INTERNATIONAL STUDIES IN POPULATION

NAVIGATING TIME AND SPACE IN POPULATION STUDIES

Edited by
Myron P. Gutmann
Glenn D. Deane
Emily R. Merchant
Kenneth M. Sylvester

International Union for the Scientific Study of Population
Union Internationale pour l'Étude Scientifique de la Population
Unión Internacional para el Estudio Científico de la Población
国际人口科学研究联盟
الاتحاد الدولى للدراسات العلمية للسكان

Springer

Navigating Time and Space in Population Studies

International Studies in Population

Volume 9

The International Union for the Scientific Study of Population (IUSSP)

The IUSSP is an international association of researchers, teachers, policy makers and others from diverse disciplines set up to advance the field of demography and promote scientific exchange. It also seeks to foster relations between persons engaged in the study of demography and population trends in all countries of the world and to disseminate knowledge about their determinants and consequences. The members of the IUSSP scientific group responsible for this book were chosen for their scientific expertise. This book was reviewed by a group other than the authors. While the IUSSP endeavours to assure the accuracy and objectivity of its scientific work, the conclusions and interpretations in IUSSP publications are those of the authors.

International Studies in Population (ISIP) is the outcome of an agreement concluded by the IUSSP and Springer in 2004. The joint series covers the broad range of work carried out by IUSSP and includes material presented at seminars organized by the IUSSP. The scientific directions of the IUSSP are set by the IUSSP Council elected by the membership and composed of:

For other titles published in this series, go to
www.springer.com/series/6944

Navigating Time and Space in Population Studies

Edited by

MYRON P. GUTMANN
University of Michigan, USA

GLENN D. DEANE
University at Albany, State University of New York, USA

EMILY R. MERCHANT
University of Michigan, USA

KENNETH M. SYLVESTER
University of Michigan, USA

Springer

Editors
Myron P. Gutmann
ICPSR
University of Michigan
PO Box 1248
Ann Arbor, MI 48106-1248, USA
gutmann@umich.edu

Glenn D. Deane
Department of Sociology
University at Albany
State University of New York
Albany, NY 12222, USA
g.deane@albany.edu

Emily R. Merchant
ICPSR
University of Michigan
PO Box 1248
Ann Arbor, MI 48106-1248, USA
eklanche@umich.edu

Kenneth M. Sylvester
ICPSR
University of Michigan
PO Box 1248
Ann Arbor, MI 48106-1248, USA
kenms@umich.edu

ISBN 978-94-007-0067-3 e-ISBN 978-94-007-0068-0
DOI 10.1007/978-94-007-0068-0
Springer Dordrecht Heidelberg London New York

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The International Union for the Scientific Study of Population's Panel on Historical Demography applies a historical perspective, such as the application of methods that allow analysis of the role of space and time in population, to promote work of contemporary relevance. New approaches to space and time in demographic analysis are now contributing in important ways to understanding how populations change. The application of these methods to historical populations helps evaluate theory and establish means for understanding demographic transformations in the contemporary world. The Panel organizes workshops that draw together scientists at the forefront of these issues to compare and discuss important new work.

Navigating Time and Space in Population Studies is intended to extend the discussions held at a *Workshop on Space and Time* cosponsored by the International Union for the Scientific Study of Population, to a broader community of demographers, geographers and historians. The contributions in this volume bring together the work of researchers from many disciplines to concentrate on the factors that accelerate and retard the spatial diffusion processes that are the theoretical norm in much demographic analysis. They show that the factors at work are complex and that the methods that are useful for understanding change across space and through time are rich and varied. Our hope is that these chapters will spur additional research that will further improve our understanding of these very important questions.

Acknowledgements

The present set of papers, focusing on the ways that time and space contribute to understanding demographic change in historical and contemporary research, were selected from the contributions to a *Workshop on Space and Time* in the fall of 2006 that was jointly sponsored by three organizations: the Scientific Committee on Historical Demography of the International Union for the Scientific Study of Population, the Inter-university Consortium for Political and Social Research and the Minnesota Population Center. The organizing committee included Myron Gutmann (University of Michigan), Glenn D. Deane University at Albany, State University of New York, Steven Ruggles (University of Minnesota) and Kenneth M. Sylvester (University of Michigan). We are grateful to all three institutions for the generous support they provided for the seminar. We are also grateful to Bree Gunter of ICPSR and to the staff of the Minnesota Population Center for logistical support. Finally, as this volume was nearing completion, the National Science Foundation of the United States provided support to Myron Gutmann for his work on editing this volume. Any opinion, finding and conclusions or recommendations expressed in this material are those of Myron Gutmann and do not necessarily reflect the views of the National Science Foundation.

Ann Arbor, MI — Myron P. Gutmann
Albany, NY — Glenn D. Deane
Ann Arbor, MI — Emily R. Merchant
Ann Arbor, MI — Kenneth M. Sylvester

Contents

Contributors

Renato M. Assunção Department of Statistics, Universidade Federal de Minas Gerais, Minas Gerais, Brazil, assuncao@est.ufmg.br

Dominique Barbero Société FCA-SIG, F-69410, Champagne-au-Mont-D'or, France, dominique.barbero@voila.fr

Andrew A. Beveridge Queens College and Graduate Center, City University of New York, New York, NY, USA, andy@socialexplorer.com

Guy Brunet UMR5190-CNRS Laboratoire de Recherches Historique Rhône-Alpes, Institut des Sciences de l'Homme, Université Lyon 2. 14, F-69363, Lyon, France, Guy.Brunet@univ-lyon2.fr

Katherine J. Curtis Department of Community and Environmental Sociology, University of Wisconsin-Madison, Madison, WI 53706, USA, kcurtis@ssc.wisc.edu

Pierre Darlu UMR7206 Eco-anthropologie et ethnobiologie, MNHN-CNRS-Université Denis Diderot, 75231 Paris Cedex 05, France, darlu@mnhn.fr

Glenn D. Deane Department of Sociology, University at Albany, State University of New York, Albany, NY 12222, USA, g.deane@albany.edu

Peter Ekamper Netherlands Interdisciplinary Demographic Institute (NIDI/KNAW), The Hague, The Netherlands, ekamper@nidi.nl

Hani A. Guend Université du Québec, INRS – Urbanisation Culture Société, Montréal, QC, Canada H2X 1E3, hani.guend@ucs.inrs.ca

Myron P. Gutmann Inter-university Consortium for Political and Social Research, University of Michigan, Ann Arbor, MI 48106-1248, USA, gutmann@umich.edu

J. David Hacker Department of History, State University of New York at Binghamton, Binghamton, NY 13902-6000, USA, hacker@binghamton.edu

Michael R. Haines Department of Economics, Colgate University, Hamilton, NY 13346, USA, mhaines@mail.colgate.edu

Kees Mandemakers International Institute of Social History, Amsterdam, The Netherlands, kma@iisg.nl

Emily R. Merchant Inter-university Consortium for Political and Social Research, University of Michigan, Ann Arbor, MI 48106-1248, USA, eklanche@umich.edu

Joseph E. Potter Department of Sociology, Population Research Center, University of Texas at Austin, Austin, TX 78712, USA, joe@prc.utexas.edu

Carl P. Schmertmann Department of Economics, Center for the Study of Population, Florida State University, Tallahassee, FL 32306-4063, USA, schmertmann@fsu.edu

Kenneth M. Sylvester Inter-university Consortium for Political and Social Research, University of Michigan, Ann Arbor, MI 48106-1248, USA, kenms@umich.edu

Frans van Poppel Netherlands Interdisciplinary Demographic Institute (NIDI/KNAW), The Hague, The Netherlands, poppel@nidi.nl

Kristine Witkowski Inter-university Consortium for Political and Social Research, University of Michigan, Ann Arbor, MI 48106-1248, USA, kwitkow@umich.edu

Chapter 1
Introduction

Myron P. Gutmann, Glenn D. Deane, Emily R. Merchant, and Kenneth M. Sylvester

People in Time and Space

Demography's task is to document, analyse and theorize population dynamics, the ways in which spatially defined populations grow, shrink or redistribute themselves over time. Historical demographers, in particular, strive to model population processes that spread across space and unfold over time, and this volume explores innovative ways in which demographers are doing so. The eight chapters that follow utilize newly available sources of historical data and implement novel methods to assess the role of time and space and their intersection in core demographic, historical and sociological theories. The theories, methods and substantive findings presented here also offer models and insights of value to demographers more broadly; space and time are critical elements in all population studies, as important for scholars attempting to understand current populations, or to make projections into the future, as for those focusing on the past.

Spatial and temporal analysis feels fresh and current, yet has a long history in the social sciences. As Voss (2007) points out, spatio-temporally explicit research agendas predate contemporary interest in all of the core topics of population research. Perhaps the most widely recognized and iconic example of the value of spatial and temporal analysis was provided by John Snow during the 1854 London cholera epidemic. His map of cholera deaths revealed clustering around the Broad Street water pump, supporting his theory that cholera was spread through the contamination of water (Snow, 1855). In another example, the dramatic maps published by the U.S. Census Bureau about population change prior to the 1890 census informed Frederick Jackson Turner's famous declaration that the American frontier had closed (Turner, 1893).

Demographers, too, have long considered space and time simultaneously. At the very beginning of modern population studies in the nineteenth century, analysts revealed that population processes operate through time in a spatial context. The

M.P. Gutmann (✉)
Inter-university Consortium for Political and Social Research, University of Michigan, Ann Arbor, MI 48106-1248, USA
e-mail: gutmann@umich.edu

M.P. Gutmann et al. (eds.), *Navigating Time and Space in Population Studies*, International Studies in Population 9, DOI 10.1007/978-94-007-0068-0_1,

demographic transition, a major trope in contemporary and historical population analysis since the 1940s (see, for example, Davis (1945)), has benefited from attention devoted to the pace and location of change. Nowhere is the mix of the spatial and temporal clearer than in the Princeton European Fertility Project, whose capstone volume (Coale & Watkins, 1986) relies on the geography of population processes to tell the story of changing fertility. Indeed, the notion of diffusion of fertility regulation continues to be salient even years after the publication of the Coale and Watkins summation (Watkins & Danzi, 1995; Chapter 2 by Haines and Hacker, this volume).

Nonetheless, social scientists' interest in 'putting people into place' (Entwisle, 2007) – their attention to how social processes are embedded within temporal and spatial contexts – has grown dramatically in recent years, from both theoretical and applied empirical perspectives. In social theory, the roles of time and space are changing. The practice of equating variation over space with temporal progress along a universal trajectory has been denounced by sociologists, geographers and historians (Thornton, 2001, 2005; Massey, 2005; McClintock, 1995) as what Thornton (2001, 2005) critically calls 'reading history sideways'. Scholars now recognize the necessity of analysing space and time together, both because these dimensions are inextricable from one another and because neither can be reduced to the other. Such theoretical notions as 'social interaction', 'social capital', 'diffusion and contagion' and 'spatial mismatch', for example, require explicit specification in spatial and temporal terms.

Likewise, the quest for analytic models of space-time dependence is not new, but social scientists continue to seek ways to overcome the inherent limit condition to any proper solution, that 'clarity in one dimension demands restriction in the other' (Knox, 1964). Elegant and powerful methods of quantifying the proximity of events in space and time have been available since well before the advent of the geographic information systems that breathed new life into spatial social science. Signature examples, resulting from efforts to define epidemicity, are found in the two-by-two cross-classification of nearness (or not) in time and in space proposed by John Knox (1964) and its generalization to a covariance measure for continuous distributions of nearness by Nathan Mantel (1967). These measures of space-time association, in turn, owe a debt to the still earlier successes of statisticians who developed the two-dimensional spatial analogues to univariate descriptive statistics known now as 'centrographic statistics' (cf. Lefever, 1926; Furfey, 1927).

Why, then, does the interest in space-time analysis remain fresh and why have the measures just described not found broader application? The answer appears to lie in the conceptual and empirical divide between research questions that ask where and when did an event occur versus those that ask whether the quantity of something is systematically related to the quantity that is near in space and prior in time. The former correspond to point patterns, which were the subject of centrographic statistics and the associational measures proposed by Knox and Mantel. When quantity, rather than occurrence, frames the research question, the measurement of dependence, or autocorrelation, is readily resolved in either time (e.g. Durbin & Watson, 1950, 1951) or space (e.g. Moran, 1950; Geary, 1954), but our ability to separate and quantify both simultaneously is almost always challenged by too little information, a problem known in econometrics as 'underidentification'. The problem is that we measure spatial dependence by the contemporary coincidence of quantity similarity with locational similarity. This 'coincidence in values-locations' is therefore already defined in temporal terms: it is not only an instant coincidence, but also the product of what happened in the past and what has subsequently spread through geographic space (Anselin, 2001a). Identification of true spatial dependence and true temporal dependence (and the separation of the two) therefore requires additional leverage. The authors of the following chapters find this leverage in a variety of ways: by introducing a time-lag in the spatial dependence term, by adapting the point pattern approach to achieve insight into questions about quantitative changes in locations over time or by gaining clarity over the spatial dimension by restricting the temporal dimension.

Such methodological advances have given historical demographers a rich new set of tools with which to re-examine received ideas and pose new questions. This volume is the result of an effort to assess recent contributions and advances, and to demonstrate how valuable this new research will be for re-imagining historical and contemporary questions. The eight chapters published here were among a larger group of papers presented at the Workshop on Space and Time organized by the International Union for the Scientific Study of Population in Minneapolis, MN, October 31–November 1, 2006 at the Minnesota Population Center (MPC) of the University of Minnesota. The workshop was sponsored by the MPC, the Committee on Historical Demography of the International Union for the Scientific Study of Population and the Inter-university Consortium for Political and Social Research (ICPSR). Three of the editors of this volume, along with Steven Ruggles of the MPC, served as scientific committee members and conveners.

The selections included in this volume cover a variety of core demographic topics and explore population change from the eighteenth to the twentieth century in the United States, Latin America, Europe and the Islamic world. The first three chapters examine mortality and fertility in the framework of demographic transition theory, and are followed by three works on migration. The volume concludes with two chapters demonstrating the broader sociological application of spatially and temporally explicit concepts and methods. The new data, innovative methods and substantive results reported here demonstrate valuable recent advances in spatio-temporal approaches to historical demographic research and suggest their striking potential for the future. This introduction begins by examining the theoretical bases on which the chapters are written. We emphasize the long-standing importance of the issues that the authors consider, arguing that the fundamental contribution of these papers is their development of new ways to analyse phenomena that have been at the heart of population studies for a long time. Our second section discusses the innovative data used in the chapters and the ways that such data have recently become available. In the third section we focus on the methods implemented by the authors to analyse their data and to answer their questions. The wide variety of methods presented in this volume illustrates the many new directions in which studies of space and time will reach in coming years. Finally, we turn to substantive findings, presenting each chapter in turn and exploring the ways they come together to provide new lenses through which to view space and time in historical population studies.

Theoretical Bases

The papers in this volume describe, test and challenge a range of theoretical formulations that are well established in demographic scholarship and, in so doing, demonstrate the power of these new approaches to historical populations. One of the most important accomplishments of the work represented by these chapters is their ability to bring new light to well-established theory by expanding upon or combining existing approaches. Some add spatial elements to traditionally temporal theories; others add temporal elements to traditionally spatial theories.

At the heart of demography – especially studies of the redistribution of population or changes in rates and determinants of population growth – is the notion of spatial diffusion: people and their demographic behaviours move through space at a regular pace over time. This core theory explicitly or implicitly animates all of the work presented in this volume. Spatial diffusion suggests that people and patterns of demographic change can flow across space in different ways, as waves spilling over barriers or discrete nodes in dense social networks, modifying other patterns of behaviour they encounter. The space over which people, ideas and behaviours diffuse is neither empty nor inert (Massey, 2005): people spread over uneven terrain that may already be inhabited; ideas spread among people who may hold competing beliefs; new norms and behaviours spread across societies that may be more or less able to support them. Diffusion is an explicitly spatial process, but also occurs over time. The social, economic, political and cultural factors that shape it also exist in and change over time. At any given moment, however, diffusion processes are identified and assessed through the location and measurement of patterns of spatial clustering.

The concept of diffusion can add a spatial element to traditionally temporal theories, such as the demographic transition. This theory posits universal change over time, from a high-mortality/high-fertility equilibrium to a low-mortality/low-fertility equilibrium. Scholars associated with the Princeton European Fertility Project began to explore spatial dimensions of the demographic transition when structural factors failed to fully explain fertility decline (Coale & Watkins, 1986). Diffusion suggests that, rather than making independent appearances in various locales when certain socio-economic thresholds are reached, end-state behaviours may start in one or more locations and then spread gradually across space and through time, reaching places that are farther away later in time. If that were the case, and there were few barriers, it would be easy to predict when in time that end-state behaviour would reach a given location. Three of the following chapters complicate this view by applying the concept of diffusion to the study of transitions between demographic regimes. They contribute to both demographic transition and spatial diffusion theory by asking whether demographic outcomes tend to spread from person to person (Schmertmann, Potter and Assunção); by exploring structural factors that may have impeded or accelerated the spatial diffusion of demographic practices (Haines and Hacker); and by situating demographic behaviour in political and ideological – in addition to geographical – space (Guend).

Classic theories of migration explicitly involve diffusion over and concentration in space: on the one hand, the dispersion of people from more to less densely settled areas (Turner, 1893); on the other, the attraction of people to established centres of population (Ravenstein, 1885). Although time often remains implicit in theories of migration, this dimension is critical, both because the movement of individuals and populations occurs over time and because migration is thought to undergo secular increase and acceleration as a result of modernization and industrialization (Lee, 1966). The three chapters in this volume dealing with migration bring time to the foreground and offer historically contextualized ways to study and theorize migration and the resulting diffusion and clustering of population in space. The authors of these chapters add an explicit temporal dimension to theories of migration by assessing barriers to the diffusion of historical populations (Darlu, Brunet and Barbero); by assessing the ways in which distance and its meaning change over time in a modernizing and nationalizing context (Ekamper, Van Poppel and Mandemakers); and by examining the effects of the environmental, economic and social features of space on the temporality and shape of population diffusion (Gutmann, Deane and Witkowski).

Social scientists typically aspire to produce ‘timeless’ and ‘placeless’ theories and inferences that can be generalized across location and era (Griffin, 1992), while historians seek to demonstrate the contingency of social processes by contextualizing them in time and place. The chapters in this volume unite the two approaches; by analysing the role of time and place in social processes, our contributors seek to further develop well-established social theories. These approaches have also proven useful for scholars working outside the core demographic themes of mortality, fertility and migration. The final chapters test propositions about urban growth in specific U.S. cities at various points in the twentieth century (Beveridge) and ask how spatial patterns of social inequality change with the dominant regime in a colonial setting (Curtis). The work presented in this volume goes well beyond the application of general social theories to particular times and places. The authors, instead, ask how time and space function within demographic and social theories, while taking seriously the ways in which ideational change can confound the simplicity of diffusion, and while paying close attention to the role of social, economic and political structures in advancing or impeding change over time and diffusion over space. The tension between the smooth spread of people and concepts and the existence of barriers to diffusion give life to the work in the book.

Data for Analyses

In a masterful survey of the renewed interest in spatial demography, Voss (2007) emphasizes the changing role of data at different units of analysis in the transformation of demographic research. He points out that, since the 1970s, individual-level data have become increasingly available while high-performance computing environments have offered improved ways to analyse them. Together, these developments have engendered extremely valuable research. However, these individual-level data are not as well-suited to spatial analysis as the aggregate-level data that formed the basis for most earlier demographic analysis. Only in the last decade have new approaches to aggregated data and newly constructed data sets significantly improved researchers' ability to analyse the role of space and time in determining demographic change in the past. Producing these data sets is often a straightforward matter of transcribing the contents of books of census results that have been published in great quantity over the past two centuries. In other instances, utility is added by merging or matching data from multiple sources, as Gutmann (2005a, 2005b) has done for the population and agriculture of the Great Plains or as Haines (2004) has done for the counties of the United States. Projects to produce improved and more useful data sets by harmonizing previously-separate raw data sources usually focus on the counties of the U.S. and its territories and have been undertaken at the University of Minnesota, Colgate University, the University of Wisconsin and ICPSR at the University of Michigan. These improved data are supplemented by increasingly useful information about the location of the spatial units whose attributes are represented in the data and by the new and sometimes reconfigured analytical methods that we describe in the next section of this Introduction.

In addition to some of the improved or newly available data sets created for the United States and its territories, contributors to this volume make effective use of spatially referenced population and mortality data for microregions of Brazil and national level data for Muslim countries. They have also found creative ways to exploit individual-level data: for example, surnames are extremely valuable in linking the geographic origins and destinations of migrants, while marriage documents provide an extremely rich way to show how patterns of migration changed over time in concert with deep-seated transformations in society, economy, political life and infrastructure. These data, when combined with new ways of thinking about underlying theory, facilitate the use of emerging methods to study important demographic issues. A large fraction of the aggregate- and individual-level data utilized in this volume is now available through the University of Minnesota's National Historical GIS repository (http://www.nhgis.org) and IPUMS-International (https://international.ipums.org/international/). These collections also include data at other spatial scales, including census tract data and an extremely accurate and useful set of digital geographical data that locate political and administrative units in coordinate space.

Innovative Methods

Although historical demography has been applauded as one of the few fields of historical research that has recognized and embraced the utility of traditional multivariate techniques (Griffin & van der Linden, 1999), it is not hard to recognize how differences in overt disciplinary objectives have constrained intellectual progress. Social scientists typically work hard to escape the spatial and temporal constraints of their studies. By definition and intent, most do not seriously ground either the theories they use or the analyses they perform in the local (temporal and spatial) contexts of the data-generating processes they study. On the other hand, most historians, including social historians, are typically indifferent to the conventions of formal social science, particularly those calling for the development and application of codified theory and the use of formal inferential techniques and methodologies. A profound skepticism about the power of formal social science to elucidate historical processes has made historians reluctant to concern themselves with methodological advances in the social sciences and to explore how emerging techniques can be fruitfully applied to historical research (Griffin & van der Linden, 1999: pp. 3–8). Given the interest of social historians in 'localizing' social processes, this is an unfortunate aversion. Recently, however, innovative approaches to historical theorizing give reason to believe that thinking of events in the past as parts of a process moving through time and across space is becoming more commonplace. A new consensus recognizes that events occurring at a given point in time are affected by earlier events, and that events occurring in one location are affected by similar events in a neighbouring area. Although the traditional methodological paradigm assumes the independence of analytic units, this new consensus tells us that interdependence (temporal and spatial autocorrelation) is the rule, requiring an explicit modelling of two-way or multi-way interactions over space.

In recent years, rapid advances in the technology of spatial data handling – such as geographic information systems, software for spatial data analysis and the explosion in available geo-coded data – have created the infrastructure for geographical analysis to become part of the standard technical toolbox of historical demographers. Our intent in publishing this collection of essays is to demonstrate the utility and promise of explicit attention to space-time dependencies in historical inquiry. Although the cases addressed by our contributors are significant on their own terms and are of clear relevance to historical demographers, the purpose of this collection is to display in an accessible manner when, why and how the application of various formal methods generates deeper, more satisfying explanations and interpretations of historical processes. To this end, the common denominator of our eight essays is their use of distribution maps, which are the natural graphical representation of geographic data.

Maps are the oldest coordinate graphs, dating back to ancient Egypt. All maps are 'distribution maps' in the sense that it is impossible to represent relative geographical location without showing the distribution of something (Robinson & Sale, 1969). Statistical maps, which illustrate the geographic distribution of quantities or of events, are substantially newer, but are still among the earliest statistical graphs. The fundamental objective of qualitative distribution maps is to present the relative geographical location of spatially rendered phenomena. Statistical maps typically graph the distribution of a variable at one point in time, but can also illustrate the distribution of change in a variable between two points in time, or can be arrayed in temporal sequence, elegantly summarizing both spatial and temporal variation. Our contributors repeatedly show how heuristically useful these most simple maps can be through the presentation of temporal sequences of maps as evidence (or counter evidence) of space-time interaction.

The most common statistical map is the choropleth map, which uses colour or shading to represent relative quantities that are tied to enumeration districts on the ground. Alternatively, the graduated symbol map employs differentially sized symbols to identify levels of a given variable at specific locations. The graduated circle is one of the oldest of the quantitative point symbols used for statistical representation. By quantifying the relative magnitude in point locations, these maps efficiently identify geographic outliers and represent temporal stability or change. Both types of maps improve upon tabular statistics, which are very convenient for many purposes, but array data alphabetically rather than geographically. It is hard to imagine an alternative to these statistical maps that could be more universally intuitive or informative for comparing magnitudes in various places. The majority of our contributors employ this method of visualization.

The visual inspection of distribution maps, however, has long been recognized by cartographers as unreliable in terms of detecting spatial clusters and patterns in the data. Human perception tends to be biased towards finding patterns, even in spatially random data. It is by now widely accepted that the map as a visualization device cannot be left to unconstrained human interpretation and needs to be augmented with tools to formally assess pattern and structure. Many of our contributors therefore apply exploratory spatial data analysis (ESDA) and its space-time extension, exploratory space-time data analysis (ESTDA). ESDA is a subset of exploratory data analysis (EDA) that focuses on the distinguishing characteristics of spatial data and, specifically, on spatial autocorrelation and spatial heterogeneity (Anselin, 1994). The point of departure in ESDA is the same as in EDA, namely to use descriptive and graphical statistical tools to discover patterns in data and suggest hypotheses while imposing as little prior structure as possible (Tukey, 1977). More specifically, ESDA is a collection of techniques to describe and visualize spatial distributions, identify atypical locations or spatial outliers, discover patterns of spatial association, clusters or hot spots and suggest spatial regimes or other forms of spatial heterogeneity. Central to this endeavour is the notion of spatial autocorrelation or spatial association (Cliff & Ord, 1981; Upton & Fingleton, 1985).

The most popular statistic of global spatial autocorrelation is undoubtedly Moran's I. Moran's I is simply the spatial analogue of a correlation coefficient; it summarizes the strength and direction of neighbouring values of a variable, Y. If we imagine a scatter plot created by the bivariate distribution of the value of Y in each locale paired with the average value of its neighbours, we can identify the form of spatial association affecting each locale according to its position in each of the four quadrants of the scatter plot: higher than average value of Y whose neighbours are also higher than average (the upper right quadrant), lower than average value of Y paired with neighbours whose average value is also lower than average (the lower left quadrant), higher than average value of Y whose neighbours are lower than average (the lower right quadrant) and lower than average value of Y whose neighbours are higher than average (the upper left quadrant). This scatter plot is a 'local indicator of spatial association', or LISA (Anselin, 1995). When mapped, it effectively diagnoses the location and form of spatial clusters. It is also useful to consider the bivariate LISA as a simple extension of the univariate LISA. Whereas the numerator of the univariate LISA gives the covariation of the (standardized) value of a variable, *Y*, at a given locale with the average value of that locale's neighbours, the numerator of the bivariate LISA gives the covariation of the (standardized) value of *Y* at a given locale with the average of neighbouring values of another variable, *X*. Another popular global statistic of spatial association is Getis and Ord's *G*. Because *G* only measures overall spatial clustering in the high-high and low-low quadrants described above, its interpretation is restricted to positive spatial autocorrelation. Specifically, if *G* is larger than its expected value, then the overall distribution of *Y* shows a preponderance of high-valued clusters. A prevalence of low-valued clusters is indicated when *G* is smaller than its expected value (Getis & Ord, 1992).

The modes of spatial analysis described so far, from distribution maps to measures of spatial association, have been facilitated by recent advances in geographic information systems technology, which map data onto the spatial locations they describe. Locating data in space, however, can impose temporal constraints on analysis, and measures of spatial autocorrelation typically capture patterns only at one point in time. The contributors to this volume, therefore, draw on innovative methods of statistically assessing the temporal dimension of spatial association in a variable of interest. For example, an effective extension of ESDA is found when the local statistic is converted into an estimator of first-order space-time dependence by combining the (spatial) distribution of *X* at time *t* with the spatially lagged distribution of *X* at time *t-1* (Chasco & López, 2004). In this instance, the LISA gives the covariation between *X* at a given locale at a given period in time with the average neighbouring *X* at a previous period in time.

Many of our contributors use these global and local statistics to formally assess patterns of geographic dispersion. However, ESDA/ESDTA as such is not an end in itself. Rather, the visualization and quantification of spatial and temporal regularity provide our contributors with a rigorous basis for model specification and hypothesis testing. For example, in accordance with Anselin, Bera, Florax and Yoon (1996), Moran's *I* will quantify evidence of residual spatial dependence in regression models and then help adjudicate evidence in support of spatial-lag vs. spatial-error model specifications. When spatial dependence is produced by the existence of a spatial interaction, spatial spillovers or spatial hierarchies in the endogenous variable of a spatial regression model, the spatial-lag is a viable solution (Anselin, 2001b). In space-time regression models, this simultaneous model of spatial dependence is problematic if our core research interest is in trying to answer questions that require separation of spatial and temporal dependencies.

Many of our contributors solve this problem by purging the regression models of correlated error through the direct incorporation of the sources of temporal or spatial dependence as independent variables. Evidence of autocorrelation in the residuals of a regression model may indicate that the model specification is incomplete – that relevant explanatory variables are missing. Thus, the inclusion of these omitted variables would fully explain – or eliminate – the autocorrelation. The following chapters demonstrate the wide choice of additional variables available to analysts. In some instances, the expanded model results from the incorporation of spatially or temporally lagged *X*s (or both), or by utilizing trend surface models, which incorporate spatial dependence via the spherical coordinate system of longitude and latitude. Trend surface and, more generally, spatial expansion models spatially detrend by regressing an outcome on location and possibly second-order and multiplicative functions of location. However, spatial expansion and spatial-temporal-lags are naïve in the sense that they offer no clear theoretical underpinning for the data generating process. Accordingly, the truly enduring contribution to social theory is often found in concluding speculation upon those additional explanatory variables.

Outline of Chapters and Their Valuable Findings

Most demographic processes are by definition both spatial and temporal. They interest us because they occur at some place and across space, and they occur at some time and over time. The originality and strength of the chapters in this book come from the authors' ability to find new ways to understand the spatial and temporal in historical demographic research. Because many analysts believe that changes in behaviours are largely a result of changes in ideas that have to be communicated from individual to individual, it makes sense to think that demographic and socio-economic change will follow spatial patterns of communication. Understanding the importance of diffusion, and the ways that our authors look at it, gives us a unifying frame for thinking about their findings and the ways that those findings come together in a series of new conclusions about the role of space and time in historical demographic research.

The first two chapters assess the extent to which the diffusion of ideas and behaviours contributed to the two interlocking elements of the demographic transition: mortality decline and fertility decline. Focusing on sustained reduction in child mortality as a signal of the onset of the demographic transition in Brazil's microregions, **Schmertmann, Potter and Assunção** employ the language of contagion, suggesting that behaviours leading to low mortality rates may have spread from person to person through space. A diffusion process of this type would lead to spatio-temporal clustering of mortality rates that cannot be explained by variables measuring socio-economic conditions and the provision of sanitary and health services. The authors, therefore, adopt and adapt a model originating in epidemiology, John Knox's (1964) cross-classification of space-time interaction. Generalizing this test to proportional hazards models allows the authors to test the null hypothesis that mortality decline was independent of its spatial and temporal context, while controlling for relevant structural and ideational covariates. Schmertmann, Potter and Assunção then use their model to predict the timing of the mortality transition and assess the possible clustering of Brazilian microregions that experienced mortality change earlier than expected, which could indicate the interpersonal transmission of low-mortality behaviours over space. That is their hypothesis for the early transition in Rio Grande do Sul, located near low-mortality Uruguay. In the north, however, where low population density and great distances make the prospect of cultural transmission of ideas unlikely, the same factors may have served as barriers to the transmission of illness. Like so many of the chapters in this volume, their research provides important answers, but also identifies significant research to be done in the future.

Turning to the spatial diffusion of fertility decline, **Haines and Hacker** take as a starting point the readily observable east-west gradient from low to high fertility levels in the United States between 1800 and 1860, which they illustrate with a time series of quantile maps of county level child–woman ratios. They proceed by contrasting and combining a spatial diffusion model with other theoretical approaches, many well established in the literature (Carter, Ransom, & Sutch, 2004). The authors, testing a wide variety of covariates while controlling for region, use these models to assess structural and ideational reasons why couples might have chosen to lower their fertility earlier than they would have done simply because the new idea 'arrived.' By effectively using Getis and Ord's *G* statistic to draw the conclusion that significant spatial clustering shifted from the low fertility counties to high fertility counties over the course of the antebellum period, Haines and Hacker show that high fertility followed the expanding western frontier, while low fertility regions were clustered near urban centres. At the same time, they demonstrate that ideational factors like the proportion of 'individualist' religious denominations and structural factors like higher literacy and better transportation connections mediated the uniform east-west progression of low to high child–woman ratios, accelerating change beyond the level predicted by an east-west spatial diffusion. Their work hints at complex simultaneous processes of fertility rise and decline, mediated by culture, transportation networks and geographic distance.

The next chapter takes a different approach to uncovering patterns of fertility change. **Guend** identifies neighbours in social, cultural and political 'space' and then determines whether proximity along these axes corresponds better to fertility outcomes than geographic proximity does. He focuses on the temporality of fertility decline in predominantly Muslim countries – a region of the world designated by other studies as lagging in the fertility transition as a result of religious barriers to behavioural change. Guend locates spatial, political and cultural regimes in the temporal transition of fertility of Muslim populations by mapping fertility outcomes and by applying hierarchical cluster analysis to the national-level proximate determinants of fertility, to national socio-economic characteristics and to indicators of a country's commitment to development and modernization. Countries are defined as neighbours by their similarity in multivariate space, and the strength of each set of determinants of Muslim fertility decline is revealed by the temporal trajectory of total fertility rates in each of the analytically defined groups. Guend concludes that, at the national scale, spatial diffusion models have less to contribute to the decline of fertility in the Muslim world than other theoretical approaches. He reveals that fertility decline is governed by the interaction between the intellectual traditions of individual countries and their adoption of – or failure to adopt – programmes of reform and development. Guend argues that these dynamic and often state-led movements of social change promote the transformation of demographic behaviour and are linked to, but not isomorphic with, economic, political and educational improvement. This chapter suggests that, while ideas may spread easily from person to person or province to province within a country, political borders and cultural and intellectual traditions may present strong barriers to diffusion. Diffusion still matters, but it is primarily the diffusion of ideas across conceptual space rather than geographical space.

The concern then turns to migration, which is the ideal spatial and temporal population process, one where we expect people to move through space over time, and where we should be able to see diffusion and spillover processes at work. The authors of the next three chapters utilize unusual data sources to track the origins and destinations of migrants in the period before systematic information about migration was collected and implement innovative methods to explicitly explore the time-dependence of spatial migration processes. **Darlu, Brunet and Barbero** assess migration patterns by following the distribution of surnames in the Savoy region of south-eastern France during the eighteenth through the mid-twentieth centuries. Their work is based on the solid knowledge that in many populations a new surname in a community reflects the arrival of a new individual or a new family; the extent of turnover of surnames therefore reflects the quantity of migration while the surnames themselves indicate where the migrants may have originated. Region- and period-specific origin/destination scaled association models allow the authors to map the distribution of kinship ties over space and time, revealing patterns in the geography and temporality of migration. Darlu, Brunet and Barbero find that surname exchanges are stronger between communes in the same sub-region than between those in different sub-regions, supporting basic diffusion processes for migration in which movement within the region is preferred. The authors also document exceptions to these patterns, demonstrating that some circumstances made people more likely to cross regional boundaries than to stay within their region. This was so in two cases, one where physical barriers made it more difficult to move within a region than to move outside, and a second where growing cities had a large demand for immigrants. These examples suggest that the physical and social characteristics of space influence the ways people move across it.

Ekamper, Van Poppel and Mandemakers focus on the social nature of space, analysing changes in the geographical distance between the birthplaces of marriage partners to assess the effects on migration of nineteenth-century processes of modernization and nationalization in the Netherlands. They illustrate migration patterns with regional- and period-specific graduated symbol maps that quantify the relative magnitude of the distance from grooms' places of birth to their places of marriage in each of five provinces over four periods spanning the years 1812–1922. After this descriptive prelude, the authors specify municipality-specific pooled cross-sectional regressions predicting the distance between groom and bride. Residual diagnostics detect evidence of spatial dependence unaccounted for by other covariates, allowing for the adjudication of evidence in support of spatial-lag or spatial-error model specifications. Ekamper, van Poppel and Mandemakers find patterns of marital migration very different than those that would appear in a world of spatial diffusion unconstrained by social, technological and cultural barriers. In the early part of the period, traditional modes of transportation limited migration distances, but less so for people with higher occupational status than for those with lower status, such as agricultural workers and farmers. This social differential in marital migration eroded over time, as social mobility increased and as national transportation and communication networks improved. These nationalizing processes diminished cultural barriers between the south and north of the Netherlands, leading to more marriage across this distance, but also strengthened the barriers imposed by national boundaries with Belgium and Germany. Through this analysis, the authors demonstrate that explicit attention to spatio-temporal patterns can offer new ways to test and validate established social and historical theories.

The chapter by **Gutmann, Deane and Witkowski** turns to the role of the physical and social characteristics of space in the settlement of the Great Plains of the United States from 1880 to 1940. Specifically, they test the generally held view that most westward migration in the U.S. gradually spread across the country as settlers in search of a farm or ranch moved from land that was already occupied to land that had not yet been claimed by the ambitious and growing European-origin population. The authors draw on earlier research indicating that the process of spatial and temporal diffusion was shaped in significant ways by other environmental, economic and social factors. Populations grew more rapidly in places where natural resources (e.g. high quality land, climate suitable for agriculture, mineral deposits) offered more potential for economic growth, suggesting that spatial diffusion theory and economic development theory work together and can be measured and tested using space-time approaches.

Gutmann, Deane and Witkowski offer their initial evidence in the form of quantile maps that display spatial variation in population density and in important environmental covariates (precipitation, temperature and elevation) across the U.S. Great Plains. Formal analysis draws on a series of nested proportional hazards models for repeated events, beginning with simple trend surface models – which incorporate spatial dependence via the spherical coordinate system of longitude and latitude – and building toward a full specification of physical environment and agricultural and industrial ecologies. Results provide strong evidence in support of east-to-west (longitude) dependencies in the temporal process of frontier settlement, but also confirm the kinds of results that others in this volume have achieved: agricultural characteristics – temperature, precipitation and elevation – shaped the diffusion process, while other economic factors, like the early development of industry and mining, caused migrants to jump over otherwise attractive agricultural areas and settle first in those regions closest to the new forms of opportunity. These findings complicate the story of east-west population movement, suggesting that prior conditions are important and offering new ways to formally specify and closely measure them.

Some of the same questions about prior conditions and the non-contiguous or ‘leap-frogging’ nature of spatial change are raised in the last two chapters. **Beveridge**’s examination of population growth in major U.S. cities tests predictions about the patterns of diffusion within urban areas generated by three schools of urban sociology, each named for the city whose experience it describes: Chicago (development in concentric rings away from the central business district), Los Angeles (dispersed, low-density, multi-centred development) and New York (continued concentrated development in the centre city). While most of the studies of population movement examined in this volume (and many of those published elsewhere) utilize a national, regional or sub-regional scale of analysis, Beveridge increases the spatial resolution to that of the census tract (a sub-metropolitan spatial unit used in the U.S. since the early twentieth century). Chloropleth maps of neighbourhood population density and density change over the twentieth century provide initial support for Beveridge's contention that the spatial patterning of urban neighbourhood growth in Chicago, Los Angeles and New York was varied enough to support three different 'schools'. Formal inference from global and local levels of spatial autocorrelation in the decadal per cent change in population over the course of the twentieth century for the three core cities and for several other large U.S. cities reveals a pattern of outward growth – as predicted by the Chicago school – during a city's early years, but also demonstrates that growth in later periods was much more complicated, confirming the conclusion that urban growth does not follow a single trajectory. Beveridge therefore suggests that social theory, though aspiring to generalizability across time and place, bears traces of the historical and geographical contexts in which it is formed. He also proposes further research into the spatial patterning of urban growth, using tract-level data on racial, ethnic and socio-economic status to explore how these distinctions mapped onto urban space.

This call to analyse change over time in spatial patterns of social inequality is taken up in the final chapter, where **Curtis** examines space-time dependencies in racial literacy differentials across Puerto Rican *municipios* after the U.S. takeover of the island (and the attendant economic restructuring) at the turn of the twentieth century. While most of the chapters in this volume emphasize diffusion processes by which populations, ideas or demographic changes spread over time and through space, this chapter is more concerned with understanding the spatial nature of socio-economic structures over time. Curtis asks whether Puerto Rico's agricultural geography and the social dimensions of production of different crops created a spatial pattern in literacy rates (a proxy for socio-economic status), and how changes in that pattern were shaped over time by prior conditions. To this end, she incorporates time and space covariates into a longitudinal growth model to assess the effects of initial conditions, change over time and spatial proximity on her outcome of interest. In addition to her conventional use of univariate and bivariate ESDA, Curtis embraces the direct assessment of space-time dependence using the methods of ESTDA. She quantifies first-order space-time dependence by combining the (spatial) distribution of literacy at time t with the spatially lagged distribution of literacy at time $t-1$. In doing so, her LISA give the covariation between literacy at a given locale at a given period in time with the average neighbouring literacy at a previous period in time. Curtis ultimately finds support for her equalizing hypothesis – that U.S. investment in Puerto Rican sugar production involved the provision of infrastructural improvements that diminished racial inequality, as proxied by literacy rates. However, she reveals that the spatial distribution of agricultural production doesn't tell the whole story, as spatial patterns in racial composition, both initially and over time, also affect the diffusion of racial equality.

The works presented in this volume explore processes of diffusion – whether of population, demographic behaviour, or other characteristics distributed over populations – as they articulate with core social and demographic theories in specific historical locales. Using creative approaches to newly available data, our contributors assess how historical instances of diffusion compare to theoretical predictions and measure the forces that accelerate or retard spatial diffusion, or that transform, divert or halt it completely. In so doing, they offer us new comparative insights into processes of modernization and cultural change and how they are manifest in time and space. Whether they examine long-lived nations and civilizations or relatively new colonial societies, all authors point to or acknowledge the manner in which population dynamics are conditioned by spatial and temporal patterning in modern times.

Taken together, the papers illustrate hierarchies of interaction, in which urban places are a key element and social stratification another layer that can be explored, that condition both the temporal and the spatial processes of change. The diffusion of change is complicated by this hierarchy and by its uneven nature. On the surface, the hierarchy is often a question of the scale of measurement. Clearly, in the Netherlands, when the story is taken to the individual level, we see how the spatial process is shaped by social class. But the hierarchy also exists at other scales and is manifest in the disruptive effects of urbanization and modernization in each of the societies examined here. All of this reminds us of the potential for this work and the need in future work to think in terms of multiple levels of analysis and nested social change.

References

Anselin, L. (2001a). Spatial econometrics. In B. H. Baltagi (Ed.), *A Companion to theoretical econometrics* (pp. 310–330). Oxford: Basil Blackwell.

Anselin, L. (2001b). Spatial effects in econometric practice in environmental and resource economics. *American Journal of Agricultural Economics*, *83*, 705–710.

Anselin, L. (1994). Exploratory spatial data analysis and geographic information systems. In M. Painho (Ed.), *New tools for spatial analysis* (pp. 45–54). Luxembourg: Eurostat.

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis*, *27*, 93–115.

Anselin, L., Bera, A., Florax, R., & Yoon, M. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, *26*, 77–104.

Carter, S. B., Ransom, R. L., & Sutch, R. (2004). Family matters: The life-cycle transition and the antebellum fertility decline. In T. W. Guinnane, W. A. Sundstrom, W. C. Whatley (Eds.), *History matters: Essays on economic growth, technology, and demographic change* (pp. 271–327). Stanford: Stanford University Press.

Chasco, C. & López, F. A. (2004). Modelos de regresión espacio-temporales en la estimación de la renta municipal: El caso de la región de Murcia. *Estudios de Economía Aplicada*, *22*, 1–24.

Cliff, A. & Ord, J. K. (1981). *Spatial processes: Models and applications*. London: Pion.

Coale, A. J. & Watkins, S. C. (1986). *The decline of fertility in Europe: The revised proceedings of a conference on the princeton European fertility project*. Princeton, NJ: Princeton University Press.

Davis, K. (1945). The world demographic transition. *Annals of the American Academy of Political and Social Science*, *237*, 1–11.

Durbin, J. & Watson, G. S. (1950). Testing for serial correlation in least squares regression I. *Biometrika*, *37*, 409–428.

Durbin, J. & Watson, G. S. (1951). Testing for serial correlation in least squares regression II. *Biometrika*, *38*, 159–178.

Entwisle, B. (2007). Putting people in place. *Demography*, *44*, 687–703.

Furfey, P. H. (1927). A note on Lefever's "Standard deviational ellipse". *American Journal of Sociology*, *33*, 94–98.

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, *5*, 115–145.

Getis, A. & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, *24*, 189–206.

Griffin, L. J. (1992). Temporality, events, and explanation in historical sociology: An introduction. *Sociological Methods and Research*, *20*, 403–427.

L. J. Griffin, M. van der Linden (Eds.) (1999). *New methods for social history*. Cambridge: Cambridge University Press.

Gutmann, M. P. (2005a). GREAT PLAINS POPULATION AND ENVIRONMENT DATA: AGRICULTURAL DATA [Computer file]. ICPSR version. Ann Arbor, MI: University of Michigan [producers]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. Available online at http://dx.doi.org/10.3886/ICPSR04254.

Gutmann, M. P. (2005b). GREAT PLAINS POPULATION AND ENVIRONMENT DATA: SOCIAL AND DEMOGRAPHIC DATA [Computer file]. ICPSR version. Ann Arbor, MI: University of Michigan [producers], 2005. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005. Available online at http://dx.doi.org/10.3886/ICPSR04296.

Haines, M. R. (2004). HISTORICAL, DEMOGRAPHIC, ECONOMIC, AND SOCIAL DATA: THE UNITED STATES, 1790–2000 [Computer file]. ICPSR02896-v2. Hamilton, NY: Colgate University/Ann Arbor: MI: Inter-university Consortium for Political and Social Research [producers]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].

Knox, G. (1964). The detection of space-time interactions. *Applied Statistics*, *13*, 25–29.

Lee, E. S. (1966). A theory of migration. *Demography*, *3*, 47–57.

Lefever, D. W. (1926). Measuring geographic concentration by means of the standard deviational ellipse. *American Journal of Sociology*, *32*, 88–94.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*, 209–220.

Massey, D. (2005). *For space*. London: Sage.

McClintock, A. (1995). *Imperial leather: Race, gender, and sexuality in the colonial contest*. New York: Routledge.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*, 17–23.

Ravenstein, E. G. (1885). The laws of migration. *Journal of the Royal Statistical Society*, *48*, 167–227.

Robinson, A. H. & Sale, R. D. (1969). *Elements of Cartography* (3rd ed.). New York: Wiley & Sons.

Snow, J. (1855). *On the mode of communication of cholera* (2nd ed.). London: John Churchill.

Thornton, A. (2001). The developmental paradigm, reading history sideways, and family change. *Demography*, *38*, 449–465.

Thornton, A. (2005). *Reading history sideways: The fallacy and enduring impact of the developmental paradigm on family life*. Chicago: University of Chicago Press.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Turner, F. J. (1893). The significance of the frontier in American history. *Annual Report of the American Historical Association*, *1*, 199–227.

Upton, G. & Fingleton, B. (1985). *Spatial data analysis by example*. New York: Wiley.

Voss, P. (2007). Demography as a spatial social science. *Population Research and Policy Review*, *26*, 457–476.

Watkins, S. C. & Danzi, A. D. (1995). Women's gossip and social change: Childbirth and fertility control among Italian and Jewish Women in the United States, 1920–1940. *Gender and Society*, *9*, 469–490.

Chapter 2
An Innovative Methodology for Space-Time Analysis with an Application to the 1960–2000 Brazilian Mortality Transition

Carl P. Schmertmann, Joseph E. Potter, and Renato M. Assunção

Introduction

The stylized demographic transition has two parts: a mortality transition and a fertility transition. Changes from high to low rates in both dimensions have been important and dramatic. There has been much debate about the factors leading up to these movements in vital rates and about the relation that each transition has to the other. One central question has been whether demographic changes respond to changing economic and social conditions, or whether they derive mainly from the spread of new ideas and behaviours.

This paper is primarily methodological and involves the development of space-time methods to investigate the role of contagion. We introduce a common spatial statistic, used to detect epidemic patterns on maps, and discuss how to modify it appropriately for demographic research. We then apply the modified method to space- and time-referenced data on Brazilian mortality, in order to assess the degree of 'contagion' in mortality transition. This study of mortality transition closely parallels the analysis of fertility transition in a companion paper (Schmertmann, Assunção, & Potter, 2010).

Diffusion/Social Interaction vs. Development

For fertility, there has been a vibrant debate over the last four decades about the factors underlying transition (e.g. Coale, 1973; Coale & Watkins, 1986; Cleland & Wilson, 1987; Guinnane, Okun, & Trussell, 1994; Brown & Guinnane, 2002; Bryant, 2007). The dominant opinion in recent years seems to be that ideas play a large role in the transition and that material conditions may be less important. Comparisons across provinces of Europe provided the original justification for

C.P. Schmertmann (✉)
Department of Economics, Center for the Study of Population, Florida State University, Tallahassee, FL 32306-4063, USA
e-mail: schmertmann@fsu.edu

M.P. Gutmann et al. (eds.), *Navigating Time and Space in Population Studies*, International Studies in Population 9, DOI 10.1007/978-94-007-0068-0_2,

rejecting a structural interpretation in which development was the central underlying factor. This conclusion rested on two findings. First, there seemed to be no verifiable development threshold that signalled the onset of fertility decline and fertility began to decline at widely differing levels of development across European provinces. Second, after constructing maps showing the spread of fertility decline through time, there seemed to be spatial clustering unrelated to development, with low fertility spreading like an epidemic from a small number of core regions through time (Coale & Watkins, 1986). These space-time inferences were based on visual impressions rather than statistical tests, but they were very influential.

Researchers have recently attempted to build a conceptual and empirical base for this ideational interpretation of the fertility transition (Bongaarts & Watkins, 1996). Some investigators have brought spatial statistics methods to bear on this question, in order to empirically verify and measure space-time interaction in the onset of fertility transition. This research has dealt with both historical and contemporary developing country populations. Some authors have been concerned with the pace of fertility decline after onset (Montgomery & Casterline, 1993; Tolnay, 1995; Weeks, Getis, Hill, Saad Gadalla, & Rashed, 2004) and several have focused specifically on the time of onset (Guilmoto & Rajan, 2001; Bocquet-Appel & Jakobi, 1998; Balabdaoui, Bocquet-Appel, Lajaunie, & Rajan, 2001; Bocquet-Appel, Rajan, Bacro, & Lajaune, 2002).

With respect to the mortality transition, there has been a parallel debate about factors underlying the onset and pace of mortality decline. One main finding is that relationships between development indicators and mortality measures have shifted through time (Preston, 1975 and 1980; Deaton, 2004; Becker, Philipson, & Soares, 2005). In addition to income per capita, other covariates appear strongly related to levels and trends in mortality across populations. Sanitation variables are important. Access to clean water and adequate sewage systems have been shown to play a large role in both national and international analyses (Preston & Haines, 1991; Cutler & Miller, 2005), as well as in several studies of Brazil (Merrick, 1985; Soares, 2007). A second prominent variable in mortality analyses has also played a large role in fertility studies: educational attainment, particularly among women. Finally, we note that religion seems to have played a role in some mortality transitions, including Brazil's (Wood, Williams, & Chijiwa, 2007).

As with fertility transition, covariates available from census or survey data (such as women's education and access to clean water) cannot fully account for mortality change. Two interesting possibilities remain. First, available variables may not adequately capture all relevant aspects of living standards, public health measures, or the quality and accessibility of health care services. Indeed, with rapid changes in medical technology and the spread of immunization and oral rehydration therapy, one would be surprised if the covariates found in censuses explained all of the variation in child mortality. However, a second possibility, paralleling the fertility debate, is that new ideas and behaviours also play an important role in mortality change. What might these ideas be?

The foremost exponent of a 'health transition' based on changes in ideas and behaviours was Caldwell (1986), who argued that there is a 'symbiosis between cultural and health inputs' including not only an 'interaction between the provision of education and that of health services' (Caldwell, 1986: p. 204), but also a change in parental attitudes towards child survival as fertility declines. Particularly important changes include the collapse of patriarchy, strengthened conjugal bonds and shared decision-making between parents, eventually leading to strengthened 'child-centredness' (Condran & Preston, 1994).

In this paper, we address the question of how to use space-time data to indirectly assess the effects of such 'unmeasured' influences on child mortality.[1] We presume that a causal mechanism such as social interaction would lead to space-time clusters of mortality transitions, much like the 'infectious' pattern exhibited by European fertility change. The logic for spatial clustering is most familiar with regard to the spread of ideas by word of mouth. If conversations regarding, say, hygiene, breastfeeding, or immunization provide important pathways towards improved child survival, then the usual covariates would fail to completely explain the onset of transition *and* areas of declining mortality at any one time would also be clustered in space. On the other hand, spatial clustering of transitions might also arise due to some unmeasured policy such as an immunization programme that was launched in a particular state or province and then later spread to the rest of a country. In either case, clusters of transitions in space and time would be informative.

The Knox Test

Notation

For a collection of demographic events that are indexed by location and time of occurrence, one can think of a *map* as either a physical drawing of when and where events happened, or as a set

$$M = \{(\mathbf{s}_i, \mathrm{t}_i, \mathbf{z}_i)\} \; \mathrm{i} = 1 \ldots \mathrm{n}$$

containing spatial locations $\boldsymbol{s}$, times t and potentially important covariates $\boldsymbol{z}$ associated with each of n events. Note that $\boldsymbol{s}_{\mathrm{i}}$ and z_{i} may be vectors, while t_{i} is scalar. Throughout this paper we number events in chronological order, so that $t_1 \le t_2 \le \ldots \le t_{\mathrm{n}}$.

As an example, map data for five events occurring in a simple space with four discrete locations ABCD might be:

$M =$

i	*s*	*t*	*z*
1	A	1	0
2	A	1	0
3	C	3	1
4	A	6	1
5	D	6	1

For this map, events happened at (space, time) combinations (A,1), (A,1), (C,3), (A,6) and (D,6); covariate values were z=0 at the first two (s,t) combinations and

z=1 at the last three. In real applications map data might indicate location $\boldsymbol{s}$ with both latitude and longitude, or include multiple covariates $\boldsymbol{z}$.

