014) Livability for all? Conceptual limits and practical implications. Appl Geogr 49:8–23

United Nations (2015) http://www.un.org/en/development/desa/population/migration/data/empirical2/docs/migflows2015documentation. Accessed 13 Jan 2013

Vining DR, Kontuly T (1978) Population dispersal from major metropolitan regions: an international comparison. Int Reg Sci Rev 3:59–73

Whisler RL, Waldorf BS, Mulligan GF, Plane DA (2008) Quality of life and the migration of the college-educated: a life-course approach. Growth Chang 39:58–94

Wilson SG, Plane DA, Mackun PJ, Fischetti TR, Goworowska J (2012) Patterns of metropolitan and micropolitan population change 2000 to 2010 (2010 census special report C2010SR-01). US Government Printing Office, Washington DC

Rachel S. Franklin is associate professor (research) of Population Studies at Brown University. Her primary research interests are population distribution and change, along with internal migration. She is Associate Director of the Spatial Structures in the Social Sciences (S4) initiative at Brown and has an appointment in the Population Studies and Training Center. Dr. Franklin earned the PhD in geography from the University of Arizona in 2004.

David A. Plane is Professor, School of Geography and Development, University of Arizona. His primary research interests are population distribution and urban development, U.S. migration and settlement, the role of the life course in shaping individuals' activity patterns, and methods for modeling temporal change in spatial interaction systems. He has previously worked as a statistician and visiting research fellow at that U.S. Census Bureau. Dr. Plane earned the Ph.D. in regional science from the University of Pennsylvania in 1981.