Epidemics and Space-Time Interaction

Contagious or epidemic processes will generate related clusters of events. Temporal or spatial clusters of events alone do not indicate an epidemic process, however. Even in a non-contagious process, events may cluster in space (for example, in more populated areas), or in time (for example, seasonally). In an epidemic, in contrast, events occurring at similar times would also be *especially* likely to occur in nearby places. Statistically speaking, epidemic processes should exhibit *space-time interactions*, with greater concentrations of events around some ($\boldsymbol{s}$,t)s on the map than one would expect from the marginal frequencies of locations $\{\boldsymbol{s}_1 \ldots \boldsymbol{s}_\mathrm{n}\}$ and times $\{t_1 \ldots t_\mathrm{n}\}$ alone.

In this chapter's application, locations are discrete regions that each experience a single event. That structure makes our marginal distribution of locations $\boldsymbol{s}$ particularly simple – there is exactly one event per region on the map. Thus, in our analysis, contagion would appear as a changing series of event locations over time – *where* a mortality transition happened would be informative about *when* it happened.

The Knox Test

Knox (1964) developed a simple statistical test for detecting epidemic patterns based on space-time clusters. The Knox test was designed for studying events such as infections that can occur many times in a single location. It requires an assumption that spatial distribution of risk is constant over time and does not include local covariates. In this paper we consider the application of the Knox test to survival processes, in which events are not repeatable and we add covariates. However, because most demographers are unfamiliar with the Knox test, we briefly defer these complications and begin by describing the standard test only.

The standard Knox test compares an observed map M to the set of all possible maps with the same event locations $\{\boldsymbol{s}_1 \ldots \boldsymbol{s}_\mathrm{n}\}$ and times $\{t_1 \ldots t_\mathrm{n}\}$, but with potentially different *pairings* of these $\boldsymbol{s}$ and t values. Intuitively, there is evidence for an epidemic pattern if the observed map shows more space-time clustering than other event maps sharing the same marginal frequencies of locations and times.

The Knox test requires binary indicators for each possible pair of events on the map

$$\delta_{ij} = I \text{ (events } i \text{ and } j \text{ are close in space)} \;;\; \tau_{ij} = I \text{ (events } i \text{ and } j \text{ are close in time)}$$

where the indicator function $I(\ )$ equals 0 if the condition in parentheses is false and 1 if true. Appropriate definitions of closeness depend on the research context and objectives. Given the indicators, the *Knox statistic* for a map is

$$X(M) = \sum_i \sum_{j>i} \delta_{ij}\, \tau_{ij}$$

which is the number of unordered pairs of events that are close in *both* space and time.

When searching for evidence of epidemic patterns of space-time interaction in a map, the important question is whether the observed X is high relative to what one would expect, given the observed marginal frequencies of events by location and by time. Answering this question requires information about the distribution of X under a null hypothesis of no space-time interaction. This distribution is analytically difficult, because the $n(n\text{–}1)/2$ elements in the sum are not statistically independent and because the distribution may depend on the exact arrangement of spatial units and the exact marginal frequencies of locations and times.

In contrast, it is easy to *simulate* the distribution of X in the absence of space-time interaction. If event locations and times are independent, events are repeatable and the spatial distribution of risk is unchanging, then all pairings of elements of $\{\boldsymbol{s}_1 \ldots \mathbf{s}_n\}$ with those of $\{t_1 \ldots t_n\}$ are equally likely. In this case one can draw an X value from the null distribution by the following method:

- generate a permutation $\pi = (1)\ldots(n)$ of indices $1\ldots n$, uniformly from the set of all $n!$ possible orderings[2]
- construct the fictitious map $M^*(\pi) = \{(\boldsymbol{s}_{(i)}, t_i)\}$
- calculate the Knox statistic for this new map $X^*(\pi) = X[M^*(\pi)]$.

This process simulates a draw from the null distribution of X by randomly shuffling the locations in the $\boldsymbol{s}$ column of a table like our simple example and then recalculating the Knox value for the new, fictitious map. Repeating the process can generate an arbitrarily good approximation to the distribution of X in the absence of space-time interaction and, therefore, allow statistical testing. If $X(M)$ is larger than 95% of many simulated X^* values, for example, then there is strong sample evidence in favour of space-time interaction.

Adapting the Knox Test for Survival Models and Changing Event Risks

In the standard Knox test, simulating X^* values from the null distribution depends critically on the fact that all permutations π are equally likely under the stated assumptions (no space-time interaction, repeatable events, unchanging spatial distribution of event risk). In many demographic applications the second and third assumptions may be violated. For example, in this paper we study a survival process for a set of discrete locations, each of which has at most one event. The assumption of unchanging spatial patterns of risk is also dubious when studying demographic processes that span decades and take place against a background of spatially patterned economic and social changes.

Even with these new complications, one can still perform a Knox test by simulating the null distribution of X. If the observed $X(M)$ is in the upper tail of the possible values for maps with the given spatial and temporal frequencies of events, then the map provides evidence of a space-time clustering and a contagious process. In survival or covariate models, however, the absence of space-time interaction no longer implies equal likelihoods for all event orders π (or equivalently, for all pairings of observed $\boldsymbol{s}$ and t values). In these more general models one cannot generate simulated maps under the null by uniform sampling from the $n!$ possible event orders.

Fortunately, proportional hazard models (Cox, 1972) were designed specifically for modelling event order in survival models with covariates. If each of n units has a constant failure hazard of μ_i, the probability that they fail in order $\pi = (1) \ldots (n)$ is

$$L(\pi) = \prod_{i=1}^{n} \left(\mu_{(i)} \Big/ \sum_{j \in R_{(i)}} \mu_j \right)$$

where $R_{(i)}=\{(i)\ldots(n)\}$ contains the units still at risk for the ith failure. If hazards vary over time but satisfy the proportionality assumptions of Cox (1972):

$$\mu(t, z_i) = \mu_0(t) \exp\left[\beta' z_i\right]$$

then the probability of failures in order $\pi = (1)\ldots(n)$ is

$$L(\pi \mid \beta) = \prod_{i=1}^{n} \left(\exp\left[\beta' z_{(i)}\right] \Big/ \sum_{j \in R_{(i)}} \exp\left[\beta' z_j\right] \right)$$

which is identical to Cox's partial likelihood function (cf. Kalbfleisch & Prentice, 1973, Eq. 2).

To extend the Knox approach to survival processes with covariates, we propose using Cox's function in two ways. First, we estimate covariate effects $\hat{\beta}$ in a standard proportional hazards regression. This set of coefficients maximizes the likelihood of the *observed* event order on the map.

Second, we use the estimated $\hat{\beta}$ to model the likelihood of different *hypothetical* event orders π (i.e. different hypothetical maps and X^* values) under the null hypothesis of no space-time interaction. Using this non-uniform distribution of π for the Knox test draws $X^*(\pi)$ values more often from pairings of event locations and times that are likely to occur, given observed patterns in covariates. Standard software makes simulation easy – in particular, the *sample* function of the *R* language (Ihaka & Gentleman, 1996) generates orderings with the frequencies implied by $L(\pi|\beta)$ above when the set of probabilities $\{\exp[\beta z_i]\}$ is included as an argument.

We discuss the mathematical rationale for this procedure more fully elsewhere (Schmertmann et al., 2010). The important point here is that in many demographic applications it is possible that covariate effects, rather than contagion, can induce space-time clusters of events. Our new procedure models the covariate effects in the Knox statistic $X(M)$ and, thus, compares the observed $X(M)$ to a null distribution of X^* values that are also affected by covariates.

Data

Brazilian Demographic Censuses and Indirect Estimates

Data for our study of mortality transition come from the Brazilian Demographic Censuses of 1960, 1970, 1980, 1991 and 2000. We define 502 *microregions* with consistent 1960–2000 boundaries. These microregions correspond closely to the microregions constructed by the Brazilian census bureau for the 1991 Census, but we aggregated some areas to maintain compatibility across changing census geographies.

For each region and census year, we used Brass indirect mortality methods to estimate a local value of q(5), the life table probability of death before exact age five (see Cavenaghi (1999) and Potter, Schmertmann and Cavenaghi (2002) for details). We also aggregated individual- and household-level data to produce several socio-economic indicators for each microregion in each census year, including data on household electrification and piped water, female labour force participation, occupational distribution and adult female education.

Definition of Mortality Transition

For this study, we define a *mortality transition* as a permanent decline of q(5) below 100 per thousand. With a series of decennial census data for each region, we operationalize the definition by requiring that once q(5) is observed below the threshold, it must stay below in all subsequent censuses. Table 2.1 provides examples. Cells

Table 2.1 Childhood mortality levels and inferred mortality transition times for selected microregions

	1,000·q(5) by census year					
Region	1960	1970	1980	1991	2000	Transition
Porto Alegre, RS	**99**	**94**	**61**	**22**	**13**	pre 1960
Salvador, BA	235	204	112	**58**	**28**	1980–1991
Belo Horizonte, MG	175	163	110	**44**	**25**	1980–1991
Rio de Janeiro, RJ	160	132	104	**40**	**22**	1980–1991
Alto Araguaia, MT	**66**	116	**86**	**56**	**7**	1970–1980
Jeremoabo, BA	388	241	198	111	104	post 2000

Table 2.2 Distribution of mortality transitions by intercensal period

	Pre 1960	1960–1970	1970–1980	1980–1991	1991–1900	Post 2000
Regions at risk	395*	483	453	270	116	7
Transitions	19	30	183	154	109	7

Note: *1960 demographic census data are unavailable from any source for 107 microregions, primarily in the North. For these regions we assume the earliest possible mortality transition was in 1960–1970.

with q(5) values below the threshold are in bold, and the inferred transition times appear in the last column. Table 2.2 shows the distribution of mortality transitions by decade across all 502 microregions.

Figure 2.1 illustrates the space-time distribution of the estimated transitions. There appears to be an extremely strong space-time interaction in this map sequence, with very clear spatial clusters appearing in different parts of the country at different times. To test this proposition formally, we turn to the Knox test for contagion.

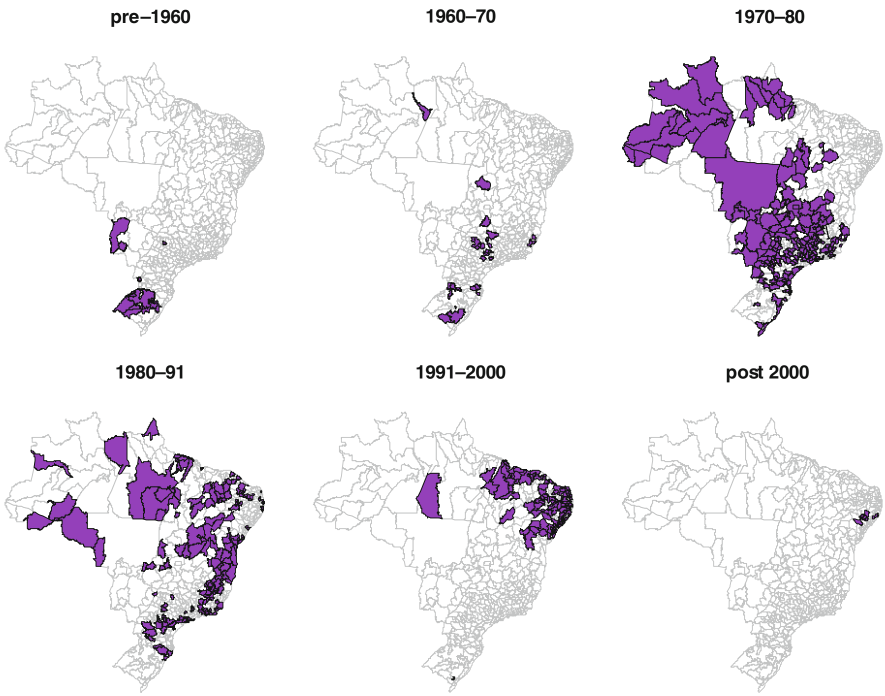


Fig. 2.1 Mortality transitions by decade. Shaded regions indicate transitions to q(5) < 100 per thousand

Knox Statistic

For the Knox test, we define indicators for the spatial and temporal proximity of event pairs as follows. Our spatial units are the 502 microregions, each of which can be represented on a physical map as a polygon with known borders and centroid and each of which has exactly one transition event. On this map there are 502∗501/2=125,751 distinct pairs of events and we define spatial proximity with the indicator[3]

$$\delta_{ij} = I\left[D_{ij} < 100\,km\ \mathbf{OR}\ i \sim j\right]$$

where D_{ij} is the straight-line distance between the centroids of microregions i and j, and $i \sim j$ represents map adjacency. Experiments show that our results are insensitive to the exact definition of spatial proximity.

Because our data are from decennial censuses, mortality transition times fall into six discrete categories – before 1960, 1960–1970, 1970–1980, 1980–1991, 1991–2000 and after 2000. We use the simplest possible definition of temporal proximity, namely

$$\tau_{ij} = I\left[i \text{ and } j \text{ had transitions in the same intercensal period}\right]$$

As with δ, changing the definition of temporal proximity has no meaningful effects on our results.

Using these definitions, the 125,751 distinct pairs of mortality transition events are cross-classified in Table 2.3:

The Knox statistic for space-time clustering is, therefore, X=1,017. The cross-classification table immediately suggests space-time interactions, because the conditional distributions within rows are different: a majority of transition pairs occurring in the same intercensal period (τ=1 row) were also close in space, but the pattern is the opposite for transitions occurring in different periods (τ=0 row). However, this pattern could have occurred at random (this is the concern of the standard Knox test), or it could have been induced by spatial patterns in relevant covariates rather than by contagion (this is the concern of our modified test). We now address these issues in turn.

Table 2.3 Knox statistic for the mortality transition map

Event pairs (i,j)		
Close in time*Close in space*	*NO* ($\delta_{ij} = 0$)	*YES* ($\delta_{ij} = 1$)
NO ($\tau_{ij} = 0$)	90,024	33,930
YES ($\tau_{ij} = 1$)	780	$X(M) = 1{,}017$

Results

Standard Knox Test for Contagion

The standard Knox test considers the map of transitions

$$M = \{(s_1, t_1) \ldots (s_{502}, t_{502})\}$$

for which we have calculated $X(M)$=1,017, simulates draws from the null distribution of X by drawing a large number of permutation maps $\pi = (1) \ldots (502)$ uniformly from the set of 502! possibilities and, finally, compares the observed $X(M)$ to the simulated null distribution to see if it is unusually large. As discussed earlier, this test does not produce the correct null distribution of X when events are non-repeatable transitions, or when covariates that affect risks change over the observation period. However, we begin the standard test for purposes of comparison.

For each simulated permutation $\pi_s = (1)\ldots(502)$ we assign simulated transition times

$$t_i^*(\pi_s) = \begin{cases} \text{pre 1960} & \text{if } i \in \{(1)\ldots(19)\} \\ \text{1960 - 1970 if } i \in \{(20)\ldots(49)\} \\ \text{1970 - 1980 if } i \in \{(50)\ldots(232)\} \\ \text{1980 - 1991 if } i \in \{(233)\ldots(386)\} \\ \text{1991 - 2000 if } i \in \{(387)\ldots(495)\} \\ \text{post 2000} & \text{if } i \in \{(496)\ldots(502)\} \end{cases}$$

That is, the first 19 simulated transitions are assumed to occur before 1960, the next 30 between 1960 and 1970 and so on.

Figure 2.2 presents results for the Knox statistic from 10,000 simulated maps, using the uniform permutation probabilities. The 99th percentile of the simulated X^*

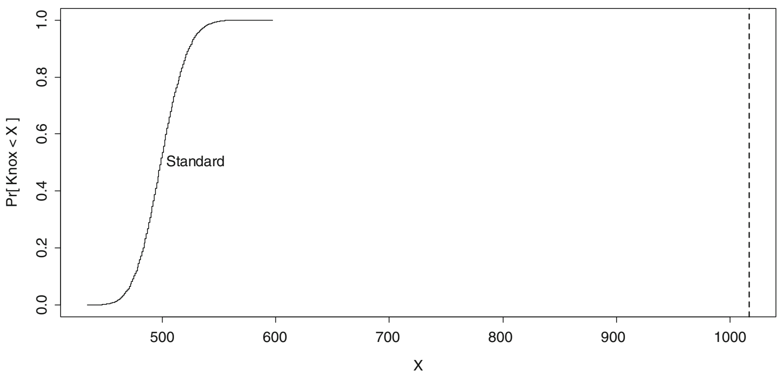


Fig. 2.2 Standard Knox test. 10,000 Monte Carlo draws. Observed X=1,017. Median X^*=499

values is 545 and none of the 10,000 simulations generated a Knox statistic exceeding 597. Thus, the observed value of 1,017 appears highly significant, providing strong evidence of space-time interaction in the mortality transition map.

Modified Test for Non-repeatable Events and Covariates

The standard test is biased, because it detects space-time interaction of any kind – whether caused by contagion, or by changing risk sets, covariates and hazards. To address this problem, we turn to the proportional hazards correction outlined earlier. We use end-of-period covariate levels as predictors for transitions/failures, allow covariates to vary within regions across periods and treat the handful of post-2000 transitions as right-censored observations. We include five covariates – per cent of households with good water supplies,[4] per cent of the labour force in the primary sector, per cent of adult women in the labour force, average education of adult women (in years) and per cent Catholic – in a proportional hazards model with time-varying covariates. That procedure yields the estimates in Table 2.4.

Our interest lies not in the specific coefficients, but rather in the model's overall power to distinguish high- versus low-risk intervals. In this regard, the regression does well. If one predicted that the 19 regions with the highest predicted risks would have a mortality transition before 1960, that the 30 regions with the next highest risk would transition during 1960–1970 and so forth, then predicted and observed transition times match for 276 of 502 regions (55%). In comparison, one would expect only 27.9% matching if simulated times were drawn in the same way without regard to predicted risk.

Socio-economic covariates and mortality transition times are clearly related. This in turn means that the space-time event clusters that contribute to the Knox statistic (for example, the concentration of early mortality transitions in the South and Southeast) may result from spatial clusters of covariates (such as greater access to clean water in the South and Southeast in early periods), rather than from a contagion process.

To account for this possibility and investigate the effects of contagion, net of covariate effects, we use the modified Knox test. As in the standard test, the first 19 simulated transitions are assumed to occur before 1960, the next 30 between 1960 and 1970, etc. However, in the modified test the sampling of π is non-uniform,

Table 2.4 Proportional hazard regression estimates for mortality transition

Covariate	$\hat{\beta}$	exp ($\hat{\beta}$)	se ($\hat{\beta}$)	t	p
Percent of HHs with good water	0.015	1.02	0.004	3.54	4.0E-04
Average years of female education	0.794	2.21	0.091	8.69	6.0E-14
Percent female LF participation	–0.014	0.99	0.009	–1.56	1.2E-01
Percent of LF in primary sector	0.036	1.04	0.005	7.93	2.1E-15
Percent Catholic	–0.024	0.98	0.006	–4.15	3.4E-05

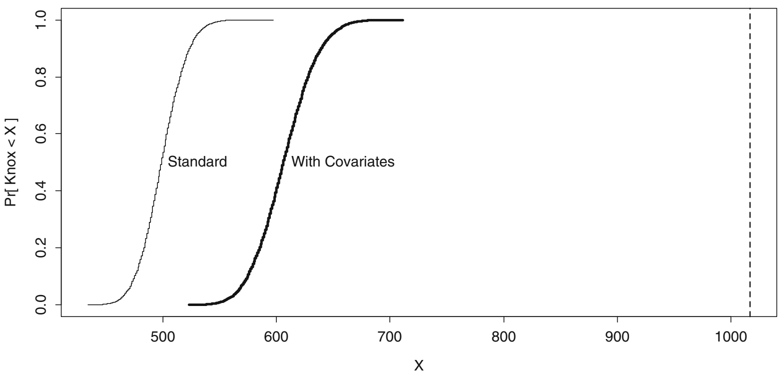


Fig. 2.3 Modified Knox test with covariate effects. 10,000 Monte Carlo draws. Observed X=1017. Median X^*=499 for standard test, 606 for modified test

with covariate levels affecting each microregion's transition probability in each time period. This approach draws samples of discrete times that match the assumptions of the proportional hazards model and also match the marginal distribution of t in the observed map.[5]

The modified null distribution of X^* appears in Figure 2.3, together with the standard null simulated earlier. Including covariates and shifting risk sets moves the null distribution higher by about 100 points. In statistical terms this is a large shift, equivalent to more than five standard deviations. Median X^* rises from 499 (standard test) to 606 (modified test), the 99th percentile goes from 545 to 666 and the maximum from 597 to 711.

Two important conclusions arise from these changes in the null distribution. First, covariate effects apparently cause some of the space-time interaction captured by the standard Knox statistic. The rightward shift of the distribution shows that attributing space-time clustering to contagion only would be a mistake. Second, the observed X=1,017 is still highly significant in the modified test. Strong evidence remains against a null hypothesis of space-time independence in events, even after controlling for this set of covariates and for changing risk sets.

Early and Late Transitions

The modified Knox test also provides another means of analysing the space-time interactions in mortality transition. After controlling for covariates, one can identify regions with unexpectedly early or late mortality transitions, given their socio-economic conditions. Specifically, we can compare each region's *observed* transition time t_i to the distribution of its *simulated* transition times $\{t_i^*(\pi_s)\}_{s=1\ldots10{,}000}$. Such comparisons, when mapped, provide information about where and when the spatial interactions detected by the modified Knox test occurred.

In order to study early and late transitions, we define indicators for each region:

$$E_i = I\left[t_i < t_i^*(\pi_s) \text{ in over } 60\% \text{ of samples}\right]$$
$$L_i = I\left[t_i > t_i^*(\pi_s) \text{ in over } 60\% \text{ of samples}\right]$$

$E_i=1$ if an area's transition occurred earlier than its covariate levels predict; $L_i=1$ if the transition was unexpectedly late. Using this definition, 91 of 502 observed transitions were early, 61 were late and 350 were neither early nor late. Table 2.5 shows the complete distribution of early and late transitions across regions and intercensal periods.

As examples, Table 2.6 below shows the distributions of simulated times $t*$ for the six microregions shown earlier in Table 2.1. Each row of Table 2.6 contains the region's distribution of simulated (covariate-adjusted) transition times across periods. Observed transition times from Table 2.1 are indicated by brackets around the number in the corresponding column. The last row of Table 2.6 shows the marginal distribution of simulated transitions (which is identical to the expected percentages across each row in the standard Knox test without covariates).

Under this definition, for example, Porto Alegre's pre-1960 transition was *early*, because 92% of simulated transitions occur later. Rio de Janeiro's 1980–1991 transition was *late* because its relatively good socio-economic conditions predicted an earlier event. The transition in Jeremoabo, Bahia should have taken place late on

Table 2.5 Distribution of early and late transitions, by intercensal period

	Observed transition time						
	Pre 1960	1960–1970	1970–1980	1980–1991	1991–2000	Post 2000	All
Early	19	29	38	5	0	0	91
Neither early nor late	0	1	145	126	78	0	350
Late	0	0	0	23	31	7	61
Total	19	30	183	154	109	7	502

Table 2.6 Simulated and observed mortality transition times for example microregions (observed times indicated with [brackets])

	Simulated transition times (% of samples)							
Region	Pre 1960	1960–1970	1970–1980	1980–1991	1991–2000	Post 2000	E	L
Porto Alegre, RS	[8]	11	64	16	1	0	1	0
Salvador, BA	2	4	48	[42]	3	0	0	0
Belo Horizonte, MG	4	5	45	[42]	4	0	0	0
Rio de Janeiro, RJ	2	9	70	[19]	0	0	0	1
Alto Araguaia, MT	3	6	[55]	34	3	0	0	0
Jeremoabo, BA	3	2	12	14	48	[21]	0	1
*All Regions**	*2*	*5*	*37*	*33*	*22*	*1*	–	–

Note: *in the standard Knox simulation without covariates, every region would have this distribution of simulated transition times

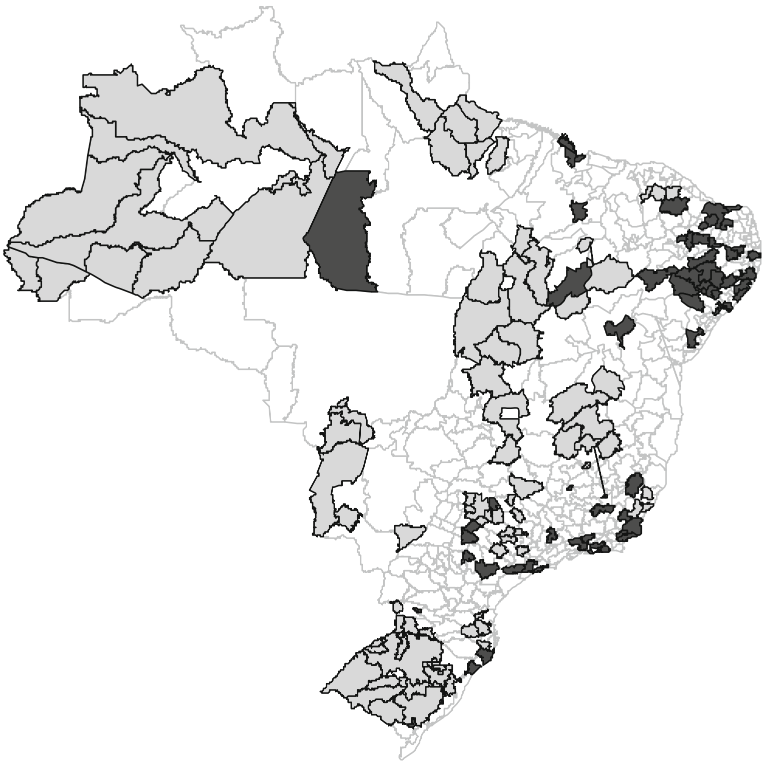


Fig. 2.4 Early (*light grey*) and late (*dark*) mortality transitions, Brazil 1960–2000

account of its covariates, but since it had not yet taken place by the 2000 census, it too is classified as late. Belo Horizonte's transition, like most, was neither *early* nor *late*.

Figure 2.4 maps the complete set of early and late transitions and, thus, illustrates some of the sites of 'contagion' in mortality transition. By (imperfectly, but effectively) filtering out effects of socio-economic covariates, the map reveals that most of the surprisingly early transitions to low childhood mortality were clustered in the South and the North and, to a lesser extent, in the Centre-West regions of the country. The surprisingly late transitions to low childhood mortality were, by comparison, less clustered. Late transitions occurred mainly in the Southeast and Northeast regions, with the greatest clustering in the interior of Pernambuco and neighbouring north-eastern states.

Discussion

Despite their wide use in epidemiology and incipient use in the study of fertility transitions, formal tests for space-time interaction have not been used in mortality research. In this paper, we have presented a test for space-time interaction that has been adapted for transitions that unfold over decades, rather than weeks or months, and that takes into account the role of changing social and economic circumstances through the use of covariates. This test is, thus, an improvement on the original Knox test for demographic applications. The modification we have introduced is small but important: we graft a familiar tool, proportional hazards modelling, onto the original method of Knox.

In addition to improving the test for space-time interaction, we have also used the modified simulations to identify the regions that, given their socio-economic conditions, had either unexpectedly early or late mortality transitions. Mapping these regions gives us a view of the transition after taking covariates into account. In applications where there is significant space-time interaction, it also shows where and when the spatial interactions detected by the modified Knox test occurred.

Space-time procedures shed new light on some important aspects of the Brazilian mortality transition over 1960–2000. The proportional hazards model provides evidence that a range of variables reflecting the sanitary and socio-economic circumstances prevailing in these microregions, as well as a variable that has to do with values and norms, per cent Catholic, are strongly associated with the timing of the transition to a child mortality rate lower than 100 per thousand. The general result and the estimated effects of the specific variables are consistent with previous studies; for example, piped water and Protestantism are both associated with a higher likelihood of transition to low mortality (Wood et al., 2007; Merrick, 1985).

Our analysis also highlights the location and timing of transitions that were less predictable and, thus, focuses attention on areas that may require other explanations for mortality change. Figures 2.1 and 2.4 show that the surprisingly early transitions were heavily concentrated in the South, especially in the state of Rio Grande do Sul. A large majority of regions reaching a child mortality rate of 100 per thousand before 1970 were located in this southernmost state. Our modified Knox procedure tells us, however, that although Porto Alegre and some other microregions had good socio-economic indicators in the 1960s, their mortality levels in 1960 and 1970 would not have been expected to have been below 100 per thousand on that basis. The unexpectedly early transitions in Rio Grande do Sul raise a provocative question about proximity to Uruguay, a neighbouring country with low rates of mortality and fertility. Perhaps ideas regarding childrearing spread across the border? Earlier work on mortality decline in Brazil has paid little attention to this unusual experience, and it merits further investigation.

The other surprising early transitions were mainly in Brazil's north region, in the 1970s and 1980s. Again, child mortality fell below 100 per thousand in these places earlier than would have been expected on the basis of the social, economic and cultural variables included in the hazards model. This finding confirms an observation that Brazilian demographers have made in previous analyses (Simões & Monteiro, 1995), but which has not been fully explained. The idea that early northern transitions resulted from the spread of new ideas is unlikely because of the region's vast distances, low population density and sparse transportation network. It seems more likely related to factors affecting disease transmission, access to clean water and, perhaps, food distribution. Again, the results of our space-time analysis provide an impetus for further work on this question.

There is less clustering of the late transitions in Figure 2.4, suggesting that the additional factors affecting laggards have weaker spatial patterns than those affecting early transitions. Space-time analysis again points to places, notably metropolitan Rio de Janeiro, whose unusual transitions merit further study.

Conclusion

The increasing availability of both recent and historical census micro-data georeferenced down to relatively small areas, along with the use of GIS to create large data sets based on other types of information, has created an unprecedented opportunity to elevate the role of spatial-temporal analysis in population research. But to fully exploit the information that is now at their disposal, demographers and other social scientists will have to learn and sometimes adapt tools that have been developed in other disciplines for quite different kinds of problems. We hope we have taken a significant step in that direction in this chapter, both by adapting the Knox test for use in demographic applications and by showing how the new procedure can illuminate a specific demographic transition. We expect that this particular method will contribute to answering long-standing questions regarding development and diffusion as determinants of changes in mortality and fertility and to identifying facets of the demographic transition that remain hidden by more traditional analytic techniques.

Notes

1. We analyse child mortality for both practical and theoretical reasons: One can estimate it with reasonable accuracy indirectly from available census data and it correlates well with the overall level of mortality as measured by life expectancy.
2. (i) denotes the index of the ith element in the permutation. For example, if 3 elements have permutation order π=[2,1,3], then (1)=2, (2)=1, (3)=3.
3. Common definitions of spatial proximity use either adjacency (a shared border or corner), or a small distance D between centroids. Brazil's combination of continental scale and high variation in population density makes either of these standard definitions problematic. An adjacency definition would categorize many pairs of microregions as 'not close' even though their centroids are very near one another: 382 non-adjacent pairs have $D < 100$ km. A centroid definition, on the other hand, would categorize many pairs as 'not close' even though they are adjacent on the map: 613 adjacent pairs, mostly in the North, have $D > 100$ km. Given these differences in scale, we opted for this compromise definition.
4. We define a 'good' household water supply as either piped water (*canalização interna*) or supply from a public system (*rede geral*).

5. Matching both the model and the observed map is possible because the proportional hazards model assigns a likelihood only to the *order* of events. Event times are arbitrary, as long as they match the ordering.

References

Balabdaoui, F., Bocquet-Appel, J. P., Lajaunie, C., & Rajan, S. I. (2001). Space-time evolution of fertility transition in India (1961–91). *International Journal of Population Geography*, *7*, 129–148.

Becker, G. S., Philipson, T. J., & Soares, R. R. (2005). The quantity and quality of life and the evolution of world inequality. *American Economic Review*, *95*, 277–291.

Bocquet-Appel, J. P., & Jakobi, L. (1998). Evidence for a spatial diffusion of contraception at the onset of the fertility transition in Victorian Britain. *Population: An English Selection, Special Issue: New Advances in Social Sciences*, *10*, 181–204.

Bocquet-Appel, J. P., Rajan, S. I., Bacro, J. N., & Lajaune, C. (2002). The onset of India's fertility transition. *European Journal of Population*, *18*, 211–232.

Bongaarts, J., & Watkins, S. C. (1996). Social interactions and contemporary fertility transitions. *Population and Development Review*, *22*, 639–682.

Brown, J. C., & Guinnane, T. W. (2002). Fertility transition in a rural, Catholic population: Bavaria, 1880–1910. *Population Studies*, *56*, 35–49.

Bryant, J. (2007). Theories of fertility decline and the evidence from development indicators. *Population and Development Review*, *33*, 101–127.

Caldwell, J. C. (1986). Routes to low mortality in poor countries. *Population and Development Review*, *12*, 171–220.

Cavenaghi, S. M. (1999). A spatial-temporal analysis of fertility transition and health care delivery system in Brazil. Unpublished Ph.D. Dissertation, University of Texas-Austin.

Cleland, J., & Wilson, C. (1987). Demand theories of the fertility transition: An iconoclastic view. *Population Studies*, *41*, 5–30.

Coale, A. J. (1973). The demographic transition reconsidered. In *International population conference, Liège* (Vol. 1, pp. 53–72). Liège, Belgium: IUSSP.

A. J. Coale, S. C. Watkins (Eds.) (1986). *The decline of fertility in Europe*. Princeton, NJ: Princeton University Press.

Condran, G. A., & Preston, S. H. (1994). Child mortality differences, personal health care practices, and medical technology: The United States 1900–1930. In L. C. Chen, A. Kleinman, & N. C. Ware (Eds.), *Health and social change in international perspective* (pp. 171–225). Cambridge: Harvard University Press.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society B*, *34*, 187–220.

Cutler, D., & Miller, G. (2005). The role of public health improvements in health advances: The twentieth-century United States. *Demography*, *42*, 1–22.

Deaton, A. (2004). Health in an age of globalization. In S. Margaret Collins, C. Graham (Eds.), *Brookings trade forum 2004* (pp. 83–110). Washington, DC: Brookings Institution Press.

Guilmoto, C. Z., & Rajan, S. I. (2001). Spatial patterns of fertility transition in Indian districts. *Population and Development Review*, *27*, 713–738.

Guinnane, T. W., Okun, B. S., & Trussell, J. (1994). What do we know about the timing of fertility transitions in Europe? *Demography*, *31*, 1–20.

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*, 299–314.

Kalbfleisch, J. D., & Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika*, *60*, 267–278.

Knox, E. G. (1964). The detection of space-time interactions. *Applied Statistics*, *13*, 25–29.

Merrick, T. W. (1985). The effect of piped water on early childhood mortality in urban Brazil, 1970–1976. *Demography*, *22*, 1–24.
Montgomery, M. R.,& Casterline, J. B. (1993). The diffusion of fertility control in Taiwan: Evidence from pooled cross-section, time-series models. *Population Studies*, *47*, 457–479.
Potter, J. E., Schmertmann, C. P., & Cavenaghi, S. M. (2002). Fertility and development: Evidence from Brazil. *Demography*, *39*, 739–761.
Preston, S. H. (1975). The changing relation between mortality and level of economic development. *Population Studies*, *29*, 231–248.
Preston, S. H. (1980). Causes and consequences of mortality declines in less developed countries during the twentieth century. In R. S. Easterlin (Ed.), *Population and economic change in developing countries* (pp. 289–341). Chicago: National Bureau of Economic Research, The University of Chicago Press.
Preston, S. H. ,& Haines, M. R. (1991). *Fatal years: Child mortality in late nineteenth-century America*. Princeton, NJ: Princeton University Press.
Schmertmann, C. P., Assunção, R. M., & Potter, J. E. (2010). Knox meets Cox: Adapting epidemiological space-time statistics to demographic studies. *Demography*, *47*, 629–650.
Simões, C. C. S., & Monteiro, C. A. (1995). Tendência secular e diferenciais regionais da mortalidade infantil no Brasil/Secular trends and regional differentials in infant mortality, Brazil. In C. A. Monteiro (Ed.), *Velhos e novos males da saúde no Brasil: A evolução do país e de suas doenças*. São Paulo: Hucitec.
Soares, R. R. (2007). Health and the evolution of welfare across Brazilian municipalities. *Journal of Development Economics*, *84*, 590–608.
Tolnay, S. E. (1995). The spatial diffusion of fertility: A cross-sectional analysis of counties in the American South, 1940. *American Sociological Review*, *60*, 299–308.
Weeks, J. R., Getis, A., Hill, A. G., Saad Gadalla, M., & Rashed, T. (2004). The fertility transition in Egypt: Intraurban patterns in Cairo. *Annals of The Association of American Geographers*, *94*, 74–93.
Wood, C. H., Williams, P., & Chijiwa, K. (2007). Protestantism and child mortality in northeast Brazil, 2000. *Journal for the Scientific Study of Religion*, *46*, 405–416.

Chapter 3
Spatial Aspects of the American Fertility Transition in the Nineteenth Century

Michael R. Haines and J. David Hacker

> *Go to the West, and visit one of our log cabins, and number its inmates. There you will find a strong, stout youth of eighteen, with his better half, just commencing the first struggles of independent life. Thirty years from that time, visit them again; and instead of two, you will find in the same family twenty-two. That is what I call the American multiplication table*
> (Indiana Democrat Andrew Kennedy, 1846).

Introduction

The American fertility transition was unusual in comparison with other currently developed nations. The decline came very early; in the nation as a whole, general fertility began to fall in 1800, with marital fertility following later (Hacker, 2003). In the north-eastern parts of the nation, marital fertility rates fell earlier and some scholars date the onset of marital fertility decline to the early nineteenth century (Wells, 1971; Smith, 1987; Main, 2006). Only France among currently developed nations had a comparably early onset of a fertility transition. Levels of fertility in the United States prior to the transition were very high, starting from a crude birth rate of 55 and a total fertility rate of about 7.0 for the white population in 1800. Moreover, the fertility transition began long before the sustained decline in mortality, dating from approximately the 1870s (Haines & Steckel, 2000: Table A.2) (See Table 3.1 for a brief overview of the demographic transition in the United States). Finally, the transition occurred in a predominantly agrarian and rural nation, although birth rates fell in both rural and urban places. The United States was only about 6% urban in 1800 and 20% urban in 1860; about 75% of the labour force derived its primary support from agriculture in 1800 and 56% in 1860

M.R. Haines (✉)
Department of Economics, Colgate University, Hamilton, NY 13346, USA
e-mail: mhaines@mail.colgate.edu

M.P. Gutmann et al. (eds.), *Navigating Time and Space in Population Studies*, International Studies in Population 9, DOI 10.1007/978-94-007-0068-0_3,

Table 3.1 Fertility and mortality in the United States. 1800–2000

	Birthrate[a]		Child–woman ratio[b]		Total fertility rate[c]		Expectation of life at birth[d]		Infant mortality rate[e]	
Approx. date	White	Black[f]	White	Black	White	Black[f]	White	Black[f]	White	Black[f]
1800	55.0		1,342		7.04					
1810	54.3		1,358		6.92					
1820	52.8		1,295	1,191	6.73					
1830	51.4		1,145	1,220	6.55					
1840	48.3		1,085	1,154	6.14					
1850	43.3		892	1,087	5.42		39.5	23.0	216.8	340.0
		58.6[g]				7.90[g]				
1860	41.4		905	1,072	5.21		43.6		181.3	
		55.0[h]				7.58[h]				
1870	38.3		814	997	4.55		45.2		175.5	
		55.4[i]				7.69[i]				
1880	35.2		780	1,090	4.24		40.5		214.8	
		51.9[j]				7.26[j]				
1890	31.5	48.1	685	930	3.87	6.56	46.8		150.7	
1900	30.1	44.4	666	845	3.56	5.61	51.8[k]	41.8[k]	110.8[k]	170.3[k]
1910	29.2	38.5	631	736	3.42	4.61	54.6[l]	46.8[l]	96.5[l]	142.6[l]
1920	26.2	35.0	604	608	3.17	3.64	57.4	47.0	82.1	131.7
1930	20.6	27.5	506	554	2.45	2.98	60.9	48.5	60.1	99.9
1940	18.6	26.7	419	513	2.22	2.87	64.9	53.9	43.2	73.8
1950	23.0	33.3	580	663	2.98	3.93	69.0	60.7	26.8	44.5
1960	22.7	32.1	717	895	3.53	4.52	70.7	63.9	22.9	43.2
1970	17.4	25.1	507	689	2.39	3.07	71.6	64.1	17.8	30.9
1980	15.1	21.3	300	367	1.77	2.18	74.5	68.5	10.9	22.2
1990	15.8	22.4	298	359	2.00	2.48	76.1	69.1	7.6	18.0
2000	13.9	13.9			2.05	2.13	77.6	71.7	5.7	14.1

[a] Births per 1,000 population per annum.
[b] Children aged 0–4 per 1,000 women aged 20–44. Taken from U.S. Bureau of the Census (1975), Series 67–68 for 1800–1970. For the black population 1820–1840, Thompson and Whelpton (1933), Table 74, adjusted upward 47% for relative under-enumeration of black children aged 0–4 for the censuses of 1820–1840.
[c] Total number of births per woman if she experienced the current period age-specific fertility rates throughout her life.
[d] Expectation of life at birth for both sexes combined.
[e] Infant deaths per 1,000 live births per annum.
[f] Black and other population for CBR (1920–1970), TFR (1940–1990), e(0) (1950–1960), IMR (1920–1970).
[g] Average for 1850–1859.
[h] Average for 1860–1869.
[i] Average for 1870–1879.
[j] Average for 1880–1884.
[k] Approximately 1895.
[l] Approximately 1904.
Source: U.S. Bureau of the Census (1975, 1985, 1997, 2004). Coale and Zelnik (1963), Coale and Rives (1973), Haines (1998), Preston and Haines (1991), Steckel (1986)

(Haines & Steckel, 2000: Table 8.1; Margo, 2000: Table 5.3). This chapter uses a set of cross-sectional county-level data sets for the period 1800–1860 to look at the transition, particularly at its spatial aspects.

Contemporary Observations

Many observers of the early United States population noted its spatial aspects. Although they paid most of their attention to geographic patterns of disease and mortality, many observers commented on the apparent lack of a Malthusian 'preventive' check on population growth in the nation's sparsely populated western regions. Timothy Flint, for example, observed large numbers of children in nearly every cabin and home along the Ohio River in 1815. 'I have never seen a country to appearance more fruitful in men', he remarked. 'The process of doubling population, without Malthus and without theory, without artificial or natural wants, goes on, I am sure, on the banks of the Ohio as rapidly as anywhere in the world' (Flint, 1827: p. 29). David Ramsay attributed the 10-year doubling of the white population and 10-year trebling of the black population in the back country region of South Carolina in the first decade of the nineteenth century to early marriage, steady industry and a wholesome diet (Ramsay, 1809: p. 582). Similarly, Daniel Drake, a physician in Cincinnati, observed that the relative ease of subsistence resulted in 'early and productive' marriages (Drake, 1815: p. 27).

Explicitly or implicitly, most observers associated high fertility with early and near universal marriage, which was in turn a product of the nation's ready availability of inexpensive land and abundant resources. Benjamin Franklin was the first to point out the association of land availability and marriage: 'Land being thus plenty in America and so cheap as that a labouring Man, that understands Husbandry, can in a short Time save Money enough to purchase a Piece of new Land sufficient for a Plantation, whereon he may subsist a Family; such are not afraid to marry; for if they even look far enough forward to consider how their Children when grown up are to be provided for, they see that more Land is to be had at Rates equally easy, all Circumstances considered. Hence Marriages in America are more general and more generally early, than in Europe' (Quoted in McCusker & Menard, 1985: p. 212). Beginning in the nineteenth century, however, observers hinted that differentials in the 'productivity' of marriages resulted from regional differentials in the practice of conscious family limitation strategies. In 1827, George Tucker, Professor of Moral Philosophy at the University of Virginia, mentioned a secret 'long practised by many in the east' that allowed couples to control fertility without abstinence (Tucker, 1827). In a subsequent statistical analysis of the 1790–1840 censuses, Tucker documented a long-term decline in the proportion of children to women in every state between 1800 and 1840, a finding he characterized as 'an interesting fact which seems never to have been suspected', and an inverse relationship between the proportion and population density. To these early observers, the dimension of space seemed the most relevant (Tucker, 1855: pp. 104–106).

Literature

A number of hypotheses have been advanced to explain the antebellum fertility transition in the United States. Beginning with Yasuba's (1962) path-breaking study of white fertility ratios in the United States between 1800 and 1860, historians have been attracted to the study of the nineteenth-century fertility transition (e.g. T'ien 1959; Forster & Tucker, 1972; Easterlin, 1976; Sundstrom & David, 1988; Leet, 1976; Schapiro, 1986; Haines & Hacker, 2006). A variety of factors have been examined to explain reductions in fertility. One hypothesis, which receives some support for the early nineteenth century, is that nuptiality adjustments accounted for much of the decline. A variety of other hypotheses, including diminished land availability, denser local labour markets, better life-cycle savings options, fundamental structural modernization and changes in ideology, all receive some support. The more traditional structural view, consistent with a shift in life-cycle saving from children and land to other forms of 'wealth' appears more prominent by 1850 and 1860.

Although the fertility decline in the United States in the nineteenth century has been extensively studied, space has been used only as a source of variation to estimate structural models. Little has been done to look at actual spatial diffusion. Indeed, much analysis in demographic and economic history has used space as a source of variation to be explained, not as a phenomenon to be accounted for in and of itself. One exception has been the work of Easterlin, Alter and Condran (1978) who divided the country into regions by the timing of settlement (from frontier to longest settled areas).

Data

A database now exists that can exploit both temporal and spatial aspects of this transition. On an ecological level, we have county-level files from 1800 to 1860 from which age-specific child–woman ratios can be calculated. The child–woman ratio is, of course, the proxy used for the American experience, since vital statistics are not available for the entire United States until 1933. It clearly suffers from the differential mortality and under enumeration of women and children and conflates the separate effects of marriage and marital fertility. Nonetheless, it is a useful indicator of differential fertility across space and time for geographic areas.

The data set used in the first part of the analysis is a compilation of (mostly) published county-level statistics for the United States from 1790 to the present (Haines, 2004). The starting point was the ICPSR data set 0003 'Historical, Demographic, Economic and Social Data: The United States, 1790–1970'.[1] To this was added the urban population of each county, obtained from the original, unpublished worksheets prepared at the U.S. Bureau of the Census in the 1930s. The basic population census data are supplemented by agricultural and manufacturing data (1840–1960), as well as additional variables on the urban population (from the original Census Bureau worksheets), area (allowing density calculations), the availability of water and/or railroad connections (1840–1860) and geographic location (latitude and longitude of county seats). Before 1900, county areas only appeared in connection with the 1880 U.S. census. Two sources were used to obtain areas at earlier dates. The first is a collection of historical atlases of counties by state (Long, 2001). We make use of these atlases for the District of Columbia, for the 21 states that have been published and for New Jersey, for which John Long kindly furnished the manuscript worksheets. For all other states and territories, we used Arc View to calculate areas from the 'Historical United States County Boundary Files' (Earle, Cao, Heppen, & Otterstrom, 1999). The data on water and rail connections in 1840 through 1860 were drawn from Craig, Palmquist and Weiss (1998).

Other modifications were made to the ICPSR data. All territories were included, as was the District of Columbia. Checks were performed for errors in the data. All data from the Censuses of Agriculture of 1840–1860 have been added, as have some additional data from the Censuses of Manufactures. The data on churches were merged with the other data. Latitudes and longitudes of the county seats were added from ICPSR Study Number 8159. An effort was made to augment the data for counties not originally included, although this was not possible for some very remote counties in early censuses. Individual level census data from the IPUMS files for 1850 and 1860 were used to estimate women's nuptiality for 1850 and 1860 and women's labour force participation for 1860. For details, see Haines and Hacker (2006).

There exists a clear pattern of spatial gradients in our measure of fertility, the child–woman ratio. Maps suggest that the distribution of child–woman ratios was not random. Figures 1 through 7 depict the child–woman ratios by county from 1800 to 1860, respectively. This pattern was, of course, the basis for the land availability hypothesis of Yasuba (1962), Forster and Tucker (1972), Easterlin (1976), Schapiro (1986) and Leet (1976), for the local labour market hypothesis of Sundstrom and David (1988) and for the life-cycle savings hypothesis of Carter, Ransom and Sutch (2004). All of these rely on some version of an intergenerational transfer model coupled with the potential risk of child default. In a society without adequate financial markets and intermediaries and with a heavy reliance on wealth holding in the form of real property (land, structures, livestock, equipment), provision for old age was much more likely in the form of children. In order to induce children to remain in the proximity of parents and provide care for them and work the farm or family business (if the parents were landowning farmers or business proprietors), inducements in the form of real property via inheritance or inter-vivos transfers were necessary. But as areas became more developed, with more urban populations and non-agricultural activity as well as transport connections, it became easier to rely on life-cycle saving in other forms, for example savings in banks and annuities.

The land availability hypothesis was originally put forward by Yasukichi Yasuba (1962: ch. 5), who found a consistent rank-order correlation between population density and the census child–woman ratio by state for the white population for the censuses from 1800 to 1860. He proposed that the rising cost of providing land or an inheritance for children caused by increased land prices (due to rising density and urbanization) led families to limit fertility. The hypothesis was refined to a correlation with availability of agricultural land by Forster and Tucker (1972), based on state-level data supplemented with county-level data for New York and Virginia. Implicit in the land availability hypothesis is the concept of intergenerational transfers of real property from parents to children in order to keep children near the parents. A further implication is that old age insurance was largely in the form of children to care for and protect aged parents. Easterlin (Easterlin 1976; Easterlin et al., 1978) further refined the concept and used microdata from the 1860 Northern Farms Sample (Bateman & Foust, 1974) to show that the gradient of fertility from the longest settled areas to the frontier was not monotonic. Children were less valuable for clearing land on the frontier, but family size increased after the settlement period. Further confirmation of this hypothesis was provided by Schapiro (1986), who extended the analysis to 1900 and used still more refined independent variables. McInnis (1977) used small area and individual level data from the manuscripts of the Canadian censuses to demonstrate that the same phenomenon was true in mid-nineteenth-century Ontario. A county-level study was undertaken for the state of Ohio by Leet (1976), who found results that supported the land availability hypothesis, although he also noted the importance of sex ratios, educational variables and the regional composition of the population.

Another explanation, not necessarily exclusive of the first, is the proximity of other alternatives for children, notably non-agricultural employment, especially in growing urban centres. This also embodies the notion that parents were seeking to reduce the risk of child default (that is, children moving far enough away so as to be unable to provide old age care). This view was put forward by Sundstrom and David (1988) in the form of an intergenerational bargaining model, which they contrasted to the land availability model as a homeostatic theory of human fertility (Smith, 1977). They argued that a more favourable ratio of non-agricultural to agricultural wages in a region would lead to a higher risk that children would leave the area close to the parents. An adaptation by the parents would be a larger 'bribe' in terms of property, both real and financial, and smaller families would be necessary to achieve that result. Although Sundstrom and David pose this as an exclusive alternative to the land availability hypothesis, it is a complement rather than a substitute for the traditional theory.

Carter et al. (2004) generally agree with this model, but also note that other life-cycle factors, such as increasing rates of school attendance, made children economically costly for farm parents. They reject the target-bequest model implied by the land availability hypothesis and stress the growing importance of alternative forms of saving and wealth accumulation over the life cycle. This hypothesis is supported by the availability of local financial intermediaries (banks), which was suggested in a paper by Steckel (1992) using state-level data for 1850. Ruggles

(2003) notes, however, that a very high proportion of elderly persons were living with children in the latter half of the nineteenth century, indicating that there was no large-scale child default. In 1850, for example, about 70% of elderly individuals were residing with a child or children (Ruggles, 2003).

Still another hypothesis stresses the ideational view of fertility transition (Lesthaeghe, 1980, 1983; Smith, 1987; Hacker, 1999). Interest in ideational causes grew out of the finding that European nations at very different levels of socio-economic development (levels of urbanization, share of non-agricultural employment in the labour force, levels of literacy) commenced their irreversible fertility transitions within a short period of time in relation to one another (roughly 1870–1920) (Knodel & van de Walle, 1979). This argues that the growing influence of secular values has changed people's willingness to control and plan family size. As an example, Lesthaeghe (1977) found that the best predictors of the timing of fertility decline in Belgium were the percentages voting socialist, liberal and communist in 1919 (positively related to early fertility decline) and the proportion of the population paying Easter dues in the Roman Catholic Church (negatively related to early fertility decline). Leasure (1982) proposed that greater adherence to religious denominations that encouraged individualism and a positive role for women in the nineteenth century (Congregationalist, Unitarians, Universalist, Presbyterians, Society of Friends) would result in earlier and more rapid fertility declines. Wells (1971) identified early fertility decline among American Quakers. Smith (1987) found support for this argument with a study of child–woman ratios in 1860 and Hacker (1999) has demonstrated similar results with individual level data drawn from the 1850 and 1880 IPUMS samples. Hacker further observed that parents' reliance on biblical names for their children – which he suggested as a possible proxy for parental religiosity – was positively correlated with marital fertility. Haan (2005) has observed a similar positive relationship between biblical names and marital fertility in a sample of the Canadian census of 1881 – which includes a direct question on religious affiliation – although he cautions that the use of biblical names does not correlate with 'strict' and 'liberal' denominations in expected ways. Gutmann (1990) and Gutmann and Fliess (1993) wrote about differences in fertility decline between Protestant and Catholic ethnic Germans in Texas, revealing significant differences in timing despite broad cultural similarities.

While stressing various hypotheses, most historians contend that traditional structural variables from standard demographic transition theory (urbanization; industrialization; increased education, especially of women; increased women's work outside the home) played a supporting role in U.S. fertility decline (Notestein, 1953). Vinovskis (1976) noted that interstate fertility differentials were well explained by urbanization and literacy in 1850 and 1860 and that the effects of these variables strengthened over the nineteenth century. He also noted that urban child–woman ratios declined in parallel with rural ratios, casting some doubt on the sufficiency of the land availability hypothesis to explain the antebellum fertility transition in the United States.

An earlier and often overlooked, paper by T'ien proposed that sex ratios (males per 100 females) could be useful in explaining differentials and changes over time

(T'ien 1959) because a surplus of males would create a more favourable marriage market for females, leading to earlier and more universal marriage. Since overall fertility was largely a function of marital fertility in the white population of the United States in the nineteenth century (and indeed much of the twentieth century), this would raise fertility, which is based on the married female population. The surplus of males arose because of migration patterns to the frontier areas selective of adult males. This is a conventional demographic explanation, which attempts to get at the problem of separating the effects of marriage and marital fertility. Using the IPUMS samples of 1850 to 1880, Hacker (2003) re-estimated total fertility rates back to 1835 and proposed that most of the fertility transition before 1840 was due to adjustments in marriage. Recent work by Haines and Guest (2008) suggests, however, that some evidence of parity-specific control of marital fertility can be found in New York State earlier in the nineteenth century.

Overall, this chapter provides a county-level study for the period 1800–1860 with a variety of independent variables. Population density, improved farm land, farm land values and urbanization are used to look at the land availability hypothesis; sex ratios and estimates of women currently married (for 1850 and 1860) are used to examine the marriage market hypothesis; composition of churches (1850 and 1860) is used to test the ideational hypothesis; and urbanization, labour force composition, transportation, illiteracy and migrant composition are used to examine both the structural and the life-cycle hypotheses.

Analysis of Fertility Ratios

Table 3.2 shows some of the variation in child-woman ratios across census regions and over time in the period between 1800 and 1860. The east-west differences are apparent throughout and it is clear that New England was the area with the lowest child–woman ratios. There is a less apparent suggestion of a north-south gradient, with the South having had higher child–woman ratios. Urban places (incorporated areas of 2,500 or more persons) had lower child–woman ratios than did rural areas, but the decline in the child–woman ratios took place in both rural and urban areas between 1800 and 1840 (the period for which we have complete census tabulations for minor civil divisions).

The variables used in the structural analysis of county-level variation in child–woman ratios are defined in Table 3.3. These variables were placed in a set of unweighted OLS multivariate regressions to account for differences in child–woman ratios across counties from 1800 to 1860. Because of significant heteroskedasticity, caused by differences in county populations, White-corrected robust standard errors were used. The results are given in Table 3.4. The regressions explain more than 50% of the variation in fertility ratios across counties, with one exception (rural counties in 1860). Urbanization was consistently and negatively related to fertility ratios. When density was in the same equation (upper panel of Table 3.4), it was not significant, although it was negative and statistically significant if urbanization was not included in the specification. In equations for rural counties, the density

Table 3.2 Number of children under 5 years old per 1,000 women aged 20–44 years, by race, residence and region. United States, 1800–1860

Year	1800	1810	1820	1830	1840	1850	1860
Region, Residence, Race							
United States, white population, adjusted	1,342	1,358	1,295	1,145	1,085	892	905
United States, black population, adjusted	–	–	–	–	–	1,087	1,072
United States, white population	1,281	1,290	1,236	1,134	1,070	877	886
United States, urban white population	845	900	831	708	701	–	–
United States, rural white population	1,319	1,329	1,276	1,189	1,134	–	–
New England, white population	1,098	1,052	930	812	752	621	622
New England, urban white population	827	845	764	614	592	–	–
New England, rural white population	1,126	1,079	952	851	800	–	–
Middle Atlantic, white population	1,279	1,289	1,183	1,036	940	763	767
Middle Atlantic, urban white population	852	924	842	722	711	–	–
Middle Atlantic, rural white population	1,339	1,344	1,235	1,100	1,006	–	–
East North Central, white population	1,840	1,702	1,608	1,467	1,270	1,022	999
East North Central, urban white population	–	1,256	1,059	910	841	–	–
East North Central, rural white population	1,840	1,706	1,616	1,484	1,291	–	–
West North Central, white population	–	1,810	1,685	1,678	1,445	1,114	1,105
West North Central, urban white population	–	–	–	1,181	705	–	–
West North Central, rural white population	–	1,810	1,685	1,703	1,481	–	–
South Atlantic, white population	1,345	1,325	1,280	1,174	1,140	937	918
South Atlantic, urban white population	861	936	881	767	770	–	–
South Atlantic, rural white population	1,365	1,347	1,310	1,209	1,185	–	–
East South Central, white population	1,799	1,700	1,631	1,519	1,408	1,099	1,039
East South Central, urban white population	–	1,348	1,089	863	859	–	–
East South Central, rural white population	1,799	1,701	1,635	1,529	1,424	–	–
West South Central, white population	–	1,383	1,418	1,359	1,297	1,046	1,084
West South Central, urban white population	–	727	866	877	846	–	–
West South Central, rural white population	–	1,557	1,522	1,463	1,495	–	–
Mountain, white population	–	–	–	–	–	886	1,051
Mountain, urban white population	–	–	–	–	–	–	–
Mountain, rural white population	–	–	–	–	–	–	–
Pacific, white population	–	–	–	–	–	901	1,026
Pacific, urban white population	–	–	–	–	–	–	–
Pacific, rural white population	–	–	–	–	–	–	–

Source: Carter et al., 2006, Table Ab315–346.
(a) Adjusted data standardized for age of women and allowance made for undercount in censuses; see text.

coefficient remained negative and significant throughout. These results give greater support to the labour market view of Sundstrom and David than to the land availability hypothesis, but the lower child–woman ratios in urban areas may also reflect higher infant and child mortality and lower nuptiality. They also support the Carter, Ransom and Sutch hypothesis of the growing importance of life-cycle saving. Nevertheless, other variables in the models bolster the land availability view. Average farm values per acre and the per cent of farmland which was improved both had negative and significant coefficients in 1850 and 1860 for all counties,

Table 3.3 Variable names and descriptions

Variable	Description
WHCWRAT	Child–woman ratio, white population: 1800–1820: children aged 0–9 years per 1,000 women aged 16–44 years 1830–1860: children aged 0–9 years per 1,000 women aged 15–49 years
WHSEXRAT	Sex ratio, white population: White males per 100 white females (all ages)
DENSITY	Population density: Persons per square mile
PCTURB	Percent urban (in places 2,500 and over)
PCTURB25	Percent of population in places 25,000 and over.
PCTNW	Percent of total population non-white
PCTNONAG	Estimated percent of the labour force in non-agricultural activity
PCTWHILL	Percent of white population aged 20 and over who were unable to read and write
PCTFOR	Percent of the total population foreign born
PCTMFGLB	Estimated percent of the white population aged 15–69 employed in manufacturing
PRELGNEW	Percent of all church accommodations Congregationalist, Presbyterian, Unitarian and Universalist
RELWAGE	Ratio of estimated monthly wages of a common labourer with board to the monthly wages of a farmhand with board (States only)
TRANSPOR	Variable = 1 if the county was on a canal, river or other navigable waterway in 1840. Otherwise = 0. For 1850 and 1860, variable = 1 if county on a railroad or navigable waterway. Otherwise = 0
WEALTHPC	1850: Value of real estate per free person 1860: Value of real and personal estate per free person
FARVALAC	Average value of farm per acre (improved and unimproved)
PCTACIMP	Percent of farm acres improved
WCURRMARR	Proportion of women currently married
WLABFORCE	Proportion of single women in the labour force
LATITUTDE	Latitude of the county seat
LONGITUDE	Longitude of the county seat

Source: See text.

consistent with the likelihood that higher land values and more settled agriculture created incentives to reduce family size. In 1850, however, the coefficient on value per acre was positive in rural counties and the same coefficient was not statistically significant for 1860. These results contradict the findings of poor performance of the land availability hypothesis by Carter et al. (2004). On the other hand, the labour market hypothesis, as well as the conventional structural view, receive some backing from the negative and significant coefficients on the per cent of the labour force in non-agricultural activities (1820 and 1840) and the strong effect of the transportation variable (1840, 1850 and 1860). The positive and significant coefficients on adult white illiteracy (1840 and 1850) also support the structuralist perspective. The relative wage variable provides no confirmation of the labour force theory and indeed is even positive (the opposite of the expected sign) and significant in 1860. It is, however, a state-level variable.[2]

Table 3.4 Regression results. White child–woman ratios as the dependent variable. counties. United States, 1800–1860

Year	1800		1810		1820		1830		1840		1850		1860	
Variable														
	(coef)	(signi)	(coef)	(signi)	(coef)	(signi)	(coef)	(signi)	(coef)	(signi)	(coef)	(signi)	(coef)	(signi)
ALL COUNTIES														
CONSTANT	448.665	*	1,472.231	***	1,502.751	***	1,094.992	***	1,310.230	***	1,039.793	***	491.895	***
WHSEXRAT	13.186	***	2.243	***	1.371	–	0.979	–	−0.028	–	−0.204	***	−0.048	–
DENSITY	−0.072	–	−0.018	–	0.017	–	−0.009	–	0.006	*	0.041	–	0.069	***
PCTURB	−5.997	***	−6.219	***	−2.322	–	−5.748	***	−2.703	***	−2.354	***	−3.304	***
PCTNW	−11.794	***	−10.433	***	−9.104	***	−7.890	***	−4.378	***	−2.799	***	−2.685	***
PCTNONAG					−6.568	***	−4.283	***						
TRANSPOR									−84.107	***	−73.457	***	−42.938	***
PCTFOR											−6.142	***	−0.354	–
PCTWHILL									2.788	***	1.144	–		
PCTMFGLB											−0.456	–	0.642	–
WEALTHPC											−0.167	**	−0.005	–
FARVALAC											−0.498	–	−0.449	**
PCTACIMP											−2.213	***	−2.478	***
PRELGNEW											−1.899	***	−2.245	***
RELWAGE											1,469.313	–	287.878	***
WCURRMAR											411.257	***	498.504	***
WLABFORC													27.411	–
REGIONS:														
New England	NI		NI		NI		NI		NI		NI		NI	
Middle Atlantic	171.475	***	320.719	***	358.403	***	295.984	***	215.754	***	193.926	***	173.863	***
East North Central	165.523	*	576.659	***	606.578	***	665.590	***	416.531	***	334.414	***	325.132	***

Table 3.4 (continued)

Year	1800		1810		1820		1830		1840		1850		1860	
Variable														
West North Central			753.566	***	767.102	***	962.821	***	592.177	***	426.945	***	379.783	***
South Atlantic	464.658	***	547.060	***	576.073	***	621.370	***	376.142	***	246.043	***	338.738	***
East South Central	727.827	***	782.528	***	779.338	***	841.779	***	588.406	***	351.642	***	406.174	***
West South Central			73.471	–	791.707	***	904.731	***	704.285	***	432.111	***	502.085	***
Mountain											79.635	–	694.100	***
Pacific											596.241	***	662.734	***
R-squared	0.6609		0.6197		0.6191		0.6186		0.6201		0.5804		0.5703	
F-ratio	77.21	***	129.00	***	172.11	***	225.32	***	200.57	***	191.10	***	178.42	***
N	417		571		753		982		1,235		1,611		2,012	
RURAL COUNTIES														
CONSTANT	964.166	***	1,775.454	***	1,728.668	***	1,550.457	***	1,450.343	***	1,093.801	***	566.353	***
WHSEXRAT	9.822	***	1.597	***	0.392	–	−0.801	**	−0.329	–	−0.194	***	−0.078	–
DENSITY	−5.438	***	−6.516	***	−5.277	***	−7.202	***	−4.082	***	−4.957	***	−2.516	**
PCTNW	−9.973	***	−8.271	***	−7.439	***	−5.963	***	3.715	***	−2.770	***	−2.760	***
TRANSPOR									−71.712	***	−59.423	***	−30.951	***
PCTFOR											−7.470	***	−0.593	–
PCTNONAG					−4.058	***			−3.536	***				
PCTWHILL									2.136	***	0.284	–		
PCTMFGLB											−0.298	–	−0.004	–
WEALTHPC											−0.134	–	−0.004	–
FARVALAC											3.793	***	−0.261	–

Table 3.4 (continued)

Year	1800		1810		1820		1830		1840		1850		1860	
Variable														
PCTACIMP											−1.560	***	−1.528	**
PRELGNEW											−1.532	***	−2.053	***
RELWAGE											1,316.368	–	273.665	**
WCURRMAR											379.561	***	479.481	***
WLABFORC													−37.624	–
REGIONS:														
New England	NI		NI		NI		NI		NI		NI		NI	
Middle Atlantic	137.887	***	234.900	***	330.997	***	281.106	***	241.849	***	209.394	***	198.077	***
East North Central	85.960	–	396.357	***	525.587	***	515.321	***	376.269	***	348.990	***	342.347	***
West North Central			497.595	***	626.466	***	717.227	***	508.558	***	400.720	***	379.437	***
South Atlantic	338.108	***	353.766	***	479.301	***	425.473	***	328.030	***	270.177	***	356.350	***
East South Central	609.344	***	595.117	***	680.179	***	647.410	***	539.973	***	369.748	***	422.093	***
West South Central			−23.765	–	633.979	***	632.418	***	594.604	***	394.016	***	512.818	***
Mountain											21.497	–	636.796	***
Pacific											518.619	***	669.294	***
R-squared	0.6638		0.6444		0.6110		0.6646		0.5876		0.5375		0.4997	
F-ratio	85.95	***	118.52	***	116.11	***	182.82	***	131.40	***	94.32	***	93.75	***
N	396		540		709		919		1,137		1,434		1,721	

Note: These regressions were estimated with White-corrected (robust) standard errors to correct for heteroskedasticity.
***: Significant at least at a 1% level; **: Significant at least at a 5% level; *: Significant at least at a 10% level; ---: not significant at least at a 1% level.

The sex ratio, a proxy for the marriage market, showed the expected positive and significant effects early in the nineteenth century, but that effect gradually diminished and even became negative by 1840. This finding supports the demographic perspective that adjustments in nuptiality had a large role in the early stages of the fertility transition (T'ien 1959).

The hypothesis that the spread of liberal and secular ideas was influential also finds some confirmation (Lesthaeghe, 1977, 1980, 1983). The religion variable (PRELGNEW) was negative and significant in a multivariate framework. Counties with a higher proportion of liberal and individualistic denominations (Congregationalist, Presbyterian, Unitarian, Universalist) were more likely to have had lower fertility ratios, holding region and economic and demographic structure constant. The level of wealth per free person seems to have had little impact on fertility ratios, but the per cent foreign born by county had a negative and significant relation to fertility ratios, even holding urbanization constant. This result is puzzling, given Spengler's (1930) finding that the foreign born often had higher birth rates, but the mass migrations from Europe in the 1840s and 1850s had some disruptive effects.

Table 3.4 also includes two additional variables, WCURRMAR and WLABFORC, constructed from the 1850 and 1860 IPUMS samples (Ruggles et al. 1997). These individual-level, 1% samples of the original manuscript census records allow estimation of women's current marital status and labour force participation at the county level or at higher levels of aggregation. For each county, we calculated the percentage of women aged 20–49 who were imputed as having a spouse present in the household (see Ruggles (1995) for details of the imputation procedure). We repeated the aggregation procedure for each state economic area (SEA, an aggregation of contiguous counties identified by the 1950 census sharing similar economic characteristics) and state. If there were enough women in the county aged 20–49 to obtain a reasonably accurate estimate (using an arbitrary cut-off of at least 30 cases), we attached the estimate to the county-level data set. If there were not enough cases, we relied on the SEA- or state-level estimates. We also constructed a variable measuring women's economic opportunity by aggregating the percentage of single women aged 20–49 currently in the paid labour force. The 'independence' and 'gains to marriage' theories contend that men and women will increasingly postpone marriage or disrupt current marriages as educational attainment and job opportunities for women improve. The relative lack of economic opportunity for young southern white women outside the home, for example, may have increased the cultural incentive for marriage in the antebellum South and helped to boost the region's fertility (Hacker, 2006). Women's nuptiality (WCURRMAR) was consistently significant, with large effects for both 1850 and 1860. This variable modestly improved the performance of the models; the adjusted R-squared for all counties in 1850, for example, increased from 0.561 to 0.575. Its inclusion, however, resulted in only modest changes to the coefficients of the existing variables. None of the coefficients changed sign and all but one of the coefficients significant at the .05 level in the earlier models remained significant with the addition of women's nuptiality. The one exception was the white sex ratio in the 1860 models, whose coefficient only remained significant at the 10% level when women's nuptiality was included, suggesting that the sex ratio was a reasonable proxy for marriage in the 1800–1840 models. Finally, women's economic opportunity, while having the expected negative sign, did not prove to be statistically significant.

So where does this analysis leave us? It is clearly difficult to exclude different hypotheses, but there are patterns over both space and time. Marriage markets and adjustments in nuptiality seem more important in the early nineteenth century, while adjustments in marital fertility appear more influential later. These are proximate causes of fertility and are probably related to the technology of family limitation, which we can think of as the supply side. What about the demand for children? The life-cycle saving model holds an important place here. Nineteenth century American adults wanted to save, either in the form of real wealth (land, structures, livestock, equipment), financial wealth (largely bank accounts, but also other instruments) or children. Children could care for ageing parents and manage the farm or business. A number of births were necessary to insure the survival of a child and wealth transfers across generations promoted willingness. Over time, the amount of wealth necessary increased because children had better alternatives to remaining with parents, for example, labour market opportunities in an increasingly dense network of urban places. The cost of land rose with rising population density, urban growth and the spread of market agriculture with transport improvements; it thus became necessary to reduce the number of children to account for the increased 'cost' of these transfers. Here the more traditional structural models provide real evidence. The spread of banks and financial alternatives was also relevant, since financial institutions provided an alternative to saving in the form of real property or children. For example, in 1800 there were 28 state banks (and the First Bank of the United States), located almost entirely in larger cities. By 1860, there were 1,562 state banks, much more widely spread across the country (Carter et al., 2006: Series Cj142). The antebellum American population was not homogeneous with respect to tastes and preferences, as indicated by the regional variety of churches. The antebellum fertility transition was the outcome of complex processes, many of which operated simultaneously and at different levels over both time and space.

Analysis of Spatial Variation

An analysis of the spatial characteristics of antebellum American fertility ratios is presented in Table 3.5. Just knowing the geographic region of the country could explain between 26 and 46% of variation in child–woman ratios across counties. Knowing longitude alone could explain between 23 and 57% of that variation. Thus, the obvious east-west gradient in fertility ratios seen in Figures 3.1–3.7 is confirmed. In contrast, the north-south gradient was rather weak. Knowing latitude alone or whether the county was in the South (i.e. in the South Atlantic, East South Central or West South Central census regions) explained very little of that variation and the fit (as given by the F-ratio) was either weak or statistically insignificant.

Table 3.5 Spatial regression results. White child–woman ratios as the dependent variable. Counties. United States, 1800–1860

Year	1800		1810		1820		1830		1840		1850		1860	
Variable														
	(coef)	(signi)	(coef)	(signi)	(coef)	(signi)	(coef)	(signi)	(coef)	(signi)	(coef)	(signi)	(coef)	(signi)
Region alone														
Constant	1,704.0	***	1,623.1	***	1,422.5	***	1,118.1	***	1,072.5	***	926.1	***	889.4	***
New England	NI		NI		NI		NI		NI		NI		NI	
Middle Atlantic	284.9	***	356.4	***	395.5	***	329.4	***	232.2	***	232.8	***	223.8	***
East North Central	583.1	***	688.0	***	753.3	***	746.0	***	561.9	***	517.5	***	453.9	***
West North Central			630.7	***	807.5	***	912.6	***	727.9	***	697.9	***	583.5	***
South Atlantic	147.2	***	208.0	***	373.1	***	371.0	***	415.9	***	409.3	***	411.0	***
East South Central	738.4	***	684.7	***	761.2	***	728.6	***	692.5	***	562.6	***	502.1	***
West South Central			–224.0	***	609.7	***	684.4	***	722.2	***	645.0	***	668.0	***
Mountain											343.1	***	739.9	***
Pacific											350.4	***	665.1	***
Adjusted R-squared	0.304		0.367		0.410		0.460		0.388		0.312		0.261	
F-ratio	46.86	***	56.82	***	89.03	***	141.92	***	135.63	***	93.02	***	92.09	***
N	422		580		761		994		1,275		1,622		2,069	
Latitude Alone														
Constant	2,486.6	***	2,450.9	***	3,025.3	***	2,542.2	***	2,731.2	***	2,078.7	***	1,763.5	***
Latitude	–14.0	**	–12.6	***	–29.2	***	–23.9	***	–30.7	***	–18.0	***	–10.5	***
Adjusted R-squared	0.008		0.012		0.078		0.062		0.136		0.063		0.028	
F-ratio	4.40	**	7.74	***	64.59	***	65.76	***	200.97	***	109.09	***	60.72	***
N	416		567		757		979		1,275		1,616		2,046	

Table 3.5 (continued)

Year	1800		1810		1820		1830		1840		1850		1860	
Variable														
Constant	1,910.6	***	1,974.5	***	1,908.0	***	1,649.6	***	1,523.1	***	1,373.9	***	1,349.8	***
South	67.4	–	–0.2	–	39.2	–	–4.3	–	105.3	***	61.6	***	38.5	***
Adjusted R-squared	0.004		–0.002		0.002		–0.001		0.029		0.011		0.004	
F-ratio	2.54	–	0.00	–	2.23	–	0.04	–	38.82	***	19.33	***	9.43	***
N	422		580		761		994		1,275		1,622		2,069	
Longitude Alone														
Constant	–2,696.5	***	–1,473.0	***	–1,595.3	***	–2,071.1	***	–1,415.8	***	327.0	***	227.8	***
Longitude	15.9	***	43.2	***	43.5	***	45.2	***	35.9	***	12.6	***	13.0	***
Adjusted R-squared	0.379		0.355		0.442		0.567		0.465		0.143		0.233	
F-ratio	254.02	***	312.76	***	600.80	***	1,278.86	***	1,106.80	***	269.85	***	620.71	***
N	416		567		757		979		1,275		1,616		2,046	
Moran's I	0.2699		0.2927		0.2776		0.2625		0.2394		0.1574		0.0844	
Z-score	65.98	***	95.90	***	121.96	***	152.29	***	183.51	***	149.03	***	34.5670	***
Getis-Ord General G	0.0042		0.0037		0.0034		0.0032		0.0030		0.0027		0.0023	
Z-score	–5.88	***	–3.01	***	2.72	***	5.68	***	10.63	***	8.09	***	1.87	**
Clustering (Values)	Low		Low		High		High		High		High		High	

***: Significant at least at a 1% level; **: Significant at least at a 5% level; *: Significant at least at a 10% level; ---: not significant at least at a 1% level

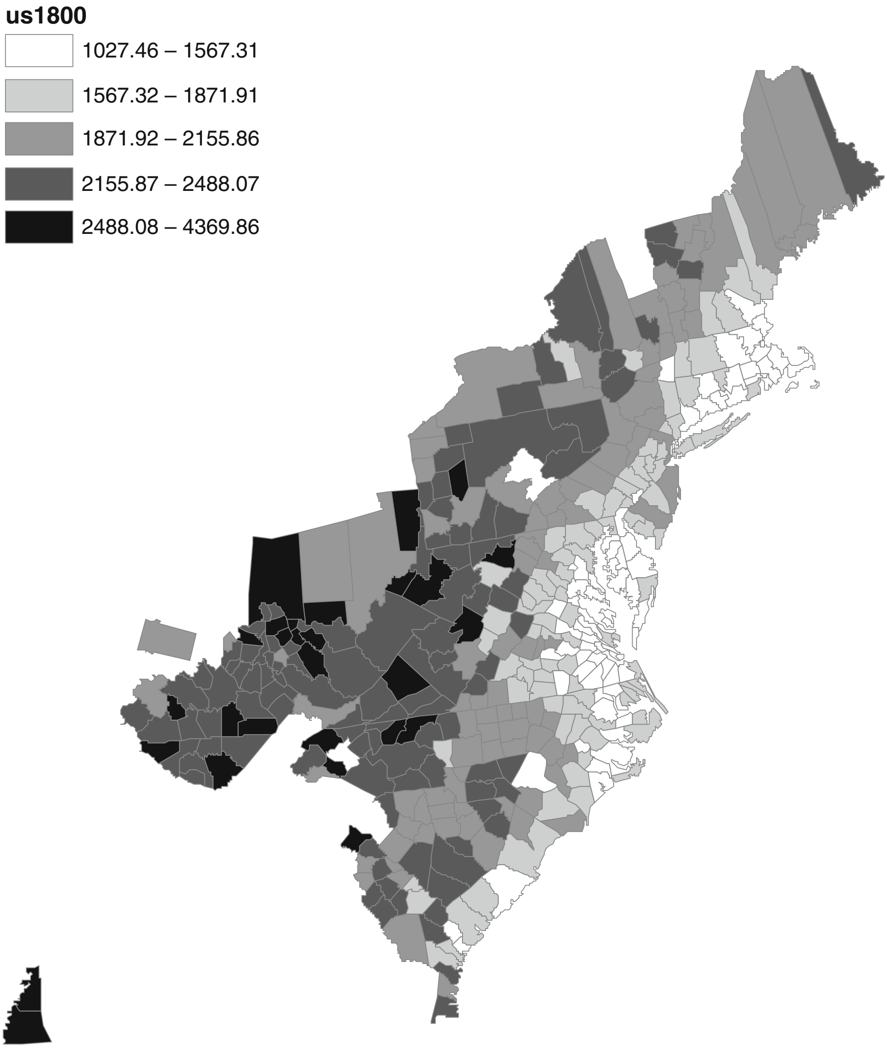


Fig. 3.1 White child–woman ratios. United States, counties, 1800

Perhaps this is not surprising, although it does confirm the higher fertility at the frontier. Basically, in the nineteenth century Americans moved from east to west along many latitudes to the frontier (Steckel, 1983). The higher fertility of the American South, which appears later in the nineteenth century (Carter et al., 2006: Table Ab315-346), is not supported for the antebellum period.

Table 3.5 also includes two measures that analyse spatial patterns: Moran's *I* and the Getis-Ord General *G*. Moran's *I* measures spatial autocorrelation (feature similarity) based not on feature locations or attribute values alone, but on both

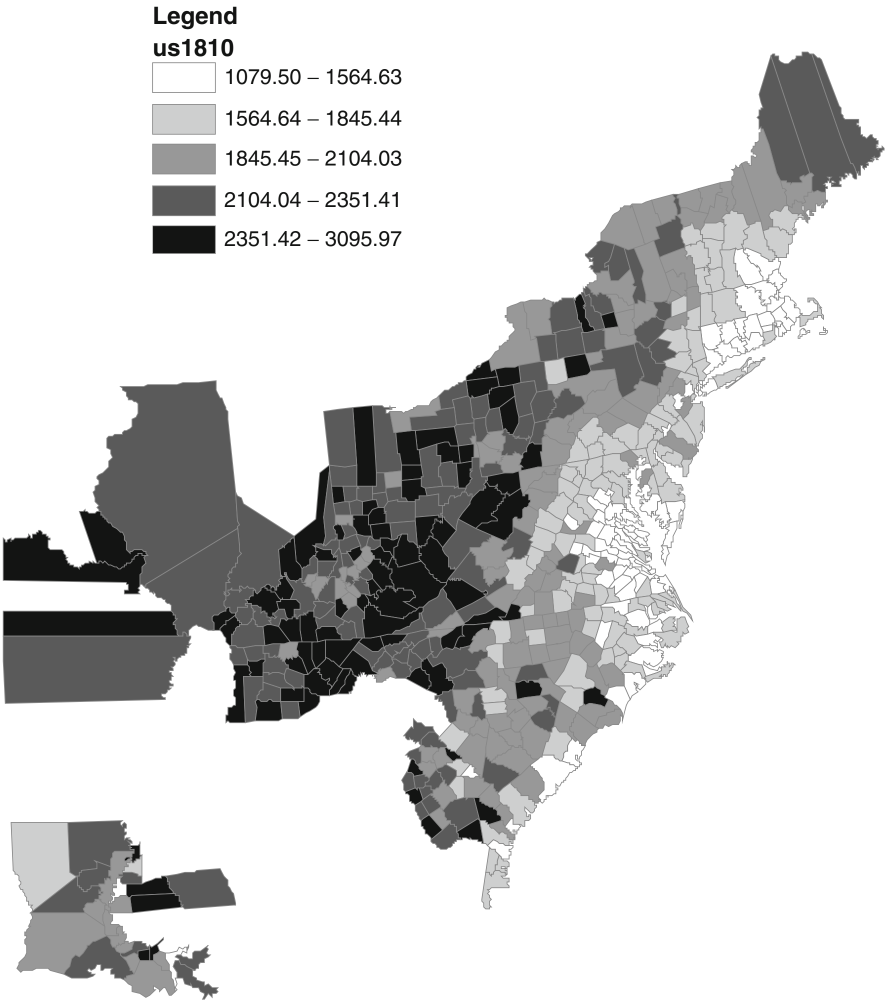


Fig. 3.2 White child–woman ratios. United States, counties, 1810

feature locations and feature values simultaneously. Given a set of features and an associated attribute, it evaluates whether the pattern expressed is clustered, dispersed or random. A Moran's Index value near +1.0 indicates clustering; an index value near –1.0 indicates dispersion. A Z score (interpreted in the same way as a t-value) is also calculated to assess whether the observed clustering or dispersion is statistically significant. The Getis-Ord General *G* (or High/Low Concentration Measure) measures concentrations of high or low values for an entire study area. A high index value indicates that high values are clustered within the study area. A low index value indicates that low values tend to cluster. A Z score on all standard deviations is calculated to determine whether the high–low index value is statistically significant.

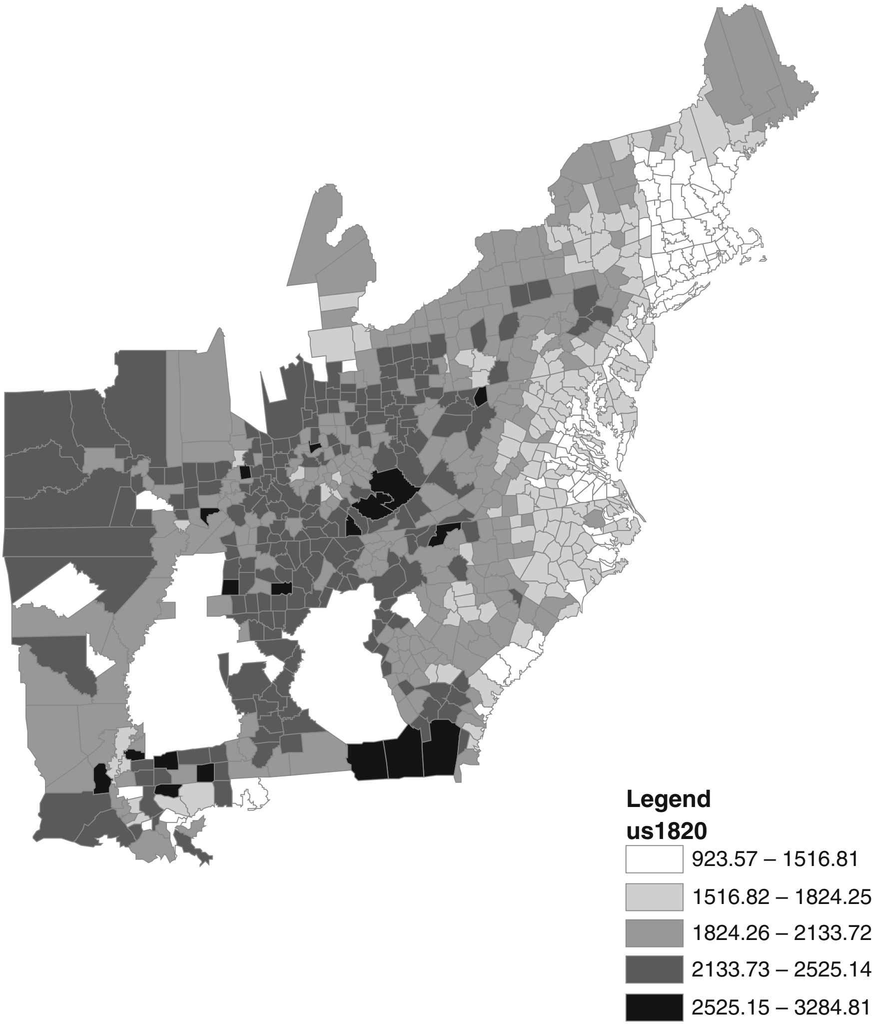


Fig. 3.3 White child–woman ratios. United States, counties, 1820

The higher (or lower) the Z score, the stronger the association. A Z score near 0 indicates no apparent clustering within the study area.

The results of Moran's *I* for this analysis clearly indicate positive spatial autocorrelation. Counties with low child–woman ratios tended to be located near counties with low ratios and counties with high ratios tended to be located near counties with high ratios. The Z-scores are quite large, indicating a fairly strong clustering, although far from perfect. The Getis-Ord General *G* points to statistically significant clustering, which shifts from the low fertility counties to high fertility counties over the course of the antebellum period. The relationship weakened as the nation moved towards more settled patterns.

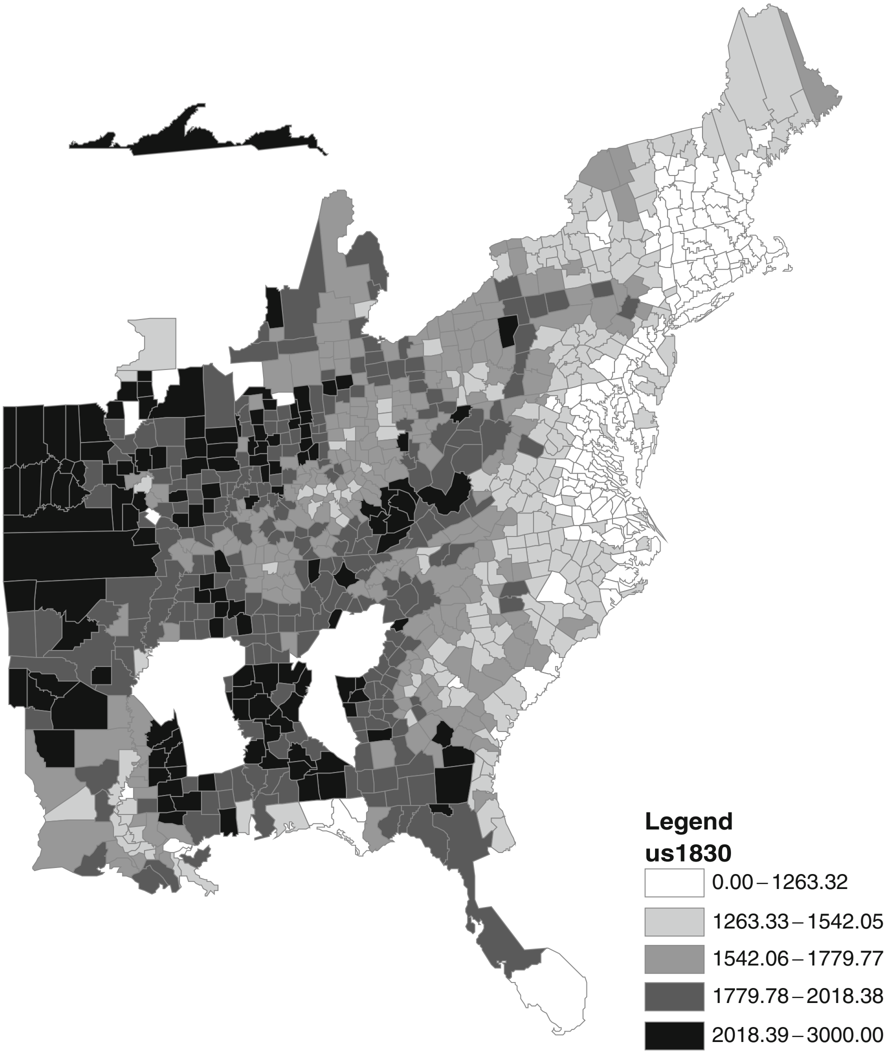


Fig. 3.4 White child–woman ratios. United States, counties, 1830

Concluding Remarks

This chapter uses structural regression models, geographic dispersion patterns of white child–woman ratios and measures of spatial autocorrelation and geographic clustering to test competing hypotheses about the fertility transition in the antebellum United States: the land availability hypothesis, the local labour market/child default hypothesis, the related life-cycle savings model, the conventional structuralist view and the ideational hypothesis. Covariates include population density (1800–1860), urbanization (1800–1860), transport connections (1840–1860), the

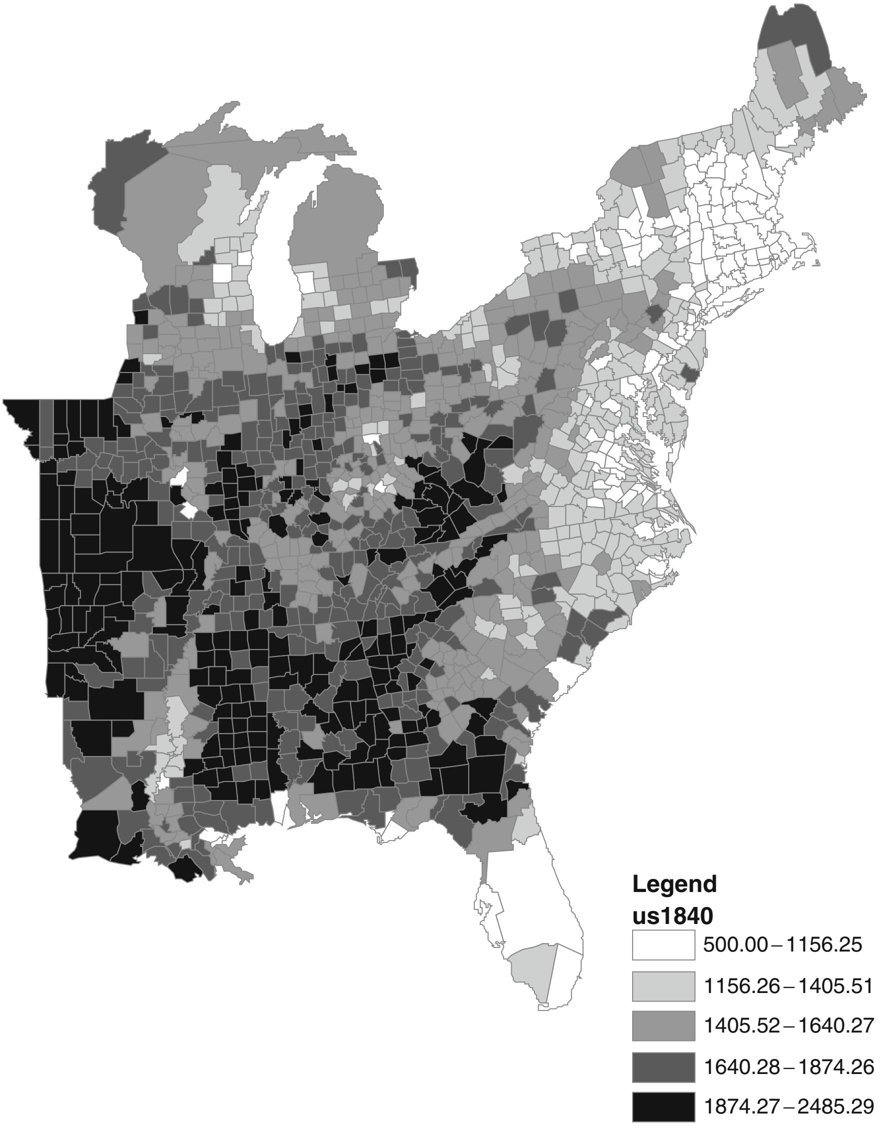


Fig. 3.5 White child–woman ratios. United States, counties, 1840

sex ratio (1800–1860), farm values (1850–1860) and wealth (1850–1860). All of the hypotheses receive some support from the analysis, confirming that the process was complex and multidimensional. Support for the land availability view is weakened by the finding that population density is not significant when urbanization is controlled. This lends more credence to the local labour market/child default hypothesis, which stresses the availability of nearby non-agricultural employment

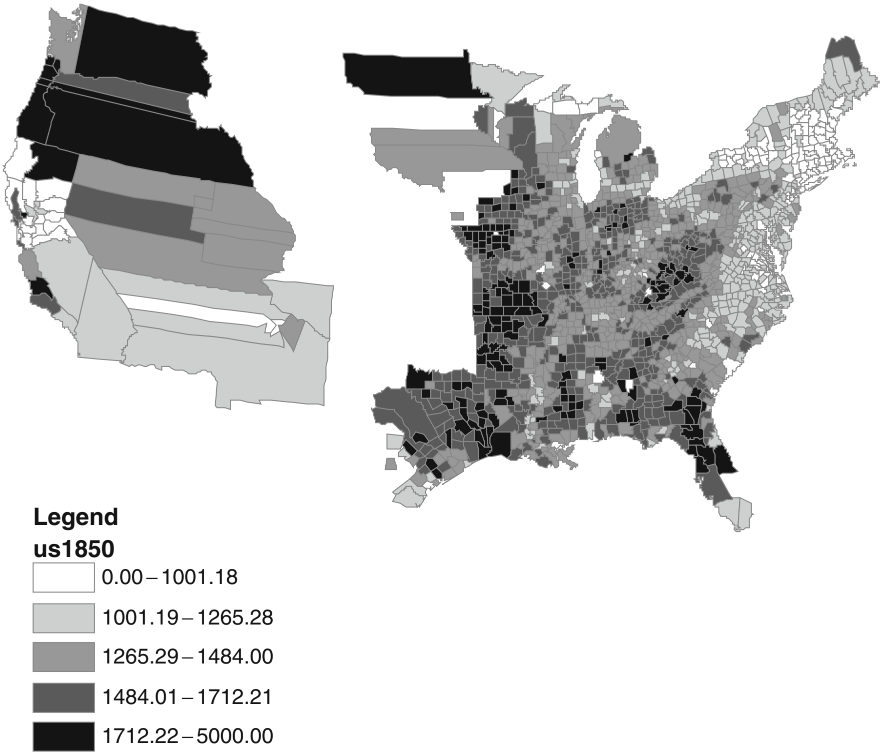


Fig. 3.6 White child–woman ratios. United States, counties, 1850

opportunities for farm children. These opportunities would be more likely found in urban places. On the other hand, structural and ideational variables (illiteracy, urbanization, transport and religion) also demonstrate some power in explaining cross sectional variation. The data also indicate that adjustments in marriage were important in the transition. These different perspectives are not necessarily mutually exclusive, however. More likely, a number of processes were underway, all of which contributed to the unusual and early fertility transition in the United States. One interesting hypothesis is that newly settled and long settled areas should exhibit substantially less variance than transitional areas. Preliminary analysis shows some support for this.

Using spatial analysis tools, Moran's *I* and the Getis-Ord General *G,* we found significant spatial autocorrelation, basically by longitude, not latitude. There was also significant clustering, of low fertility counties in 1800 and 1810, but of high fertility counties thereafter. The country was becoming more heterogeneous over time, but the distribution of counties by their fertility ratios was spatially far from random.

It is important that space in and of itself be considered in demographic analysis. Many long term analyses have understandably focused on the dimension of

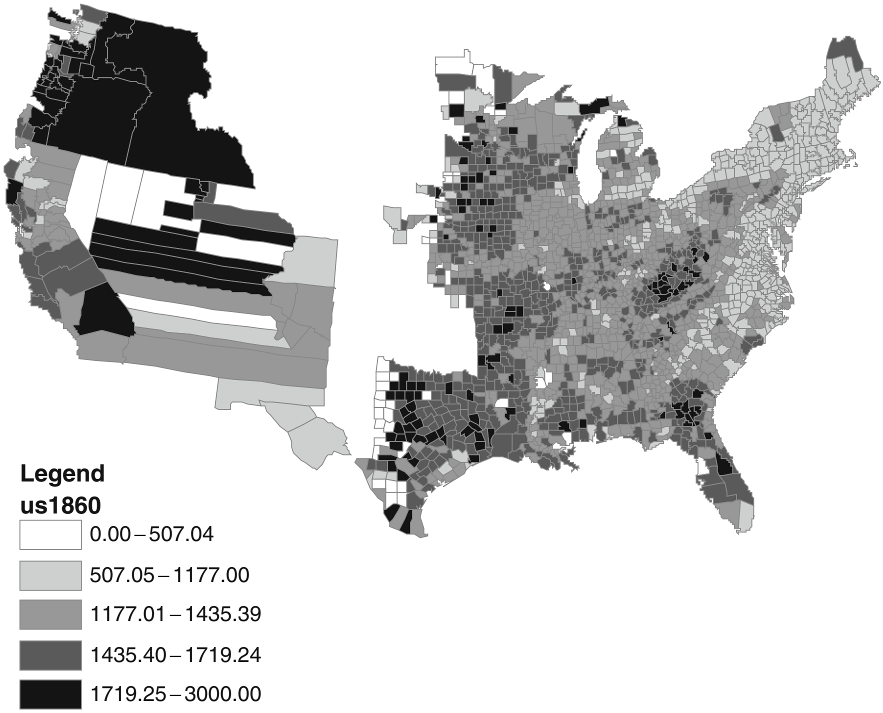


Fig. 3.7 White child–woman ratios. United States, counties, 1860

time and have used space merely as a source of variation for statistical estimation. But the spatial patterns were of considerable interest to contemporaries and should furnish additional insights into historical demographic processes and social and economic change. An important economic history of the period by Taylor (1951) identified the salient characteristic as transportation improvement. The noted political figure John C. Calhoun, despite his strong sectional interests, was a supporter of 'internal improvements' (i.e. roads, canals and later railroads), which he spoke of as 'conquering space'. He stated in 1817: '(T)o what can we direct our resources and attention more important than internal improvements? What can add more to the wealth, strength and the political prosperity of our country?' (U.S. Congress, 1817). The study of space should be a worthy subject for our collective historical inquiry.

Acknowledgements The authors wish to thank the IUSSP, the Inter-university Consortium for Political and Social Research and the Minnesota Population Center for sponsoring the conference. Some of the data originated with the Inter-University Consortium for Political and Social Research and the Integrated Public Use Microdata project. The authors wish to acknowledge those data archives.

Notes

1. The augmented data set used here is, in part, now available as ICPSR Study Number 2896 (Haines and ICPSR, 2004).
2. There exist county-level data for some states on wages from the Censuses of Social Statistics for 1850 through 1870, compiled from the manuscripts by Robert Margo. Since they do not cover all states, they are not used here.

References

Bateman, F. & Foust, J. D. (1974). A sample of rural households selected from the 1860 manuscript Censuses. *Agricultural History*, *48*(1), 75–93.

S. B. Carter, S. S. Gartner, M. R. Haines, A. L. Olmstead, R. Sutch, G. Wright (Eds.). (2006). *Historical statistics of the United States: Earliest times to the present*. *Millennial Edition* (Vol. 5). New York: Cambridge University Press.

Carter, S. B., Ransom, R. L., & Sutch, R. (2004). Family matters: The life-cycle transition and the antebellum fertility decline. In T. W. Guinnane, W. A. Sundstrom, W. C. Whatley (Eds.), *History matters: Essays on economic growth, technology, and demographic change* (pp. 271–327). Stanford, CA: Stanford University Press.

Coale, A. J. & Rives, N. W., Jr. (1973). A statistical reconstruction of the black population of the United States, 1880–1970: Estimates of true numbers by age and sex. *Population Index*, *39*(1), 3–36.

Coale, A. J. & Zelnik, M. (1963). *New estimates of fertility and population in the United States: A study of annual white births from 1855 to 1960 and of completeness of enumeration in the censuses from 1880 to 1960*. Princeton, NJ: Princeton University Press.

Craig, L., Palmquist, R. B., & Weiss, T. (1998). Transportation improvements and land values in the antebellum United States: A hedonic approach. *The Journal of Real Estate Finance and Economics*, *16*, 173–190.

Drake, D. (1815). *Natural and statistical view, or picture of Cincinnati and the Miami country*. Cincinnati, OH: Looker and Wallace.

Earle, C., Cao, C., Heppen, J., & Otterstrom, S. (1999). *The historical United States county boundary files 1790–1999 on CD-ROM*. Louisiana State University: Geoscience Publications.

Easterlin, R. A. (1976). Population change and farm settlement in the northern United States. *Journal of Economic History*, *36*(1), 45–75.

Easterlin, R. A., Alter, G., & Condran, G. (1978). Farms and farm families in old and new areas: The northern states in 1860. In T. K. Hareven, M. A. Vinovskis (Eds.), *Family and population in nineteenth-century America* (pp. 22–84). Princeton, NJ: Princeton University Press.

Flint, T. (1827). *Recollections of the last ten years*. Boston: Cummings, Hilliard (reprinted Da Capo, New York, 1968).

Forster, C. & Tucker, G. S. L. (1972). *Economic opportunity and white American fertility ratios, 1800–1860*. New Haven, CT: Yale University Press.

Gutmann, M. P. (1990). Denomination and fertility decline: The Catholics and Protestants of Gillespie County, Texas. *Continuity and Change*, *5*, 391–417.

Gutmann, M. P. & Fliess, K. H. (1993). The determinants of early fertility decline in Texas. *Demography*, *30*, 443–457.

Haan, M. (2005). Studying the impact of religion on fertility in nineteenth-century Canada: The use of direct measures and proxy variables. *Social Science History*, *29*(3), 373–411.

Hacker, J. D. (1999). Child naming, religion, and the decline of marital fertility in nineteenth-century America. *The History of the Family: An International Quarterly*, *43*, 339–365.

Hacker, J. D. (2003). Rethinking the 'early' decline of fertility in the United States. *Demography*, *40*(4), 605–620.

Hacker, J. D. (2006). Economic, demographic, and anthropometric correlates of first marriage in the mid nineteenth-century United States. *Center for Population Economics,Working Paper Series*. Chicago: University of Chicago.

Haines, M. R. (1998). Estimated life tables for the United States, 1850–1910. *Historical Methods*, *31*(4), 149–169.

Haines, M. R. and The Inter-university Consortium for Political and Social Research. (2004). Historical, Demographic, Economic, and Social Data: The United States, 1790–2000 [Computer file]. ICPSR02896-v2. Hamilton, NY: Colgate University/Ann Arbor: MI: Inter-university Consortium for Political and Social Research [producers], 2004. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005–04–29.

M. R. Haines, R. H. Steckel, (Eds). (2000). *A population history of North America*. New York: Cambridge University Press.

Haines, M. R. & Guest, A. M. (2008). Fertility in New York State in the pre-Civil War era. *Demography*, *45*, 345–361.

Haines, M. R. & Hacker, J. D. (2006). The puzzle of the antebellum fertility decline in the United States: New evidence and reconsideration. *Working paper number 12571*, (September). Cambridge, MA: National Bureau of Economic Research.

Knodel, J. & van de Walle, E. (1979). Lessons from the past: Policy implications of historical fertility studies. *Population and Development Review*, *5*(2), 217–245.

Leasure, W. J. (1982). La baisse de la fécondité aux Etat-Unis de 1800 a 1860. *Population*, *37*(3), 607–622.

Leet, D. R. (1976). The determinants of fertility transition in antebellum Ohio. *Journal of Economic History*, *36*(2), 359–378.

Lesthaeghe, R. J. (1977). *The decline of Belgian fertility, 1800–1970*. Princeton, NJ: Princeton University Press.

Lesthaeghe, R. J. (1980). On the social control of human reproduction. *Population and Development Review*, *6*(4), 527–548.

Lesthaeghe, R. J. (1983). A century of demographic and cultural change in Western Europe: An exploration of underlying dimensions. *Population and Development Review*, *9*(3), 411–435.

Long, J. H. (2001). *Building longevity into the design of a historical geographic information system: The atlas of historical county boundaries*. Paper presented at the annual meetings of the Social Science History Association, Chicago.

Main, G. L. (2006). Rocking the cradle: Downsizing the New England family. *Journal of Interdisciplinary History*, *37*(1), 35–58.

Margo, R. (2000). *Wages and labor markets in the United States, 1820–1860*. Chicago: University of Chicago Press.

McCusker, J. J. & Menard, R. R. (1985). *The economy of British America, 1607–1789*. Chapel Hill, NC: University of North Carolina Press.

McInnis, R. M. (1977). Childbearing and land availability: Some evidence from individual household data. In R. D. Lee (Ed.), *Population patterns in the past* (pp. 201–227). New York: Academic Press.

Notestein, F. W. (1953). The economics of population and food supplies: The economic problems of population change. In *Proceedings of the eighth international conference of agricultural economists*. London: Oxford University Press.

Preston, S. H. & Haines, M. R. (1991). *Fatal years: Child mortality in late nineteenth century America*. Princeton, NJ: Princeton University Press.

Ramsay, D. (1809). *The history of South Carolina* (Vol. 2). Charleston, SC: Longworth.

Ruggles, S. (1995). Family interrelationships. *Historical Methods*, *28*, 52–58.

Ruggles, S. (2003). Multigenerational families in nineteenth-century America. *Continuity and Change*, *18*, 139–165.

Ruggles, S., Sobek, M., et. al. (1997). Integrated public use microdata series: Version 2.0. Minneapolis: Historical Census Projects, University of Minnesota. Available on-line at http://www.ipums.org.

Schapiro, M. O. (1986). *Filling up America: An economic-demographic model of population growth and distribution in the nineteenth-century United States*. Greenwich: JAI Press.

Smith, D. S. (1977). A homeostatic demographic regime: Patterns in West European family reconstitution studies. In R. D. Lee (Ed.), *Population patterns in the past* (pp. 19–51). New York: Academic Press.

Smith, D. S. (1987). Early fertility decline in America: A problem in family history. *Journal of Family History*, *12*(1–3), 73–84.

Spengler, J. J. (1930). The fecundity of native and foreign-born women in New England. *Brookings institution pamphlet series (*Vol. II (1)). Washington, DC: The Brookings Institution.

Steckel, R. H. (1986). A dreadful childhood: Excess mortality of American slaves. *Social Science History*, *10*(4), 427–465.

Steckel, R. H. (1983). The economic foundations of east-west migration during the 19th century. *Explorations in Economic History*, *20*(1), 14–36.

Steckel, R. H. (1992). The fertility transition in the United States: Tests of alternative hypotheses. In C. Goldin, H. Rockoff (Eds.), *Strategic factors in nineteenth century American economic history* (pp. 351–375). Chicago: University of Chicago Press.

Sundstrom, W. A. & David, P. A. (1988). Old-age security motives, labor markets, and farm family fertility in antebellum America. *Explorations in Economic History*, *25*(2), 164–197.

Taylor, G. R. (1951). *The transportation revolution, 1815–1860*. New York: Holt, Rinehart, and Winston.

Thompson, W. S. & Whelpton, P. K. (1933). *Population trends in the United States*. New York: McGraw-Hill.

Tucker, G. (1827). *A voyage to the moon*. New York: Elam Bliss.

Tucker, G. (Joseph Atterly, pseudo). (1855). *Progress of the United States in population & wealth in fifty years as exhibited by the decennial census from 1790 to 1840, with an Appendix containing an abstract of the census of 1850*. New York: Press of Hunt's Merchant Magazine.

T'ien, H.Y. (1959). A demographic aspect of interstate variations in American fertility, 1800–1860. *The Milbank Memorial Fund Quarterly*, *37*, 49–59.

U.S. Bureau of the Census. (1975). *Historical statistics of the United States from colonial times to 1970*. Washington, DC: Government Printing Office.

U.S. Bureau of the Census. (1985). *Statistical abstract of the United States, 1986*. Washington, DC: Government Printing Office.

U.S. Bureau of the Census. (1997). *Statistical abstract of the United States, 1997*. Washington, DC: Government Printing Office.

U.S. Bureau of the Census. (2004). *Statistical abstract of the United States, 2004–2005*. Washington, DC: Government Printing Office.

U.S. Congress (1817). *Debates (*pp. 851–858). Washington, DC: 14th Congress, 2nd Session. House of Representatives.

Vinovskis, M. (1976). Socioeconomic determinants of interstate fertility differentials in the United States in 1850 and 1860. *The Journal of Interdisciplinary History*, *6*(3), 375–396.

Wells, R. V. (1971). Family size and fertility control in eighteenth century America: A study of Quaker families. *Population Studies*, *25*(1), 73–82.

Yasuba, Y. (1962).*Birth rates of the white population of the United States, 1800–1860: An economic analysis*. Baltimore: The Johns Hopkins University Press.

Chapter 4
Spatial and Temporal Patterns of Fertility Transition in Muslim Populations

Hani A. Guend

A Conceptual Framework for Muslims' Fertility

Context and Objectives

The motivation for this project is essentially a theoretical one. Its ambition is to provide an explanation, grounded in sound theory and backed by empirical analyses, of the similarities and peculiarities of fertility changes in Muslim populations. I document the spatial patterns and the temporal trends of fertility transition in Muslim populations in a comprehensive way within the limits of available data. The goal is to expand the breadth of the *demographic transition* theory through the testing of new hypotheses on populations of non-European stock. The area (Fig. 4.1) stretching from Morocco, in Northern Africa, to Indonesia, in South East Asia, is home to more or less a billion people with a remarkable homogeneity in reproductive behaviour and known, yet unexplained, peculiarities in the timing of onset and the pace of fertility decline. The countries that make up this area provide the study's frame of reference.

In his presidential address to the Population Association of America, Rupert Vance (1952) pointed to the lack of high theory in demography and claimed that the framework for an integrated theory of high order was emerging: the theory of demographic transition (Notestein, 1945). Vance envisioned such a theory to constitute a sort of a binder that holds together the findings of demographic research and, as in the physical sciences, supports a repetitive testing of hypotheses that resist falsification. It should (1) be dynamic (2) take account of demographic interrelations as between countries and groups within nations and (3) require a multi-science approach. Since Vance's call for a theory of high order, convincing evidence of the onset of fertility transition in most countries of the world has accumulated and the downward trend of fertility has continued everywhere (Bongaarts, 2002a, b, 2005). Five decades of fertility research in a sustained effort to articulate explanatory

H.A. Guend (✉)
Université du Québec, INRS – Urbanisation Culture Société, Montréal, QC, Canada H2X 1E3
e-mail: hani.guend@ucs.inrs.ca

M.P. Gutmann et al. (eds.), *Navigating Time and Space in Population Studies*,
International Studies in Population 9, DOI 10.1007/978-94-007-0068-0_4,

Fig. 4.1 A reference map of the Muslim world, source: http://www.lib.utexas.edu/maps/world_maps/muslim_distribution.jpg

models of fertility change, however, still fall short of providing a comprehensive explanation that matches the level of universality of fertility decline.

Empirical validations of the theory with respect to populations of diverse origins contribute to building such a comprehensive theory: the *binder theory* able to explain the global trend as well as the observed peculiarities. Three historical moments in empirical demographic research contributed substantially to this continued effort towards a comprehensive explanation. The *European Fertility Project* concluded that timing of the 'onset of fertility decline was only weakly linked to provincial levels of socio-economic modernization but unmistakably linked to language and ethnic or religious cleavages' (United Nations, 2002). Analyses of the *World Fertility Survey* also hinted to the importance of the relationship between ethnicity and religion on the one hand and reproductive behaviour on the other (Freedman, 1987) and highlighted the major role played by modern contraception in the onset of fertility decline in developing countries (Caldwell & Ruzicka, 1987). Furthermore, analyses of the *Demographic and Health Surveys* (DHS) documented recent rapid fertility declines in countries still at the lowest levels of economic development, such as Bangladesh (Cleland, Phillips, Amin, & Kamal, 1993) and in countries with diverse levels of wealth and human development.

The historical fertility transition in Europe unfolded as a component of a global socio-economic change that resulted in the spread of secular individualism, which translated into individual behaviour motivated by the 'pursuit of personal goals devoid of references to a cohesive and overarching religious or philosophical construct' (Lesthaeghe, 1983). This work builds on the long term efforts of empirical testing of theories of fertility transition. It examines a specific area of the world that is homogenous enough to allow inference of the findings across countries and regions, yet sufficiently heterogeneous to permit the necessary comparisons required in causal analyses. The European experience provided a model of reference to this study. However, using that experience as a reference does not imply belief in any kind of historical determinism, nor does it invalidate a counterfactual reasoning to be applied to other areas of the world and, I may say, to the European historical experience itself. The homogeneity of the field of investigation stems from a common cultural background. Yet, the populations of the countries selected as units of analysis provide quantifiable differences that allow the identification of the causal determinants of fertility change. One might ask: what makes the fertility of Muslims stand apart with respect to testing the demographic transition theory?

This study investigates long range historical change in the Muslim world. It deals with demographic regimes of relatively homogeneous human populations as single units of analysis for analytic and heuristic reasons. First, following Caldwell's (1995) general statement, I argue that fertility transition within the Muslim world can be understood only as a global phenomenon, not as a series of separate national transitions. This kind of postulate provided the rationale for launching the landmark study of fertility transition in historic Europe known as the European Fertility Project. Second, partitioning the developing world along cultural lines made empirical sense (Beaver, 1975: p. 3–8), and statistical analysis had found that *region*, with Islam identified as one, was an important multivariate predictor of natality

(Janowitz, 1971). Finally, the only major cultural area besides sub-Saharan Africa where fertility decline had been lagging until recent times (Chesnais, 1986) is the 'belt of Muslim populations from Morocco to Indonesia', as Caldwell (1995) referred to the Muslim world.

While the persistence of high fertility is commonly noted and the effect of religion in the process taken for granted, the identification of possible factors (Kirk, 1996; Cleland, 1985) has remained sketchy. Some analysts have turned to the scriptures to investigate the nature of the relationship between Islam and reproductive behaviour (Kouaouci, 1983; Omran, 1992; Guend, 1993; Obermeyer, 1994) and found little formal doctrine in Islam relating to reproductive choices of Muslim women. Thus, the homogeneity of pre-transition reproductive behaviours of Muslims has remained an open question. Yet, the pattern of change across the Muslim world since 1945 has followed the broad outlines of the transition theory. In particular, as I explain below, the diffusion of the ideas of reformist movements embedded in the indigenous cultural realities, as well as those of political movements that advocated the developmental thinking in their nation building efforts, have played crucial roles in shaping the character of the transition underway. I refer intuitively to diffusion of new ideas as causal factors of change, but this work is not an attempt to implement a specific diffusion model, although my reasoning is more or less grounded in that theory (see Palloni (2001) for a discussion of the topic). Before moving from general theory to the background on Muslim populations, basic elements of the geography and history of Islam may be helpful.

I began this research with a geographic frame of reference that encompasses all fifty-seven countries that are members of the Organization of Islamic Conference. As the investigation progressed, I reduced the area of reference to countries with populations at least 55% Muslim. That included populations living in the area stretching across Northern Africa, the Middle East, Central Asia and South East Asia. Hence, the investigation has to do with all the states of the Arab region, Turkey, Iran, Afghanistan and the Muslim republics of the former USSR, without distinction between Shiites and Sunnis. The lack of data limited the sample to twenty-six countries, but the theoretical framework still applies to this whole area. For practical and methodological reasons, I restricted the study to Muslim populations living within the boundaries of a national entity. The practical reason has to do with the demographic and socio-economic data that are organized on national levels. The methodological reason is related to the nature of the study, which is concerned with the investigation of long range historical change rooted in demographic regimes of culturally homogenous populations. The study is concerned with Islam as a dominant culture, rather than with the religious beliefs of individuals immersed in non-Muslim cultures.

Key Concepts

There is consensus that the first civil war in the history of Islam – what the historiography of Islam calls *al-Fitna al-Kubra* – set into motion processes devoid of the spirit that was open to change in early Islam. These processes suffused the Islamic faith, laws and traditions with the spirit and the characteristics of tribal Arab society. A setback to the momentum of change and at least a partial return to the old social order occurred after that decisive war.

The turning point is usually identified with the battle of *Siffin* (657), near the upper Euphrates river, between Ali (died 661), then the elected supreme commander of the faithful (Caliph) and Muawiyya (died 680), then the governor of Syria. Muawiyya rejected Ali's rule as Caliph and ascended to power by military force subsequent to a mitigated victory at Siffin. Ali's caliphate was a continuation of the prophet's legacy as extended by his three companions and immediate successors (Abu Bakr, Ummar and Othman). The historiography of Islam refers to these four rulers as the *well guided caliphs*.

The main branches of Islam emerged after Siffin in a context of unrest and violence. Besides the legitimists who supported Ali and later became the Shiites, three major factions emerged as forces of opposition to Ali's rule and claimed to have a say on how the political system of Islam should be reorganized: (1) the forces of Arabic conservatism supported Muawiyya and their views became dominant in the Sunni world; (2) another group modern sociology would identify as *anarchist* evolved into the Ibadite branch of Islam; and (3) a short lived group lead by Aisha, the Prophet's youngest and beloved wife, emerged and was supported by a few of the Prophet's most prominent companions. The institutionalization of Islam following al-Fitna al-Kubra had produced three embedded legacies of the social system of classical Islam: the *arabization of Islam*, *patrimonialism* and *Sufism*. These key concepts are of particular interest to understanding why Muslim societies remained resilient to social change for centuries.

Arabization of Islam meant that the tension between the tribal social structure of Arab society and the new order Islam introduced was resolved at Siffin to the advantage of the ancient Arab order. This resolution gave the subsequent dominant political system its definitive form. In such a context, none of the constraints placed on the family, for example by the church and the modern state, were present in traditional Islam. Christianity had evolved into an organized religion within the context of a highly hierarchical and bureaucratic empire. Individual families were accountable to a central authority for every new mouth to feed and for the tiniest needs a newborn creates. Even securing a new space on church benches could become an issue for families. Whereas Christianity tied individual families to the imperatives of organized religion and a hierarchical and bureaucratic state, the Arab social order organized 'almost every important relationship . . . in terms of kinship' (Bellah, 1970). Ironically, this state of affairs developed despite the spirit of pristine Islam characterized as inherently modern, in that it liberated individuals from the web of duties and obligations to local self-contained social entities (Bellah, 1970). This early retrogression encompassed the Islamic family.

> The Islamic family as it took shape under the Prophet's guidance was to reflect the reduced significance of extended kinship in the new Muslim community and to express the enhanced dignity of the individual, including women and children, as persons standing in direct relation to God . . . The main tendencies of Koranic thinking about the family, however, were not only not developed in the early centuries of Islam, but they suffered a serious retrogression (Bellah, 1970: p. 154).

Muawiyya founded the Umayyad dynasty, which shaped the character of the political system in classical times with patrimonialism as a salient feature. Patrimonialism (Turner, 1974) refers to that traditional political system in Islam wherein the exercise and transfer of power are intertwined with the patriarchal, tribal system. In the patrimonial political system, large families were instrumental in running the political and administrative institutions of Muslim state hierarchies. Göçek's (1996) study of the collapse of the Ottoman Empire and the birth of the modern republic of Turkey provided an illustration of this household-based administration of the Empire. Kinship-based values, beliefs and behavioural patterns became intimately intertwined with Islamic faith, laws and morality as a result of processes set in motion after Siffin.

The Sufi ethos dominated ideologically and culturally and came to be seen as the orthopraxy of Islam until the emergence of the modern reformist movements. Up to the eleventh century, Sufism was much more of an elite spiritual preoccupation. It later developed into more organized and permanent associations representing all levels of society and with an increasingly international scope in membership and influence (Denny, 1994). The Sufi ethos sustained a worldview that leaves birth control beyond the 'realm of the calculus of conscious choice', to borrow Coale's expression, thus leading to unrestrained, one might say, irrational reproductive behaviour. In practice, the Sufi ethos translates into rituals, customs and other social manifestations affecting marriage, pregnancy and birth, childcare in health and sickness and death.

This blending of Islamic law and doctrine with the cultural patterns of Arab tribal society still affects the status of women, the value of children, the social roles assigned to brothers and sisters as well as family and clan solidarity. Anthropological studies of Muslim societies showed that it is not uncommon for social customs to bypass the Koran's injunctions when they are perceived as threats to a kinship-based social structure. For example, Tillon (2001) provided an illustration of such customary adaptations by showing how endogamy was used to preserve the material foundations of the tribe and the clan in Northern Africa. Endogamy was practiced as a way to bypass the rules of division of inheritance among heirs as the Koran's teachings wanted it; the aim was to keep wealth in the clan.

Reform awaited the tide of new ideas that emerged after World War I, decolonization and the reorganization of the international order. The process of nation building in the newly independent Muslim states shook the foundations of the system of classical Islam. *Developmental idealism*, as epitomized by Ataturk's reforms in Turkey and *Islamic reformism*, as promoted by such social reformers as Abduh in Egypt or Ben Badis in Algeria, emerged as twin forces aiming to reshape the classical system. Fertility decline, in turn, became part of general social change. The specific drive for reform in many countries was part of a larger global process of ideational change that can be linked to developmental idealism, whose central idea is the *developmental paradigm*, defined as:

> A model of change that has been applied at the individual, organizational, and societal levels. In this paradigm, change was pictured as natural, uniform, necessary, and directional . . . At the societal level, many versions of the model used a biological metaphor whereby

> societies were compared to individuals and viewed as developing through the same relatively uniform and necessary life cycle stages. In most versions of the paradigm it was also recognized that each society had its own individual circumstances, which produced variations in the trajectory. An important variant of societal development stripped the biological metaphor of decline and left only permanent improvement in the developmental trajectory (Thornton, 2001: p. 450).

The most visible manifestation of this intellectual ferment in the Muslim world was a pattern of change labelled the *Turkish paradigm* (Richards & Waterbury, 1990), referring to the experience of nation building in modern Turkey, described in these terms:

> The foundation of the modern Republic ... signified a radical departure from the previous social formation. A modern constitution was introduced, the Sultanate and Caliphate were abolished, religious schools and courts were closed, Western headgear and dress were adopted, Islamic Law was abandoned and replaced with modified versions of the Swiss and Italian Civil and Penal Codes, and the Arabic alphabet was replaced with an alphabet based on Latin characters. In short, the direction of change, led by Ataturk, was one from a religious, oriental Empire to a modern, Westernized, and secular Republic (Hancioğlu, 1994: p. 1)

The Turkish experience provided a model many other countries in the Muslim world have emulated. Independent Egypt, Syria, Iraq, Libya, Algeria, Tunisia, Iran, the People's Republic of Yemen before unification, Sudan and Indonesia have had a common blueprint for the radical transformation of their societies and economies. Arab socialism, popular democracy and *pancasila* (the 5-pillar state ideology of the republic of Indonesia) were national ideologies that showed some country-specific differences, but adhered to the tenets of developmental idealism in its Turkish version: the state, backed by the army, is the principal agent of development and the goal is to reach levels of modernization similar to those seen in Europe, but within a shorter time frame. Marxist states, as in the Muslim republics of Central Asia in the former USSR, also shared these characteristics.

This is not a claim that the remaining countries were totally immune to the influence of developmental idealism, or that they made no effort to develop. Most implemented policies to strengthen their economies and to raise the standards of living of their populations, but development and developmental idealism refer to two different realities, which may or may not be intertwined. That is, adoption of developmental idealism as an ideology does not necessitate achievements in the realm of social and economic development. Likewise, many conservative countries rejected outright that philosophy of social change, yet achieved quite advanced levels of social and economic development.

The second force of social change, Islamic reformism, attempted to promote alternative trajectories of development by reference to worldviews inspired by indigenous patterns of thinking. It focused on education of the masses with the explicit goal of wiping away the influence of the Sufi ethos. Advocates of this social movement, known as the modernist *Ulema* (scholars) or the advocates of

Islah (reformism), sought to integrate elements of modernity into the social system in a way that warrants preservation of the essential features of the traditional institutions and the essence of the Islamic culture.

The Sufi ethos was an impediment in the way of both forces of change. As Turner (1974) remarks, the enemy of both Islam and modern society is a set of attitudes (fatalism, imitation and passivity) that was brought into the Islam of the *Salaf* (elders) by Sufism. Islamic reformism clashed, to various degrees of intensity, with Sufism in Northern Africa, the Middle East, Malaysia and Indonesia.

Note that the current Islamic movement known as Islamic fundamentalism is yet another expression of Islamic praxis that is less relevant to the purpose of this study. Its emergence is too recent to fulfil the time order required for a causal effect on change in reproductive behaviour.

The Hypotheses

As stated above, the goal of this project is to expand the breadth of demographic transition theory through the testing of hypotheses about fertility change in Muslim populations. The study's ambition is to provide an explanation, grounded in sound theory and backed by empirical analyses, of the similarities and peculiarities of reproductive behaviours. Developmental idealism affects the collective perception of progress and wellbeing and the means to reach them. A small family norm is perceived as a marker of progress. Islamic reformism, for its part, aimed to transform the traditional system of belief infatuated with the Sufi ethos towards a more rational conception of religiosity.

The dissemination of developmental idealism throughout the Muslim world unfolded in many ways and with different strengths. The most intense diffusion took place in countries where packaged national policies driven by the belief in the developmental paradigm were implemented. Two components of this belief stipulate that Westernization is the fastest path to progress and that it is possible to telescope the stages of development, which took a couple centuries for Western Europe to achieve, in a generation or two, through centrally planned economies.

The two forces of change and their interaction constitute the engines of processes that triggered the onset of fertility transition, impacted the differential pace of fertility decline and might determine the expected levels of stabilization. Their effects on fertility were mediated through ideational and behavioural changes that created the social, cultural and political contexts that make meaningful the micro economic calculus of the costs and benefits of childbearing. It is within such contexts that family planning programmes are likely to operate efficiently. These processes have been at work long enough for fertility to start declining, but none of the experiences has yet run its full course.

Overall, the study claims that spatial and temporal patterns of fertility decline in Muslim populations are more closely linked to the strength of the pathways of dissemination of developmental idealism and Islamic reformism than to such attributes as ethnicity, language or levels of wealth.

John Caldwell (1982: p. 229) predicted that the rate at which family relations are westernized will determine the timing and speed of fertility decline. He further stated:

> This is not merely a function of the level of cultural flow from the West. It is also very much affected by the receptivity or opposition of the receiving cultures. The areas of the world where high fertility persists longest will not necessarily be the poorest. It is likely that they will be the Islamic World where the Koranic injunction serves to insulate the existing family structure, and sub-Saharan Africa where the lineages retain a surprising amount of their strength even while welcoming Western imports which do not impinge on family relations.

This statement provides a platform for further specification of this study's hypotheses. First, in previous discussion, I introduced new concepts to contextualize the effect of the *Koranic injunctions* and showed how lineages too are intertwined with the classical Islamic legacy. Second, developmental idealism is a more specific concept than westernization and one does not necessarily imply the other. Then, in light of the narrative about the background of Muslim populations, Caldwell's statement can be distilled into four formal hypotheses:

- First, in countries where the two factors of change – developmental idealism and Islamic reformism – were strong and, consequently, the Sufi ethos was weakened, receptivity to the westernization of family relations is likely to be the highest. Therefore, the onset of fertility transition is likely to take place relatively early and a rapid fertility decline is more likely.
- Second, in countries where strong developmental idealism operated against a system of classical Islam marked by the Sufi ethos, receptivity to the westernization of family relations is likely to remain low despite – or because of – poorly timed state policies. As a consequence, the onset of fertility decline will be delayed.
- Third, in countries where traditional Islam was not significantly altered by Islamic reformism, and developmental idealism disseminated through non-induced pathways such as trade and aid to development, receptivity to the westernization of family relations is likely to develop moderately. In such contexts, the onset of fertility decline might occur before substantial changes affect the traditional family structure. The international context is particularly important in this case and contraception within marriage plays a crucial role.
- Fourth, in countries where the pathways of developmental idealism were weak, and state sponsored Islamic reformism altered the system of belief of classical Islam towards an even more radical religious interpretation, receptivity to the westernization of family relations is likely to be very low. As a consequence, it is likely that fertility transition remains stalled for a long time.

Empirical Analyses

Ansley Coale (1973) summarized in a compact formula three pre-conditions that must be fulfilled simultaneously for the onset of fertility decline to occur. These are

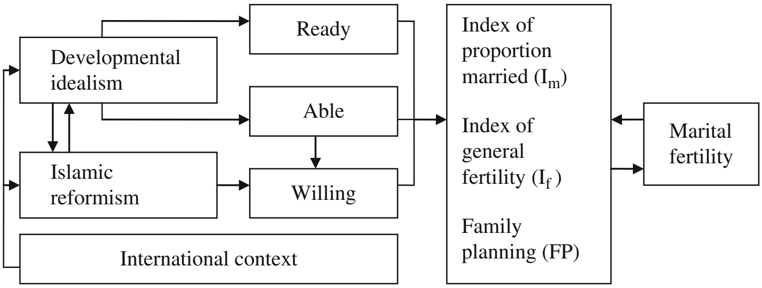


Fig. 4.2 Hypothetical causal paths of developmental idealism and Islamic reformism on marital fertility

for women to be *ready*, *willing* and *able* to control births. The first pre-condition refers to the structural change necessary to make contraceptive use possible, the second hints at the ideational change that creates the desire to control births and the third points out the availability of the means of contraception. The hypothetical causal paths that link developmental idealism and Islamic reformism to marital fertility integrate these preconditions as intermediate variables to the outcome, as Figure 4.2 illustrates.

The Universe of Investigation and Inference

I selected a subset of twenty-six countries from among the fifty-seven members of the Organization of the Islamic Conference. A country was included in the sample if its population was at least 55% Muslim and if available census and vital statistics data allowed for the calculation of Coale's comparative index of general fertility (I_f) at two or more points in time, with an acceptable time lag.

Exploratory spatial analysis of the data showed that the universe of investigation can be divided into six regions along the lines of language and ethnicity. This classification is instrumental for testing hypotheses that include ethnicity, religion and language as key factors. Besides the Muslim populations of sub-Saharan Africa, there are five major ethnic-linguistic groups. The populations of Indian ethnic background represent about 31% of the sample and are distributed between Pakistan and Bangladesh; the remaining populations of Asian stock constitute a second major group, which represents 26% of the total population of the sample of countries. The Persian, the Arabic and the Turkic populations stand as distinct ethnic-linguistic blocs. Iran, Egypt and Turkey predominate demographically within each bloc with a population size of about seventy million, while the rest of the countries are young nations born in the wake of decolonization.

The availability of demographic information is itself a function of how deeply ingrained developmental idealism may be. Indeed, a national population census and a national system of vital statistics covering most, if not all, of the population in a country are two major institutions of the modern state. The more a country has been exposed to westernization during the colonial past, and the more it has embraced policies driven by developmental idealism, the more likely it is that the country will have a well-developed population census and system of vital statistics. Therefore, it is important to characterize the fourteen cases left out because of lack of data.

Lack of data may be related to institutional instability, as is the case in Afghanistan, Iraq, Kazakhstan, Lebanon, Palestine and Sudan. These countries were neither alien to developmental idealism, nor were their respective institutional instabilities unrelated to the nature of their developmental and nation building experiences. Nonetheless, there are surrogates for these cases in the sample. In other countries, such as Djibouti, The Gambia, Mauritania, Niger, Oman, Saudi Arabia, Senegal and the United Arab Emirates, the lack of data might be related to a lack of a developmental legacy. That is, the post-colonial nation state did not embrace developmental idealism and, in many cases, opposed westernization.

The populations of three countries displayed characteristics that are uniquely relevant to the main thesis of this work and there are no surrogates for them in the sample. Each case is associated with a specific brand of classical Islam: *Ibadism* in Oman, *Wahabism* in Saudi Arabia and *Sufism* in Senegal. All three brands are forces of conservatism. Ibadism is one of the main branches of Islam that emerged after Siffin as a distinct school of thought, as discussed in the previous section. Its adepts survived in distinct communities distributed all over the Muslim world. Nowadays, the Sultanate of Oman is the only standing Ibadite state. Wahabism, the official doctrine of the Saudi Kingdom, is more of a modern reformist movement. However, its brand of reformism advocates an interpretation of the scriptures that leaves little room for social change. As such, it stands in stark contrast to the Islamic reformism (Islah) discussed earlier. Senegal is another case where Sufism still impacts social life and the course of political events in, or despite, a modern democratic state.

The Data

As measures of the outcome, I calculated Coale's standardized comparative indices of proportion married (I_m) and of total fertility (I_f). These serve as the primary measures of nuptiality and fertility levels; their average annual changes between two points in time provide comparative measures of changes in fertility and nuptiality. I also used time series of Total Fertility Rates (TFRs) collected from secondary sources as secondary measures of the patterns of fertility decline. Table 4.1 shows the figures for the outcome and for related demographic indicators.

I further defined the explanatory variables by reference to the *conceptual framework for Muslims' fertility* discussed above. Hence, I created indices that summarize historical experiences, contemporary developmental policies, national achievements on specific aspects of human development and the strength of the channels of international exchange. Table 4.2 displays, for each country, the levels of the covariates.

Table 4.1 Fertility and related demographic indicators

Country	Population In 2000 (Millions)	Period of obser-vation I_f	Most recent figure of I_f	Change in I_f Over period	Average Annual change I_f	
Azerbaijan	8.2	1989–1999	0.14	46	4.59	
Albania	3.1	1955–1989	0.25	52	4.53	
Tajikistan	6.1	1989–1993	0.36	20	4.91	
Iran	67.2	1986–1996	0.21	57	5.74	
Turkey	69.3	1960–2000	0.23	56	1.40	
Uzbekistan	25.3	1989–1999	0.24	36	3.61	
Algeria	30.7	1966–1998	0.26	63	1.97	
Indonesia	214.0	1964–1997	0.21	35	1.05	
Turkmenistan	4.7	1989–1995	0.28	28	4.61	
Brunei	0.3	1971–1992	0.26	50	2.40	
Tunisia	9.6	1966–1997	0.19	67	2.17	
Bahrain	0.7	1981–2000	0.21	40	2.09	
Malaysia	23.5	1957–2000	0.25	53	1.24	
Bangladesh	141.0	1981–1991	0.38	13	1.31	
Kyrgyzstan	5.0	1989–2000	0.20	41	3.69	
Morocco	29.6	1982–2000	0.17	11	0.59	
Egypt	69.1	1975–1996	0.30	30	1.44	
Qatar	0.6	1982–1986	0.38	–4	–1.08	
Libya	5.3	1973–1991	0.52	29	1.60	
Kuwait	2.4	1965–1998	0.22	64	1.95	
Syria	17.0	1981–1994	0.37	31	2.41	
Jordan	5.2	1979–2000	0.28	54	2.57	
Pakistan	146.0	1968–1998	0.38	15	0.51	
Mali	12.3	1976–1987	0.55	–19	–1.71	
Yemen	18.7	.	.	.	.	
Maldives	0.3	1985–1995	0.32	47	4.66	
Country	**Period of observation I_m**	**Change in I_m over period**	**Most recent figure of I_m**	**Average Annual change I_m**	**Use FP %**	**TFR year**	**TFR**
Azerbaijan	.	.	0.63	.	55	2000	1.9
Albania	1955–1989	10	0.67	0.28	58	2000	2.1
Tajikistan	.	.	0.74	.	34	2000	2.4
Iran	.	.	0.75	.	74	2000	2.5
Turkey	1960–1990	16	0.70	0.52	64	1998	2.6
Uzbekistan	.	.	0.73	.	67	2000	2.7
Algeria	1966–1987	27	0.57	1.28	64	2000	2.8
Indonesia	1964–1997	10	0.70	0.31	57	1997	2.8
Turkmenistan	.	.	0.67	.	62	2000	2.9
Brunei	1971–1986	9	0.55	0.61	.	1992	3.1
Tunisia	1966–1984	25	0.56	1.37	60	1995	3.2
Bahrain	1981–1991	6	0.59	0.63	62	1995	3.2
Malaysia	1957–2000	24	0.57	0.56	55	1994	3.3
Bangladesh	1981–1991	2	0.85	0.20	54	2000	3.3
Kyrgyzstan	.	.	0.71	.	60	1997	3.4

Table 4.1 (continued)

Country	Period of observation I_m	Change in I_m over period	Most recent figure of I_m	Average Annual change I_m	Use FP %	TFR year	TFR
Morocco	.	.	0.62	.	58	1997	3.5
Egypt	1975–1996	3	0.67	0.15	56	2000	3.5
Qatar	1982–1986	2	0.70	0.54	43	1998	3.9
Libya	1973–1984	30	0.56	2.75	45	1995	4.1
Kuwait	1965–1996	27	0.58	0.88	52	1999	4.2
Syria	1981–1994	9	0.61	0.69	40	1993	4.2
Jordan	1979–1994	16	0.57	1.07	53	1997	4.4
Pakistan	1968–1976	4	0.78	0.47	28	2001	5.5
Mali	1976–1987	2	0.81	0.14	8	2001	6.8
Yemen	.	.	.	.	21	1997	7.6
Maldives	1985–1990	3	0.73	0.56	.	.	.

In order to operationalize the causal paths specified in Figure 4.2, I selected social, economic and political variables whose meanings can be derived from the key concepts of developmental idealism and Islamic reformism (UNDP, 2003). I also created variables that summarize the country's historical experiences and cultural attributes relevant to the purpose of the study. Some variables are self-explanatory, but others are synthetic indices whose meanings need specification.

Experience with European colonization is particularly relevant to the study's purpose. The length of the period of colonial rule provides a proxy for the intensity of European colonization as a pathway of dissemination for developmental idealism. The variable *per cent Muslim* measures the weight of minorities, especially Christian minorities, who are likely to provide a meaningful pathway for developmental idealism. The year women acquired the right to vote indicates how early social change triggered by developmental idealism and Islamic reformism took place.

The type of contemporary developmental policy, especially during the post colonial era, is an important factor in classification. I used linear regression of TFR with potential covariates to identify the relevant explanatory variables. This approach proved particularly useful in defining the variable *policy*. Several trials of regressing TFR on independent variables returned non-significant relationships. I then created a variable *policy* that tells whether a country experienced a Turkish paradigm type of development policy or not. Controlling for *policy* in the regression changed the significance of the other covariates in a positive direction. *Population size*, *per cent urban*, *index of proportion married*, *per cent using family planning*, *ratio of female to male literacy rate*, as well as the variable *polic*y itself all became statistically significant at the 90% confidence level or above.

The United Nations (UN) data fit the purpose of the study particularly well because developmental idealism constitutes the underlying perspective of the UN actions and policies as discussed in Thornton's (2005) book. In particular, the meaningfulness of gender sensitive indices cannot be overstated in the socio-cultural

Table 4.2 Economic, Social and Cultural Indicators, circa 2000

Country	Percent urban	Percent Muslim	Girl/boy school enrollment ratio[b]	FPPEI[b]	NFDI [2].	HDI[b]	ODA[b].	Women voting year[a]	GDP[b] per capita
Azerbaijan	52	93	96	.	4.0	0.74	27.5	1921	3090
Albania	43	70	94	.	0	0.74	86.1	1920	3680
Tajikistan	28	90	90	.	2.0	0.68	25.9	1924	1170
Iran	65	99	91	44	–0.3	0.72	1.7	1963	6000
Turkey	66	99	89	37	0.5	0.73	2.4	1930	5890
Uzbekistan	37	88	96	.	0.6	0.73	6.1	1938	2460
Algeria	58	99	88	31	2.2	0.70	5.9	1962	6090
Indonesia	42	88	95	71	1.0	0.68	7.0	1945	2940
Turkmenistan	45	89	.	.	2.5	0.75	15.2	1927	4320
Brunei	73	67	90	.	.	0.87	1.0	None	.
Tunisia	66	98	91	63	0.6	0.74	39.2	1957	6390
Bahrain	93	100	96	.	.	0.84	25.8	1973	16060
Malaysia	58	52	95	58	.	0.79	1.9	1957	8750
Bangladesh	26	83	96	52	0.2	0.50	7.3	1972	1610
Kyrgyzstan	34	75	95	.	0.3	0.73	37.7	1918	2750
Morocco	56	99	84	42	0.6	0.61	17.5	1963	3600
Egypt	43	94	89	48	1.7	0.65	18.2	1956	3520
Qatar	93	95	95	.	.	0.83	1.7	None	28132
Libya	88	97	97	.	.	0.78	1.9	1964	6453
Kuwait	96	85	96	.	.	0.82	1.5	None	18700
Syria	52	90	89	26	0.6	0.69	9.0	1949	3280
Jordan	79	92	95	22	0.9	0.74	83.3	1974	3870
Pakistan	33	97	55	41	0.6	0.50	17.5	1947	1890
Mali	31	90	71	24	2.6	0.34	83.2	1956	810
Yemen	25	90	60	17	–2.7	0.47	22.8	1967	790
Maldives	28	99	95	.	2.6	0.75	83.2	1932	2082

Table 4.2 (continued)

Country	Women to men literacy rate[b]	Islamic reformism score[a]	Turkish paradigm policy[a] score	Ottoman rule[a] score	European colonial[a] time years[b]	Colonial legacy[a]	GDI[b]	NFDI per capita[b]
Azerbaijan	.	1	1	1	72	Russia	.	4.0
Albania	0.97	1	1	1	0	None	0.73	0.0
Tajikistan	1.00	1	1	1	50	Russia	0.67	2.0
Iran	0.95	1	1	0	0	Britain	0.70	–0.3
Turkey	0.95	1	1	1	0	None	0.73	0.5
Uzbekistan	1.00	1	1	1	100	Russia	0.73	0.6
Algeria	0.90	1	1	1	130	France	0.69	2.2
Indonesia	0.99	1	1	0	129	Dutch	0.68	1.0
Turkmenistan	.	1	1	1	106	Russia	.	2.5
Brunei	1.00	0	0	0	75	Britain	0.87	0.0
Tunisia	0.92	1	1	1	76	France	0.73	0.6
Bahrain	1.00	0	0	0	0	Britain	0.83	0.0
Malaysia	1.00	0	1	0	129	Britain	0.78	0.0
Bangladesh	0.70	0	1	0	180	Britain	0.50	0.2
Kyrgyzstan	.	1	1	0	127	Russia	.	0.3
Morocco	.	0	0	0	44	France	0.59	0.6
Egypt	0.83	1	1	1	63	Britain	0.63	1.7
Qatar	1.05	0	0	0	25	Britain	.	0.0
Libya	0.94	0	1	1	40	Italy	.	0.0
Kuwait	1.00	0	0	1	62	Britain	0.81	0.0
Syria	0.83	0	1	1	26	France	0.67	0.6
Jordan	1.00	0	0	1	27	Britain	0.73	0.8
Pakistan	0.60	0	0	0	180	Britain	0.47	0.6
Mali	0.54	0	0	0	62	France	0.33	2.5
Yemen	0.58	0	0	1	20	Britain	0.42	–2.7
Maldives	1.00	0	0	0	32	Britain	.	2.5

Sources:
[a] Author generated variables
[b] UNDP (2003). *Human Development Report*

context with which the study deals. They speak to cultural patterns that have an explanatory power for reproductive change. Therefore, I gave preference to gender specific indices, such as the ratio of girls to boys in primary education enrollment, over non gender weighted ones such as the total rate of school enrollment. There is an obvious relationship between this gender sensitive index and the statements of developmental idealism: more education is better and gender equality is a goal of development. It is also related to Islamic reformism, which champions female education, as opposed to social settings where the Sufi ethos is ingrained and female education is not taken seriously, to say the least.

I measured overall national development using indices that reflect the position of each country on a scale of actual achievements. The most comprehensive one is the *Human Development Index* (HDI), which measures a country's average achievement in three basic dimensions of human development: (1) a long healthy life (2) knowledge (3) a decent standard of living (see the formal definition in Table 4.3).

The strength of the channels of international exchange is captured using indices collected from the UNDP published material. Foreign aid and international economic exchange might affect the outcome as modifiers, or as stand-alone factors of change in reproductive behaviour. Their intensity is captured through the use of the average official development assistance (ODA) received in 2001 and the net foreign direct investment inflows (NFDI) as per cent of GDP. These indices are proxies meant to measure a development process that took decades to unfold.

Family planning programmes can affect reproductive behaviour directly. They can also operate as pathways of dissemination for developmental idealism, thus

Table 4.3 Definitions and mathematical formulae of indices

Coale's indices: At a point in time t the following symbols are defined, f_i as the births per women age i; w_i the number of women age i; m_i the number of married women age i; and F_i the standard fertility schedule (age-specific marital fertility rates of Hutterite women, for the period 1921–1930). The following standardized indices can be calculated:

I_f, a comparative index of general fertility $= \Sigma f_i w_i / \Sigma F_i w_i$
I_m, a comparative index of proportion married $= \Sigma F_i m_i / \Sigma F_i w_i$

United Nations' Human Development Indices (HDI) are summary measures of a country's achievements in human development. *HDI*, the human development index, measures a country's average achievement in three basic dimensions of human development: (1) a long healthy life as measured by life expectancy at birth (2) knowledge as measured by adult literacy rate (with 2/3 weight) and the combined primary, secondary and tertiary gross enrolment ratio (with one third weight) (3) a decent standard of living, as measured by GDP per capita (PPP US$). Before HDI itself is calculated, an index needs to be created for each of the three dimensions. Performance in each dimension is expressed as a value between 0 and 1. It represents the ratio of the difference between the actual value and the minimum value to the difference between the maximum value and the minimum value of the dimension. HDI is then calculated as a simple average of the dimension indices. *GDI*, the Gender-related Development Index, adjusts the average achievement measured by HDI to reflect the inequalities between men and women in the three dimensions: long and healthy life, knowledge and a decent standard of living. *FPPEI*, the Family Planning Policy Evaluation Index, is a score attributed to each country on the basis of its government's efforts in the field of family planning activities.

affecting reproductive behaviour indirectly. The family planning policy evaluation index (FPPEI) provides a country-specific assessment of a government's efforts in support of family planning. In order to create a summary index that measures the consistency of a government's support for family planning during times prior to the onset of fertility decline, I averaged the estimates of this index over 1972, 1982, 1989 and 1994, but dropped the estimates for 1999.

Only 14 governments had reported their national figures of FPPEI. If we exclude Uzbekistan due to a misclassification caused by an artifact of the data, the front runner was Indonesia followed by Tunisia, with average scores of 72 and 62. Mali, Yemen, Syria and Jordan provided the least governmental support. Most governments, in fact, provided weak support corresponding to scores of about 40 and, against expectation, several developmental states fell into this category.

The Methods

The empirical analyses of the spatial and temporal patterns aimed to verify whether the structure of the data supports the narratives developed above and, more specifically, the four formal hypotheses. For this purpose, the analysis involved three approaches: (1) a visual examination of the spatial patterns using GIS mapping tools (2) more formal statistical procedures of multivariate cluster analyses (3) examination of the temporal patterns through time series of TFRs.

First, I used maps generated in Arc GIS to visually examine the spatial patterns of the levels of fertility, the pace of fertility change and the major factors contributing to that change. This visual examination focused on the most recent estimates of TFRs, indices of general fertility (I_f) and proportion married (I_m) and on the average annual decline in I_f and I_m (Fig. 4.3). This way of mapping the data is equivalent to a one-dimensional cluster analysis. Indeed, using the so-called *natural breaks* option in Arc GIS defines the cut-off values in the classification variable according to an algorithm that minimizes the intra-group variances and maximizes the inter-group variance. I restricted to four the user-defined number of classes, but relied on the *natural breaks* option for class magnitude, except for TFR, where I forced the creation of one class for below-replacement fertility (TFR = 2.1). This first analysis classified the countries according to the measures of the outcome, thus giving a sense of fertility levels and the rates of fertility decline.

Second, I applied a multivariate quantitative cluster analysis using hierarchical procedures of classification in order to test whether the structure of the data supports the ideal-types that reflect the four hypotheses specified above. For this classification, quantitative summary indices are used to characterize the socio-economic level of development with special emphasis on its human dimension. The general model includes fertility measures such as I_f, I_m and the average annual change in these measures, TFR and per cent using family planning. The model also includes socio-economic and cultural indices such as per cent Muslim, per cent of the population in urban dwellings, ratio of girls to boys in school enrolment, HDI, purchase parity power of the gross domestic product and per head ODA received. I contrasted this

(A)

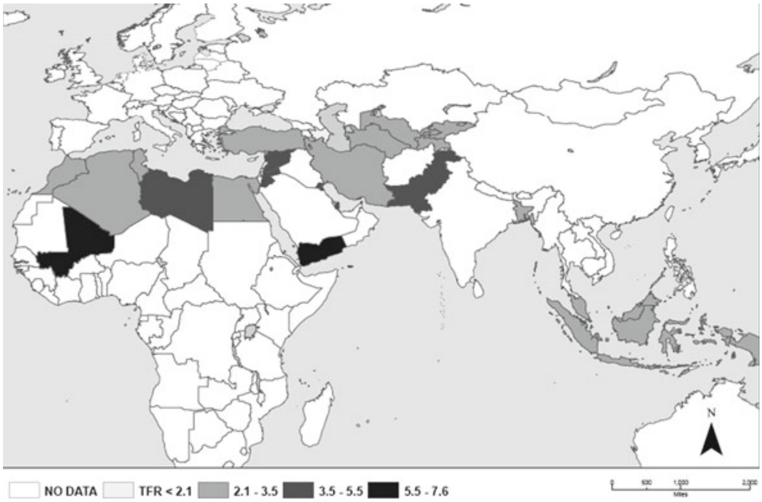


Fig. 4.3 Fertility levels, pace and major factors of the decline **A** Total fertility rates (TFRs)

(B)

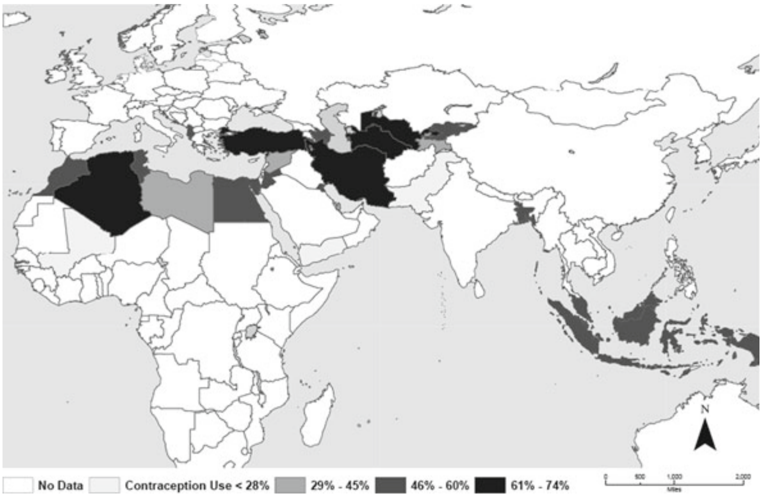


Fig. 4.3 **B** Per cent of women using family planning

socio-economic model with a second model based solely on demographic characteristics. The aim was to check the role of variables related to developmental idealism, versus other general socio-economic variables, in shaping the patterns of fertility transition. If the socio-economic characteristics of the countries were the causal factors of fertility change, the clusters generated in the socio-economic model would match those generated in the demographic model.

(C)

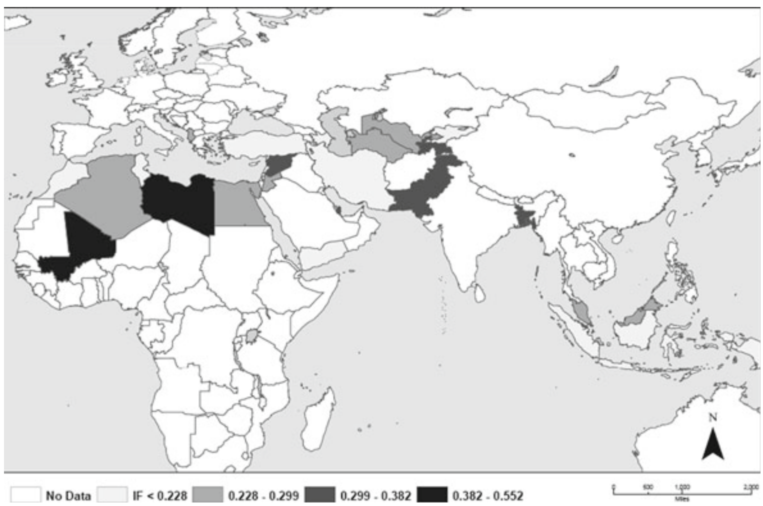


Fig. 4.3 **C** Index of general fertility (I_f)

(D)

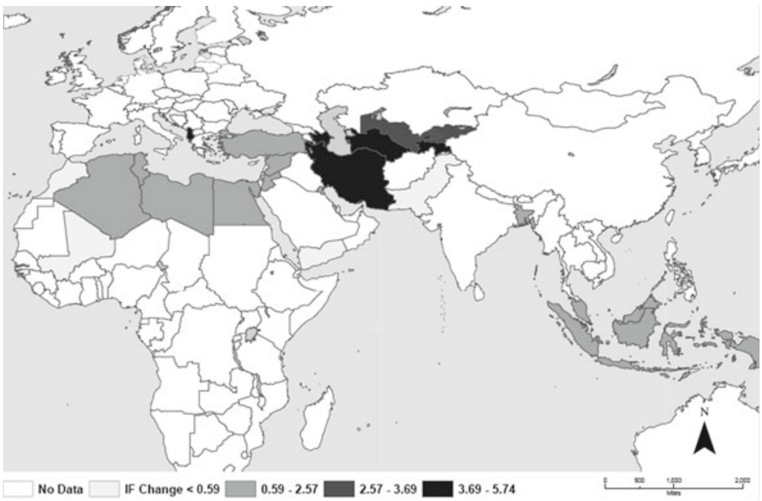


Fig. 4.3 **D** Average annual change in the index of general fertility

For this same purpose, I further developed a cluster analysis model based on qualitative variables and constrained it to return four clusters. The qualitative classification model aimed to test the explanatory power of binary variables assessing the presence or absence of a family planning policy, experience with Islamic reformism, a Turkish paradigm type of policy, the Ottoman legacy and European colonization. The last two variables hint at two major historical episodes that had meaningful

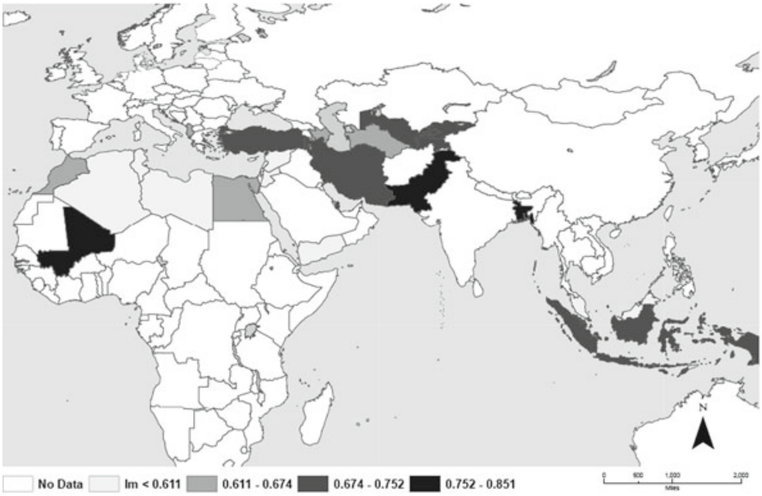
(E)


Fig. 4.3 **E** Index of proportion of women married (I_m)

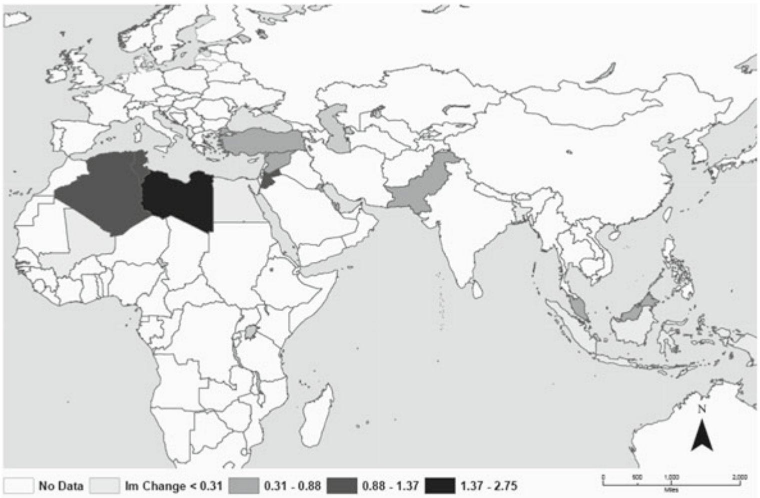
(F)


Fig. 4.3 **F** Average annual change in the index of proportion married

impacts on the social structures in which the forces of societal change (family planning, developmental idealism and Islamic reformism) operated. The expectation was that the results would reflect the key concepts underlining the whole approach through the salient characteristics of the clusters. For example, one cluster was expected to be composed of countries where developmental idealism was strongest, a second composed of countries where the Sufi ethos was still predominant, a third composed of countries with mixed influences and a fourth cluster was expected to represent countries with little to no acceptance of developmental idealism.

Third, I compared and contrasted temporal patterns of fertility decline using time series of TFRs in each of the clusters that resulted from the previous three classification models. The goal was to reconstitute time series of TFRs covering at least the last 50 years, but this proved impossible for most countries. Consequently, the truncated time series failed to provide conclusive evidence about pre-transition fertility levels or locate the timing of the onset of fertility decline.

Results and Discussion

Spatial Patterns of Fertility, Nuptiality and Contraception

The maps in Figure 4.3 provide a first level of exploration of the spatial patterns of fertility levels, the different rates of decline and the two major contributors to the decline (family planning and proportion married). Panel A shows the spatial patterns of recent estimates of TFRs (from 1995 to 2001, except for Syria and Brunei).

Only Albania and Azerbaijan achieved replacement level or below. On the highest end stand Mali and Yemen with TFRs above 5.5 births per woman. The remaining countries form two groups. The first group shows evidence of a well engaged transition with TFRs ranging from 2.1 to 3.5. This group fits the prediction of the first hypothesis. It is composed of countries that implemented post-colonial national development programmes either in a centrally planned economy or with a more liberal Turkish Paradigm policy, thus providing strong pathways for the dissemination of developmental idealism. This cluster stretches over the Iranian, the Asian, the Turkic and the Arabic ethno-linguistic groups. Hence, it underscores the significance of developmental idealism, not ethnicity or language, as a causal factor of change. Bangladesh, Brunei, Bahrain and Morocco also belong to this group, but they more closely fit the third hypothesis, developmental idealism disseminated through non-induced means in a social structure marked by classical Islam. These countries, however, were comparatively more open to international influence.

In the second group, fertility transition is either at an early stage, with TFRs ranging from 3.9 to 4.4 in Qatar, Libya, Kuwait, Syria and Jordan, or the onset of fertility decline is still doubtful with TFRs of 5.5 and above in Pakistan, Mali, Yemen and the Maldives. These countries share two features: an intact traditional system of classical Islam and non-induced pathways of developmental idealism.

Contraception within marriage is one of two major contributors to fertility decline, given that marriage is quasi-universal in the Muslim context. Panel B in Figure 4.3 displays a classification on the basis of the contraception prevalence rate. In 18 of the 24 countries, more than half of women of reproductive age were current users of family planning. Again, the highest rates (darkest areas) belong to the core developmental states, in compliance with the first hypothesis, which predicts a rapid change in reproductive behaviour in countries where both Islamic reformism and developmental idealism were strong and operated against a weakened Sufi ethos. It is precisely the case in Turkey, Algeria, Iran and its northern neighbours, the former Muslim republics of central Asia.

Panel C in Figure 4.3 displays a classification according to the most recent estimates of the index of general fertility (I_f). The earliest transitions took place in Albania and Malaysia with an I_f of 53 in 1955 and in Turkey with an I_f of 52 in 1960; however, it is doubtful that these figures represent pre-transition levels. Albania was identified in the European Fertility Project as a laggard with regard to fertility transition. It should have never been included in the European Fertility Project. Misclassification put aside, it belongs, rather, to the earliest transitions in the Muslim world because of an early exposure to westernization. Like Albania, Turkey also experienced an early transition and for similar reasons.

Mali in 1987 and Libya in 1991 had the highest levels of I_f yet, positioned below 55% of the Hutterites' fertility, a figure close to European pre-transition levels. Although the data at hand give only truncated views of fertility transitions, making it difficult to assess with certainty pre-transition fertility levels, the highest figures of I_f, such as in Libya (74 in 1973), Pakistan (70 in 1976) and Algeria (69 in 1966), provide acceptable approximations of these levels. The populations of Kuwait, Jordan and Tunisia also experienced levels of I_f of about 60%, from 1965 to 1971. Given the late onset of the decline in these countries, these figures lend support to the previous approximations of the pre-transition levels.

Following the lead of the European Fertility Project, the pace of fertility decline was measured using the average annual change in I_f (Table 4.1). Among countries with suitable data, general fertility declined at an average annual rate higher than 20% for periods of observation ranging from 4 years in Tajikistan to 43 years in Malaysia. This average rate of decline is twice the 10% considered an indicator of the onset of the secular decline in the European Fertility Project (Coale & Watkins, 1986).

Five countries departed from the general trend. Fertility rose in Qatar and Mali during short periods of observation: a 4% increase in I_f over 4 years in the first case and a 19% increase over 11 years in the second case. It declined at a moderate pace in Morocco, Bangladesh and Pakistan, with 11, 13 and 15% declines in I_f over the periods 1982–2000, 1981–1991 and 1968–1998. Although at times truncated, the series of TFRs provide further evidence of substantial decline in Morocco and Bangladesh. Again, with regard to the pace of decline, these cases fit well the third hypothesis.

Panel D of Figure 4.3 displays the spatial distribution of the countries according to the average annual per cent change in $I_{f,}$ which informs about the rate of fertility decline. Aside from Albania, the most rapid pace of decline occurred in Iran and its northern neighbours: Azerbaijan, Turkmenistan, Uzbekistan, Tajikistan and Kyrgyzstan. This group fits the second hypothesis with some nuances: developmental idealism disseminated through strong pathways in the context of centrally planned economies in the former Muslim republics of central Asia within the political context of the USSR. It seems that the influence of the state-sponsored developmental policies outweighed the effect of the traditional system marked by the Sufi ethos.

Iran fits even better the second hypothesis. It implemented a Turkish paradigm model of modernization under the Shah's rule against an intact traditional Islam. Not only was the onset of fertility decline delayed, but the developmental policy backfired for a time with regard to fertility control. What then explains the rapid fertility decline in post-revolution Iran? I contend that the impact of the Shah's modernization continued to affect post-revolution Iran, and the new Islamic republic is, in a sense, a reformation of the classical system of Islam. It seeks to transform the system of classical Islam that is oblivious to social progress. On the other end of the spectrum are the populations of Pakistan (0.51%) and Morocco (0.59%). The onset of fertility decline is yet to be seen in Mali and Qatar, where fertility was increasing at average annual rates of 1.71 and 1.08%. These two cases also fit well the third hypothesis. Yet, they might not have been sufficiently exposed to developmental idealism, thus the lagging transition.

Overall, notwithstanding the challenges presented by the unequal periods of observation, most Muslim populations are experiencing an annual rate of fertility decline ranging from 1.05% in Indonesia to 3.69% in Kyrgyzstan.

The index of proportion married (I_m) provides a standardized measure of the intensity of nuptiality, relative to the Hutterites'. Panel E in Figure 4.3 displays a classification of the countries according to the scores of I_m. The upper limit of the highest cluster is only 15% below that of the Hutterite standard, while the upper limit of the lowest cluster is about 40% below that standard. It comes as no surprise that Mali and Pakistan, both laggards with respect to fertility decline, had the highest proportion married.

The figure for Bangladesh shows an interesting contrast with Mali and Pakistan. Even though Bangladesh's estimate of I_m is more recent (1991), it is comparable to that of Mali (1987) and Pakistan (1976), and we know that fertility declined rapidly in Bangladesh. This is further evidence of the important role contraception within marriage played in Bangladesh's well documented fertility decline (Cleland et al., 1993, 1994). It fits the third hypothesis, underscoring international influence, which promotes fertility decline even within the context of a traditional family structure.

A substantial structural change would necessarily affect the patterns of nuptiality. It might lower the proportion married significantly, as is the case in Tunisia, Libya, Algeria or Malaysia, where the index of proportion married (I_m) reached below 60% of the standard. Nonetheless, overall Muslim nuptiality remained high over the periods of observation, with a proportion married that ranged from 60 to 75% of the Hutterites'.

Panel F in Figure 4.3 displays the annual average decrease in I_m. Given the problems associated with the unequal reference periods, one must be cautious about these figures, without completely dismissing their significance. They show a substantial downward change, which predicts rapid fertility decline in the future. Libya, Algeria and Tunisia are frontrunners, with Jordan having a comparable rate of decline but referring to more recent estimates of I_m.

Two major regimes of nuptiality seem to bifurcate the area. The first, characterized by a slow change in the intensity of marriage, predominates in Mali, Egypt, Bangladesh, Indonesia and the states of the Arab Gulf. The second, characterized by a faster change, predominates in Albania, Turkey, Syria, Pakistan and Malaysia.

Multivariate Cluster Analysis of Fertility Levels and Trends

The character of this study is heuristic and structure-seeking. The study aims to generate knowledge that will lead to further research questions and to bring structure to existing knowledge and data. Therefore, cluster analysis is a particularly suitable approach since most cluster methods are heuristic (Aldenderfer & Blashfield, 1984). It is applied here in a way that minimizes structure-imposing operations such as a pre-defined number of clusters, level of clustering and choice of variables. Nonetheless, this study is a theory-driven investigation, not a blind exploratory data analysis. The key to using cluster analysis is to know when the groups are *real* and not merely imposed on the data by the method. Keeping this in mind, I selected the variables and operationally defined them by reference to an explicitly stated theory.

Each of the three models of cluster analysis sought to uncover the inner structure of the universe investigated from a different perspective. The first model (Fig. 4.4) aimed to uncover the general structure of the data on the basis of both demographic and socio-economic characteristics of the populations. The second is a pure fertility transition model that aimed to uncover the structure of the data on the sole basis of measures of fertility level and pace of change. The third model (Table 4.3) is based on author-generated qualitative variables that characterize the populations' exposure to factors of progress and conservatism likely to speed up or slow down social change in general and change affecting reproductive behaviour in particular.

Socio-economic Cluster Analysis

The socio-economic based cluster analysis aimed to check whether the empirical data support or refute the hypothetical classification deduced from the theory, along the lines of the key concepts discussed above. The dendrogram – the tree-like diagram generated by the cluster procedure and representing the hierarchy of clusters based on degree of similarity – in panel A (Fig. 4.4) illustrates the results of the first model, where three clusters emerge clearly. One is composed of the atypically wealthy countries (Kuwait, Bahrain and Qatar) of the Arabic Gulf. It emerges at the first level of classification wherein the lengths of the two branches testify to the singularity of this group of countries with small populations enjoying standards of living comparable to those of the most economically advanced countries. Despite high standards of living, fertility transition is at an

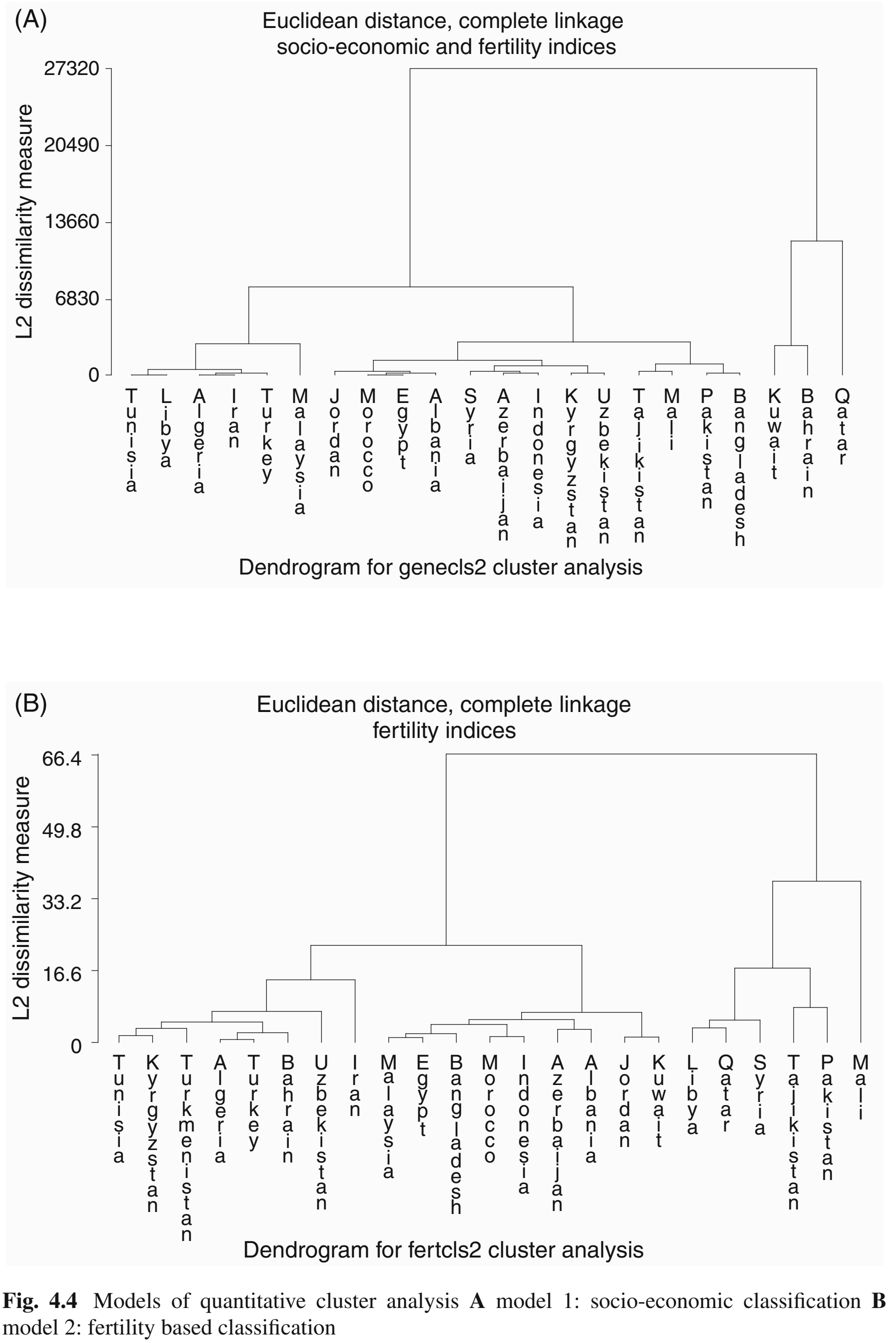


Fig. 4.4 Models of quantitative cluster analysis **A** model 1: socio-economic classification **B** model 2: fertility based classification

early stage in all three countries. Qatar's fertility level is comparable to the levels observed in such poor countries as Pakistan and Mali. This singularity is reflected in the substantial distance that separates it from the other two neighbouring countries.

A second cluster composed of Tunisia, Libya, Algeria, Iran, Turkey and Malaysia emerges at the second level of classification. It can be characterized as the group of developmental states. Its internal structure differentiates the Turkish Paradigm countries from Malaysia, which shares most but not all tenets of the Paradigm. The rest of the countries are clustered in a third group. These are populations whose development and fertility changes occurred under the influence of several forces. This aspect is clearly illustrated by the internal structure of this big cluster, if one overlooks minor features of individual units. In one case the salient characteristic can be labelled non-induced diffusion of developmental idealism (Jordan, Morocco, Albania and Egypt to some extent); in a second case, it is centrally planned development (Syria, Azerbaijan, Indonesia, Kyrgyzstan and Uzbekistan); while in a third case, targeted programmes of development and family planning constitute, more or less, the privileged pathways for the diffusion of developmental idealism (Tajikistan, Mali, Pakistan and Bangladesh).

Fertility Cluster Analysis

Likewise, the fertility-based cluster analysis aimed to assess the extent to which fertility data support the structures implied in the four hypotheses discussed previously. The dendrogram in panel B of Figure 4.4 illustrates the results of the fertility model (model 2). Two clusters emerge at the first level of classification. The first cluster is composed of a group of laggards, including countries enjoying high levels of wealth (Libya and Qatar) alongside countries with significantly lower standards of wealth (Pakistan and Mali). The internal structure of the cluster contrasts Mali, a natural fertility country, from the rest of the units. The structure of this cluster discounts the socio-economic factors as determinants of fertility transition.

The second cluster is composed of units unmistakably divided into one group of hard core Turkish paradigm countries and a group of countries with more moderate national policies. Diffusion of developmental idealism was state induced in the hard core Turkish paradigm group either by way of centrally planned economies (Kyrgyzstan, Turkmenistan, Algeria and Uzbekistan), or through nation building policies inspired by the Turkish model (Tunisia, Turkey and Iran). The internal structure of the group of moderate national policies itself highlights the proximities of countries where family planning was an important factor (Malaysia, Egypt and Bangladesh), those with centrally planned economies (Albania and Azerbaijan), those with targeted policies (Morocco and Indonesia) and those countries with non-induced diffusion of developmental idealism (Jordan and Kuwait). The structure of this cluster underscores the importance of developmental policies and family planning programmes as causal factors in fertility transition.

Table 4.4 Clusters based on qualitative variables

Cluster A	Cluster B	Cluster C	Cluster D	
Jordan	Pakistan	Bahrain	Egypt	Tajikistan
Syria	Bangladesh	Maldives	Tunisia	Algeria
Kuwait	Morocco	Qatar	Azerbaijan	Kyrgyzstan
Yemen	Iran	Brunei	Turkmenistan	Albania
Libya	Malaysia	Mali	Turkey	
	Indonesia		Uzbekistan	

Qualitative Cluster Analysis

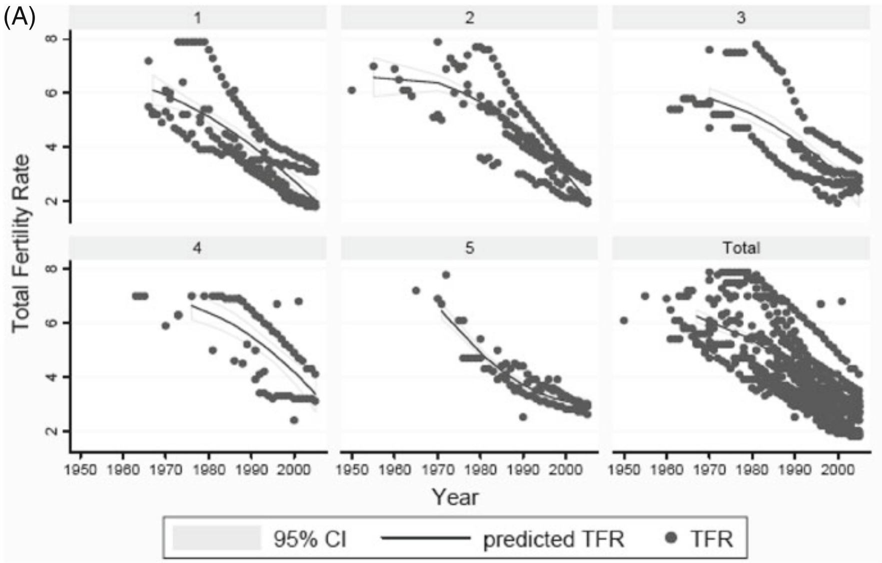
The cluster analysis based on qualitative variables was applied to check further whether the theory-based classification makes sense. The aim was to capture important qualitative attributes of the units of analysis that might be missed by the quantitative cluster analysis. Unlike the quantitative cluster analysis, the cluster analysis based on qualitative variables does not return a neat graphical illustration of the results. It creates instead a user-defined number of clusters to which it assigns numbers for identification purposes only (Table 4.4). Units assigned to the same group are the most homogenous according to binary variables I created to account for the most powerful pathways of diffusion of developmental idealism and for historical experiences that predispose the populations to acceptance of the tenets of developmental idealism.

Cluster D confirms the characterization obtained from the previous two models as a group of pure developmental states. Cluster C sits at the other end of the developmental scale; it regroups countries that are oblivious to developmental idealism. Clusters A and B are hybrid groupings with regard to the five qualitative variables used to build the model (family planning, Islamic reformism, a Turkish paradigm type of policy Ottoman legacy and European colonialism).

The geographic distribution of the four clusters suggests that spatial proximity has some influence but it is not the decisive factor. For example, despite geographic distance, Morocco – a non developmental state with no Ottoman legacy – is closer to Iran, Pakistan or Indonesia than it is to the cluster of developmental states with an Ottoman legacy, such as neighbouring Algeria and Tunisia. Similarly, Mali, a poor African country, clusters with such wealthy countries as Brunei, Bahrain or Qatar. Although Libya, Yemen, Jordan and Syria share some characteristics of developmental states and Ottoman legacy, their nation building policies emphasized Arab identity. A possible effect of this emphasis could have been the strengthening of traditional social structures.

Temporal Patterns of Fertility Decline

Figure 4.5 displays the plots of time series of TFRs over the last four to five decades for the twenty-six countries. Each country is represented by a dashed line; the solid



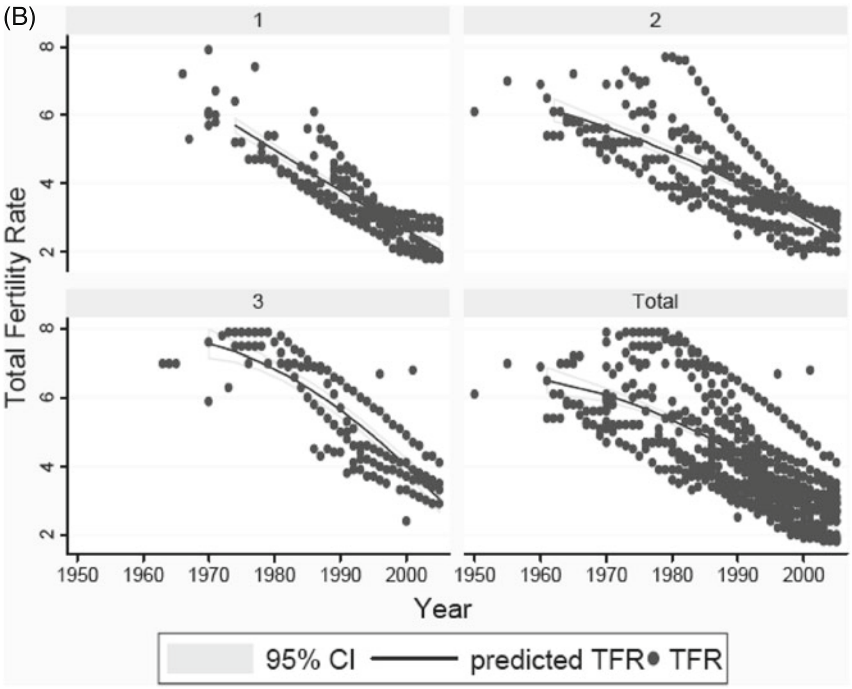


Fig. 4.5 Temporal patterns of fertility decline in Muslim populations **A** Socio-economic classification **B** Fertility classification **C** Qualitative variables classification

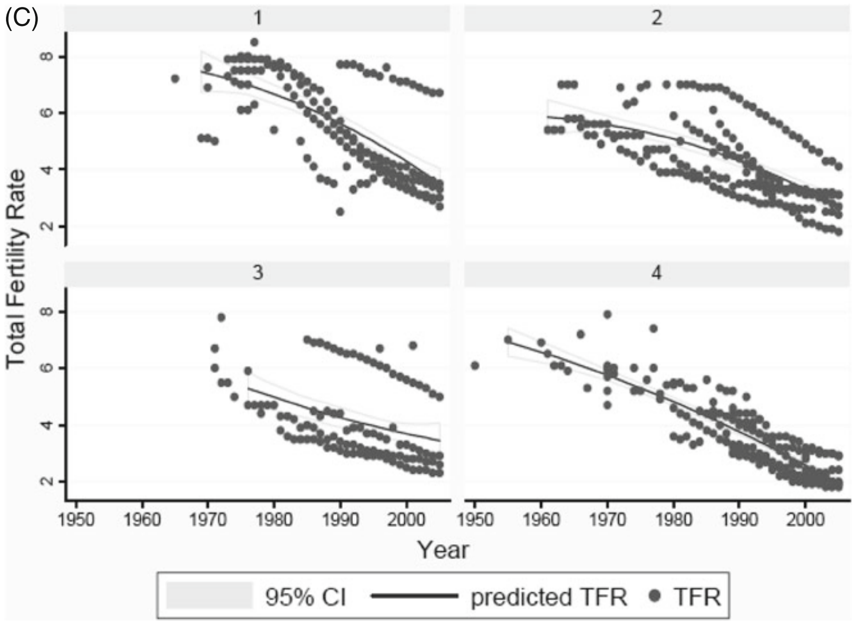


Fig. 4.5 (continued)

line represents the global fertility trend in the cluster with a specification of the 95% confidence interval. Each panel in the figure represents the temporal patterns of the clusters generated by one model of the cluster analysis. The aim is to test the hypotheses with regard to the pace of fertility decline.

The plot for all countries (total in panels A and B) shows that overall fertility decline followed a short transition model. Once it started declining, fertility fell rapidly from quite high pre-transition levels that reached six children per woman and above. The slope of the curve denotes important variation in the pace of the decline, as well as significant differentials in current and in pre-transition levels.

The demographic classification (panel B) underscores the consistency of the temporal pattern. The general trend in the cluster of the hard core Turkish paradigm countries is characterized by a straight line and a steep slope of the curve of TFRs that reached replacement-level fertility at the end of the century (panel B.1, Fig. 4.5). The temporal pattern of fertility decline in this group is different from all other groups regardless of the type of analysis applied. Indeed, cluster D (Table 4.4 and panel C.4), in which the common denominator is developmental idealism, displays a fairly homogenous fertility trend: while its units are spread over very dissimilar clusters in the socio-economic model of classification, they regroup in cluster 1 of the fertility model characterized as the cluster of developmental states. This observation discounts the claim that levels of wealth and human development are the causal factors of fertility decline. It rather lends strong support to the first hypothesis of this study.

Conclusion, Limitations and Recommendations

The story of fertility transition is first and foremost a story of secular social change. The European experience served as a launch pad to the construction of fertility transition theories for reasons known to students of population. Now that fertility decline is becoming universal and a critical mass of international demographic data is available, there are good opportunities to take the theory to new levels of inference.

I began this research with a holistic view of the Muslim world that hides the difficulty of stating what it really is. In order to overcome this identification problem, I started with an operational definition that equates the Muslim world with membership in the Organization of Islamic Conference. This definition was subsequently adapted by excluding from the universe of investigation countries whose populations did not fit the criterion of a Muslim demographic majority.

An exploratory analysis using the mapping tools of Geographic Information System software identified six sub-regions, providing delineation of sub-cultural areas within the realm of Islam. This specification in itself advances knowledge about the Muslim world in that it provided an entry to more specific analyses at the outset, rather than starting the study with Islam as a black box. It further proved useful in showing that patterns of fertility transition were linked to cleavages along the lines of the key concepts of this study, rather than along the lines of language and ethnicity.

Further analyses using multivariate cluster analysis based on quantitative and qualitative data lent support to the main hypotheses of the study. They underscored the importance of developmental idealism as an engine of change in reproductive behaviour and, consequently, in fertility decline.

According to the criterion of the European Fertility Project for the onset of secular fertility decline, there is ample evidence that fertility transition is well advanced in the Muslim world. However, unlike in historical Europe, fertility transitions in the Muslim world began at high levels of general fertility (more than six children per woman) and a high intensity of marital fertility that reached 70% of the Hutterites' standard. Furthermore, contraception within marriage was the most used mean of fertility control. The temporal patterns of fertility transition underscore a rapid decline with noticeable differentials between clusters. The hard core developmental countries took the lead with the fastest declines.

Marriage remained almost universal even in the Muslim populations most exposed to developmental idealism. Changes in nuptiality regimes towards later marriage and more egalitarian gender relationships have had some bearing on the ideational and behavioural changes towards small family size. However, these changes did not seem to be a condition for the adoption of the norms of small family size or for the practice of family planning, as Bangladesh's case amply demonstrates.

The study supports the conclusion of previous major research projects that fertility decline is only weakly linked to levels of economic development. It also discounts the claim that markers of ethnicity and languages are causal determinants of fertility differentials and trends. It rather establishes that the causal factors are more voluntary societal choices. These choices are mainly related to the project of society as implemented through nation building efforts and policies, with adoption of developmental idealism and reformation of the classical system of Islam as the crucial factors.

This research underscores the complexity of studying Muslim fertility, not only because of the scarcity and low quality of the data, but also because of the complexity of the realm of Islam itself. Ironically, this complexity contrasts with the simplicity of the approaches hitherto used to investigate it. The conclusions of this study suffer from the data predicament. Nonetheless, the study broke new ground with welcome additions to the theory and methods of studying fertility transitions.

The study suffers from limitations related to data quality and availability and the small number of previous studies on Muslim fertility. Lack of fertility data for Saudi Arabia excluded a case where the Sufi ethos was wiped out not by the tide of developmental idealism, but by the tide of Wahabism, an Islamic reformism skeptical of Western patterns of thinking and lifestyle and seeking a return to the Islam of the founders, purified of all accretions. As such, this country would have been the appropriate case study to test the fourth hypothesis. Because it had no substitute in the sample, we may have missed an important source of insight.

I also did not fully exploit survey data to calculate time series of TFRs for the few countries that participated in the Demographic and Health Surveys or in other DHS-like national surveys. Such data sources provide excellent opportunities to pursue this research, even though only a limited number of Muslim countries participated in these major international data collection programmes.

Developmental idealism as an ideology of social change and Islamic reformism as reformation of the indigenous system of belief provided powerful concepts for explaining the dynamics of fertility decline, and of social change in general, in the Muslim world. A better understanding of the dynamics of these high order concepts requires better-researched historical comparative studies. Such studies should compare and contrast the trajectories of national populations and sub-national populations, as well as cross-national regions within the realm of Islam, a task only partially accomplished here.

More effort towards census and vital statistics data collection will also be needed to create a richer data bank that will allow analysts to build longer time series of measures of fertility. Longer series of indices are a *sine qua non* condition for more advanced comparisons of the patterns of fertility decline at the country and sub country levels. This study only broke the ground for such investigations.

Acknowledgement I began this research at the CDE, UW-Madison as part of my PhD dissertation. The present text was completed at the INRS-Urbanisation Culture Société. Support from both institutions made this research possible. The Social Sciences and Humanities Research Council of Canada provided partial financial support through an institutional grant to the INRS-UCS for the completion and the diffusion of this document. I thank my dissertation committee at UW-Madison for comments on early versions of the paper and Hanne Bamba for assistantship at the INRS-UCS. I thank the editors, in particular Myron Gutmann and Ken Sylvester, as well as anonymous readers whose comments and editorial suggestions helped in improving the text readability. I remain of course solely responsible for the content and any remaining imperfection.

References

Aldenderfer, M. S. & Blashfield, R. K. (1984). *Cluster analysis. Sage University paper series on quantitative applications in the social sciences 07-044*. Beverly Hills, CA: Sage.

Beaver, S. E. (1975). *Demographic transition theory reinterpreted*. Toronto, ON: Lexington Books.

Bellah, R. N. (1970). *Beyond belief: Essays on religion in a post-traditional world*. New York: Harper & Row.

Bongaarts, J. (2002a). The end of the fertility transition in the developing world. Policy Research Division Working Paper no. 161. Washington, DC: Population Council.

Bongaarts, J. (2002b). The end of the fertility transition in the developed world. *Population and Development Review*, *28*, 419–443.

Bongaarts, J. (2005). The causes of stalling fertility transitions. Policy Research Division Working Paper no. 204. New York: Population Council.

Caldwell, J. C. (1982). *Theory of fertility decline*. London: Academic Press.

Caldwell, J. C. (1995). What do we know about fertility transition? Paper presented at the John Caldwell Seminar, 14–17 August, 1995. Canberra: The Australian National University, National Center for Epidemiology and Population Health.

Caldwell, J. C. & Ruzicka, L. T. (1987). Demographic levels and trends. In J. Cleland, C. Scott (Eds.), *The world fertility survey: An assessment (*pp. 741–772). London: Oxford University Press.

Chesnais, J. C. (1986). *La transition démograhique, étapes, formes, implications économiques*. Paris: Presses universitaires de France.

Cleland, J. (1985). Marital fertility decline in developing countries: Theories and the evidence. In J. Cleland, J. Hobcraft (Eds.), *Reproductive change in developing countries: Insights from the World Fertility Survey* (pp. 222–252). London: Oxford University Press.

Cleland, J. G., Phillips, J. F., Amin, S., & Kamal, G. M. (1993). Fertility decline in post-Independence Bangladesh: Evidence from recent demographic surveys. Policy Research Division Working Paper no. 54, New York: Population Council.

Cleland, J. G., Phillips, J. F., Amin, S., & Kamal, G. M. (1994). *The determinants of reproductive change in Bangladesh: Success in a challenging environment*. Washington, DC: World Bank.

Coale, A. J. (1973). The demographic transition reconsidered. *In IUSSP, Proceedings of the IUSSP International Population Conference, Liège, Belgium*, Ordina Edition. Liège: IUSSP.

A. J. Coale, S. C. Watkins (Eds.) (1986). *The decline of fertility in Europe*. Princeton, NJ: Princeton University Press.

Denny, F. M. (1994). *An introduction to Islam* (2nd ed.). New York: Macmillan.

Freedman, R. (1987). Fertility determinants. In J. Cleland, C. Scott (Eds.), *The world fertility survey: An assessment* (pp. 773–795). London: Oxford University Press.

Guend, A. (1993). Discours politique, discours religieux et transition de la fécondité en Algérie. Paper presented at the IUSSP International Population Conference, Aug. 24–Sept. 1, Montreal, Canada.

Göçek, F. M. (1996). *Rise of the bourgeoisie, demise of empire*. New York: Oxford University Press.

Hancioğlu, A. (1994). *Turkish demographic and health survey*. Ankara: Ministry of Health, Hacettepe University Institute of Population Studies, and Macro International Inc.

Janowitz, B. S. (1971). An empirical study of the effects of socioeconomic development on fertility rates. *Demography*, *8*, 319–330.

Kirk, D. (1996). Factors affecting Moslem natality. In B. Berelson (Ed.), *Family planning and population programs: A review of world developments*. Chicago: University of Chicago.

Kouaouci, A. (1983). *Normes familiales islamiques et fécondité en Jordanie, en Indonésie et au Pakistan* (Recherches Démographiques Cahier no. 5). Louvain-la-Neuve, Belgium: UCL, Département de Démographie.

Lesthaeghe, R. (1983). A century of demographic and cultural change in Western Europe: An exploration of underlying dimensions. *Population and Development Review*, *9*, 411–435.

"Muslim Distribution" [map]. 1:40,000,000. "Perry-Castañeda Library Map Collection-UT Library Online." http://www.lib.utexas.edu/maps/world_maps/muslim_distribution.jpg (February 13, 2010).

Notestein, F. W. (1945). Population – The long view. In T. W. Schultz (Ed.), *Food for the world* (pp. 36–57). Chicago: University of Chicago Press.

Obermeyer, C. M. (1994). Reproductive choice in Islam: Gender and state in Iran and Tunisia. *Studies in Family Planning*, *25*, 41–51.

Omran, A. R. (1992). *Family planning in the legacy of Islam*. London: Routledge.

Palloni, A. (2001). Diffusion in sociological analysis. In J. Casterline (Ed.), *Diffusion processes and fertility transition: Selected perspectives* (pp. 66–114). Washington, DC: The National Academies Press.

Richards, A. & Waterbury, J. (1990). *A political economy of the Middle East*. Boulder: Westview.

Thornton, A. (2001). The developmental paradigm, reading history sideways, and family change. *Demography*, *38*(4), 449–465.

Thornton, A. (2005). *Reading history sideways: The fallacy and enduring impact of the developmental paradigm on family life*. Chicago: The University of Chicago Press.

Tillon, G. (2001). *Le Harem et les cousins*. Paris: Seuil.

Turner, B. S. (1974). *Weber and Islam: A critical study*. Boston: Routledge & Kegan Paul.

UNDP. (2003). *Human development report 2003*. United Nations, New York: Oxford University Press.

United Nations. (2002). *Demographic yearbook: Natality statistics CD-ROM*. New York: United Nations.

Vance, R. B. (1952). Is theory for demographers? *Social Forces*, *31*, 9–13.

Chapter 5
Spatial and Temporal Analyses of Surname Distributions to Estimate Mobility and Changes in Historical Demography: The Example of Savoy (France) from the Eighteenth to the Twentieth Century

Pierre Darlu, Guy Brunet, and Dominique Barbero

Introduction

The distribution of surnames is widely used to describe the genetic structure of populations, based on the analogy between genes and surnames. Surnames are quite strictly neutrally transmitted through the male line in most European countries (see reviews in Crow (1983) and Lasker (1985)), allowing some inference on the demographic evolution and movement of male members of a population and – since males usually choose their spouse from within a limited distance – of the whole population. The number of surnames is usually large and their regional distribution mostly local, such that the person's name can quite precisely signify his or her geographic origin. Moreover, thanks to the possibility of obtaining surname registers, in some cases from as early as the eighteenth century, surname turns out to be an efficient tool to describe past migration patterns that cannot be otherwise described quantitatively (see review in Darlu, Degioanni and Zei (1996)). Indeed, reliable statistics on the movement of populations before the creation of the first census are unfortunately lacking, particularly for investigating local migrations. Moreover, historians have had to resign themselves, for lack of information, to focus on the movements of the nobility and the rich upper class, rather than on those of rural people (Nicolas, 1978).

Two different strategies are generally used, both coming from population genetics approaches. The first is used when the surnames are recorded only once at a given time. It is initially based on F_{ST} statistics, first developed by Wright (1951), a variance analysis intended to describe the distribution of genes into subdivided populations and to infer migration flux (Slatkin & Barton, 1989; Cockerham & Weir, 1993). Another model, proposed by Karlin and McGregor (1967), has been used

P. Darlu (✉)
UMR7206 Eco-anthropologie et ethnobiologie, MNHN-CNRS-Université Denis Diderot, 75231 Paris Cedex 05, France
e-mail: darlu@mnhn.fr

M.P. Gutmann et al. (eds.), *Navigating Time and Space in Population Studies*, International Studies in Population 9, DOI 10.1007/978-94-007-0068-0_5,

in different geographical contexts such as Italy (Zei, Guglielmino, Siri, Moroni, & Cavalli-Sforza, 1983; Barrai et al., 1991, 1992; Rodriguez-Larralde et al., 1993), France (Darlu & Ruffié, 1992a, b) and England and Wales (Sokal, Harding, Lasker, & Mascie-Taylor, 1992). This more detailed model is based on the discrepancy between an observed surname distribution and the expected distribution under an assumption of neutrality and provides an estimate of the immigration rate.

The second strategy requires the data to be recorded at least twice at the same location. When demographic parameters such as mortality and fertility rates at each place are known, the migratory flux from one place to another during the period between two recordings can be quantitatively inferred (Wijsman, Zei, Moroni, & Cavalli-Sforza, 1984; Darlu et al., 1996). Finally, other methods are based on the Bayesian probability that new immigrants settling in a given geographic area come from a specific geographic origin (Degioanni & Darlu, 2001; Darlu & Degioanni, 2007), or methods are based on the proportions of names 'imported from abroad' and classified according to their ethnic origin (Longley, Webber, & Lloyd, 2007)

When applied to historical demography, several difficulties in handling surname data can be encountered, particularly because of their scarcity, in time and space and because of the mutability or changeability of the spelling of names when going back in time. Most of these problems have been discussed in relevant research carried out by the French School working on personal names from their medieval origin in Europe to the present (Bourin 1990; Bourin & Chareille 1992, 1995, 2002a, b; Beech, Bourin, & Chareille, 2002; Chareille, 2003).

The aim of this chapter is to take advantage of the availability of an exceptional corpus of surnames in order to exemplify the efficiency of various statistical approaches and to draw new findings about population mobility and past migration during the last three centuries, in a rural region whose scarce historical documents cannot provide quantitative conclusions.

The Data and the Space-Time Sampling Strategy

This study was carried out in Savoy, France (see Fig. 5.1), a region for which the question of migration has not been tackled in an historical perspective, even by Rousseau in his comprehensive synthesis (1960), mainly because the material information usually handled by the historian is missing (Barbero, 1979, 1980). Three different sources were available for this study. The first consists of parish birth registers maintained by Catholic clergy in this region when it was part of the Kingdom of Piedmont-Sardinia, before it was incorporated into France by plebiscite of the Savoyard people in 1860. At that time, the clergy registered baptisms rather than births, which for our purposes are essentially equivalent, even if a slight proportion of non-Catholic births were missing. Depending on the parish, the conscientiousness of the priest and the conditions of conservation, registers were more or less well preserved. Because of the large size of the region and the variable quality of the data, our strategy was to focus on a few parishes for which surname records were

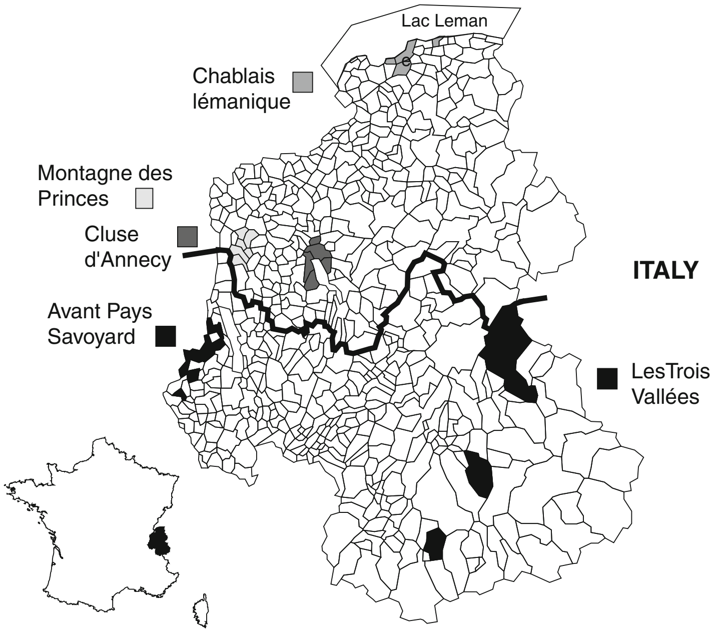


Fig. 5.1 Location of the five regions (north: Haute-Savoie; south: Savoie)

consistently well kept from the seventeenth to the nineteenth century. We made use of registers from two successive periods of 20 years: 1710–1729 (*T*1) and 1810–1829 (*T2*).

The second historical source of data was the birth register provided by INSEE (Institut National de la Statistique et des Etudes Economiques 1985). This register gives, for each of the 36,000 communes in France and for each surname, the number of infants born during two periods, 1891–1915 (*T3*) and 1916–1940 (*T4*) and still living in 1970. From this register, we extracted data covering the communes selected for the previous period.

The third source is the land registry drawn up between 1728 and 1730 at the instigation of Duke Victor-Amadeus II. This is one of the oldest land registries in Europe (Guichonnet, 1955). For each parish, two kinds of documents are available: registers and maps. One register lists each plot of land and the nature of its cultivation, which are noted on the maps. Other registers include names of landowners and have also been exploited in this study when necessary, particularly to correctly lemmatize, or standardize, the spelling of some ambiguous names.

Finally, twenty-six communes were selected for this study, taking into account several criteria. Communes must vary in terms of population size and subsistence economy and must have experienced different evolutions from the eighteenth to the nineteenth century. They also have to be located in contrasting geographic areas in terms of landscape and kind of settlement. Consequently, they show a large range of heterogeneity in time as well as in space. These communes are grouped into five regions: Chablais lémanique and Cluse d'Annecy include important cities (Thonon and Evian and Annecy, respectively) within a rural environment (at least during the first two or three periods); Montagne des Princes is a cluster of small rural communes in a hilly region; Avant Pays Savoyard is also a rural region located at the border between Savoy and France (until 1860); Trois Vallées is formed by three small cities, each deeply settled in its own valley.

Figure 5.1 maps the five regions from which communes were sampled, while Table 5.1 provides the names of these regions and communes and the number of births by year and by region, from the first to the last period, as an indicator of variation in population size over time.

Because the old documents usually present personal names with large orthographic variations, even for persons belonging to the same family line in the same or successive generations, a process of lemmatization was necessary. This must be done carefully because too extensive a lemmatization could mix up names that originated from different lines or from different areas, provoking a loss of geographic information. Conversely, too parsimonious a lemmatization could cut the transmission of a name through related families or from one generation to the next. Lemmatization is therefore a delicate and subtle exercise requiring a very good knowledge of all documents and of the history of families. As an example, if one finds seven births registered as 'EXCOFFIER' and only one as 'EXCOFIER' in a given commune, this last name will be rewritten as 'EXCOFFIER'. Conversely, if 'EXCOFIER' is the most frequent form in another commune, we keep it untransformed, considering that these two geographically unconnected variants designate two different lineages.

Table 5.1 The five regions, their communes and the number of births by year for the four successive periods (*T1*: 1710–1729; *T2*:1810–1829; *T3*:1891–1915; *T4*:1916–1940)

Region	Commune	Nb of births/year *T1,T2,T3,T4*
Montagne des Princes	Clermont, Crempigny, Lornay, Menthonnex, Val-de-Fier, Versonnex	9.0, 14.7, 4.4, 3.7
Chablais lémanique	Allinges, Anthy, Evian, Thonon	31.4, 55.2, 28.8, 64.6
Cluse d'Annecy	Annecy, Annecy-le-Vieux, Argonay, Naves, Sevrier, Veyrier	37.7, 37.6, 24.6, 64.9
Avant Pays Savoyard	Billième, Gresins, La Bame Loisieux, Lucey, Ontex, Yenne	19.3, 28.4, 5.2, 5.2
Les Trois Vallées	Bourg-Saint-Maurice, Saint-Bon-Tarentaise, Saint-Michel-de-Maurienne	39.4, 62.4, 24.5, 31.5

Methods

Distribution of Surnames

Several correlated statistical indicators can provide information about the degree of isolation of populations and about their proportion of in-migrants at a given time. The first is derived from the distribution of the number of neutral genes in a population (Karlin & McGregor, 1967). Under certain conditions, such as a stable population, with a large number of individuals compared to the number of surnames, ν, an indicator of the *rate of in-migration*, whatever the origin, is expressed in term of S and N, respectively the number of surnames and the size of the population, as follows (see Zei et al. (1983)):

$$\frac{S}{N} = \frac{\nu}{(\nu - 1)} \ln(\nu)$$

This parameter ν is highly correlated to the ratio *S/N*. Two other parameters are the proportion of hapax (a surname present once in the population), also correlated to ν, and the isonymy parameter I, which represents the probability of drawing by chance two identical surnames in the same population and is related to the kinship coefficient observed in a population (Piazza, Rendine, Zei, Moroni, & Cavalli-Sforza, 1987; Darlu & Ruffié, 1992b):

$$I = \sum_k \left(\frac{S_k}{N}\right)^2$$

the sum being over all k surnames.

This isonymy index increases with the degree of isolation of the populations and decreases when the proportion of arriving migrants and their original and specific surnames increases. Figure 5.2 shows the evolution of these parameters with time (standardized for the length of the periods: 20 years for *T1* and *T2* and 25 for *T3* and *T4*). An ANOVA with repeated measures confirms a significant evolution of the ν and the proportion of hapax with time ($p < 0.001$), with a significant increase between *T2* and *T3* ($p < 0.001$, after Bonferroni correction), except in Chablais lémanique, where the proportion of hapax seems quite stable with time. This region, which includes two cities (Thonon and Evian) with large populations and as a result large diversity of surnames, may serve as a 'buffer'. On the other hand, the isonymy index stays stable, with a slight tendency to decrease, as expected, except in Avant Pays Savoyard, which shows some non-significant increase of this parameter, suggesting a possible recent isolation or a drastic reduction of the population size, as suggested by the decreasing number of births by year (Table 5.1).

Quantification of Surname Similarity and Tree Representation

To represent the global diversity of surnames and how these surnames are structured within the space and time investigated, we have calculated the surname similarity

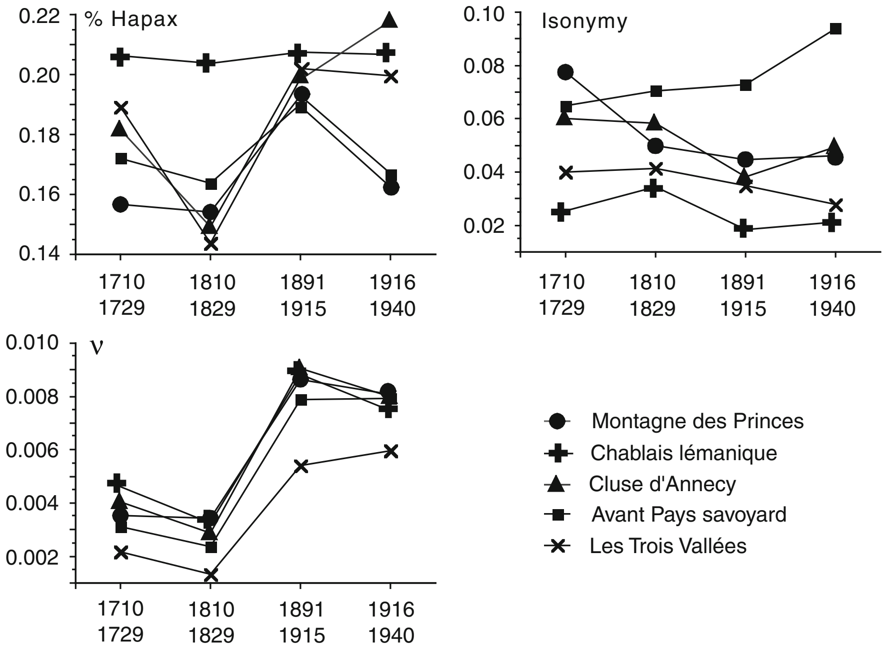


Fig. 5.2 Evolution over time of the proportion of hapax, the rate of immigration (by year) and the isonymy according to the regions

between two communes i and j by the standardized coefficient proposed by Chen and Cavalli-Sforza (1983):

$$\phi_{ij} = \frac{\sum_{k=1}^{S} N_{ik} N_{jk}}{\left(\sum_{k=1}^{S} N_{ik}^2 \sum_{k=1}^{S} N_{jk}^2 \right)^{1/2}}$$

where N_{ik} is the number of surnames k in the area i. However, as this number and the frequency of surnames could be quite biased for the oldest period, we retain only information about the presence or absence of each surname in a given area and the number of surnames shared by pairs of communes (Table 5.2). Doing so, ϕ_{ij} is equivalent to the similarity index defined by Ochiai (1957; in Sokal and Sneath (1963)). This coefficient is then transformed to obtain a distance, as suggested by Nei (1973)

$$\delta_{ij} = -\ln(\phi_{ij})$$

Once the surname distance matrix was obtained, a tree was constructed by the neighbour joining method (Saitou & Nei, 1987), with bootstrap resampling to estimate robustness at nodes of the tree.

Table 5.2 Rank correlation of surname distance matrices between the 26 communes, calculated by period (all $p < 0.001$)

	T1	*T2*	*T3*	*T4*
1710–1729 (*T1*)	1			
1810–1829 (*T2*)	0.66	1		
1891–1915 (*T3*)	0.57	0.65	1	
1916–1940 (*T4*)	0.63	0.70	0.69	1

Distance matrices can be calculated following several procedures: (i) By including all communes and each commune being present four times, one for each period (*T1* to *T4*). In this case, the four representations of the same commune are usually clustered together in the tree, with a high value for the bootstrap proportion (not shown). This means that surname distances are shorter among the same commune at different times than between two different communes at the same time or at different times. (ii) By calculating four matrices, one per period. These matrices turn out to be well correlated (between 0.57 and 0.70, depending on the periods, Table 5.3; Mantel test, Legendre and Lapointe (2004)). All these results confirm that the structures between communes are well preserved over time, each commune keeping its own surname profile with time. (iii) By summing up over time all the data leading to a single matrix (Fig. 5.2). This last figure underlines the individuality of each region, meaning that there was little population displacement between these regions during the last two centuries.

Probability of Origin of Migrants: A Bayesian Approach

Since migration of people involves migration of their surnames (or at least the surnames of their children quoted in the birth registers), the movement of people – usually males because name is paternally transmitted – can be reasonably inferred

Table 5.3 *N* is the number of surnames during 1710–1729. Diagonal: number of surnames particular to each region (not found elsewhere among the other regions); below diagonal: number of shared surnames between two regions (over all periods)

	N	Montagne des Princes	Chablais lémanique	Cluse d'Annecy	Avant Pays Savoyard	Trois Vallées
Montagne des Princes	174	*99*	28	58	21	22
Chablais lémanique	596	28	*448*	108	57	49
Cluse d'Annecy	998	58	108	*784*	89	63
Avant Pays Savoyard	446	21	57	89	*327*	40
Trois Vallées	351	22	49	63	40	*249*
All	9414					

from the movements of their surnames, although with some limitations (Darlu, 2004). The origin of migration between two areas can be estimated by looking at surnames that are absent at time *t* and present at time *t+1* in a given area. These newly arriving surnames necessarily come from somewhere else, a place where they are already present. Therefore it becomes possible to detect the geographic origin of these migrants, using the following Bayesian approach.

For the area under investigation, called the 'recipient area', the probability that the surname s_k, which is present at time *t+1* and absent at time *t*, originated from another area, a_i called the 'source area' *i*, is, according to Bayes' Theorem:

$$p\left(a_i\,|s_k\right) = \frac{\pi\ (a_i)\,p\,(s_k\,|a_i\,)}{\sum_i \pi\ (a_i)\,p\,(s_k\,|a_i\,)}$$

where $p\,(s_k\,|a_i\,)$ is the probability of observing the surname s_k within the a_ith area. It can be estimated by the observed frequency of the k_{th} surname in the a_ith area. $\pi\ (a_i)$ is the *a priori* probability of emigration from the geographic area a_i to any other areas, whatever the surname. The sum is over all geographic areas included in the analysis. Thus, all surnames are presumed to originate exclusively from these areas, except those that are not present in any studied area at time *t*. They consequently come from outside.

As this probability of origin of surnames is estimated for each surname s_k, one obtains a more accurate estimate by summing all surnames and then calculating the weighted mean probability of geographic origin, pgo_i, of any surname newly arriving between two periods in a given recipient area *i* as:

$$pgo_i = \frac{1}{\sum_k \omega_k} \sum_k \omega_k p\left(a_i\,|s_k\,\right)$$

where ω_k is a weight taking into account the fact that several persons could share the same surname. When this weight is set to 1, one obtains the probability of geographical origin of the newly arriving *surnames* and when this weight is set to the frequency of the surnames within the recipient area, at time *t+1*, one estimates the probability of the origin of *migrants* (for extensive discussion, see Degioanni and Darlu (2001)). Once these probabilities are obtained, they are used as a new estimate of the *a priori* probability $\pi\ (a_i)$ and are replaced into the Bayesian formula, which is then recalculated. This iterative process is carried on until a convergence criterion is met. The estimation of pgo_i converges whatever the *a priori* probability.

One must underline that this method is not intended for estimating migration flux, but only for specifying the geographic origin of this flux. Of course, the accuracy of each *pgo* depends on the number of surnames taken into account and is naturally weaker in small areas or communes, such as within the Montagne des Princes region, than in larger groups of communes as Cluse d'Annecy or Chablais lémanique.

This method was applied to the successive time periods to estimate the probability of migrant origins in each region (Table 5.4). For example, most of the

Table 5.4 Bayesian estimation of the probability of geographic origins (*pgo*) of the surnames/migrants newly arriving from the regions indicated in first column into a given region (column 3–6) at a given span of time (*T1*: 1710–1729; *T2*: 1810–1829: *T3:* 1891–1915*; T4:* 1916–1940). In the diagonal, the same legend applied to *pgo* between communes among each region

From	To ->	Montagne des Princes	Chablais lémanique	Cluse d'Annecy	Avant Pays Savoyard	Trois Vallées
Montagne des Princes	*T1->T2*	0.152	0.002	0.013	0.014	0.006
	T2->T3	0.196	0.037	0.037	0.008	0.003
	T3->T4	0.098	0.003	0.006	0.000	0.001
Chablais lémanique	*T1->T2*	0.092	0.279	0.070	0.038	0.026
	T2->T3	0.074	0.256	0.087	0.030	0.027
	T3->T4	0.061	0.293	0.036	0.073	0.054
Cluse d'Annecy	*T1->T2*	0.303	0.091	0.409	0.182	0.209
	T2->T3	0.250	0.098	0.320	0.089	0.044
	T3->T4	0.336	0.107	0.402	0.263	0.094
Avant Pays Savoyard	*T1->T2*	0.037	0.025	0.027	0.241	0.048
	T2->T3	0.036	0.037	0.036	0.283	0.074
	T3->T4	0.024	0.006	0.023	0.192	0.045
Trois Vallées	*T1->T2*	0.008	0.040	0.025	0.022	0.024
	T2->T3	0.032	0.028	0.014	0.019	0.070
	T3->T4	0.025	0.053	0.011	0.004	0.007
Outside	*T1->T2*	0.409	0.562	0.456	0.503	0.686
	T2->T3	0.412	0.544	0.507	0.571	0.782
	T3->T4	0.456	0.538	0.522	0.468	0.800

immigrants (40.9%) to Montagne des Princes between the first and second periods came from regions not under study and 30.3% came from the Cluse d'Annecy. During the same time period, 15.2% of surnames/persons moved from one commune to another commune but remained within the Montagne des Princes region, signifying intra-regional migration. The results are rather different when focusing on Cluse d'Annecy, which includes both small rural villages and Annecy, an important urban centre. Between the first and second periods, most of the migrants settling in this region came from outside (45.6%) and resulted from movement between communes belonging to the same area (40.9%).

Another limitation of this method comes from the fact that the different surnames used to estimate the *pgos* do not necessarily originate exclusively from the investigated area. They can also be found in other 'source areas'. However, since the most frequent surnames in a given area are usually quite particular to this area, and since they bring the largest contribution to the *pgo*, the loss of information can be assumed to be reduced. Of course, it is difficult to assess that it is entirely correct without an exhaustive corpus of names for all the periods and areas investigated. We only know that lists of surnames are almost exhaustive for the periods 1891–1940 for all French communes, but not presently for the previous periods.

Based on these exhaustive surname lists, one can observe that surnames particular to a region for period 1710–1729 still have their centre of gravity within the same area two centuries later, reinforcing our assumption.

Migrations Between Regions by Logistic Regression

By logistic regression, with surnames coded present or absent in each region as the dependent variable, we have tried to explain the presence or absence of surnames in a given area at time $t+1$ as follows:

$$\log\left(\frac{p_{x_{i,t+1}}}{1-p_{x_{i,t+1}}}\right) = b_{0,t} + \sum\nolimits_i b_{i,t}x_{i,t}$$

where the probability of presence or absence of surnames in the area i at time $t+1$, $p_{i,t+1}$, is inferred from the presence or absence of these surnames $x_{i,t}$ in all the i different areas in the previous time t. A significant coefficient attached to the area i at time t means that the profile of surnames in this area is to some extent predictive of what will be the profile of surnames of the given area at time $t+1$. Results are shown in Table 5.5. This method provides a way to estimate the relationships between areas through time. It clearly appears that the largest coefficient predicting the presence or absence of surnames in a given area at time $t+1$ is the presence or absence in the same area at time t. This finding clearly demonstrates that each region has its own stock of surnames, quite stable over time, although some areas are more stable than others. For example, the logistic coefficients are larger for Montagne des Princes (between 3.83 and 4.22 depending on the periods) than for Cluse d’Annecy (between 0.92 and 1.90), suggesting that the turnover of the population of Montagne des Princes is less than that of Cluse d’Annecy.

Discussion

The different tools proposed in this work allow us to draw up an evolution of the migration patterns occurring in this part of Savoy since the eighteenth century. The relationships between communes and regions are demonstrated to be quite stable, from the beginning of the eighteenth century until now, as stated by the various analyses, by both the tree representation and the correlations between surname distance matrices (Table 5.2). As described earlier, the tree (Fig. 5.3) clearly delineates the five regions, particularly Montagne des Princes, with its six villages showing a reasonably high value of the bootstrap coefficient (77%) and Avant Pays Savoyard with a slightly weaker bootstrap value (69%). The other bootstrap values are quite low, even though all five regions are easily distinguished. However, it does not appear that there is any convincing additional clustering between some of these five regions, leading to the conclusion that their historical relationships since the eighteenth century have been largely independent. This point is strengthened by the logistic

Table 5.5 Logistic regression coefficients inferring presence/absence of surnames at a given time t + 1 in a given region, covariates being the presence/absence of surnames observed at the previous time t in the different regions

	Montagne des Princes	Chablais lémanique	Cluse d'Annecy	Avant Pays savoyard	Trois Vallées
Independent, 1710–1729	Dependent, 1810–1829				
Montagne des Princes	4.18***	0.23*	0.883***	0.71*	0.27 ns
Chablais lémanique	0.54+	2.22***	0.28+	0.57**	0.55**
Cluse d'Annecy	0.98***	0.39***	1.90***	0.24 ns	0.367+
Avant Pays Savoyard	0.66*	0.23 ns	0.24 ns	3.31***	0.51*
Trois Vallées	0.70*	0.36+	0.29 ns	0.095 ns	2.99***
R^2	0.263	0.096	0.096	0.207	0.154
Independent, 1810–1829	Dependent, 1891–1915				
Montagne des Princes	3.84***	0.54**	1.13***	0.01 ns	0.33 ns
Chablais lémanique	0.43 ns	1.64***	–0.02 ns	0.45+	–0.16 ns
Cluse d'Annecy	0.70***	0.26+	1.31***	0.75***	–0.01 ns
Avant Pays Savoyard	0.24 ns	0.00 ns	0.27+	3.38***	0.01 ns
Trois Vallées	0.84**	0.03+	0.04 ns	0.19 ns	2.46***
R^2	0.255	0.055	0.043	0.240	0.079
Independent, 1891–1914	Dependent, 1916–1940				
Montagne des Princes	4.22***	0.56***	0.80***	0.86**	0.31 ns
Chablais lémanique	0.57**	1.36***	–045***	0.28 ns	–0.20 ns
Cluse d'Annecy	1.10***	–.37***	0.92***	0.42+	–0.19 ns
Avant Pays Savoyard	0.26 ns	0.27+	0.01 ns	4.20***	0.43+
Trois Vallées	0.54 ns	–53***	–0.60***	0.42 ns	2.19***
R^2	0.332	0.048	0.031	0.300	0.084

(***: p<0.001; **: p<0.01; *: p<0.05; +: p<0.10; ns: non significant)

regression, which shows that the best predictors of presence and absence of surnames in a given period in a given area is the presence and absence of these surnames in the previous period in the same area. The marked distinction between regions is reinforced by the fact that surname similarities and surname exchanges are stronger among communes of the same region than between communes belonging to different regions. Indeed, the intra-regional probability of geographic origin of migrants is much higher than the inter-regional probability (Table 5.3). All these results indicate

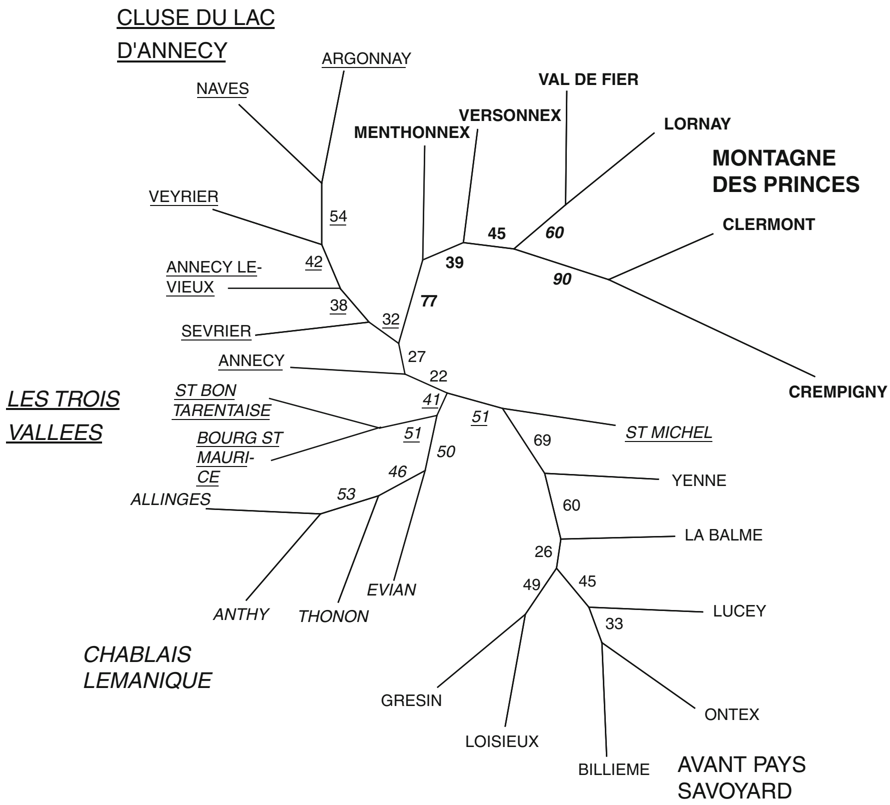


Fig. 5.3 Neighbour joining tree representation of the surname distance matrix between communes (see text). Values are the bootstrap proportions (%) estimated on 1,000 replicates (Avant Pays Savoyard: standard; Chablais lémanique: italics; Les Trois Vallées: italics and underlined, Cluse d'Annecy: standard and underlined; Montagne des Princes: bold). Branch lengths are proportional to the Nei's distance. Angles between two branches are arbitrary

that moving from one commune to another within the same region was far more frequent than migration from one region to another, although the geographic distances between these five regions are not very great.

The various analyses and results also lead us to underline some regional idiosyncrasies. For instance, Trois Vallées seems to be quite a heterogeneous region, including three communes that are not clustered together in the tree analysis. Moreover, the probability of geographical origin of in-migrants shows that exchanges of surnames (or people) between the three communes were very rare, these communes being geographically distant and separated by mountain ranges. This region received migrants preferentially from outside (between 0.69 and 0.80 of the *pgos*) and, to a lesser extent, from Cluse d'Annecy (at least between the beginning of the eighteenth and the nineteenth century (*pgos* equal to 0.21)), which is located where the valleys open out. The low value of the isonymy index, the rise of the hapax proportion with time (Fig. 5.2) and details about the origins of these hapax confirm that the renewal of this population most likely drew on migrants from neighbouring departments (Rhône and Isère), from Paris and from the north of France.

Chablais lémanique shows a quite different pattern because of its two important cities (Evian and Thonon), which were expanding throughout the nineteenth century: low isonymy, a large and stable proportion of hapax over time and a large proportion of immigration from outside (*pgos* between 0.54 and 0.56). This region preferentially welcomed migrants from outside the area and to a lesser extent, from the only region that also included a city (Annecy). Conversely, Chablais lémanique sent relatively few migrants to other regions, due to rural exodus and urban development at the end of the nineteenth century.

Cluse d'Annecy has a specific profile based on its location at the centre of the other regions, which all received migrants from it, particularly Montagne des Princes (*pgos* around 0.30), Avant Pays Savoyard and Les Trois Vallées at the beginning of the nineteenth century. This is also the region for which the presence or absence of surnames for one period is not based solely on the presence or absence of surnames in the previous period at the same place, but also on their presence in the other places. Obviously, Cluse d'Annecy played a role as hub for the movement of people in this area, because it had very active traditional factories before the French Revolution and an administrative centre after the incorporation of Savoy into France after 1860.

Montagne des Princes is a rural area with small communes that received few migrants from outside (external *pgos* between 0.41 and 0.46) compared to other regions. Here, the presence or absence of surnames at one time is significantly dependent on the presence or absence in the previous time period within the region itself. However, the number of surnames observed in this region was small, as was the size of the population, which decreased from one period to the next (Table 5.1), making results rather imprecise.

Avant Pays Savoyard is also a rural region with small towns, but with more surnames than Montagne des Princes. However, as this region is bordered by the Rhône river and located at the limit of the study area, most of the exchange could be turned towards the departments to the west (Isère and Rhône), complicating this analysis. Despite this restriction, we have already shown that Avant Pays Savoyard received surnames and migrants from Cluse d'Annecy (*pgos* between 0.09 and 0.26 depending on the period).

Finally, beyond this particular example dealing with the mobility of people in Savoy since the eighteenth century, which turns out to be limited and local, the main purpose of this paper is to show how one can compensate for the lack of accurate historical documentation by taking advantage of the rich and valuable corpus of surname data covering large periods of time. From a methodological point of view, it also emphasizes that the large diversity of methods allows significant quantitative insight into the dynamics of past populations.

Acknowledgement This work was granted by the ACI (Action Concertée Incitative #ET0063) from the French Government (Ministère délégué à l'Enseignement supérieur et à la Recherche)

References

Barbero, D. (1979). *Paroisses et communes de France. Dictionnaire d'histoire administrative et démographique: Savoie*. Paris: CNRS Edition.

Barbero, D. (1980). *Paroisses et communes de France. Dictionnaire d'histoire administrative et démographique: Haute-Savoie*. Paris: CNRS Edition.

Barrai, I., Formica, G., Scapoli, C., Beretta, M., Mamolini, E., Volinia, S., et al. (1992). Microevolution in Ferrara: Isonymy 1890–1990. *Annals of Human Biology*, *19*, 371–385.

Barrai, I., Scapoli, C., Canella, R., Formica, G., Barale, R., & Beretta, M. (1991). Isonymy in records of births and deaths in Ferrara. *Annals of human biology*, *18*, 395–404.

G. T. Beech, M. Bourin, P. Chareille, (Eds.). (2002). *Personal names studies of medieval Europe. Studies in medieval Culture XLIII*. Kalamazoo, MI: Medieval Institute Publication, Western Michigan University.

M. Bourin, (Ed.). (1990). *Genèse médiévale de l'anthroponymie moderne*. Tours: Publications de l'Université de Tours.

M. Bourin, P. Chareille, (Eds.). (1992). *Genèse médiévale de l'anthroponymie moderne. Persistance du nom unique*. Tours: Publication de l'Université de Tours.

M. Bourin, P. Chareille, (Eds.). (1995). *Genèse médiévale de l'anthroponymie moderne. Enquêtes généalogiques et données prosopographiques*. Tours: Publications de l'Université de Tours.

Bourin, M. & Chareille, P. (2002a). Colons et Serfs dans le polyptique d'Irminon: Quelles différences anthroponymiques? In M. Bourin, P. Chareille, (Eds.), *Genèse médiévale de l'anthroponymie moderne, V-1 (*pp. 31–132). Tours: Université de Tours.

M. Bourin, P. Chareille, (Eds.). (2002b). *Genèse médiévale de l'anthroponymie moderne. Intégration et exclusion sociale: Lectures anthroponymiques*. Tours: Publications de l'Université de Tours.

Chareille, P. (2003). *Le nom: Histoire et statistiques. Quelles méthodes quantitatives pour une étude de l'anthroponymie médiévale. Thèse d'Histoire*. Paris: Université de Paris I- Panthéon-Sorbonne.

Chen, K. & Cavalli-Sforza, L. L. (1983). Surnames in Taiwan: Interpretations based on geography and history. *Human Biology*, *55*, 367–374.

Cockerham, C. C. & Weir, B. S. (1993). Estimation of gene flow from F-statistics. *Evolution*, *47*, 855–863.

Crow, J. F. (1983). Discussion. *Human Biology*, *55*, 383–397.

Darlu, P. (2004). Patronymes et démographie historique. *Annales de démographie historique*, *2*, 53–65.

Darlu, P. & Degioanni, A. (2007). Localisation de l'origine géographique de migrants par la méthode patronymique: Exemple de quelques villes de France au début du XXème siècle. *Espace géographique*, *3*, 251–265.

Darlu, P., Degioanni, A., & Zei, G. (1996). Patronymes et migrations dans les populations humaines. In J. P. Bocquet-Appel, D. Courgeau, D. Pumain, (Eds.), *Spatial analysis of biodemographic data* (pp. 225–253). Paris: John Libbey/INED.

Darlu, P. & Ruffié, J. (1992a). L'immigration dans les départements français étudié par la méthode des patronymes. *Population*, *3*, 719–734.

Darlu, P. & Ruffié, J. (1992b). Relationships between consanguinity and migration rate from surname distributions and isonymy in France. *Annals of Human Biology*, *19*, 133–137.

Degioanni, A. & Darlu, P. (2001). A Bayesian approach to infer geographical origins of migrants through surnames. *Annals of Human Biology*, *28*, 537–545.

Guichonnet, P. (1955). Le cadastre savoyard de 1736 et son utilisation pour les recherches d'histoire et de géographie sociale. *Revue de Géographie Alpine*, *43*, 255–298.

INSEE. (1985). *Registre français des noms patronymiques*.

Karlin, S. & McGregor, J. (1967). The number of mutant forms maintained in a population. *Proceedings of the 5th Berkeley Symposium on Mathematics, Statistics, and Probability*, *4*, 415–438.

Lasker, G. W. (1985). *Surnames and genetic structure*. Cambridge: Cambridge University Press.
Legendre, P. & Lapointe, F. J. (2004). Assessing the congruence among distance matrices: Single malt Scotch whiskies revisited. *Australian and New Zealand Journal of Statistics*, *46*, 615–629.
Longley, P. A., Webber, R., & Lloyd, D. (2007). The quantitative analysis of family names: Historic migration and the present day neighborhood structure of Middlesbrough, United Kingdom. *Annals of the Association of American Geographers*, *97*(1), 31–48.
Nei, M. (1973). Genetic distance between populations. *American Naturalist*, *106*, 283–292.
Nicolas, J. (1978). *La Savoie au XVIII° siècle, noblesse et bourgoisie*. Paris: Maloine.
Ochiai, A. (1957). Zoogeographic studies on the soleoid fishes found in Japan and its neighbouring regions (in Japanese, summary in English). *Bulletin of the Japanese Society of Scientific Fisheries*, *22*, 526–530.
Piazza, A., Rendine, S., Zei, G., Moroni, A., & Cavalli-Sforza, L. L. (1987). Migration rates of human populations from surname distributions. *Nature*, *329*, 714–716.
Rodriguez-Larralde, A., Scapoli, C., Beretta, M., Mamolini, E., & Barrai, I. (1993). Microevolution in Perugia: Isonymy 1890–1990. *Annals of Human Biology*, *20*, 261–274.
Rousseau, R. (1960). *La population de la Savoie jusqu'en 1861*. Paris: SEVPEN Ed.
Saitou, N. & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*, 406–425.
Slatkin, M. & Barton, N. H. (1989). A comparison of three methods for estimating average levels of gene flow. *Evolution*, *43*, 1349–1368.
Sokal, R. R., Harding, R. M., Lasker, G. W., & Mascie-Taylor, C. G. (1992). A spatial analysis of 100 surnames in England and Wales. *Annals of Human Biology*, *19*, 445–476.
Sokal, R. R. & Sneath, P. H. A. (1963). *Principles of Numerical Taxonomy*. San Francisco: W.H. Freeman and Co.
Wijsman, E., Zei, G., Moroni, A., & Cavalli-Sforza, L. L. (1984). Surnames in Sardinia. II. Computation of migration matrices from surname distribution in different periods. *Annals of Human Genetics*, *48*, 65–78.
Wright, S. (1951). The genetic structure of populations. *Annals of Eugenics*, *15*, 323–354.
Zei, G., Guglielmino, C. R., Siri, E., Moroni, A., & Cavalli-Sforza, L. L. (1983). Surnames as neutral alleles: Observations in Sardinia. *Human Biology*, *55*, 357–365.

Chapter 6
Widening Horizons? The Geography of the Marriage Market in Nineteenth and Early-Twentieth Century Netherlands

Peter Ekamper, Frans van Poppel, and Kees Mandemakers

Introduction

For men and women looking for a partner the geographical distance to a potential mate is still an important consideration. One look at the personal ads in the *Acquaintance* column of one of the most important Dutch newspapers (*NRC Handelsblad*) makes clear that the horizon of the seekers might vary a lot: 'Woman, demanding, slim and more than worth to look at. (. . .) Is looking for a (. . .) man with humour (. . .) preferably at walking distance of the Concert Hall'. Another says, '50-year old Aquarius (. . .) living in the eastern part of the country and for that reason having not enough opportunity to make the acquaintance of a definitely civilized Gentleman with sense of humour'.[1] The geographical criteria in these advertisements not only express the wish to find a mate who can easily be reached, but equally strongly articulate a preference for cultural proximity. Such a preference for a spatially close and, for that reason, a more or less familiar spouse has historically found expression in numerous proverbs and sayings (Vandenbroeke, 1986: p. 24; Van der Molen, 1961). 'Court the boy next door, so you know what you get' and 'Lovers coming from far away are to be feared' are among two of the many examples. As marriage is not commonly undertaken without regular and frequent meeting of the two parties during several months of courtship, marriage indicates the absence of spatial isolation and the presence of contact between the regions where the spouses come from. Changes over time in marital distances, in the direction of marital choices and in differentials in distances between social groups thus can be used as an indicator of the degree of contact between people from different regions.

In the latter part of the nineteenth century the development of the Dutch transportation and communication infrastructure increased the opportunities to meet potential spouses from outside the region of residence or birth. Although there is no doubt that at the same time the cultural integration of the country increased as

F. van Poppel (✉)
Netherlands Interdisciplinary Demographic Institute (NIDI/KNAW), The Hague, The Netherlands
e-mail: poppel@nidi.nl

M.P. Gutmann et al. (eds.), *Navigating Time and Space in Population Studies*, International Studies in Population 9, DOI 10.1007/978-94-007-0068-0_6,

well, there is not much evidence about the consequences of this process for the population at large. The evidence that does exist is scattered and relates to restricted time periods and small regions only. In this chapter we argue that changes in the experiences and attitudes of the population brought about by changes in the transportation/communication infrastructure and in cultural terms can be studied in a unique way by analysing marital distances between spouses. This information is of a more or less continuous character and includes information from more than one million marriage certificates for five provinces of the Netherlands in the period 1812–1922. With the help of simple descriptive measures, maps and more sophisticated spatial models, we first study whether the geography of the marriage market in the Netherlands has expanded over the nineteenth century and thus testifies to an increase in cultural integration of the country. Second, we examine whether there were any class differences in marital distance and whether these differences changed over time. The results will allow us to study how cultural integration and modernity developed in different segments of society, characterized by differences in access to transportation and communication networks and by a culture that could be more or less open to the outside world. The results of our study have relevance for the study of national integration as well as for the historical study of migration, geographical isolation, modernization, family structure and ideology and the use of infrastructural networks.

We start with a brief overview of the relevance of the study of marital distances, not only from a demographic but also from a geographical, historical and sociological perspective. Next we present the mechanisms affecting the degree of spatial homogamy between spouses and describe how the operation of these mechanisms changed over the nineteenth and early twentieth centuries and how they differed by social class. We then introduce the GENLIAS database and present the methods that we use to analyse marital distances. On the basis of descriptive indicators and maps we discuss time trends in marital distances, changes in the direction of the preferences of grooms and brides and time trends in marital distances by social class. Spatial regression techniques are applied using both individual explanatory variables and regional socio-economic explanatory variables. We conclude with a brief discussion of the relevance that our outcomes have for some key issues in the history of modernization, family life and technology.

Perspectives on the Geography of the Marriage Market

Historians and social scientists have used information on the geographical proximity of spouses for a variety of purposes.[2] Geographers assume that the distribution of marital distances in a community approximates the distribution of social contacts and social knowledge arising through day-to-day life; they therefore use information on the birthplaces of spouses to study the geographical range of social and economic activities and as a measure of the geographical knowledge of people and places around the home base (the mean information field) (Coleman & Haskey, 1986). Information on geographical proximity (or what might be called geographical endogamy or homogamy) has been used by social scientists as a means to define and identify social communities; it is interpreted as an index of the social distance and contact between communities that might shed light on the heterogeneity of communities and on their survival as minorities (Coleman, 1979: p. 416; Snell, 2002).

Sociologists and historians have used time trends in the marriage horizon of communities as an index of the modernization of Western societies. In pre-modern settings, the local milieu is supposed to have been the place where social relations developed and the small spatial scale of these relationships provided for their solidity. The local community was the focus of and an important base for the ontological security of people (Giddens, 1990: pp. 100–103). Crucial to modernity was a process of 'disembedding', the lifting out of social relations from local contexts of interaction, offering the possibility of change by breaking free from the restraints of local habits and practices. The advent of modernity fostered relations between people 'locationally distant from any given situation of face-to-face interaction' (Giddens, 1990: pp. 16–20). 'Modernization' was also characterized by a shift from 'ascription' to 'achievement' as the measure of a person's status in society. This was supposed to result in a broadening of people's horizons, which in its turn led to a decline in people's preference for members of their own group. Modernization, therefore, is expected to lead to increased contact and a high frequency of intermarriage between people from different regions and to a reduction of the cultural distance between them (Van de Putte, 2003). Bonvalet and Maison (2007) have pointed to the modernization of family life as a consequence of residential choices; the independence and degree of contact between generations has been strongly affected by the practice of generations living in different villages, a process that is heavily influenced by marrying outside one's own community.

Although trends in marital distances have been studied since the 1930s, existing studies have serious limitations.[3] Most of the sociological literature on geographical proximity does not focus on long-term changes (Van Leeuwen & Maas, 2005) and although much is known about the geography of marriage in rural villages in the past for various European countries, there have been relatively few studies of long-term changes in the geography of marriage in towns. Small-scale studies have dominated the field, making it impossible to study the interplay between social position, time and geographic horizons. When data are broken down into regions, time periods and socio-economic position, some sub-groupings become quite small, introducing problems of representativeness (Pooley & Turnbull, 1998: p. 51).

In this chapter, we study the changes in the geographic horizon of the Dutch population in the nineteenth and early decades of the twentieth century. The Netherlands is an interesting case for the study of marriage horizons. It was highly urbanized and densely populated from the beginning of the nineteenth century, geographical distances between communities were rather small, there were few physical barriers relative to other parts of Europe and the Netherlands was an easy territory to traverse. We make use of a very large database that allows us to study long term trends in marriage horizons for five of the eleven Dutch provinces, covering a large part of both the area and population of the nation. The size of the database allows us to compare differentials in the spatial horizon and changes in these differentials between various social groups. The first question we want to answer is whether changes in communication and transport, in opportunities for meeting and in cultural diversity led to a geographically more scattered and more extensive choice of marriage partners and to an increase in marital distances. We will study this question by calculating distances between the birthplaces of groom and bride and by studying the specific geographical preferences for spouses from certain regions over the period 1812–1922. The second question we want to study is whether there were social class differences in geographic horizon and whether, due to improvements in the transport system, economic developments and cultural changes, these social class differences changed over time. Historical studies of the relationship between social class and geographic homogamy have been rare and usually do not focus on changes in the relationship over time (Prost, 1981; Vandenbroeke, 1986; Meurkens, 1984: pp. 34–39, p. 171; Ogden, 1980; Perry, 1969; Pooley & Turnbull, 1998). We study both of these questions by using simple descriptive measures and maps and by applying spatial regression techniques to analyse the effect of individual and regional explanatory variables on changes in the spatial location of brides and grooms.

Mechanisms Affecting Spatial Homogamy

Patterns of marital homogamy, be they educational, religious or geographical, arise from the interplay between three social forces: the preferences of the marriage candidates, the influence of the social group and the structure of interaction opportunities (Kalmijn, 1991).

Geographical homogamy primarily depends on the opportunities people have to meet and to stay in touch with future spouses from a given region. The universal tendency to choose a partner living nearby, rather than choosing geographically at random, can be understood in terms of models of human behaviour such as 'the principle of least effort': 'individuals will almost always prefer not to seek far afield to satisfy needs which can be provided close at hand with minimum inconvenience' (Coleman, 1979: p. 415). When people live near each other, they tend to meet more frequently, increasing the chances of meeting a potential partner. Proximity thus increases meeting probabilities and thereby has a direct effect on partner choice.

The level of geographical homogamy between spouses is also strongly influenced by the preferences that people have for cultural resources in future spouses: norms about marriage and the family, religious preferences, shared language and dialect, etc. As similarity of values and opinions leads to mutual confirmation of each other's behaviour and worldviews and similarity of taste increases opportunities to participate in joint activities, people prefer to marry spouses who share similar attitudes and expectations. Such a preference makes it more likely that a partner will come from the same community or at least from a region that is culturally related. The preference for a spouse from the same community is also affected by more banal factors. People usually marry someone with the same social and religious background and are more likely to meet a partner who satisfies this criterion through family parties, voluntary associations, religious organizations and the friendship network. These meeting spheres are strongly locally oriented. Geographically assortative mating is thus fostered by socially assortative meeting.

Geographical homogamy was also determined by third parties, individuals who were not directly involved in the marriage, specifically the parents of the spouses and the peer group, but who may have applied social sanctions to individuals who chose a spouse who did not meet their approval.

Opportunities to meet potential spouses from outside the individual's own region on a regular basis were rather limited in the Netherlands until the middle of the nineteenth century. Particularly in the eastern and southern parts of the country, connections between one settlement and another were relatively poor and the means of transport limited. For the majority of people, 5 km – an hour's walk for a young rural dweller – was perhaps the greatest distance that the average man was prepared to walk and return at fairly frequent intervals. Those who had to travel daily found it difficult to spend more than an hour on such a trip, given the long working day. On foot in accessible terrain one could perhaps occasionally walk 20 km to another village and walk the same distance back on the same day and villages that were this far apart did not present serious obstacles to a courtship. Problems could arise, however, if the partner lived further away. Assuming that most people walked or used a means of transport that had the same velocity – a regular or market ferry, a dogcart or horse and cart – one could cover 20 km at the most to meet a potential spouse and return home the same day. Staying in touch with a partner living further away required spending the night in the partner's neighbourhood or carrying on a frequent correspondence. The 20 km radius not only had practical but also symbolic meaning. Communities further away from one's own place of residence did not stir feelings of familiarity and intimacy: when one traveled a distance between 20 and 30 km, it normally meant that one left the belt of more or less familiar communities and entered an area with which one was less acquainted (Rosental, 2004; Renard, 1984).

Cultural norms of endogamy were another important mechanism in partner selection in nineteenth century society. Personal identity was strongly based on membership in the local community. One distinguished oneself from people from other regions who had competing and sometimes conflicting beliefs, values and aspirations and who were not likely to be seen as 'like us'. People coming from the same known universe could be trusted and spouses were favoured from regions where the preferred cultural characteristics were dominant. Cultural distinctiveness thus stimulated geographical endogamy.

Third parties played a role in maintaining the preference for culturally known partners with the desired identical characteristics. There was often a strong opposition to men and women from outside the community who were looking for a partner. Women were a 'local perquisite', so their courtship by outsiders often met with suspicion or violence (Snell, 2002). Village youngsters sometimes protected the stock of unmarried women by beating up outsiders, by parading people who were courting outside their village sitting back to front on a donkey, or by depositing a trail of rotting vegetables or manure around the house of the deviant lover. Because the visibility of outsiders was relatively strong – due to their different dress, dialect, customs and habits – this opposition could easily be expressed (Van den Berg, 1941: pp. 72–73). This enforced geographical endogamy was more than an expression of local patriotism: it also served the interest of the youngsters as inheritances remained within the village and the number of nubile girls was safeguarded (De Jager, 1981; Kuipers, 1976: p. 75).

The tendency to choose a partner living nearby rather than geographically at random, all other things being equal, has the effect that distance leads to less frequent interaction. Yet, there may be factors at work that lead to a preference for partners from areas that are more distant from each other, leading to a deviation in the direction and volume of contacts between people from different regions (directional bias or prevalence of greater contact in certain directions (Morrill & Pitts, 1967)). Areas that are close to one another on a map might be very distant from each other in terms of access, leading to a deviation in the direction and volume of contacts between people from different regions. This could be due to physical barriers, such as rivers or swamps, but also to mental barriers (traditional antipathy between villages, differences in religion or ethnicity, in dialect or language, the presence of a state border), dividing the population into closed groups and hampering circulation (Segalen, 1979; Johnston & Perry, 1972: p. 23). For this reason we also explicitly focus on the regional origin of the spouses.

Changing Opportunities and Preferences

In the past 150 years, The Netherlands experienced a series of changes that fundamentally transformed cultural preferences and the opportunities to meet potential spouses. Many of the factors that until the middle of the nineteenth century had a restraining influence on the choice of a partner from outside one's own community have, in this process, lost their effect.

First of all, the Netherlands became culturally more and more unified. Distinctive local accents, phrases and vocabulary decreased in importance, enhancing the degree to which members of the national community could communicate with each other. The distribution of goods across the country implied that people in a given region could eat the same foods and wear the same clothes as those in other areas. Local costume, customs and habits disappeared. As a result, the visibility of the fact that a person came from outside the community decreased. National newspapers and political and economic integration caused an ever-growing connectedness, emotionally as well as cognitively, between different regions and their inhabitants, stimulating cultural homogenization. Identification with the national state created a sense of membership in a national community, thereby decreasing the role of regional identities (Knippenberg & de Pater, 1988: pp. 90–91, pp. 178–179, p. 202, p. 206). As a consequence, persons from other regions were no longer perceived as strangers and began to be seen as potential partners.

The process of cultural integration was strongly intertwined with two other processes that enormously increased the opportunities to meet potential spouses from outside the region: a process of market integration, state formation and nation building (Watkins, 1991: pp. 118–138) and a radical transformation of the communication and transport systems. The three processes ran parallel to each other and are distinguished here only for analytical reasons.

The expansion of the modern economy broke down local economic boundaries and incorporated labor and capital into larger exchange networks. National labour market integration increased the possibility of direct interaction involving people from different communities. General conscription led to regular displacement of individuals and to contacts with inhabitants from a variety of regions, with marriage as one result (De Vos, 1984). Educational expansion at the secondary and university levels beginning in the last quarter of the nineteenth century brought with it social intercourse between youngsters from a wide variety of areas. The growth of the national community was not confined to the economy and the state. Local sporting clubs, political parties and labour unions became part of tightly organized national networks. The general amelioration of the conditions of the working class – the increase in wages and the reduction in working hours – made it possible for members of this class to extend their spheres of contact. A new form of mass mobility – tourism – developed and introduced people (especially from the middle classes) to areas they hitherto had never been able to visit. The geography of the country became part of the curriculum of secondary schools, leading to an increased knowledge of all parts of the Netherlands (Van der Woud, 2006: p. 135). Old social institutions of the village, such as balls and fairs, which had traditionally played a vital role as places where local residents might meet their future spouses, lost their importance (Ogden, 1980).

At the same time, there was a major improvement in transportation, facilitating mobility, and in methods of communication, which made it easier to keep in touch (telegraph, telephone and postal services). The 1850s witnessed the start of this development in the Netherlands (Van der Woud, 2006). The increase in the means and speed of transportation brought about by new and improved roads and canals and by new means of transport such as the train, the bicycle and the tram brought a wider range of potential spouses within reach. These new forms of transport increased the distance one could travel during the same day and thus expanded the geographical marriage horizon. Equally important, increased income gave larger numbers of people access to these faster means of travel (Pooley & Turnbull, 1998).

The main element in the Dutch transportation system from the middle third of the seventeenth century until the 1830s was the *trekschuit*, a system for the cheap movement of a large volume of passengers in barges, drawn by horses along specially constructed waterways over an intensive, interconnected network of routes. The system mainly served the low-lying part of the country: Holland, Groningen and Friesland (De Vries, 1978). In the 1820s and 1830s the *trekschuit* service lost ground to coaches that used the expanding network of paved roads. Although between 1810 and 1848 construction of paved roads had started, the network of paved thoroughfares remained very meager until the 1850s, outside the province of Holland. It was only around 1880 that the paving of the main roads was completed (van der Woud, 2006: p. 367). Such roads could be easily used in all seasons and all weather conditions and linked urban centres within the Netherlands to one another and to main towns in neighbouring countries (Horsten, 2005).[4] From 1850, an intricate structure of local roads that could be used almost every day became the essential link between the citizen's home, the station and the workplace. For a large part of the country, this new system signified 'a revolution in daily life': 'it offered the possibility to give up the isolation, and the desolation and to connect to the new, modern world' (Van der Woud, 2006: p. 368).

From the middle of the nineteenth century, coaches and *trekschuiten* started to lose their importance in Holland and Utrecht because of the expansion of the railways. Distances travelled by train increased markedly from the 1860s. At the end of the 1870s, the main railroad system was more or less complete. After 1880, the national railroad system was supplemented by a system of local railways and a very dense system of steam tramways (Jonckers Nieboer, 1938; Veenendaal, 1993). The railway and tram system made the largest contribution to the reduction in travel time over the period of our study. In 1850, the geographical centre of the country could be reached by people living in Groningen, most parts of Limburg and large parts of Zeeland in around 12.5 hours. In 1870, travel time to the centre had been reduced to, at most, 7.5–10 hours from the more isolated parts of the provinces and in 1920 people from almost all parts of the country were able to reach the centre of the country within 2.5–5 hours (Thurkow et al., 1984; see also Schot, 2002).

The automobile appeared on Dutch roads in the late 1890s. It was initially considered primarily a sport and recreational toy for the well-to-do and did not have much effect on transport improvement until the 1920s (Mom, Schot, & Staal, 2002). Much more important was the introduction and growing popularity of the bicycle. The first bicycles appeared in the Netherlands in 1867.[5] Cyclists in this period constituted a small elite group who had the time and money to buy bicycles, which were mainly used for leisure purposes. However, from the 1890s, the bicycle became more than a toy for the well-to-do. Considerably cheaper production of bicycles in Germany, England and the Netherlands after 1900 and increased wages brought the bicycle within reach of more people. Tax data indicate that from 1895 onwards, the demand for bicycles increased significantly. After 1920, the bicycle became the most popular mode of transport of the Dutch.

Means of communication, allowing a relationship to be maintained by correspondence when distance made frequent meeting difficult, improved as well (Van der Woud, 2006: pp. 43–45). The increased ability to write letters, the reduced costs of sending letters and improved distribution caused an increase in the number of letters and postcards written per head of the population. A national telegraph system developed in the early 1850s and by about 1855 all important towns were connected by telegraph. From the end of the 1860s, the number of telegrams received and sent increased exponentially (De Wit, 1993). The telephone system developed from the 1880s, but the number of subscribers was very restricted before the end of the nineteenth century. In 1895, access to an intercity telephone network in the eastern, northern and southern provinces was still almost non-existent (De Wit, 2002).

To summarize, between 1850 and 1880 the basis of the modern transport system was laid. A detailed network of national, regional and local paved roads was completed, the local tramway system was established and started to open up the countryside and more and more communities became connected by the primary railway system through a growing number of secondary railway lines. In addition, communication via the telegraph system and the postal service became faster, cheaper and much easier.

Before analysing whether and how people in different parts of the Netherlands, each characterized by its own transport and communication network, involvement in the national economy and regional identity, adapted the width and direction of their marriage horizons, we formulate some hypotheses on the differences that we expect between social classes in their reactions to the changed opportunities.

Social Class, Opportunities and Preferences and Their Effect on Marital Distances

There are strong indications that opportunities to meet potential spouses from more distant areas and preferences for spouses from nearby regions differed by social class. Marriage among people who earned their living in agriculture was geographically much more tightly circumscribed than in the rest of society. People working 'in industrial, commercial or service occupations had greater opportunities and need to travel both locally and further afield and, except on relatively rare social occasions or trips to the local market, the peasant tended to be rooted to the soil' (Ogden, 1980: p. 174). Given the strong tendency to marry within a particular occupational or social group, the opportunity to do so was naturally restricted by the small size of non-agricultural groups in the nineteenth-century village, which might have led to the search for a partner over longer distances. Agricultural labourers were also constrained by low wages and the operation of local and regional labour markets. Within agriculture, differences could be expected in the degree to which partners from outside the area were sought. Dairying tied its labour force 'to the cow's tail' (Perry, 1969), seven days a week and every day of the year, but the labour force in arable agriculture operated a 6-day week with a shorter working day in winter. These workers could meet more outsiders and strangers at their work and have more time to themselves outside their hours of employment than the dairy worker (Perry, 1969).

The upper and middle classes were able to select their marriage partners from a much wider area than the working class. They were more mobile as they possessed the time and money needed to travel far and often, and had more knowledge of further-off areas, giving them an advantage in getting in touch with areas further from their birthplaces. Changing transport technology may have caused a gradual reduction in average travel times in the second half of the nineteenth century, but travel remained expensive; for most ordinary working people travel by rail or steam tram remained for a long time a major and only occasional expense. The upper classes also had a more universalistic value orientation and their lives were more organized by structures that facilitated large zones of contact. They disposed of wider means of communication (including letter-writing) and participated in a geographically more extensive political, economic and friendship network. Many of them experienced a longer educational and training period, partly spent outside their own region, which offered them the chance to meet partners from other regions. For part of the group – army officers, the professions and higher civil servants – the effect of national labour markets stimulated mobility (Pooley & Turnbull, 1998: p. 69; Van de Putte, 2003). Furthermore, for the elite the desire to preserve the family's property and social status was an overriding issue, making it necessary to broaden marriage horizons when potential partners with comparable economic and cultural qualities were scarce in their own communities.

We expect the marital horizon of the skilled and unskilled manual workers to be wider than that of the agrarian workforce. Yet, the fact that they earned low wages, operated on local and regional labour markets and had long working hours and low levels of education made them less apt to participate in modern communication networks. These characteristics also made it less necessary for them to spend part of their training outside their region of upbringing and might have led to a regionally more restricted choice of spouse. Another important factor, which could lead to marriages over characteristically shorter average distances, is that this social group met their spouses in circumstances that differed from those of the upper and middle classes. Casual encounters in the street or in local public houses were more typical of working-class courtship and were more likely to be affected by neighbourhood knowledge, the maturing of childhood friendships and the bringing together of couples who were or had been close neighbours (Coleman, 1979: pp. 418–419; Perry, 1969). The mobility of the upper and middle classes was shared to some extent by those who were 'in service', such as footmen, coachmen, lady's maids and many more (Perry, 1969; Pooley & Turnbull, 1998: p. 69). For those in the armed services, movement around the country as one was transferred from camp to camp could have produced more extensive marriage horizons.

The labour market for a large part of the middle classes was not a local one, but had at least a regional character. The income of this group made it possible for them to make use of the newly developed transport and communication networks much earlier than could people from the working class. We therefore expect this group to have had a much wider horizon than the working class and the group working in agriculture.

Many of the factors that caused differences between social classes in the width of their geographic horizons changed in the period that we selected to study. Rising real incomes and technological developments brought personal travel within reach of people from all social classes. Other forms of communication enabled information to reach all social groups. Access to secondary and tertiary education where participants might begin to look for a partner generally increased, encouraging some people to attend schools far from their birthplaces. Labour markets became national for almost all occupations. The mechanization of agriculture reduced the degree to which agricultural labourers and farmers were restricted in their movements and the greater emphasis on production for the market increased the need to get in touch with the outside world.

To summarize, we will test the following hypotheses:

- What time trends are visible in the average marital distance between spouses and in the width of the zones in which they found each other and how did these trends vary between provinces?
- What deviations in the direction and volume of contacts between people from different provinces can be observed and did this directional bias change over time?
- What effect did the position in the social class structure (work in agriculture versus in other sectors, upper and middle classes versus lower social classes) have on marital distances and was there a change in that effect over time?

Data

The data that we use include all marriages contracted in the period 1812–1922 in five (of the eleven) provinces of the Netherlands. Data from marriage certificates have been entered into a database by dozens of volunteers and staff within the framework of the GENLIAS project. GENLIAS is a joint initiative of the National Archive Services and the Regional Historical Centers and officially started in 2004. The aim of this project is to build up a database containing genealogical information on all marriages, deaths and births that took place in the Netherlands from the introduction of the vital registration system until the date these data are no longer in the public domain. Marriage records enter the public domain only after 75 years. At the time of writing, the data set contained 36% of all marriages contracted in the period 1812–1922 in the Netherlands as a whole. Until the 1860s, almost 40% of Dutch marriages are covered, but after 1860 that percentage decreases gradually to around 30% in the 1920s.

The quality of the data in the marriage certificates was for a large part ensured by the rules governing marriage and marriage registration as laid down in the Civil Code. Marriages had to be contracted in the *de jure* place of residence of one of the partners. The birthplace of both parties can be considered as reliably registered as bride and groom had to present copies of their baptism or birth certificate to the vital registration officer.

The entry of the records in the GENLIAS database has been done in the participating archives by hundreds of volunteers. It stands to reason that a process of data entry on this scale cannot be completed faultlessly. Many marriage certificates are difficult to read and typing errors are common. Errors in data entry are frequently reported to the participating archives by the hundreds of thousands of people using the database and are corrected on a regular basis. One might expect that the volunteers were more familiar with names of places within the region and that as a consequence places further away will be more frequently misspelled, which increases the chance that their geographic location will not be specified exactly.

As a rule, marriages were contracted in the place of residence of the bride. A larger proportion of the brides thus lived in the municipality in which the marriage took place and a larger percentage had been born there than was the case for grooms. The distance between the place of marriage and the place of residence of the groom, therefore, generally tended to be greater than the distance between place of marriage and the place of residence of the bride. Information on place of residence of bride and groom was unfortunately not included in the database. Birthplaces of bride and groom and place of marriage are our crucial geographical indicators. Using the birthplace of the groom or bride (instead of their place of residence) generally suggests a wider marriage field (Leboutte & Hélin, 1986: p. 438; Blanchet & Kessler, 1992: pp. 346–353).

A problem in using the distance between birthplaces of grooms and brides to measure geographic horizons is that the information principally concerns people aged between 20 and 40 years; a move, then, might have taken place at any age between birth and the age at marriage and the exact time of the move is not known.

The five selected provinces in our sample each have their own particular ecological, social and economic structure and include larger and smaller cities as well as rural areas.

Gelderland is located in the central eastern part of the country, extending from the German border westwards to the Zuyder Zee. The south-western section was a long, narrow westward extension along the Rhine River with brickyards and dairy farming. Small regional market places and several larger towns such as Arnhem and Nijmegen hosted industrial activities and administrative services. Farms in Gelderland were relatively small, the infrastructure not well-developed.

Groningen, situated in the extreme northeast of the Netherlands, can be roughly divided into two regions: a northern area of clay soils and a southern one of sand and peat. The peat districts became an area of important industrial development in the second half of the nineteenth century. A common feature of the agriculture of both areas was the high degree of commercialization.

The province of Limburg is in the southeast of the Netherlands and adjoins the Belgian province of the same name. The population always had close ties with the German and Belgian cities close to the borders. The inhabitants of each municipality spoke their own, distinct dialects: more Germanic towards the German border, with a more French tone in the Meuse valley. From the end of the nineteenth century, coal mining became an important economic activity in Limburg. The capital city of Maastricht was the first place where large-scale industries developed in the Netherlands.

Overijssel is located in the central-eastern part of the country. In the western part, Overijssel is low-lying and covered with fertile pastures. Cattle-rearing and butter- and cheese-making were important occupations. Cotton-spinning, together with bleaching works and machine manufacturing, became very prominent in the Twente district in the east of the province beginning in the 1860s.

Zeeland forms the south-western part of the coastal zone and consists of a strip of the Flanders mainland, bordering Belgium and six former islands, all of them now connected to each other or to the inland provinces by dams and bridges. Zeeland was for a long time a rural area of which the towns of Middelburg and Vlissingen were the administrative and industrial centres.

Together, the five regions cover a large part of the economic and cultural landscape of nineteenth-century Netherlands. Unfortunately, we do not have information on the economic heartland, the (urban) part of the provinces of Zuid- and Noord-Holland (Van Zanden & van Riel, 2004: pp. 53–64). Figure 6.1 gives an overview of the location of the five selected provinces.

In the analysis, we make use of information on age at marriage, year of marriage, marital status at time of marriage, birthplaces of bride and groom, the place in which

Fig. 6.1 Map of the Netherlands by province around 1920

the marriage took place and the occupation of the husband. The occupation of the husband is used to classify marriages by social class. The social class categorization that we applied is based on a recently developed coding scheme called HISCO (Historical International Standard Classification of Occupations) (Van Leeuwen, Maas, & Miles, 2002). HISCO translates occupational descriptions into a common code, compatible with the International Labor Organization's International Standard Classification of Occupations (ISCO68) scheme. These historical occupational titles were classified into a social class scheme recently proposed by Van Leeuwen and Maas (2005).

Van Leeuwen and Maas called their classification scheme HISCLASS. Twelve classes are distinguished. In view of the fact that some of the categories included relatively few cases, we merged categories and adopted the following classification in our analyses: upper class (higher managers and higher professionals), middle class (lower managers, lower professionals and clerical and sales personnel, lower clerical and sales personnel and foremen), skilled workers, farmers, lower skilled workers, unskilled workers and farm workers.

Marriages were also classified according to the rural or urban character of the place in which the marriage was contracted, based on the number of inhabitants, the population density, the percentage of the population working in agriculture in 1889 (approximately the mid-point of the period studied) and the historical designation of a municipality as either 'town' or 'village'.

Table 6.1 presents descriptive statistics for the relevant variables in the database. The total number of marriages in the database was 1,080,700, of which there were 344,300 in Gelderland, 189,000 in Groningen, 176,000 in Limburg, 213,700 in Overijssel and 157,700 in Zeeland. About 1,005,500 of all marriages took place in the period 1812–1922. Around 65% of all marriages were contracted by men aged between 20 and 30 years; for women this was around 73%. People living in a rural area at the time of marriage predominated. The distribution of men by social groups shows that a large majority of the grooms belonged to the labouring classes: semi-skilled and unskilled workers in agriculture constituted 23% of all grooms, workers outside agriculture 26% and skilled workers another 16%. Around 18% of the grooms were working as farmers. Members of the upper and middle classes made up less than 13% of the total.

Table 6.1 Descriptive statistics for control and dependent variables (Source: GENLIAS and ISIS database)

	Number of marriages	Percentage
Province		
Gelderland	344,286	31.9
Groningen	189,032	17.5
Limburg	176,048	16.3
Overijssel	213,693	19.8
Zeeland	157,655	14.6
Total	1,080,714	100.0

Table 6.1 (continued)

	Number of marriages	Percentage
Period of marriage		
1812–1819	43,659	4.3
1820–1829	58,773	5.8
1830–1839	63,407	6.3
1840–1849	68,238	6.8
1850–1859	78,680	7.8
1860–1869	86,644	8.6
1870–1879	92,405	9.2
1880–1889	90,807	9.0
1890–1899	106,536	10.6
1900–1909	123,552	12.3
1910–1922	192,847	19.2
Total	1,005,548	100.0
Age at marriage of men		
14–19	10,324	1.1
20–24	245,996	26.8
25–29	352,539	38.4
30–34	175,542	19.1
35–39	71,975	7.8
40–44	32,030	3.5
45–49	15,825	1.7
50–54	8,212	0.9
55–59	4,473	0.5
60–64	2,193	0.2
Total	919,109	100.0
Age at marriage of women		
14–19	58,740	6.4
20–24	374,340	40.7
25–29	296,640	32.3
30–34	115,887	12.6
35–39	42,120	4.6
40–44	17,746	1.9
45–49	8,486	0.9
50–54	3,639	0.4
55–59	1,460	0.2
60–64	616	0.1
Total	919,674	100.0
Place of marriage		
Urban	293,617	29.2
Rural	712,290	70.8
Total	1,005,907	100.0
Social class groom		
Higher managers	10,659	1.1
Higher professionals	8,544	0.9
Lower managers	17,617	1.8
Lower professional and clerical, sales	61,744	6.3
Lower clerical and sales	25,925	2.7

Table 6.1 (continued)

	Number of marriages	Percentage
Foremen	1,885	0.2
Skilled workers	161,330	16.5
Farmers	179,373	18.3
Lower skilled workers	111,886	11.4
Lower skilled farm workers	11,352	1.2
Unskilled workers	147,214	15.1
Unskilled farm workers	217,739	22.3
Not given	17,866	1.8
Explicit without	4,116	0.4
Unclear	549	0.1
Total	977,799	100.0

Methods

To study the distances between birthplaces of bride and groom and place of marriage, we first had to determine the exact name and location of the birthplaces of bride and groom as given in the marriage certificate. In many cases this proved difficult, usually due to errors and omissions committed during the process of data entry. In particular the names of the birthplaces of persons born in Belgium or Germany were not always familiar to those doing the data entry.[6] Determining the location of these places was also hampered by the fact that the vital registration officers themselves applied different spelling rules. Furthermore, municipalities with the same name could be found in more than one province and the vital registration officers did not always consider it necessary to mention the name of the province in the marriage certificate. Municipalities were also sometimes difficult to identify, particularly in countries outside the Netherlands.

The next step in the research involved the identification of the exact geographic location of the birthplace. This was done by assigning geographic coordinates according to the Netherlands National Coordinate System (*Rijksdriehoeksmeting*), the Netherlands' geodetic or surveying reference system that constitutes the standard for the determination of locations. This system was developed between 1885 and 1904. The reference frame was based on traditional triangulation of circa 5,500 assigned points in the country, using monuments such as church towers and other triangulation points, together with specifications of those points by their X- and Y-coordinates (De Bruijne, Van Buren, Kösters, & Van der Marel, 2005). To determine the relative location of an administrative unit (a municipality or a known part of such an administrative unit) in the reference system, use was made, in rank order, of a series of databases with geographic coordinates: the official Netherlands National Coordinate System points (*Rijksdriehoeksmetingpunten* or RD-points) of the Dutch Land Registry Office (*Topografische Dienst Kadaster*), the Kloeke-tabel of the Meertens Institute, the GEOnet Names Server (GNS) of the National

Geospatial-Intelligence Agency (NGA) and the Topographic Names Register of the Dutch Land Registry Office (*Topografisch Namenregister Kadaster*). For Belgian and German municipalities, coordinates were determined on the basis of the official Netherlands National Coordinate System points for municipalities near the Dutch frontier and the GEOnet Names Server (see Annex). The GEOnet Names Server data contain approximate longitude/latitude coordinates, which were first transformed into the Dutch coordinates system. All geographic names of the birthplaces of bride and groom have been coded according to the situation that applied in the year in which the marriage took place.

Finally, the distance between the municipality in which the marriage was contracted and the municipality of birth (or a known part of this administrative unit) of bride and groom was calculated as the distance in kilometres in a straight line. It was impossible to collect systematic information on actual journey times involved in particular contacts, so linear distances have to suffice as a reasonable surrogate. We realize that for most people the absolute distance over which movement took place was not as important as the travel time involved. Contacts were influenced as much by perceived distances as they were by actual distances. Perceptions of distance were affected by factors such as mode of transport, costs of transport, levels of information and social, cultural and physical differences. Distances were also calculated for administrative units outside the Netherlands, but only for those brides or grooms who were born in a neighbouring part of Germany (one of the neighbouring *Länder*) or Belgium.[7] For brides and grooms who were born in the place in which the marriage was contracted the distance was fixed at 0 km, though we recognize that this method may underestimate the distance between birthplaces and residence in municipalities whose surface area increased over time. We calculated average distances travelled and also grouped distances in categories in order to study separately changes in long- and short-distance migration.

To study whether there was a preference for partners from areas that are more distant from each other, leading to a deviation in the direction and volume of contacts between people from different regions (directional bias), we mapped the number of grooms who married in a given province according to their birthplace for four periods: 1812–1829, 1830–1859, 1860–1889 and 1890–1922. The number of grooms and brides coming from a specific place is partly a function of the size of the population in that place. By definition only a limited number of grooms could be selected from a municipality with a small population. Various authors have suggested calculating the ratio between the number of grooms and the population of the region from which the grooms originated (Leboutte & Hélin, 1986: p. 432; Bonneuil, 1992: p. 112). As we had no information available on the population size of the Belgian and German municipalities, we decided to refrain from this procedure. In constructing the maps, the municipal boundaries of 1922 have been used, with data for discontinued municipalities added to those of the municipalities of which they were part in 1922.

Since our procedure results in as many as twenty different maps (five provinces times four periods), we summarized the maps for readability by visualizing spatial statistical descriptors of the original mapped data, using the weighted mean centre and the standard deviation ellipse (see Ebdon, 1988). The weighted mean centre (or centre of gravity) is calculated as the mean of the X and Y coordinates of the birthplaces of the grooms weighted by the number of grooms from these birthplaces. The standard deviation ellipse calculates the standard deviations of the X coordinates and the Y coordinates from the mean centre of the birthplaces of the grooms (all weighted by the number of grooms) in order to define the axes of the ellipse to see both the dispersion and the orientation of the spatial distribution. Both measures were calculated with CrimeStat, a spatial statistics program (Levine, 2007).

In a third stage of the research we applied spatial regression techniques (see for instance, Anselin & Bera, 1998). Spatial regression techniques are able to deal with spatial dependency in regression analysis and allow us to include individual explanatory variables such as age and social class, (regional) socio-economic explanatory variables and spatial location (according to the Dutch geographical coordinate system). In the analysis, socio-economic variables at various regional levels were used, such as urban/rural, religious composition and transport availability. The *GeoDa* software package (Anselin, 2004, 2005) contains the functionality for spatial regression modelling. *GeoDa* includes standard ordinary least squares regression with the basic diagnostics for spatial autocorrelation, heteroskedasticity and non-normality implemented; estimation of spatial lag and spatial error models is supported by means of the maximum likelihood method (Anselin, Syabri, & Kho, 2004). *GeoDa* regression routines can be applied to large data sets, such as the ones at hand in our chapter.

The Widening of the Horizon Over Time

Figures 6.2 and 6.3 present geometric mean distances between the birthplace of the groom and the birthplace of the bride and between the birthplace of the groom and the place where a couple married, by period of marriage and province of marriage.

In all provinces, there is a clear trend visible in the distance between the birthplace of the groom and that of the bride. Setting aside the deviations from the trend, it is evident that, from the period 1880–1889, the horizon of young men and women started to widen. This trend gained momentum in Overijssel (and, although to a much lesser degree, in Groningen) after 1890–1899 and in Limburg and Zeeland after 1900–1909. Grooms came from regions further and further away and the average distance increased from around 15–20 km during the first three quarters of the nineteenth century to around 25–30 km in the second decade of the twentieth century. There are clear differences visible in the horizon of people who contracted a marriage in the various provinces. Until the last decades of the nineteenth century the horizon of people in Gelderland and up to 1910, the horizon of grooms and brides in Limburg and Zeeland, stretched less far than those of people in Overijssel and Groningen; after the first decades of the twentieth century Limburg grooms found their brides further away than grooms in any other province. Grooms in

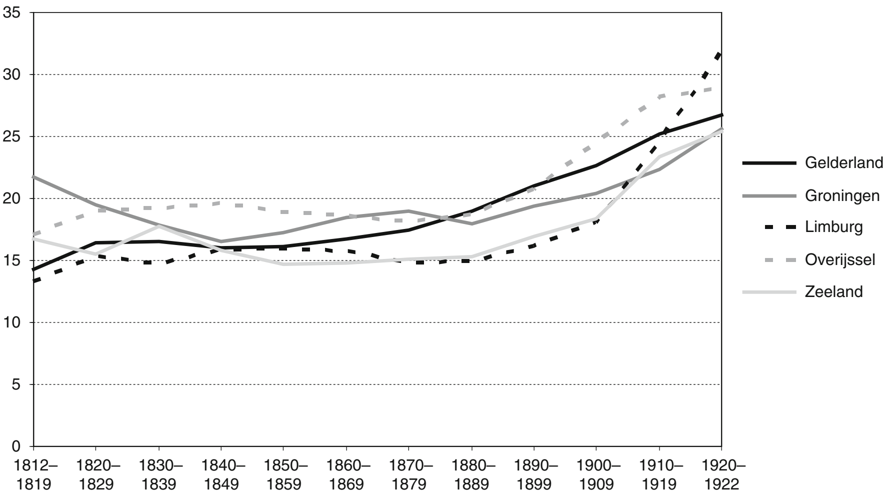


Fig. 6.2 Mean average distance (in km) between place of birth of bride and groom, by province and period of marriage

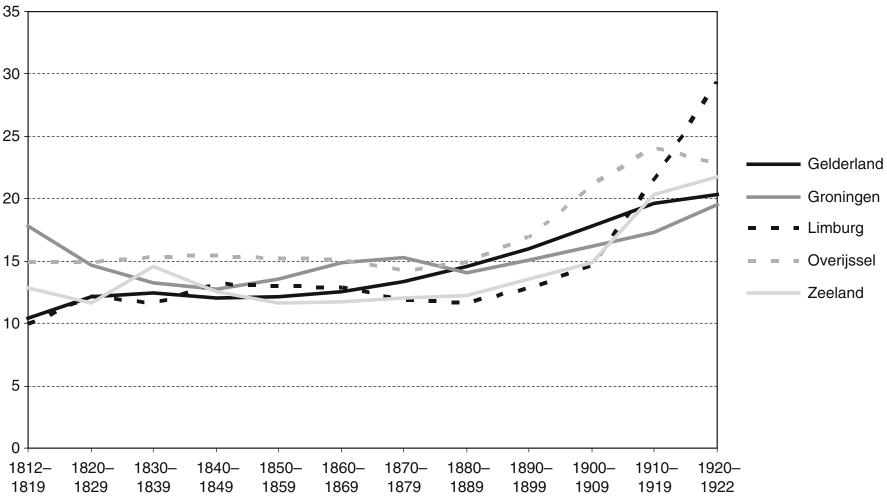


Fig. 6.3 Mean average distance (in km) between place of birth of groom and place of marriage, by province and period of marriage

Groningen now found their brides as far away as those in Gelderland and Zeeland. Differences in mean distances between provinces were around 5 km until the 1890s. In the second decade of the twentieth century, the horizon of grooms in Limburg had widened by some 15 km, that of grooms in Gelderland, Overijssel and Zeeland by some 10 km and that of grooms in Groningen only by 5 km.

Figure 6.3 shows the average distances between the birthplace of the groom and the place where the marriage was celebrated (usually the place of residence and very often the birthplace of the bride). Although distances are a little bit lower, the trends and the regional differences remain almost the same. For that reason, in the following we only make use of information on birthplaces of bride and groom.

In a second step in the analysis of the data, all distances have been grouped together in zones. The first covers contacts with neighbouring villages (walking distance, up to about 5 km, including traversing the community in which the individual lived and involving travel time of up to 2 hours back and forth). A second zone identifies contacts with communities directly contiguous to the first zone (a radius of 5–20 km), the supposed maximum walking (and later on, cycling) distance, enabling regular face-to-face contact to be maintained. The third zone includes interaction up to about 40 km and the fourth, contacts beyond 40 km. The chosen scales of interaction have already been used by other authors (Morel, 1972: p. 63; Millard, 1982). We only present here information on the two most extreme categories: marriages between brides and grooms whose birthplaces were no more than 5 km apart, and marriages in which more than 40 km separated bride and groom at the time of birth.

Figures 6.4 and 6.5 confirm the trends in marriage horizons suggested in Figures 6.2 and 6.3. There were strong differences between provinces in the degree of interaction at the local scale. Before 1900 in Limburg more than half of all marriages were contracted between partners who were born within a distance of less than 5 km; in Groningen, only 40% of marriages involved partners less than 5 km

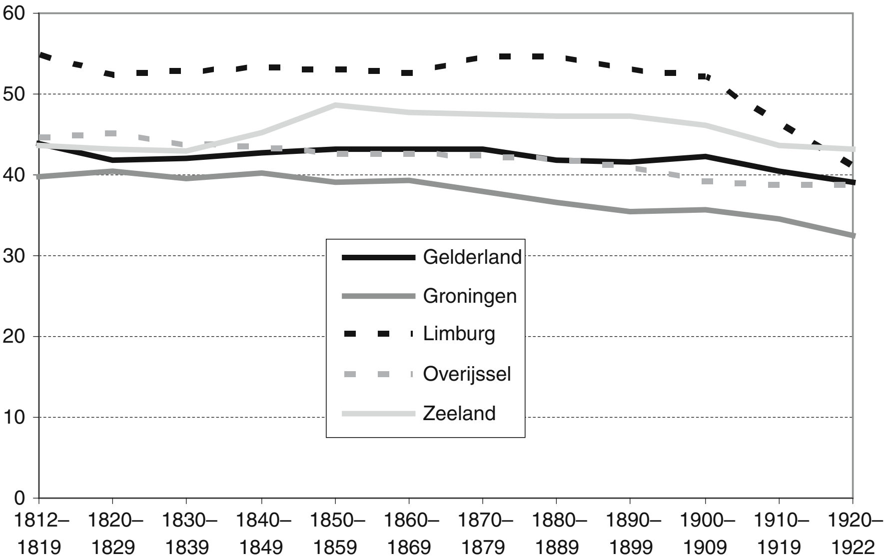


Fig. 6.4 Percentage of all marriages with a distance of less than 5 km between places of birth of bride and groom, by period of marriage and province

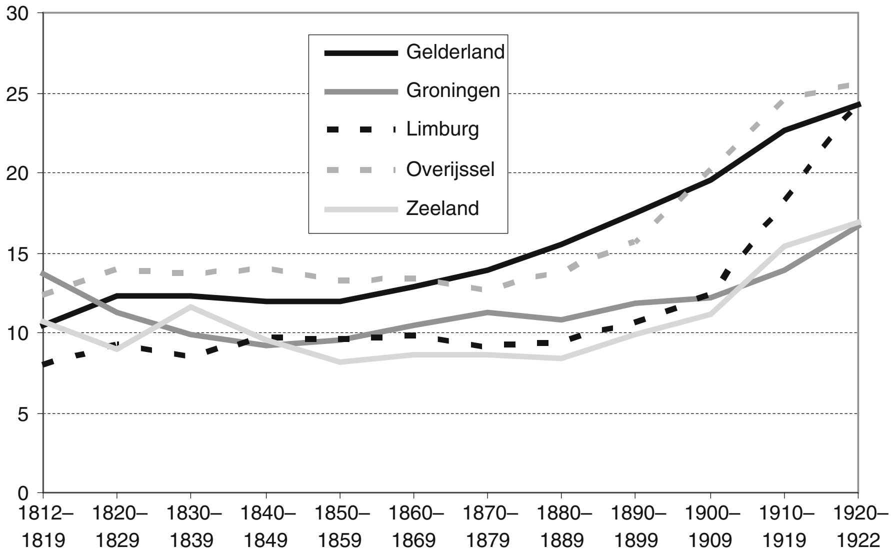


Fig. 6.5 Percentage of all marriages with a distance of more than 40 km between places of birth of bride and groom, by period of marriage and province

apart, with the percentage falling further after the 1860s. Limburg was for a long time almost completely isolated from the other provinces of the country (Van der Woud, 2006: p. 320). The strong local orientation decreased in Limburg after 1910, a period during which mining started to attract very large numbers of migrants and local communities dissolved. In all provinces, the interaction over long distances increased at first gradually (in the 1880s) and later much more strongly, in particular after 1910. For brides and grooms marrying in Zeeland it was much more difficult, given the isolation of the various islands and peninsulas that composed Zeeland, to get in touch with spouses coming from a distance beyond 40 km.

Of those coming from a distance of more than 40 km, a non-negligible proportion came from neighbouring countries. All five provinces bordered one or more other countries and this is evident in the choice of spouses (Fig. 6.6).

Figure 6.6 shows that in Limburg, located in a remote corner of the Netherlands between Belgium and Germany, more than 14% of all grooms who married in the period 1810–1819 were born outside the Netherlands. For Zeeland, the percentage was only a little bit lower. Limburg remained for a long time much more strongly oriented towards the south than towards far-away Holland. In the eastern part of southern Limburg, German was until the First World War the main language, while in Maastricht French was the first language after the local dialect. There were frequent contacts with friends and families in towns and villages across the border. However, particularly in the first half of the nineteenth century, the borders between the Netherlands and Belgium and Germany became stronger barriers. This nationalization of the border regions was a consequence of the Belgian Revolution (1830),

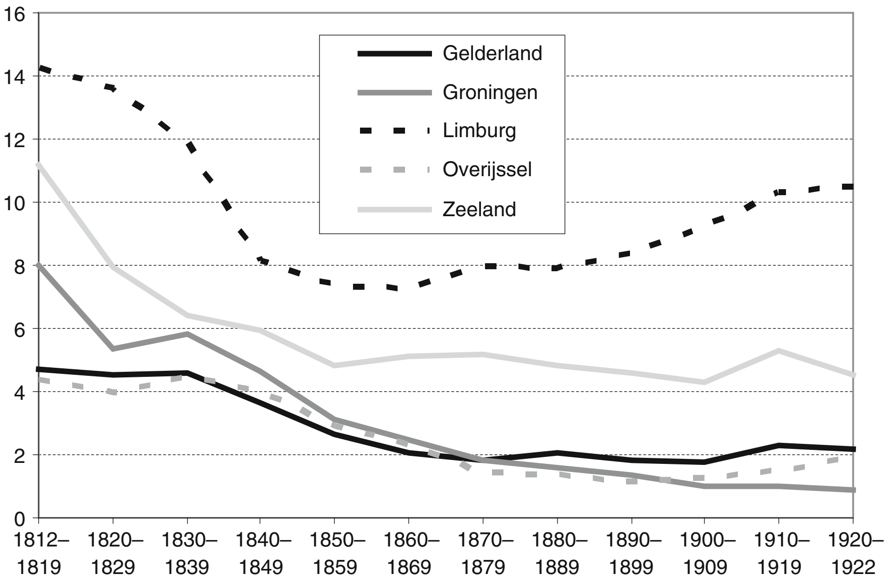


Fig. 6.6 Percentage foreign-born grooms per province, by period of marriage

during which Limburg was divided into Belgian and Dutch parts, separated by a border (Rutten, 1989). As a result, places where young people met each other became more and more organized within a national framework (schools, sporting clubs etc.). It was only from the end of the nineteenth century, when coal mining and the transport system developed, that the percentage of foreign-born grooms increased again.

To a certain degree the same development took place in Zeeland, where Zeeuws-Vlaanderen in the south had been part of what became after 1830 the Belgian region of Flanders and was for a long time much more oriented to the Belgian hinterland of Eastern Flanders, partly as a consequence of better transport connections with Belgium than with the Netherlands.

The Direction of Preferences

Possible directional bias in the spatial interaction between people from the five provinces is depicted in Figure 6.7. This figure gives the weighted mean centres for the four consecutive periods and the standard deviation ellipses for the first and the last period by province of marriage.

Brides and grooms in Gelderland were strongly focused on the province itself, but there were also clear deviations from a pattern that one could expect when only distance played a role. These deviations might first of all be understood by referring to the location of the big cities (Amsterdam, Rotterdam and The Hague) and

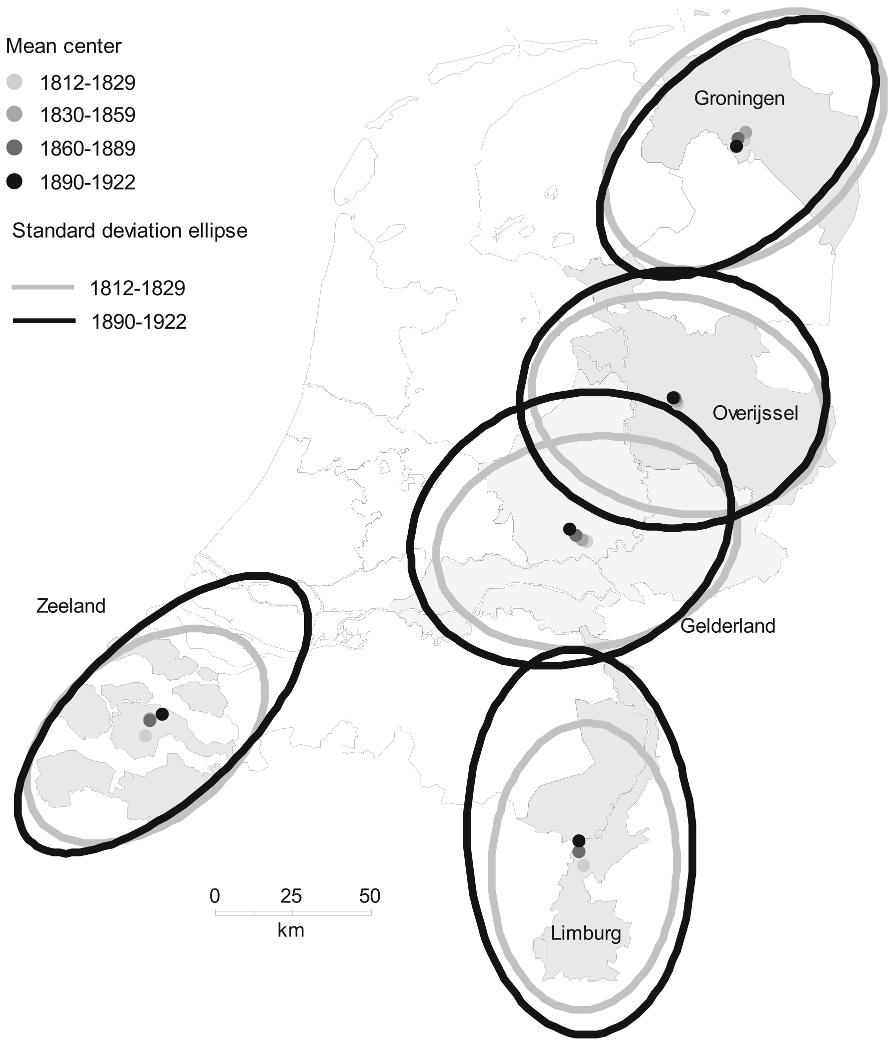


Fig. 6.7 Weighted mean centre points and standard deviation ellipses for birthplaces of grooms by period of marriage and province of marriage

the other urbanized parts of the country in the west that are related to the density and distribution of the population. Gelderland also had a relatively high level of interaction with Germany, particularly the Land of Cleves in the Lower Rhine area. With respect to the German grooms, the 1890–1922 data show a shift to the Ruhr Area. Gelderland started to attract grooms from all parts of the Netherlands from 1830. In the period 1812–1829 grooms originated from around 900 different Dutch places. This figure increased to around 1,460 different places in the period 1890–1922. Figure 6.7 shows that the mean centre of the birthplaces of the grooms married in Gelderland shifted more to the west. The same applies to the standard deviation ellipse: the area of the ellipse and thus the dispersion, increased to the (north)west.

As expected, in Groningen the most frequent contacts were indeed with places that were closest to the municipalities in question (other municipalities in the province itself and nearby Friesland). Yet, there were also clear deviations from a pattern that one could expect when only distance played a role. These deviations might again be understood by referring to the location of the big cities and the other urbanized parts of the country in the west. The same factor also explains the small number of grooms coming from nearby, but scarcely populated, Drenthe (the province to the south of Groningen). The channels of communication of Drenthe also were more directed to the south than to the north. In the course of time, grooms marrying in Groningen came from an ever-increasing number of municipalities throughout the Netherlands (from around 640 in the period 1812–1829 to around 1,310 in the period 1890–1922). There remained over time two areas from which only small numbers of grooms were coming: the eastern part of Noord-Brabant and the northern part of Limburg. Both were strictly Catholic regions (whereas Groningen was predominantly Dutch Reformed) and similarly sparsely populated. Small numbers originated in the Veluwe, an area with low population densities. Relatively large numbers of grooms originated from Zeeland, a province that was far away and physically isolated, but with a comparable agricultural system and religion. In Groningen, the mean centre of the birthplaces of the grooms shifted more to the southwest from the period 1830–1859 (see Fig. 6.7). The area of the standard deviation ellipse actually decreased in 1830–1859 and from then on increased and shifted to the southwest again.

Brides and grooms in Limburg were strongly focused on the province itself, and much less on nearby and likewise Catholic Brabant. Relatively more grooms from Gelderland, the province to the north of Limburg, married in Limburg. With Drenthe, Overijssel and the Veluwe, interaction was very restricted, and that applied to Zeeland as well. Large parts of these areas were Orthodox Calvinist and cultural barriers might have been strong. Marriage fields stretched far out into Belgium and Germany. Germany was very popular especially in the first decades of the twentieth century, when mining in Limburg became an important economic activity: the number of places where German grooms originated increased from around 390 places in the period 1860–1889 to 635 in the period 1890–1922 (the similar number of Dutch places in the same period increased from around 730 to around 1,020). Former barriers with the north and southwest lost their importance in that period as well. Figure 6.7 shows that the mean centre of the birthplaces of the grooms married in Limburg shifted more to the north. The same applies to the standard deviation ellipse, which increased in size and thus dispersion and shifted towards the northwest.

Overijssel also had a high level of interaction with Germany. This province started, like Gelderland, to attract grooms from all parts of the country from 1830 on. In general, the interaction with the southern provinces of Noord-Brabant and Limburg and with Zeeland was much less intense than one would expect and certainly less than with the north of the country. Figure 6.7 shows that the mean centre of the birthplaces of the grooms married in Overijssel shifted (to the north and west), but much less so than in the other provinces. The same applies to the standard deviation ellipse: the dispersion increased to a lesser extent than elsewhere.

In the beginning of the nineteenth century, the situation in Zeeland was comparable to that in Limburg. As a consequence of its isolated position, almost all grooms had been born in this province. Interaction with the northern part of eastern Flanders was rather frequent. The number of places from which Belgian grooms originated, however, declined from around 300 in the period 1812–1829 to 200 in the period 1830–1859. From the period 1860–1889 to 1890–1922, the number increased from around 180 to 290. Contacts with Catholic Brabant were relatively few, with Orthodox Calvinist parts of Zuid-Holland and Gelderland much more common. Figure 6.7 shows that the mean centre of the birthplaces of the grooms married in Zeeland shifted to the northeast (in the direction of Zuid-Holland). The same applies to the standard deviation ellipse, which increased in size and thus dispersion and shifted mainly towards the northeast and only slightly towards the direction of Eastern Flanders.

The strong directional component in migration could thus be explained in part by regional variations in levels of urbanization and in part by cultural differences, connected with differences in religion.

Distances and Social Classes

Table 6.2 gives the average distance between the birthplace of the groom and his wife for the various social classes and for each time period. In the five provinces there are common trends visible but also clear differences. Common to all provinces is that the higher social classes had, as expected, the widest horizon; on average, members of these classes were born 40–70 km away from their spouses. Only the middle class came near the distances travelled by the higher social classes, but nonetheless the distance between the birthplaces of the spouses remained on average some 20 km less. In the five provinces there were clear differences visible in the trends over time in the higher social class: in Overijssel distances increased regularly after 1830–1839, from around 57 km to almost 70 in the first decade of the twentieth century. In Zeeland, the distance between the birthplaces of the spouses increased even more markedly, whereas distances travelled by the higher social class in Gelderland and Groningen increased much more gradually. In Limburg, on the other hand, the circle within which partners were found narrowed considerably after the middle of the nineteenth century. It was only from the first decade of the twentieth century that the horizon of the upper class seemed to widen again.

As for the middle class, Gelderland, Overijssel and Zeeland showed comparable trends, indicating a widening of geographic horizons beginning in the middle of the nineteenth century and continuing until the First World War (and beyond in the case of Zeeland). Stability of distances between birthplaces of bride and groom was the main characteristic of the middle class in Limburg and Groningen.

Table 6.2 Mean average distance (in km) between place of birth of bride and groom, by social class of groom, period of marriage and province

Province	Period	Social class of groom							Total
		Higher	Middle	Skilled workers	Farmers	Semi-skilled workers	Farm workers	Unskilled workers	
Gelderland	1812–1819	52.9	31.8	19.5	6.7	22.5	9.9	15.9	14.3
	1820–1829	53.8	40.2	19.6	7.2	24.9	10.4	20.3	16.4
	1830–1839	60.5	39.5	20.1	6.3	25.3	9.8	18.0	16.5
	1840–1849	55.7	40.0	19.6	6.5	20.4	9.4	19.8	16.1
	1850–1859	55.2	42.4	19.9	7.0	21.6	9.3	17.9	16.1
	1860–1869	52.7	45.7	20.6	7.0	22.9	9.0	17.6	16.7
	1870–1879	57.2	44.9	20.6	7.3	24.4	8.9	15.9	17.5
	1880–1889	61.0	46.0	21.2	7.6	24.7	9.4	15.9	18.9
	1890–1899	66.0	47.7	24.8	7.7	26.5	9.8	18.2	21.0
	1900–1909	65.3	48.6	26.2	7.8	26.9	10.5	18.9	22.7
	1910–1919	61.9	48.6	27.9	9.8	28.9	14.1	21.0	25.2
	1920–1922	61.7	47.8	28.1	10.0	28.4	15.4	20.5	26.8
Groningen	1812–1819	46.8	35.9	25.5	8.2	35.4	13.5	27.0	21.7
	1820–1829	49.5	32.9	24.9	9.0	28.5	11.9	22.7	19.5
	1830–1839	53.9	33.3	21.7	9.5	23.5	10.8	20.3	17.8
	1840–1849	46.7	32.3	18.8	9.1	21.4	9.8	19.0	16.5
	1850–1859	52.2	37.8	18.3	9.1	23.8	9.6	21.8	17.2
	1860–1869	47.3	40.2	18.6	9.1	25.6	10.5	24.5	18.5
	1870–1879	53.5	37.1	19.1	10.2	25.8	10.6	23.6	18.9
	1880–1889	57.3	34.6	18.6	10.0	24.3	10.4	17.7	18.0
	1890–1899	51.7	38.9	20.9	10.9	22.2	10.4	16.6	19.4
	1900–1909	55.6	38.4	21.6	12.1	22.2	9.7	18.1	20.5
	1910–1919	57.8	41.4	23.3	13.5	25.1	10.2	19.2	22.3
	1920–1922	61.3	41.1	28.8	17.8	27.0	11.3	20.5	25.6
Limburg	1812–1819	50.3	37.5	14.5	5.9	21.3	7.9	11.6	13.3
	1820–1829	55.7	47.7	13.3	5.9	23.9	9.6	12.5	15.4
	1830–1839	58.6	42.2	12.0	6.0	25.9	8.7	11.6	14.7
	1840–1849	58.8	47.1	13.6	5.3	24.5	9.1	14.2	15.8
	1850–1859	53.0	44.6	14.5	6.1	19.8	9.1	16.2	16.0
	1860–1869	48.6	47.7	14.4	5.7	19.7	9.3	14.4	15.7
	1870–1879	51.5	46.0	12.9	5.9	16.2	8.4	12.8	14.7
	1880–1889	48.7	42.8	13.6	5.8	13.2	9.7	11.6	14.9
	1890–1899	42.4	45.3	15.0	6.6	13.5	11.5	14.2	16.1
	1900–1909	45.4	44.2	17.3	7.2	17.5	11.9	16.1	18.1
	1910–1919	50.4	47.7	23.3	8.6	27.5	18.5	21.0	24.6
	1920–1922	57.9	48.9	27.1	9.7	36.2	21.9	28.0	31.8
Overijssel	1812–1819	57.0	36.2	25.4	7.6	22.8	10.9	20.0	17.1
	1820–1829	63.0	39.4	25.4	6.5	20.5	12.6	22.6	19.0
	1830–1839	61.1	38.3	24.4	6.4	20.3	13.2	22.9	19.2
	1840–1849	56.4	39.7	24.7	6.9	18.0	14.8	24.3	19.6
	1850–1859	61.5	39.4	22.9	7.4	19.2	13.5	23.0	18.9
	1860–1869	59.7	41.8	22.8	7.1	22.6	12.2	19.1	18.6
	1870–1879	62.0	41.8	21.6	7.3	21.9	11.8	16.9	18.1

Table 6.2 (continued)

		Social class of groom							
Province	Period	Higher	Middle	Skilled workers	Farmers	Semi-skilled workers	Farm workers	Unskilled workers	Total
	1880–1889	64.4	42.4	23.3	6.9	22.5	11.2	16.0	18.7
	1890–1899	66.8	45.7	26.3	7.2	21.9	10.8	19.1	20.7
	1900–1909	68.1	49.5	29.3	7.4	25.0	12.1	24.2	24.5
	1910–1919	66.7	52.0	31.3	8.8	29.1	14.5	28.6	28.2
	1920–1922	69.9	48.2	30.9	8.5	27.5	16.2	29.5	28.9
Zeeland	1812–1819	46.7	32.6	20.6	8.3	16.9	9.8	22.3	16.8
	1820–1829	49.5	31.0	19.7	7.3	19.4	9.0	21.0	15.5
	1830–1839	55.2	36.3	21.5	7.6	31.2	8.6	23.4	17.7
	1840–1849	48.3	34.2	18.4	7.2	24.2	8.2	20.7	15.9
	1850–1859	55.1	36.1	19.0	7.7	23.3	6.8	16.5	14.7
	1860–1869	63.7	35.7	19.4	7.1	22.9	6.6	15.8	14.8
	1870–1879	63.5	35.9	19.0	7.6	18.5	6.5	17.0	15.1
	1880–1889	66.8	32.2	19.6	7.2	23.8	6.6	14.9	15.3
	1890–1899	68.7	38.2	21.0	7.5	20.7	6.4	15.6	16.9
	1900–1909	71.6	39.3	22.8	8.5	24.7	6.6	15.5	18.4
	1910–1919	71.6	42.4	28.5	11.8	33.1	8.6	22.2	23.4
	1920–1922	75.9	45.5	28.3	12.9	31.7	8.2	21.6	25.4

At the other end of the scale, two groups that had their basis in farming can be distinguished. Both were characterized by the fact that grooms found their brides close to their own birthplaces. Until the middle of the nineteenth century, the birthplaces of farmers and those of their wives were on average no more than 4–6 km apart. Distances increased gradually at first, but much more strongly after 1910. In the early twentieth century in particular, farmers in Groningen married brides who were born further away than had been the case earlier on.

Farm workers were also very strongly locally oriented. In four of the five provinces, this local orientation hardly changed during the century that we studied and in some cases the data suggest that the horizon of this group became further restricted. There was one exception to this trend: farm workers in Limburg. Here, after 1890 a very strong increase in their horizon is apparent.

Between those working in agriculture and the high and middle classes were the various groups of workers outside agriculture. Although the numbers are relatively large in all three groups, we find strong fluctuations over time that are hard to explain. Among these groups, unskilled workers were in general characterized by a narrower geographic horizon than that of the others. The distances separating the birthplaces of skilled workers and their brides in Gelderland, Groningen, Limburg and Zeeland were relatively stable over the period 1810–1890. Later, a clear increase took place. In Overijssel the horizon narrowed initially, but this trend came to an end after 1880, when, as in other provinces, the distances between the birthplaces of skilled workers and those of their brides started to widen. Semi-skilled and unskilled workers had more or less the same distances between their own and their bride's birthplace. In Zeeland, both groups at first witnessed a clear narrowing of the geographic range within which they found a bride and after 1900 an increase. In Limburg and Overijssel, a more or less comparable pattern was found (although in the first half of the nineteenth century stability, rather than a clear decrease, was observed). In both groups the distances between birthplaces of bride and groom increased after 1900 by some 15–20 km. This increase started earlier in Overijssel than in Limburg. Groningen had a different pattern: here the distances decreased as well, but did reach higher levels again later in the century.

Finally, we compared again the two groups of couples whose birthplaces were nearest and furthest apart. For all provinces and social classes this involved a comparison of the percentages found in 1812–1819 with those observed one century later in 1910–1919 (see Table 6.3).

In all provinces the higher class had already in the first decades of the nineteenth century a relatively high percentage of marriages (between 40 and 55%) in which brides and grooms were born more than 40 km apart. Usually, this percentage increased in the century that followed, but it did not do so in Groningen and did so only slightly in Limburg. Members of the higher classes relatively rarely found their spouses near their birthplaces; this happened in less than one quarter of all marriages (but in just over a quarter in Limburg in both periods). In the middle class, a relatively even share of local and far-away interactions is evident, never falling below 20% or exceeding 48% for any combination of region or time period. Over time, in all provinces a strong increase occurred in marriages between spouses who had been born far apart. The only major exception was again Groningen. Farmers and farm workers were very locally oriented: marrying someone who was born more than 40 km away was very rare. Over time, the marked preferences of farmers for brides born in the direct neighbourhood of their own community of birth declined, in particular in Groningen. A considerable proportion of the various groups of workers outside agriculture found a spouse nearby; nonetheless, marriages with partners who were born rather far away were relatively frequent as well (on average around a quarter). The development over time is far from consistent; in Gelderland, Overijssel and Limburg, more interaction took place with regions further away, leading to increases in the number of brides born outside the 40-km zone. In Zeeland, this tendency was observed among almost all groups of workers, whereas in Groningen the marriage market of workers was more contracted in the 1920s than it was in the first decades of the nineteenth century.

Spatial Regression

The previous sections have analysed the spatial interaction and widening geographic horizons in a descriptive way. This section will extend those analyses with a first attempt to apply spatial regression techniques (see, for instance, Anselin (1988); Anselin and Bera (1998)). Spatial regression techniques allow us to deal with spatial

Table 6.3 Percentages of all marriages with a distance of less than 5 or more than 40 km between places of birth of bride and groom, by province and social class of groom, 1812–1819 and 1910–1919

	Gelderland		Groningen		Limburg		Overijssel		Zeeland	
	1812–1819	1910–1919	1812–1819	1910–1919	1812–1819	1910–1919	1812–1819	1910–1919	1812–1819	1910–1919
Higher class										
<5 km	16.8	14.4	21.6	19.2	28.9	27.3	19.4	15.8	22.8	16.8
40+ km	54.5	60.5	40.0	39.7	40.8	41.1	51.5	62.0	39.7	53.2
Middle class										
<5 km	30.6	23.6	29.1	23.0	36.2	30.0	30.5	24.1	33.8	31.0
40+ km	28.7	46.6	28.7	28.5	29.8	36.9	31.8	48.2	27.1	30.7
Skilled workers										
<5 km	40.8	37.6	35.0	32.5	52.1	47.1	35.4	37.7	43.8	41.6
40+ km	16.3	25.1	17.8	13.7	9.2	18.3	22.1	28.6	14.2	19.8
Farmers										
<5 km	55.9	52.9	50.7	34.6	65.9	55.4	55.5	52.3	48.7	46.4
40+ km	2.1	5.1	2.1	4.7	2.1	3.5	2.9	3.5	2.0	4.8
Lower skilled workers										
<5 km	38.2	38.8	36.0	34.4	47.0	47.4	41.1	39.8	47.5	35.5
40+ km	19.5	26.5	24.8	15.2	13.8	21.2	17.1	26.4	10.0	23.5
Farm workers										
<5 km	41.1	52.3	39.6	47.3	56.7	45.3	42.4	43.6	46.6	54.6
40+ km	4.1	9.6	6.8	3.8	2.9	12.6	5.0	9.4	4.0	3.6
Unskilled workers										
<5 km	44.1	42.0	41.8	36.9	49.9	48.8	41.7	36.2	39.6	45.5
40+ km	12.5	17.0	17.8	11.1	6.0	16.9	15.5	26.3	16.1	15.7

autocorrelation in regression analysis by including individual explanatory variables such as age and social class, (regional) socio-economic explanatory variables and spatial location (in our case identified according to the Dutch geographical coordinate system). Spatial autocorrelation occurs when a value at any one point in space is dependent on values at the surrounding points. Spatial autocorrelation can either be positive or negative: positive means that similar values tend to be near each other, whereas negative means that different values tend to be near each other. The *GeoDa* software package (Anselin, 2004, 2005) contains the functionality for spatial regression modelling, including the standard ordinary least squares regression and basic diagnostics for spatial autocorrelation (Anselin et al., 2004). Since the data sets are very large, estimation of spatial regression models requires intensive and time-consuming computational effort from the software (for example matrix inversion and computing of spatial weights). *GeoDa* regression routines can be applied to quite large data sets, but ours appeared to be too large to include in a single spatial regression analysis. We therefore had to split and limit our spatial regression to separate 7.5% random samples of each of the five data sets of grooms married in the provinces of Gelderland, Groningen, Limburg, Overijssel and Zeeland.

In the spatial regression analysis we included independent variables at the level of the individual (such as age, social class and birthplace from the marriage certificates), the municipality (such as classification of the area as either rural or urban and the predominant religious denomination) and the nation (national income and transport availability, i.e. length of railway tracks, number of bicycles). The dependent variable in our analysis is the distance between birthplace of groom and birthplace of bride. See Table 6.4 for an overview of the variables.

Tables 6.5-6.9 present the summary outcomes of the regression models in *GeoDa* for each province. The regression outcomes confirm our earlier findings. The fit of the models are not that impressive, with adjusted R^2s ranging from 0.115 to 0.155, but most of the variables are highly significant with the expected sign except for the national variables.

Table 6.4 Variables used in the spatial regression analysis

Dependent	– Distance between place of birth of groom and place of birth of bride (in kilometres)
Independent	– Social class/occupation of the groom (in dummy variables):
	– Higher class
	– Middle class
	– Skilled workers
	– Farmers
	– Lower skilled workers
	– Farm workers
	– (Unskilled workers – the reference category)
	– Age of the groom at marriage
	– Age of the bride at marriage
	– Marriage rank number (first/second marriage)
	– Working status of the bride (working or non-working)

Table 6.4 (continued)

- Groom is born abroad (or not)
- Rural or non-rural area
- Municipality of marriage is a (large) city (or not)
- Religion: percentage of Catholics in municipality of marriage
- National income per capita in the Netherlands (index 1920 = 1.00)
- Total length of the railway tracks in the Netherlands (index 1920 = 1.00)
- Number of bicycles in the Netherlands (index 1920 = 1.00)
- Observation period (in dummy variables):
  - Period from 1830 to 1860
  - Period from 1860 to 1890
  - Period from 1890 to 1922
  - (Period before 1830 – the reference category)
- Municipality of marriage is (or is not) located on an island; (applies to Zeeland only)

Table 6.5 Summary output of spatial regression analysis for province of Gelderland

Regression summary of output	*Ordinary Least Squares estimation*			
Dependent variable	Distance between places of birth of groom and bride (in km)			
Number of observations	27,355			
Mean dependent var	19.842	Number of variables		21
S.D. dependent var	33.4764	Degrees of freedom		27,334
R-squared	0.120605	F-statistic		187.437
Adjusted R-squared	0.119962	Prob(F-statistic)		0
Sum squared residual	2.70E+07	Log likelihood		–133,096
Sigma-square	986.266	Akaike info criterion		266,234
S.E. of regression	31.4049	Schwarz criterion		266,407
Sigma-square ML	985.509			
S.E of regression ML	31.3928			
Variable	Coefficient	Std. error	t-Statistic	Probability
Constant	14.32465	1.404252	10.20091	0.0000000
Higher class	30.05507	1.91527	15.69235	0.0000000
Middle class	32.6492	1.831569	17.8258	0.0000000
Skilled workers	0.195748	0.5442099	0.359692	0.7197760
Farmers	–11.52719	0.5122099	–22.5048	0.0000000
Lower skilled workers	1.923424	0.6109526	3.148237	0.0016446
Farm workers	–2.959372	1.423229	–2.07934	0.0376131
Age at marriage (groom)	0.1915213	0.03036064	6.308213	0.0000000
Age at marriage (bride)	0.2601479	0.03756227	6.925776	0.0000000
Marriage rank number	–3.720393	0.859088	–4.33063	0.0000149
Working status of bride	–3.053386	0.4394115	–6.94881	0.0000000
Groom is born abroad	28.13122	1.668276	16.86245	0.0000000
Rural area	–7.605051	0.4937407	–15.4029	0.0000000
City	12.04401	0.7208599	16.70783	0.0000000
Percentage of Catholics	–0.04532509	0.006920918	–6.549	0.0000000
National income	3.698576	2.584235	1.431207	0.1524076
Length of railway tracks	2.839017	1.77351	1.60079	0.1094008
Number of bicycles	1.301872	1.500747	0.867483	0.3856220

Table 6.5 (continued)

Period 1830–1859	–0.4479198	0.8561976	–0.52315	0.6008555
Period 1860–1889	–0.4474313	1.169629	–0.38254	0.7017782
Period 1890–1922	1.775433	1.659024	1.070167	0.2845239
Regression diagnostics				
Multicollinearity condition number			27.4001	
Test on normality of errors				
Test	*DF*	*Value*	*Prob*	
Jarque-Bera	2	87,288.08	0	
Diagnostics for heteroskedasticity				
Random coefficients				
Test	*DF*	*Value*	*Prob*	
Breusch-Pagan test	20	6,024.552	0	
Koenker-Bassett test	20	1,287.277	0	
Specification robust test				
Test	*DF*	*Value*	*Prob*	
White	230	N/A	N/A	
Diagnostics for spatial dependence				
Test	*MI/DF*	*Value*	*Prob*	
Moran's *I* (error)	0.021486	16.5322564	0.0000000	
Lagrange multiplier (lag)	1	62.8893805	0.0000000	
Robust LM (lag)	1	3.479866	0.0621198	
Lagrange multiplier (error)	1	75.7716067	0.0000000	
Robust LM (error)	1	16.3620923	0.0000523	
Lagrange multiplier (SARMA)	2	79.2514727	0.0000000	

Table 6.6 Summary output of spatial regression analysis for province of Groningen

Regression summary of output	*Ordinary least squares estimation*		
Dependent variable	Distance between places of birth of groom and bride (in km)		
Number of observations	15,617		
Mean dependent var	19.7528	Number of variables	21
S.D. dependent var	36.6919	Degrees of freedom	15,596
R-squared	0.116204	F-statistic	102.531
Adjusted R-squared	0.115071	Prob(F-statistic)	0
Sum squared residual	1.86E+07	Log likelihood	–77,456.1
Sigma-square	1,191.45	Akaike info criterion	154,954
S.E. of regression	34.5174	Schwarz criterion	155,115
Sigma-square ML	1,189.85		
S.E of regression ML	34.4942		

Variable	Coefficient	Std. error	t-Statistic	Probability
Constant	11.1	1.964313	5.65083	0.0000000
Higher class	23.91941	2.096008	11.41189	0.0000000
Middle class	12.71445	1.015652	12.51851	0.0000000
Skilled workers	–0.8991385	0.8756594	–1.02681	0.3045506
Farmers	–6.984254	1.054539	–6.62304	0.0000000
Lower skilled workers	1.401745	1.184578	1.183329	0.2367258

Table 6.6 (continued)

Farm workers	–5.019742	0.8380199	–5.99	0.0000000
Age at marriage (groom)	0.2072703	0.05032128	4.11894	0.0000383
Age at marriage (bride)	0.1037542	0.05661012	1.832785	0.0668652
Marriage rank number	–1.59047	0.9815981	–1.62029	0.1051719
Working status of bride	–1.889122	0.6475232	–2.91746	0.0035346
Groom is born abroad	36.21525	2.267746	15.96972	0.0000000
Rural area	–4.482644	0.8677672	–5.16572	0.0000002
City	11.60837	1.070393	10.84496	0.0000000
Percentage of Catholics	0.1908934	0.05322492	3.586541	0.0003362
National income	3.277446	2.320512	1.412381	0.4345110
Length of railway tracks	–2.221491	2.487614	–0.89302	0.3718160
Number of bicycles	3.157791	4.042126	0.781221	0.1578218
Period 1830–1859	–1.100149	1.180849	–0.93166	0.3514439
Period 1860–1889	1.46925	1.627176	0.902945	0.3665729
Period 1890–1922	1.433851	2.325026	0.616703	0.5374982
Regression diagnostics				
Multicollinearity condition number			25.98997	
Test on normality of errors				
Test	*DF*	*Value*	*Prob*	
Jarque-Bera	2	201,107.2	0	
Diagnostics for heteroskedasticity				
Random coefficients				
Test	*DF*	*Value*	*Prob*	
Breusch-Pagan test	20	6,885.588	0	
Koenker-Bassett test	20	759.4399	0	
Specification robust test				
Test	*DF*	*Value*	*Prob*	
White	230	N/A	N/A	
Diagnostics for spatial dependence				
Test	*MI/DF*	*Value*	*Prob*	
Moran's *I* (error)	0.004292	2.688123	0.0071856	
Lagrange multiplier (lag)	1	1.114525	0.2911005	
Robust LM (lag)	1	1.702834	0.1919178	
Lagrange multiplier (error)	1	1.726262	0.1888890	
Robust LM (error)	1	2.314571	0.1281667	
Lagrange multiplier (SARMA)	2	3.429095	0.1800452	

Table 6.7 Summary output of spatial regression analysis for province of Limburg

Regression summary of output	*Ordinary least squares estimation*		
Dependent variable	Distance between places of birth of groom and bride (in km)		
Number of observations	14,226		
Mean dependent var	17.787	Number of variables	21
S.D. dependent var	37.4594	Degrees of freedom	14,205
R-squared	0.155807	F-statistic	131.087
Adjusted R-squared	0.154619	Prob(F-statistic)	0

Table 6.7 (continued)

Sum squared residual	1.69E+07	Log likelihood		–70,525.5
Sigma-square	1,186.33	Akaike info criterion		141,093
S.E. of regression	34.4431	Schwarz criterion		141,252
Sigma-square ML	1,184.58			
S.E of regression ML	34.4177			
Variable	Coefficient	Std. error	t-Statistic	Probability
Constant	75.07561	5.924068	12.67298	0.0000000
Higher class	26.51208	2.071121	12.80083	0.0000000
Middle class	23.70225	1.176396	20.1482	0.0000000
Skilled workers	–0.06968816	0.9893649	–0.07044	0.9431800
Farmers	–7.021315	0.9070016	–7.74124	0.0000000
Lower skilled workers	2.686111	1.013469	2.650412	0.0080474
Farm workers	–1.53047	1.167306	–1.31111	0.1898460
Age at marriage (groom)	0.1452882	0.04657861	3.119204	0.0018175
Age at marriage (bride)	0.007675968	0.05203559	0.147514	0.8820525
Marriage rank number	0.2077095	1.012734	0.205098	0.8369867
Working status of bride	–1.456864	0.6271511	–2.32299	0.0201914
Groom is born abroad	17.09421	1.042231	16.40155	0.0000000
Rural area	–8.123077	0.7612207	–10.6711	0.0000000
City	2.682365	0.9701267	2.764963	0.0056988
Percentage of Catholics	–0.6422265	0.05742621	–11.1835	0.0000000
National income	0.1354192	4.102846	0.033006	0.9767870
Length of railway tracks	–5.740627	2.824298	–2.03259	0.0421097
Number of bicycles	11.85288	2.404787	4.92887	0.0000008
Period 1830–1859	2.508045	1.06387	2.357473	0.0184149
Period 1860–1889	3.64413	1.674722	2.175961	0.0295826
Period 1890–1922	4.074876	2.524424	1.61418	0.1065007
Regression diagnostics				
Multicollinearity condition number			86.54396	
Test on normality of errors				
Test	*DF*	*Value*	*Prob*	
Jarque-Bera	2	106,372.2	0	
Diagnostics for heteroskedasticity				
Random coefficients				
Test	*DF*	*Value*	*Prob*	
Breusch-Pagan test	20	8,402.116	0	
Koenker-Bassett test	20	1,199.498	0	
Specification robust test				
Test	*DF*	*Value*	*Prob*	
White	230	N/A	N/A	
Diagnostics for spatial dependence				
Test	*MI/DF*	*Value*	*Prob*	
Moran's *I* (error)	0.018105	9.565514	0.0000000	
Lagrange multiplier (lag)	1	22.13177	0.0000025	
Robust LM (lag)	1	0.001073	0.9738713	
Lagrange multiplier (error)	1	27.97997	0.0000001	
Robust LM (error)	1	5.849269	0.0155835	
Lagrange multiplier (SARMA)	2	27.98104	0.0000008	

Table 6.8 Summary output of spatial regression analysis for province of Overijssel

Regression summary of output	*Ordinary least squares estimation*			
Dependent variable	Distance between places of birth of groom and bride (in km)			
Number of observations	16,549			
Mean dependent var	21.2401	Number of variables		21
S.D. dependent var	36.1079	Degrees of freedom		16,528
R-squared	0.139846	F-statistic		134.358
Adjusted R-squared	0.138805	Prob(F-statistic)		0
Sum squared residual	1.86E+07	Log likelihood		–81,588.7
Sigma-square	1,122.87	Akaike info criterion		163,219
S.E. of regression	33.5093	Schwarz criterion		163,381
Sigma-square ML	1,121.45			
S.E of regression ML	33.4881			

Variable	Coefficient	Std. error	t-Statistic	Probability
Constant	19.62607	1.780918	11.0202	0.0000000
Higher class	28.72034	2.570128	11.17467	0.0000000
Middle class	11.1134	1.544397	7.195944	0.0000000
Skilled workers	–1.579172	0.7834699	–2.01561	0.0438552
Farmers	–15.05616	0.6862466	–21.9399	0.0000000
Lower skilled workers	–3.218326	0.8815811	–3.65063	0.0002624
Farm workers	–6.680784	2.70659	–2.46834	0.0135880
Age at marriage (groom)	0.2385199	0.04397825	5.423589	0.0000001
Age at marriage (bride)	0.06586974	0.04978295	1.323138	0.1857959
Marriage rank number	–0.7520247	0.8626111	–0.8718	0.3834719
Working status of bride	–2.009069	0.5716142	–3.51473	0.0004415
Groom is born abroad	53.6953	1.977575	27.1521	0.0000000
Rural area	–10.23132	0.5863851	–17.4481	0.0000000
City	3.66751	0.9827259	3.731976	0.0001906
Percentage of Catholics	0.01983569	0.01112928	1.782298	0.0747227
National income	5.095453	3.729718	1.366177	0.1719437
Length of railway tracks	5.916897	2.368952	2.497685	0.0125114
Number of bicycles	2.370952	2.144382	1.105658	0.2688562
Period 1830–1859	–1.403677	1.101914	–1.27385	0.2027344
Period 1860–1889	–5.040092	1.541246	–3.27014	0.0010773
Period 1890–1922	–3.843945	2.200736	–1.74666	0.0807344
Regression diagnostics				
Multicollinearity condition number			25.24618	
Test on normality of errors				
Test	*DF*	*Value*	*Prob*	
Jarque-Bera	2	64,405.3	0	
Diagnostics for heteroskedasticity				
Random coefficients				
Test	*DF*	*Value*	*Prob*	
Breusch-Pagan test	20	4,789.851	0	
Koenker-Bassett test	20	927.3398	0	
Specification robust test				
Test	*DF*	*Value*	*Prob*	
White	230	N/A	N/A	

Table 6.8 (continued)

Diagnostics for spatial dependence			
Test	*MI/DF*	*Value*	*Prob*
Moran's *I* (error)	0.032003	19.58624	0.0000000
Lagrange multiplier (lag)	1	128.6497	0.0000000
Robust LM (lag)	1	29.23055	0.0000001
Lagrange multiplier (error)	1	101.6977	0.0000000
Robust LM (error)	1	2.27864	0.1311669
Lagrange multiplier (SARMA)	2	130.9283	0.0000000

Table 6.9 Summary output of spatial regression analysis for province of Zeeland

Regression summary of output	*Ordinary least squares estimation*			
Dependent variable	Distance between places of birth of groom and bride (in km)			
Number of observations	12,337			
Mean dependent var	17.5516	Number of variables	22	
S.D. dependent var	38.0578	Degrees of freedom	12,315	
R-squared	0.141784	F-statistic	96.8825	
Adjusted R-squared	0.14032	Prob(F-statistic)	0	
Sum squared residual	1.53E+07	Log likelihood	–61,457.9	
Sigma-square	1,245.26	Akaike info criterion	122,960	
S.E. of regression	35.2882	Schwarz criterion	123,123	
Sigma-square ML	1,243.04			
S.E of regression ML	35.2568			
Variable	Coefficient	Std. error	t-Statistic	Probability
Constant	15.34991	2.30747	6.65227	0.0000000
Higher class	41.17853	2.40102	17.15044	0.0000000
Middle class	19.16284	1.218985	15.72032	0.0000000
Skilled workers	1.884111	1.16816	1.612888	0.1068071
Farmers	–6.776013	1.387297	–4.88433	0.0000011
Lower skilled workers	5.329122	1.441063	3.698049	0.0002182
Farm workers	–5.047125	1.030641	–4.89707	0.0000010
Age at marriage (groom)	0.1126482	0.05656801	1.991376	0.0464520
Age at marriage (bride)	0.1126415	0.06302906	1.787137	0.0739363
Marriage rank number	–1.099182	1.109783	–0.99045	0.3220347
Working status of bride	–3.13019	0.7442065	–4.20608	0.0000262
Groom is born abroad	6.169954	1.622154	3.803557	0.0001433
Rural area	–7.643262	1.021237	–7.48432	0.0000000
City	10.47486	1.191457	8.791643	0.0000000
Percentage of Catholics	–0.01005445	0.01179407	–0.8525	0.3939605
National income	6.942392	4.901479	1.416387	0.1566584
Length of railway tracks	1.123544	2.959703	0.379614	0.7044684
Number of bicycles	1.371071	2.784965	0.492312	0.6226053
Period 1830–1859	–0.4006058	1.271391	–0.31509	0.7529237
Period 1860–1889	–2.17712	1.852233	–1.1754	0.2398512
Period 1890–1922	–1.326655	2.69381	–0.49248	0.6226053
Island	–1.739572	0.8148644	–2.1348	0.0328000
Regression diagnostics				
Multicollinearity condition number			24.32744	
Test on normality of errors				

Table 6.9 (continued)

Test	*DF*	*Value*	*Prob*
Jarque-Bera	2	174,614.4	0
Diagnostics for heteroskedasticity			
Random coefficients			
Test	*DF*	*Value*	*Prob*
Breusch-Pagan test	21	8,233.752	0
Koenker-Bassett test	21	865.4642	0
Specification robust test			
Test	*DF*	*Value*	*Prob*
White	252	N/A	N/A
Diagnostics for spatial dependence			
Test	*MI/DF*	*Value*	*Prob*
Moran's *I* (error)	0.020453	9.964293	0.0000000
Lagrange multiplier (lag)	1	36.41522	0.0000000
Robust LM (lag)	1	7.717704	0.0054682
Lagrange multiplier (error)	1	30.96581	0.0000000
Robust LM (error)	1	2.268295	0.1320451
Lagrange multiplier (SARMA)	2	38.68351	0.0000000

The geographic horizon of grooms from higher social classes and to a lesser extent from the middle class is much larger than that of other social classes. In particular, farmers and farm workers had a much smaller geographic horizon. The ages of the partners indicate that the older the partners, the wider their geographic horizon. People marrying for the first time tend to find their partner further away than those marrying for a second time, but this effect is only significant in Gelderland. If the bride was employed at the time of marriage, a negative coefficient resulted, meaning that the geographic horizon was smaller. Both the coefficients of the variables indicating the transportation availability in the Netherlands have the expected sign in Gelderland, Overijssel and Zeeland. However, the (increasing) number of bicycles is significant in Limburg only. National income per capita is also positively correlated, but is not significant. The time variables show a mixed picture and are not significant in most cases. Table 6.9 presents the summary outcomes for the province of Zeeland including one extra variable, i.e. whether the municipality of marriage is located on a (less accessible) island. The additional 'island' variable indeed confirms the negative effect on marriage horizons of less accessible regions.

The first statistic to determine spatial dependence is Moran's *I* (Moran, 1950). Moran's *I* is a spatial autocorrelation measure for variables with interval and ratio scales. The value of Moran's *I* ranges from -1 for negative spatial autocorrelation (nearby units have dissimilar values or characteristics) to +1 for positive spatial autocorrelation (adjacent units have similar values or characteristics). If no spatial autocorrelation exists there is no particular systematic spatial pattern. Tables 6.5-6.9 reveal that the Moran's *I* statistics show slightly positive values and are highly significant. The Moran's *I* statistic is less helpful in suggesting which alternative specification should be used. To this end, the so-called Lagrange Multiplier (LM) test statistics can be used to determine whether the spatial autocorrelation is due to spatial autocorrelation in the residuals (spatial errors) or spatial autocorrelation in the dependent variable (spatial lags) (see Anselin, 2005). Spatial autocorrelation in the residuals of the regression model indicates that the model specification is incomplete and additional independent variables that would fully explain the spatial patterns in the residuals are missing. Spatial autocorrelation in the dependent variable might incorporate spatial effects of unmeasured independent variables but also an additional effect of neighbouring dependent variables, i.e. the lagged dependent variable (as in diffusion processes). In Tables 6.5-6.9 the first two Lagrange Multiplier test statistics (LM lag and Robust LM lag) refer to the spatial lag model as the alternative and the next two (LM error and Robust LM error) refer to the spatial error model as the alternative. However, the Robust Lagrange Multipliers are only relevant if the others are significant (see Anselin (2005) for an extensive discussion on this issue).

Both the LM lag and the LM error are significant in the spatial regression models for Gelderland, Limburg, Overijssel and Zeeland. In the case of Groningen, neither the LM lag nor the LM error are significant, suggesting that no spatial dependency is left unexplained. For Gelderland and Limburg, the Robust LM lag is significant and seems to point to a spatial lag model more than to a spatial error model. For Overijssel and Zeeland the Robust LM error is significant, suggesting a spatial error model. These analyses and models need further exploration, which is beyond the scope of this chapter.

Conclusion

This chapter is the first attempt to study the widening of the geographic horizon for a large part of a whole country over a long time period.

As we expected, clear time trends were visible in the average marital distance between spouses and in the width of the zones in which they found each other. From the 1880s, the geographic horizon of young men and women in the Netherlands started to increase, an increase that gained momentum after the 1890s. The average horizon increased from around 15–20 km to around 25–30 km in the second decade of the twentieth century. At the same time, there were large differences between provinces, the horizon of people in Gelderland, Limburg and Zeeland being smaller before 1880 than those of people in Overijssel and Groningen. When a distinction is made between contacts within different zones, it appears that there were in particular strong differences between provinces in the degree of interaction at the local scale. In Limburg more than half of all marriages before 1900 were contracted between partners born fewer than 5 km apart; in Groningen, the proportion marrying locally never exceeded 40%. The strong local orientation eased in Limburg after 1910. In all provinces the interaction with regions further away increased, at first gradually, (around 1890, but earlier in Gelderland and Overijssel) and later much more strongly, particularly after 1910. Limburg has been depicted as a region with less intensive contacts with the outside world due to the less developed transportation network and agricultural activities that were less market-oriented. As a consequence, the province was touched to a lesser degree by an individualistic orientation (Wichers, 1965). It was only when mines were opened and transport and communication with the rest of the Netherlands improved that the province became better integrated into the country, its cultural peculiarities lessened and its geographic horizon opened.

Social status had the expected strong effect on marital distances. The higher social classes had a much wider horizon than any other social class. Those working in the agricultural sector, whether as farmers or farm workers, had very restricted geographic horizons. The geographic horizons of the various groups of workers outside agriculture usually extended much further than those of the farmers and rural labourers, but much less so than those of the high and middle classes. There was thus a close relationship between an individual's social position and the degree to which he interacted with the world outside his own community – the lower classes focusing on the locality or at the most on the region, the higher classes having a inter-regional or national orientation. Members of the lower socio-economic groups often spoke only dialect, had a low level of education, read few newspapers and had no free time and even less money with which to travel outside their own community. The middle class and higher class – particularly those working in trade and as entrepreneurs – had, because of their work, frequent contacts with people and regions outside the community, had received better education, were more open to the world outside, and had better access to all means of communication and transport.

During the nineteenth and early twentieth centuries, the differences in geographic horizon between the higher and middle classes and the agricultural workers and farmers increased in most provinces. The same trend over time, an accentuation of the social differences as a consequence of a constant widening of the marriage distances for the non-agricultural population and few changes for the agricultural one, has been observed elsewhere (Ogden, 1980).

Over time, the marriage horizons of the working classes widened, but nonetheless remained much smaller than those of the upper and middle classes. Even directly after the end of the First World War, the various social classes still varied strongly in their interaction with the outside world. How the geographic horizons of the various social classes evolved over time differed by province. In particular, the marriage horizons of working class grooms marrying in Gelderland and Groningen scarcely expanded at all in the last decades of the nineteenth century. Both agricultural provinces were hit very hard by the agricultural crisis. The real disposable incomes of farmer-households fell steadily and it was only at the beginning of the twentieth century that the economic position of the farmers improved again because of increasing agricultural prices and rising productivity. The differences in trends over time for the various regions make clear that local studies do not provide a comprehensive picture of marriage horizons and reveal the need to make use of large databases covering significant portions of a country.

The time trends we observed and the social differentials we found more or less parallel those reported by authors studying these issues in the UK and France.

Previous work on temporal trends in the nineteenth century generally showed increases in mobility after 1870 (Coleman, 1979), after 1886 (Perry, 1969) and after 1890 (Morel, 1972), depending on the area. The various provinces of the Netherlands fit into this time frame.

As mentioned in this chapter, the GENLIAS database does not yet contain information on marriages contracted in the economic heartland of the country, the provinces of Zuid- and Noord-Holland. Earlier studies, based respectively on the cities of Delft and Gouda, both situated in the province of Zuid-Holland (Van Poppel and Ekamper, 2005; Van Poppel, 1994), showed for the same time period comparable trends and differences as observed in the five provinces studied here. In both cities there was an increase in marriages in which one of the spouses came from contact zones not enabling regular face-to-face contact. This trend started in Gouda earlier than in the five provinces. A clear social gradient was also observed in both cities, corresponding with the one found in the GENLIAS data.

Our study does not provide a complete answer to all the questions that one might pose concerning changes to the geographic horizon of the Dutch population. Although marriage was one of the most frequently mentioned reasons for geographical mobility, we cannot interpret this form of spatial contact as representative of all moves: the mobility pattern of those who remained single is different from those of the majority who married (Pooley & Turnbull, 1998: p. 49, p. 69). Individuals who never married are excluded from our study, but marriage registers do record a very large segment of the population. In most European countries during the nineteenth century only a small proportion of the population remained unmarried.[8] Marriage registers principally measure the geographic proximity of younger adults, not that of persons in late middle age or the elderly. Moreover, the registers record only the overall horizon of brides and grooms, without indicating the exact date at which contacts were made. Mobility, then, is defined in a more restricted way than is the case for more general mobility measures. Local studies of marriage registers reveal nothing about how many people leave their home village to contract a marriage elsewhere. It will be clear that this issue plays a less serious role in our case, as our study uses information on hundreds of localities in five provinces. Apart from the fact that it is difficult to assign a precise date to changes in the geographic horizon (we know only that migration from the birthplace took place between the date of marriage and 20–30 years before marriage), a further problem is that the data do not allow us to know whether the horizon of grooms and brides in a given region widened because people from that province started to migrate over longer distances to other regions, or whether this was caused by the fact that more and more migrants born outside the province started to live in that province. The establishment of a link between changes in the horizon and a specific change in the propensity of the inhabitants of that province to migrate is therefore not fully justified.

Our study convincingly shows that changes in the transportation and communication network and the closely linked processes of national and cultural integration had far-reaching consequences for human behaviour. It caused in the most literal sense of the word a large increase in the emotional and cognitive connectedness between inhabitants of the various parts of the country. In its turn, this process of increasing marital distances further stimulated cultural homogenization and decreased the importance of regional identity.

The process of national integration not only narrowed cultural distances within the Netherlands: at the same time it widened cross-border cultural distances, as could be concluded from the decrease in marriages between people born in the Netherlands and those in Germany and Belgium.

Our study also has wider implications for the study of historical migration processes, family life and family ideology, modernization and technology. The results of the spatial analysis confirm that the long-term evolution of migration distances in the nineteenth and twentieth centuries is difficult to link to specific processes and events, such as the increase in rail travel, the introduction of the bicycle and economic development. The regularity of the increase in distance travelled by couples therefore indicates the existence of much more complex processes, for which a large number of factors are responsible and which are difficult to control or validate. Instead of a definitive explanation, the best we can do at present is to suggest how the relevant variables might have operated (Rosental, 1999: pp. 18–19).

Our study provides some new perspectives on issues that have been studied by social scientists and historians and that had previously been based on rather restricted and scattered sources. Our study has demonstrated that, judging by the information on marriage horizons, there were strong differences in the geographical range of activities and in geographical knowledge of different social classes until at least the second decade of the twentieth century, with higher social classes having the widest horizons and agricultural workers and farmers the narrowest. From the last decade of the nineteenth century, these horizons increased markedly over time. Our study also showed that in the nineteenth and early twentieth centuries the Netherlands still contained a series of strongly isolated regional entities; the distance between the south and the north of the country proved to be large not only in geographical terms, but also from a cultural and social perspective. At the same time, contacts with the core area of the Netherlands, the three large cities in the west, was frequent in all the provinces included in the study. The process of what Giddens (1990: pp. 100–103) calls ‘disembedding’, the withdrawal from the local social contexts of interaction, a process crucial to modernity, can on the basis of our study also be dated much more precisely; in the first half of the nineteenth century, this process had hardly made any progress in the eastern and southern provinces, but was more advanced in the coastal regions. From the fourth quarter of the nineteenth century the eastern and southern provinces took the lead in this process. As far as the modernization of family life is concerned, based on the choice of spouses, it could be argued that by marrying within one’s own community, the degree of contact between generations remained very strong in the eastern and southern provinces until at least the beginning of the twentieth century. This phenomenon was less common in Zeeland and Groningen. This conclusion is in line with the vast body of more qualitative literature that has been published on the differences in family ideology, family structure and family relations between the western and northern part of the Netherlands (the coastal areas) and the south and the east (Wichers, 1965;

Lesthaeghe & Wilson, 1986). In the latter regions, households were larger and of a more complex structure (Bulder, 1993), more importance was attached to the family, and solidarity between generations and parental authority was strongly stressed.

We also think that the results of our study might function as a useful addition to recent scholarship on the history of technology and infrastructure in the Netherlands. These studies postulate changes in the living conditions of the population as a consequence of the new technologies, but very few authors have been able to describe in detail how the experiences, attitudes and worldview of the population changed (Van der Woud, 2006: pp. 11–19). How the new technologies and infrastructure changed the geographical knowledge of the average citizen and whether or not this change led to a massive increase in knowledge of – and closer personal relations with – people from other cities and regions, can at least in part be answered with the data presented here (Van der Woud, 2006: p. 135, pp. 327–328).

Acknowledgement We wish to thank the *Gelders Archief* (province of Gelderland), the *Groninger Archieven* (province of Groningen), the *Rijksarchief Limburg* (province of Limburg), the *Historisch Centrum Overijssel* (province of Overijssel) and the *Zeeuws Archief* (province of Zeeland) for putting their data at our disposal.

Notes

1. *NRC Handelsblad* 21 May 1994 and 7 February 2004; Zeegers, 1998.
2. The distances between marriage partners of course also determine the genetic structure of communities and the topic has therefore attracted the attention of physical anthropologists and geneticists as well.
3. Geographic endogamy has been the subject of many historical studies in the Netherlands. Those focusing on the nineteenth century (Rutten, 1989; Peeters, 1967; Boekholt, 1979; Boekholt, 1981; Meurkens, 1984; Van Poppel, 1994; Van Poppel & Ekamper, 2005 and Kok, 1998) have for the most part been based on small, mostly strictly local samples.
4. Lesger (2006) convincingly showed how important easy access by water was for the choice of spouses in seventeenth century Holland.
5. The historical description is based on Directorate General for Passenger Transport (1999). We are indebted to Frank Veraart (Department of Technology Management, Eindhoven University of Technology) for sharing this and other information with us.
6. The total number of administrative units mentioned in the database for the province of Gelderland (including all spelling variants) was 12,963, for Groningen 10,154, for Limburg 23,761, for Overijssel 8,450 and for Zeeland 7,477, giving a total (including double counting) of 62,805. The largest number of spelling variants (86) was found for the Belgian municipality of Sint-Gillis-Waas in East Flanders.
7. For Gelderland, Groningen, Limburg and Overijssel one or more neighbouring German Länder were involved, for Limburg and Zeeland the whole of Belgium. Distances to places in France, the UK and other countries were not calculated.

8. For the Dutch population, the percentage never married females and males among cohorts born between 1800 and 1900 varied between 14 and 16 for women and between 11 and 14 for males (Frinking & van Poppel, 1979: pp. 77–85).

Annex: Geographic Coordinates Databases

- Topographical Service of the Dutch Land Registry Office (Topografische Dienst Kadaster):
  - Official Netherlands National Coordinate System points (Rijksdriehoeks metingpunten or RD-points) [www.rdnap.nl]
  - Topographic Names Register (Topografisch Namenregister) [http://www.kadaster.nl/namenregister/]
- Meertens Institute
  - Kloeke-tabel [http://www.meertens.knaw.nl/projecten/mand/CARTkartografiefie.html]
- National Geospatial-Intelligence Agency (NGA)
  - GEOnet Names Server (GNS) [http://earth-info.nga.mil/gns/html/]
    - Belgium [http://earth-info.nga.mil/gns/html/cntyfile/be.zip].
    - Germany [http://earth-info.nga.mil/gns/html/cntyfile/gm.zip]
    - Netherlands [http://earth-info.nga.mil/gns/html/cntyfile/nl.zip]

References

Anselin, L. (1988). *Spatial econometrics. Methods and models*. Dordrecht: Kluwer Academic.

Anselin, L. (2004). *GeoDa 0.95i release notes*. Urbana, IL: Center for Spatially Integrated Social Science.

Anselin, L. (2005). *Exploring spatial data with GeoDa: A workbook*. Urbana, IL: Center for Spatially Integrated Social Science.

Anselin, L. & Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ulah, D. E. Giles, (Eds.), *Handbook of applied economic statistics* (pp. 237–289). New York: Marcel Dekker.

Anselin, L., Syabri, I., & Kho, Y. (2004). GeoDa: An introduction to spatial data analysis. *Geographical Analysis*, *38*, 5–22.

Blanchet, D. & Kessler, D. (1992). La mobilité géographique, de la naissance au marriage. In J. Dupâquier, D. Kessler, (Eds.), *La société française au XIXe siècle, Tradition, transition, transformations* (pp. 343–378). Paris: Fayard.

Boekholt, P. Th. F. M. (1979). Duitsers in Drenthe. *Driemaandelijkse Bladen*, *31*, 43–68.

Boekholt, P. Th. F. M. (1981). Grenscontacten. Huwelijksbetrekkingen over en weer van de Nederlands-Duitse grens. *Driemaandelijkse Bladen*, *33*, 54–66.

Bonneuil, N. (1992). Démographie de la nuptialité au XIXe siècle. In J. Dupâquier, D. Kessler, (Eds.), *La société française au XIXe siècle, tradition, transition, transformations* (pp. 83–120). Paris: Fayard.

Bonvalet, C. & Maison, D. (2007). The family circle: Spatial proximity and personal affinity. In C. Bonvalet, A. Gotman, Y. Grafmeyer, (Eds.), *Family, kinship and place in France (*pp. 27–68). London: Southern Universities Press.

Bulder, E. (1993). *The social economics of old age: Strategies to maintain income in later life in the Netherlands 1880–1940*. Amsterdam: Thesis Publishers.

Coleman, D. A. (1979). A study of the spatial aspects of partner choice from a human biological viewpoint. *Man, New Series*, *14*, 414–435.

Coleman, D. A. & Haskey, J. C. (1986). Marital distance and its geographical orientation in England and Wales, 1979. *Transactions of the Institute of British Geographers, New Series*, *11*, 337–355.

De Bruijne, A., Van Buren, J., Kösters, A., & Van der Marel, H. (2005). *De geodetische referentiestelsels van Nederland. Definitie en vastlegging van ETRS89, RD en NAP en hun onderlinge relaties*. Delft: Netherlands Geodetic Commission.

De Jager, J. L. (1981). *Volksgebruiken in Nederland. Een nieuwe kijk op tradities*. Utrecht: Het Spectrum.

De Vos, L. (1984). De smeltkroes. De Belgische krijgsmacht als natievormende factor, 1830–1885. *Belgisch Tijdschrift voor Nieuwste Geschiedenis- Revue Belge d'histoire Contemporaine*, *15*, 421–460.

De Vries, J. (1978). Barges and capitalism: Passenger transportation in the Dutch economy (1632–1839). *AAG-Bijdragen*, *21*, 33–354.

De Wit, W. O. (1993). Telegrafie en telefonie. In H. W. Lintsen, et al. (Eds.), *Geschiedenis van de techniek in Nederland: De wording van een moderne samenleving 1800–1890*: Vol. 4. Delfstoffen, machine- en scheepsbouw; stoom; chemie; telegrafie en telefonie (pp. 273–298). Zutphen: Walburg Pers.

De Wit, W. O. (2002). Het communitiecatielandschap in de twintigste eeuw: De materiële basis. In J. W. Schot, et al. (Eds.), *Techniek in Nederland in de twintigste eeuw*: Vol. 5. Transport, communicatie (pp. 161–178). Zutphen: Walburg Pers.

Directorate-General for Passenger Transport. (1999). The Dutch bicycle master plan. Description and evaluation in an historical context. The Hague: Ministry of Transport, Public Works and Water Management.

Ebdon, D. (1988). *Statistics in Geography* (second edition with corrections). Oxford: Blackwell.

Frinking, G. A. B. & van Poppel, F. W. A. (1979). *Een sociaal-demografische analyse van de huwelijkssluiting in Nederland*. The Hague: Staatsuitgeverij.

Giddens, A. (1990). *The Consequences of Modernity*. Cambridge: Polity Press.

Horsten, F. H. (2005). *Doorgaande wegen in Nederland, 16e tot 19e eeuw: Een historische wegenatlas*. Amsterdam: Aksant.

Johnston, R. J. & Perry, P. J. (1972). Déviation directionnelle dans les aires de contact. Deux exemples de relations matrimoniales dans la France rurale du XIXe siècle. *Etudes rurales*, *46*, 23–33.

Jonckers Nieboer, J. H. (1938). *Geschiedenis der Nederlandsche spoorwegen: 1832–1938*. Rotterdam: Nijgh & Van Ditmar.

Kalmijn, M. (1991). From family origins to individual destinations: the changing nature of homogamy in the United States. University of California, LA: Ph.D. dissertation.

Knippenberg, H. & de Pater, B. (1988). *De eenwording van Nederland. Schaalvergroting en integratie sinds 1800*. Nijmegen: SUN.

Kok, J. (1998). 'Vrijt daar je zijt'; Huwelijk en partnerkeuze in Zeeland tussen 1830 en 1950. *Zeeland*, *7*, 131–143.

Kuipers, G. (1976). *Volksleven in het oude Drenthe*. The Hague: Boekencentrum.

Leboutte, R. & Hélin, E. (1986). Le choix du conjoint: À propos de l'usage d'indicateurs quantifiables pour apprécier l'endogamie. In *Chaire Quetelet. Au-delà du quantitatif. Espoirs et limites de l'analyse qualitative en démographie* (pp. 415–465). Louvain-La-Neuve: CIACO Editeur.

Lesger, C. (2006). Informatiestromen en de herkomstgebieden van migranten in de Nederlanden inde Vroegmoderne Tijd. *Tijdschrift voor Sociale en Economische Geschiedenis*, *3*, 3–23.

Lesthaeghe, R. & Wilson, C. (1986). Modes of production, secularization and the pace of the fertility decline in Western Europe, 1870–1930. In A. J. Coale, S. C. Watkins, (Eds.), *The decline of fertility in Europe* (pp. 261–292). Princeton, NJ: Princeton University Press.

Levine, N. (2007). *CrimeStat: A spatial statistics program for the analysis of crime incident locations (v 3.1)*. Houston, TX; Washington, DC: Ned Levine & Associates, the National Institute of Justice.

Meurkens, P. C. G. M. (1984). *Sociale verandering in het oude Kempenland (1840–1910), Demografie, economie en cultuur van een preindustriële samenleving*. Bergeijk: Stichting Eicha.

Millard, J. (1982). A new approach to the study of marriage horizons. *Local Population Studies*, *28*, 10–31.

Mom, G. P. A., Schot, J. W., & Staal, P. E. (2002). Werken aan mobiliteit: De inburgering van de auto. In J. W. Schot, et al. (Eds.), *Techniek in Nederland in de twintigste eeuw*: Vol. 5. Transport, communicatie (pp. 45–74). Zutphen: Walburg Pers.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*, 17–23.

Morel, A. (1972). L'espace social d'un village picard. *Etudes Rurales*, *45*, 62–80.

Morrill, R. L. & Pitts, F. R. (1967). Marriage, migration and the mean information field: A study in uniqueness and generality. *Annals of the Association of American Geographers*, *57*, 401–422.

Ogden, P. (1980). Migration, marriage and the collapse of traditional peasant society in France. In P. White, R. Woods, (Eds.), *The Geographical Impact of Migration* (pp. 152–179). London: Longman.

Peeters, K. C. (1967). Trouwen in de Brabantse grensstreek. De resultaten van een dubbele enquête. *Brabants Heem*, *19*, 225–234.

Perry, P. J. (1969). Working-class isolation and mobility in rural Dorset, 1837–1936: A study of marriage distances. *Transactions of the Institute of British Geographers*, *46*, 121–141.

Pooley, C. & Turnbull, J. (1998). *Migration and mobility in Britain since the 18th century*. London: University College London Press.

Prost, A. (1981). Mariage, jeunesse et société à Orléans en 1911. *Annales ESC*, *36*, 673–701.

Renard, J. P. (1984). De l'intérêt du dépouillement d'actes de mariage dans le cadre général de l'étude des limites et frontières. *Espace Populations Sociétés*, *1*, 125–130.

Rosental, P. A. (1999). *Les sentiers invisibles. Espaces, familles et migrations dans la France du 19ème siècle*. Paris: Editions de l'école des hautes études en sciences sociales.

Rosental, P. A. (2004). La migration des femmes (et des hommes) en France au XIXe siècle. *Annales de Démographie Historique*, *1*, 107–136.

Rutten, W. J. M. J. (1989). Huwelijk en partnerkeuze in een grensstad (circa 1830–1910). Verschuivingen in de geografische herkomst van huwelijkspartners te Maastricht vóór en ná de scheiding van de beide Limburgen. In *Eenheid en scheiding van de beide Limburgen* (pp. 95–124). Leeuwarden-Maastricht: Eisma.

Schot, J. W. (2002). De mobiliteitsexplosie in de twintigste eeuw. In J. W. Schot, et al. (Eds.), *Techniek in Nederland in de twintigste eeuw*: Vol. 5. Transport, communicatie (pp. 13–18). Zutphen: Walburg Pers.

Segalen, M. (1979). Mating in French industrial rural areas. In J. Cuisenier, (Ed.), *Europe as a cultural area* (pp. 147–158). Paris: Mouton.

Snell, K. D. M. (2002). English rural societies and geographical marital endogamy, 1700–1837. *Economic History Review*, *55*, 262–298.

Thurkow, A. J., Harten, J. D. H., Knippenberg, H., Prins, L., Renes, J., Schuyf, J., et al. (1984). *Bewoningsgeschiedenis. Atlas van Nederland*. 's-Gravenhage: Deel 2. Staatsuitgeverij.

Van Leeuwen, M. H. D. & Maas, I. (2005). Endogamy and social class in history: An overview. In M. H. D. van Leeuwen, I. Maas, A. Miles, (Eds.), *Marriage choices and class boundaries: Social endogamy in history*: Vol. 50. *International Review of Social History* (pp. 1–24). Cambridge: Cambridge University Press.

Van Leeuwen, M. H. D., Maas, I., & Miles, A. (2002). *HISCO: Historical international standard classification of occupations*. Leuven: Leuven University Press.

Van Poppel, F. (1994). Verbreding van de horizon? Veranderingen in de geografische herkomst van huwelijkspartners. *Acta Geographica Lovaniensia*, *34*, 79–88.
Van Poppel, F. & Ekamper, P. (2005). De Goudse horizon verruimd. Veranderingen in de herkomst van Goudse bruiden en bruidegoms. In J. Kok, M. H. D. van Leeuwen, (Eds.), *Genegenheid en gelegenheid. Twee eeuwen partnerkeuze en huwelijk (*pp. 181–212). Amsterdam: Aksant.
Van de Putte, B. (2003). Homogamy by geographical origin: Segregation in nineteenth-century Flemish cities (Gent, Leuven, and Aalst). *Journal of Family History*, *28*, 364–390.
Van den Berg, P. W. J. (1941). *Het karakter der plattelandssamenleving*. Assen: Van Gorcum.
Van der Molen, S. J. (1961). *Levend volksleven. Een eigentijdse volkskunde van Nederland*. Assen: Van Gorcum.
Van der Woud, A. (2006). *Een nieuwe wereld. Het ontstaan van het moderne Nederland*. Amsterdam: Uitgeverij Bert Bakker.
Van Zanden, J. L. & van Riel, A. (2004). *The strictures of inheritance: The Dutch economy in the nineteenth century*. Princeton, NJ: Princeton University Press.
Vandenbroeke, C. (1986). *Vrijen en trouwen. Van de Middeleeuwen tot heden. Seks, liefde en huwelijk in historisch perspectief*. Amsterdam: Elsevier.
Veenendaal, A. J., Jr. (1993). Spoorwegen. In H. W. Lintsen, et al. (Eds.), *Geschiedenis van de techniek in Nederland: De wording van een moderne samenleving 1800–1890*: Vol. 2. Gezondheid en openbare hygiëne, waterstaat en infrastructuur, papier, druk en communicatie (pp. 129–164). Zutphen: Walburg Pers.
Watkins, S. C. (1991). *From provinces into nations. Demographic Integration in Western Europe, 1870–1960*. Princeton, NJ: Princeton University Press.
Wichers, A. J. (1965). *De oude plattelandsbeschaving: Een sociologische bewustwording van de 'overherigheid'*. Assen: Van Gorcum.
Zeegers, W. (1998). *Op zoek naar een kanjer: Contactadvertenties in Nederland*. Amsterdam: Uitgeverij Bert Bakker.

Chapter 7
Finding Frontiers in the U.S. Great Plains from the End of the Civil War to the Eve of the Great Depression

Myron P. Gutmann, Glenn D. Deane, and Kristine Witkowski

In an earlier example of spatial analysis of long-term historical processes, Gutmann and Sample (1995) examined the factors that drove the settlement and development of the rural Texas population from the mid-nineteenth century until the recent past. In doing so, they went beyond classic formulations by Turner (1893), Webb (1931, 1957) and Worster (1987) to show that the generally acknowledged east-to-west settlement of the U.S. was also a process shaped by the availability of resources, not just water for farming but also other natural resources (oil, for example), and eventually the development of the transportation infrastructure.

This chapter takes the fundamental questions and findings in earlier work and extends them to the whole natural region of the Great Plains of the U.S. Characterized by semi-arid and arid grasslands, the Great Plains region was settled between the 1860s and the 1920s. Defining settlement as the process by which each of the roughly 500 counties in the region reached a specific population density threshold, we hypothesize that the most important forces at work were a combination of structural attributes of the national process (population moving from east to west), climate (precipitation, temperature), other resources inside the region (irrigation, transportation, energy, employment in industry) and developments outside the region, such as the need to supply food to gold and silver miners working in the mountainous region to the west (Fite, 1966).

The data and cartography for this paper are drawn from the Great Plains Population and Environment project's databases (Gutmann, 2005a, b). They include county-level demographic, agricultural and environmental information about all counties in ten Great Plains states from 1870 until the present. Our outcome of interest is whether and when population density exceeded a threshold of four persons per square mile.

M.P. Gutmann (✉)
Inter-university Consortium for Political and Social Research, University of Michigan, Ann Arbor, MI 48106-1248, USA
e-mail: gutmann@umich.edu

M.P. Gutmann et al. (eds.), *Navigating Time and Space in Population Studies*, International Studies in Population 9, DOI 10.1007/978-94-007-0068-0_7,

Our analytic method is a proportional hazards (Cox regression) model for repeated events (Allison, 1995). Prior to including theoretically relevant indicators of geophysical conditions and developmental potential, we explore competing baseline models of spatial expansion, testing the explanatory power of longitude and latitude coordinates of county centroids, the distance from a point of origin along the eastern fringe of the Great Plains, and measures of temperature, precipitation and elevation. This method offers significant opportunities for more sophisticated analysis of spatial population change than other methods and our analysis shows how valuable this approach can be.

One of the major challenges of this research and an important contribution of this chapter derives from the fact that county boundaries in the region changed significantly prior to 1920, making many types of conventional econometric approaches, including standard discrete time event history analysis, inappropriate. We solve the problem of new counties coming into risk of settlement and their dependence on precursor counties via the same structure that one uses in an event history analysis with repeated events. Newly formed counties are at risk of settlement, but that risk is correlated with factors associated with the settlement risk of their precursor counties. Put simply, our method clusters just-founded and precursor counties into a hierarchy of strata and then weights those strata as part of the final analysis. This novel approach simultaneously eliminates the concern for spatially autocorrelated error and the rationale for relying on time-invariant county cluster boundaries, which define an inappropriate spatial scale as the unit of analysis.

Our unconventional analysis leads to valuable new results about the process of settlement, as well as to valuable methodological insights, both about how to work with changing county boundaries and in the form of innovative statistical analyses. We do not have the detailed data about resources and transportation for this larger region that Gutmann and Sample had when they analysed population change in Texas, which imposes some limits on what we can do. Nevertheless, we draw important conclusions. Population did move from east to west, but its movement was definitely shaped by the availability of natural resources. Driven largely by agricultural concerns, people preferred areas where there was more moisture, cooler temperatures and lower elevation, but they also followed opportunities for employment in industry and mining.

The Region and Its Settlement

This chapter is about the Great Plains of the United States (see Fig. 7.1). The Great Plains region, as we define it, occupies a substantial portion of ten states in the west-central U.S. and is currently made up of approximately 450 counties (Gutmann, Parton, Cunfer, & Burke, 2005; Cunfer, 2005). We define its boundary on the east as roughly the point where average rainfall is less than 700 mm per year and on the west at the point where counties generally have an elevation greater than 1,500 m. On the south, we bound the Great Plains at the southern end of the Texas Panhandle,

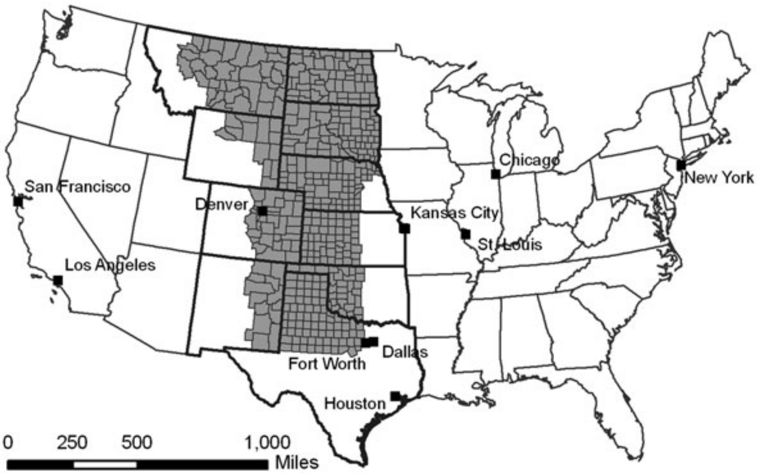


Fig. 7.1 The 10-state great plains region in context

where the high plains end, and on the north at the Canadian border. This set of environmental limits gives us a region made up of flat and rolling terrain that rises from east to west and is semi-arid and arid. Its natural vegetation is short and mixed-grass steppe, with some tall-grass steppe in the east. Temperature, precipitation and elevation vary significantly across the region, as shown in Figures 7.2, 7.3, and 7.4.

Prior to the middle of the nineteenth century, the Great Plains region had a small native North American population, along with a tiny European-origin population engaged in trading and trapping. Beginning just before the U.S. Civil War of the 1860s, European-origin settlers came to the region in ever-larger numbers, attracted by opportunities for mining and for agricultural development and encouraged by the passing of the 1862 Homestead Act, which helped settlers acquire land at little cost. Figure 7.5 illustrates the time-frame within which an agriculturally based and European-origin population came to the region. The figure shows the year that population density in each county reached four persons per square mile, the settlement threshold that we use in the statistical analysis in this chapter.

What Figure 7.5 shows is that by 1880–1890, when the first detailed information about the region is readily available from Census sources, a few counties in the east had already reached a density of four persons per square mile. The population subsequently spread westward, with the most rapid expansion along a line west through northern Kansas and southern Nebraska to Denver and the Colorado section of the Front Range of the Rocky Mountains. Oklahoma, which was not opened for settlement until the 1890s, was one area of retarded development, not reaching post-frontier densities until 1900. In addition, a number of counties never reached post-settlement densities during the period described in this chapter, 1880–1940, probably because their agriculture didn't require higher population densities and their location did not attract other economic activities that would have produced larger populations.

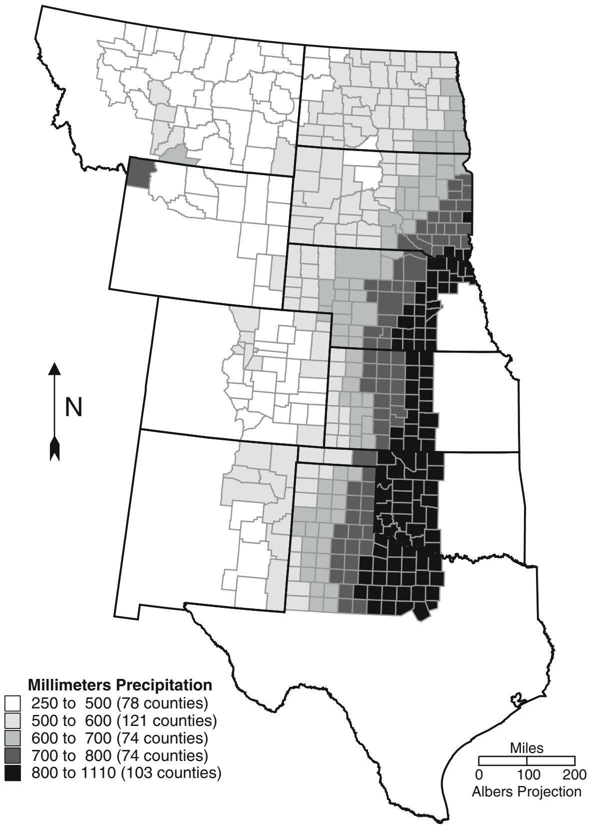


Fig. 7.2 Average annual precipitation, 1961–1990

Figure 7.6 shows how the process of settlement played out by 1940. Except for urban areas, little of the region was densely populated in that year. A few counties along the eastern edge of the region, plus counties in the Texas panhandle and along the Front Range, had population densities greater than 15 persons per square mile. Everywhere else, densities were still quite low in 1940 and in most places would not increase afterwards. In a significant number of counties, densities were very low: areas coloured white had densities ranging from one-half to four persons per square mile. We can put this in context by asking how much land a traditional six-person family would farm at these densities, if population were evenly distributed. At one-half person per square mile, this is twelve square miles per family (48 Homestead-Act farms of 160 acres); at four persons per square mile, it is 1.5 square miles, or six of those farms.

How and why did this pattern of settlement come about? That's the question we ask in this chapter. The maps of settlement and population density suggest that the story is more nuanced than simply saying that settlement moved from east to west. A straightforward diffusion story would be consistent with the history of the

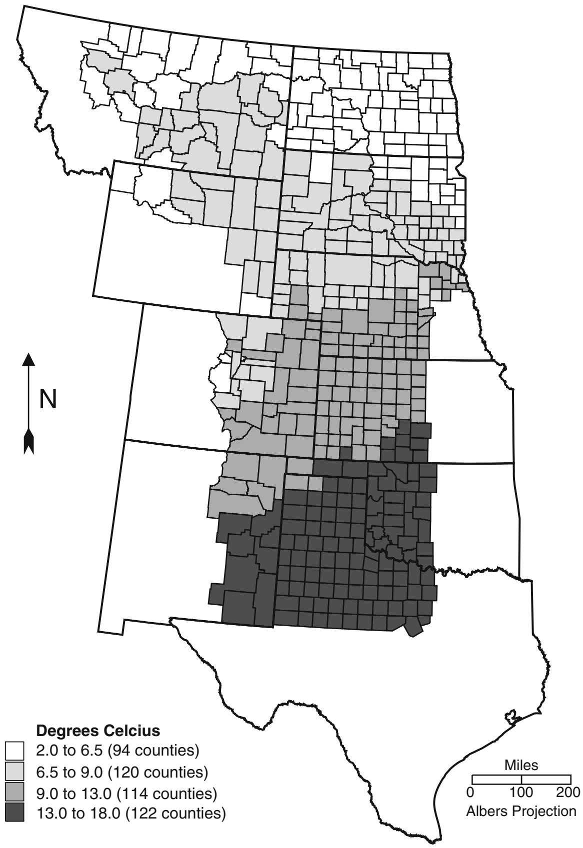
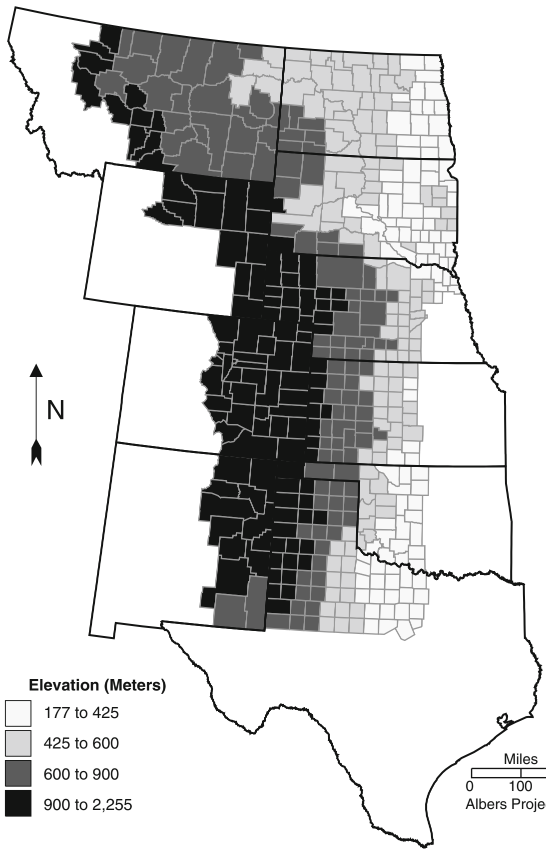


Fig. 7.3 Average annual temperature, 1961–1990

United States, with the technology of the era and with the notion shared by many geographers. But it is not at all clear from Figures 7.5 and 7.6 that the settlement process of the Great Plains was a simple diffusion process, where people slowly moved from areas of higher population density in the east to areas of lower population density in the west. The maps reveal that certain areas in the west quickly became attractive destinations, especially agriculturally viable areas in the eastern plains and more generally in Nebraska and Kansas and the communications and mining hub around Denver. The picture everywhere else is mixed.

This description of a more nuanced set of explanations is in keeping with the findings in Gutmann and Sample (1995), which builds on Webb's (1931, 1957) work to emphasize the role of precipitation, transport and mineral resources in the settlement of Texas. Most of the process of historical settlement was driven by the resources provided by the natural environment, as shaped by a social and political world, rather than being driven mainly by economic processes. In this finding we differ, for example, from Harley (1978), who argues that the most important determinant of settlement was the local price of wheat.

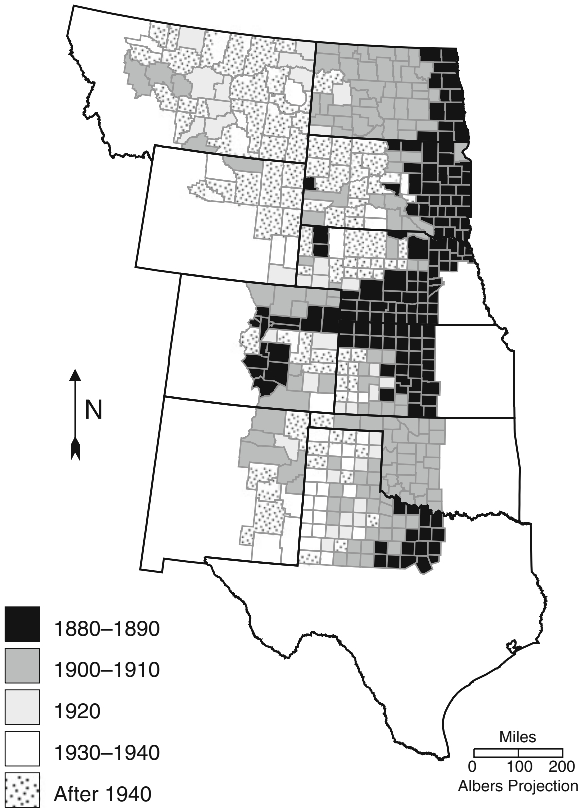
Fig. 7.4 Minimum county elevation, 1940 boundaries



In order to understand the settlement process, this chapter examines six hypotheses:

1. Settlement followed a spatial expansion process from east to west. We also consider a number of variations on this theme to assess the directionality of settlement.
2. Settlement radiated from a point of origin, a gateway or central node, on the eastern fringe, so that areas farther away, whether to the west, north or south, settled later.
3. Settlement was more rapid in areas that were well-suited to crop-based agriculture, with either optimal precipitation, suitable temperature, moderate elevation, access to irrigation or relatively good soils. This model of geophysical determinants is based on Gutmann, Pullum-Piñon, Baker and Burke (2004) and Gutmann, Deane, Lauster and Peri (2005).

Fig. 7.5 Year population density reached 4 persons/square mile



4. Settlement was more rapid in areas where transportation was more available, which, because of the lack of navigable water, we associate primarily with the development of the railroad network, as indicated by the size of the population employed in transportation in 1930.
5. Settlement was more rapid where there were extractable natural resources, as indicated by the proportion of the population employed in mining in 1930.
6. Finally, settlement was more rapid where other aspects of economic development were most advanced, as indicated by the proportion of the population employed in manufacturing in 1930.

We define settlement in our analysis as a county reaching a threshold population density of four persons per square mile. We have evaluated various alternatives for this analysis and choose a density of four because it gives us the best gauge of when counties reach a significant level of population. A level of two persons per square mile has resonance as the threshold established by Turner (1893) in his essay on the closing of the frontier and also allows more counties to reach the threshold during the time period we are studying. Still, it has the disadvantage of being so low

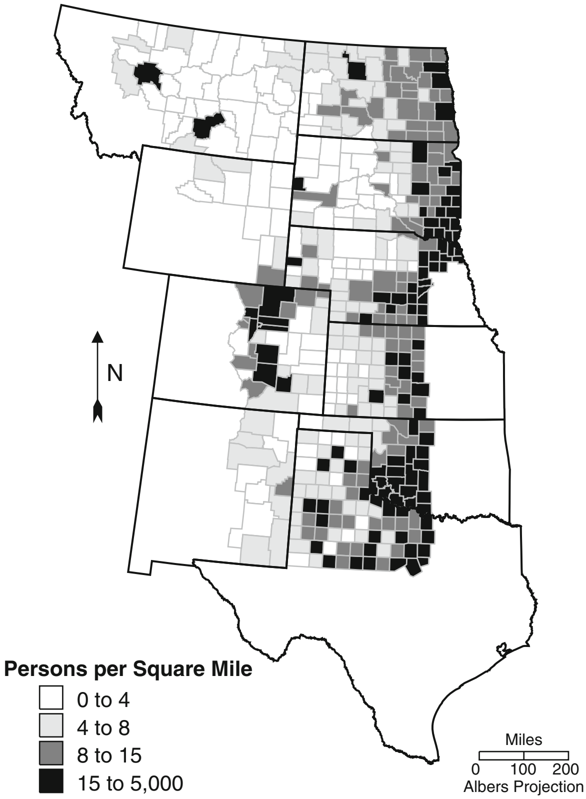


Fig. 7.6 Population density, 1940

that it is difficult to imagine that counties with such a small population density had viable populations. Moreover, such a large number of counties reached a density of two before we are able to begin our analysis with the 1880–1890 time period, that our analysis would have been impossible. Four persons per square mile, on the other hand, is low enough to be achievable even in sparsely settled areas while still large enough to give us something to measure. We tried much of our analysis with alternative densities and the results are substantially the same for all specifications of the dependent variable between two and ten persons per square mile.

We detail our approach to this analysis in the sections that follow. We begin with a discussion of how difficult it is to study settlement processes over time for individual counties because their boundaries changed. After that we describe our data and methods and then turn to results. The first part of our results deals with choosing a single baseline model, because several alternate specifications of the baseline exist. The second part explains the models that evaluate our hypotheses. We conclude with the implications of our methods and results for future research on spatial and temporal processes in historical demography.

The Challenge of Boundary Changes

County-level analyses that span time are plagued by problems of boundary comparability because counties split or merge, or new counties are founded. We show an example of how this kind of change happens in Figure 7.7 with a map of the county boundaries of Montana from 1880 to 1920. Just looking at the northeast corner of the state we can see that what is one county in 1880 is two counties by 1910 and ten counties in 1920. Moreover, in 1920, three of those ten counties spill over into territory that was part of two other precursor counties as recently as 1910. The typical solution to boundary changes in longitudinal analyses is to group counties involved in any kind of shared boundary shifts into county clusters (Horan & Hargis, 1995). In the case we've just mentioned, we would have wound up combining twenty or more counties into a single unit in order to overcome the challenges of county divisions and mergers.

There are two problems with this solution. First, it would result in a grossly inappropriate spatial scale for our research question. For example, in any given snapshot in time (the cross-sectional perspective), if the population is not randomly distributed across space, population densities will be dependent on the degree of

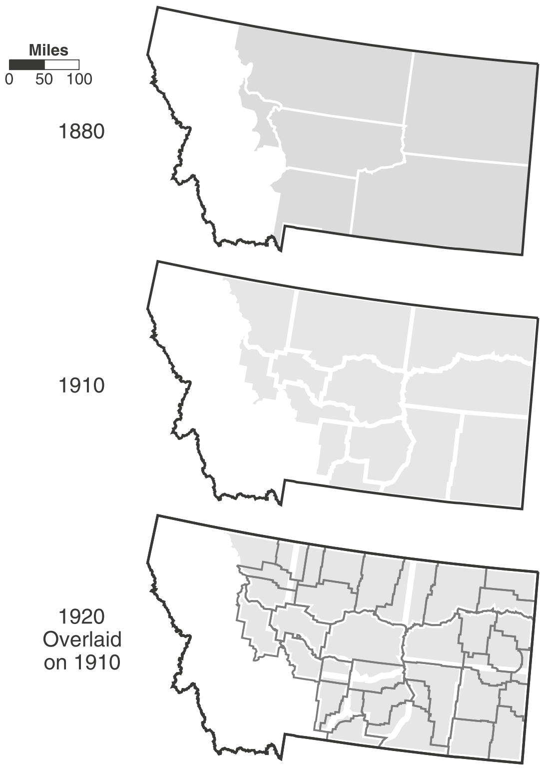


Fig. 7.7 Montana county boundaries, 1880–1920

spatial aggregation. The same is true of the independent variables. The coarser the spatial resolution, the more difficult it is to judge the importance of differences in precipitation, or the access to transportation networks, or the proportion of the population employed in manufacturing. Counties are a natural unit of analysis because they are meaningful geopolitical units; county clusters are not.

The second problem is that counties themselves are endogenous to the very issue we are attempting to measure. In the years from 1880 through 1940 the number of counties in the Great Plains, as well as their sizes and boundaries, changed substantially as new counties were created to reflect growing population and local political pressures (Gutmann, Pullum-Piñon, Cunfer, & Hagen, 1998; Gutmann, Deane, Lauster, & Peri, 2005). The particular issue of interest to us derives from the need to know characteristics of counties for our analysis for a time that might be prior to their formation. If, for example, County A exists in 1890 but is divided into two counties (County A and County B) before 1910, certain attributes of County B (for example the date when it reached a threshold population density) would be wrongly assigned if we assumed that there was no population in the area prior to 1910. Our solution is to use the settlement time of the precursor county (in this case County A) if it has already crossed the population threshold before County B is created and if County B appears to have crossed the threshold at the time of its creation. We also have the potential to draw information from precursor counties for independent variables as a way to understand the attributes of successor counties, without having to undertake unnecessary aggregation. This has implications for our statistical analysis, which we will describe later.

We show another example in Figure 7.8, which presents county boundary changes for northeast Colorado between 1880 and 1910. An area that begins with two counties (Weld and Arapahoe) in 1880 becomes eleven counties in 1910. Moreover, counties don't merely subdivide. While most of the ultimate counties are produced by division, Washington and Yuma counties include sections of both Weld and Arapahoe. Finally, the whole process is strongly shaped by the presence of the city of Denver (Colorado's capital) at the western end of Arapahoe County in 1880, which largely determines that county's density in every year.

Table 7.1 shows the impact of this history on our assessment of when settlement occurred in these counties. The top panel in Table 7.1 shows the population density of each county in each year, as well as the overall population density ('Total') for all counties together. The lower panel shows the year that an analyst could determine that each county had reached the threshold of four persons per square mile. The first column is the analysis used in this chapter. The second column is the result of relying on current-year population, which would mean that Adams and Denver counties would only be counted as reaching the threshold in 1910 (the year they are formed), rather than in 1880, the year their precursor county (Arapahoe) reached that density level.[1] The last column shows that if all the counties were aggregated as a single entity, the 'settlement' date would be 1890, reflecting the fact that the area as a whole had a population density of 10.2 in 1890.

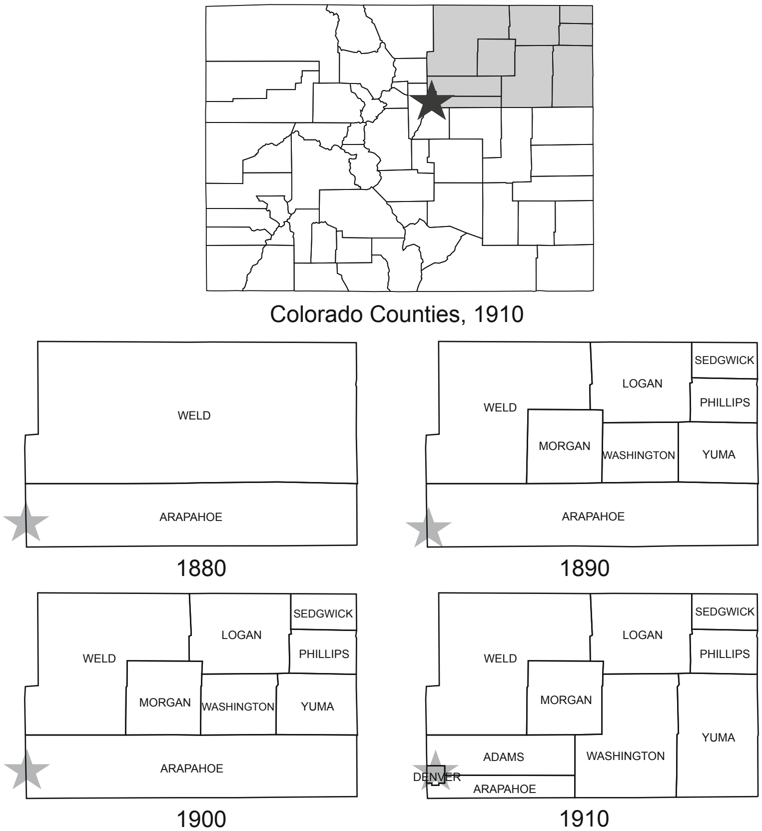


Fig. 7.8 County boundaries, Northeast Colorado: 1880–1910. Note that what were two counties in 1880 became eleven counties in 1910. *Shaded area* in upper map is area represented in four detail maps. *Star* represents location of Denver, capital of Colorado

Data

The data used in this chapter reflect the social, demographic, economic, agricultural and environmental attributes of the Great Plains from 1880 to 1940.[2] The economic, demographic and agricultural data are all drawn from the U.S. Censuses of Population, Manufacturing, Mining and Agriculture and their compilation has been published as part of the Great Plains Population and Environment Project (Gutmann, 2005a, b).[3]

We have already described the rationale for our choice of dependent variable. Because we are visualizing the settlement process through the lens of an event-history analysis, we have constructed our dependent variable as the first decade when the population density in each county exceeds a settlement threshold.

Table 7.1 Population density and settlement timing in Northeast Colorado, 1880–1920

	1880	1890	1900	1910	1920
A. Population densities, northeast Colorado counties					
Adams				6.6	10.7
Arapahoe	8.0	27.1	31.6	13.7	18.4
Denver				2,771.2	3,331.1
Logan		1.7	1.8	5.3	10.2
Morgan		1.2	2.5	7.5	12.6
Phillips		3.9	2.3	4.7	8.1
Sedgwick		2.4	1.8	5.7	7.8
Washington		2.2	1.2	2.3	4.4
Weld	0.5	2.9	4.2	9.7	13.4
Yuma		2.3	1.5	3.7	6.0
Total	2.9	10.2	11.8	20.2	26.5

	This paper	Current	Clustered
B. Year population density reached four persons per square mile, under three different assumptions (see text for explanation)			
Adams	1880	1910	1890
Arapahoe	1880	1880	1890
Denver	1880	1910	1890
Logan	1910	1910	1890
Morgan	1910	1910	1890
Phillips	1910	1910	1890
Sedgwick	1910	1910	1890
Washington	1920	1920	1890
Weld	1900	1900	1890
Yuma	1920	1920	1890

The environmental data were collected as part of the same project, but come from other sources. Our climate history data are drawn from the VEMAP database (Kittel, Rosenbloom, Painter, Schimel, & VEMAP Modelling Participants, 1995; 1997), which divides the U.S. into a half-degree by half-degree grid and assigns daily or monthly weather characteristics to each of the cells of the grid. We have averaged monthly grid values for each county, then summed monthly precipitation values into annual totals and averaged monthly minimum and maximum temperature values into annual averages. The precipitation and temperature variables are long-term averages for the period 1895–1940, which allow us to focus our analysis on the effects of climate, rather than on short-term weather variations. We begin these averages with 1900 because the source data for long-term weather estimates for the United States are rarely available before 1895.

We characterize soil with county-level averages of the depth of topsoil, based on the Statsgo databases (U.S. Dept. of Agriculture, no date), which collect recent soil survey data from the 1970s and 1980s. In general, we expect deep topsoil to lead to more cropping because crops planted on deeper topsoil are more successful and farmers were aware of the benefits of topsoil depth for cropping. If settlement follows cropping, then areas with deep topsoil would have been settled first. To the extent that soils had a temporal impact, areas with shallower soils would have been settled later, which is also in keeping with our hypothesis.

Finally we ascertain each county's lowest level of elevation from the U.S.G.S. Digital Elevation Model Data, as defined by their 1900–1940 boundaries.

The analysis in this chapter makes use of a very limited number of variables drawn from these sources. The variables are listed in Table 7.2 and their role in the models is explained when necessary later in the chapter.

Table 7.2 Variables used in the analysis

Variable	Source	Notes
Population density: 4 or more persons	Population and area from U.S. census (GP data)	1880–1940 observation years (1870 data used to identify left-truncation)
Settlement year: 4 or more persons	Population and area from U.S. census (GP Data)	1880–1940 observation years (1870 data used to identify left-truncation)
Long-Term average temperature	VEMAP data sources, transformed	Average of 1895–1940 observation years
Long-Term average precipitation	VEMAP data sources, transformed	Average of 1895–1940 observation years
Elevation	U.S.G.S. digital elevation model data	Low point in county, based on 1900–1940 boundaries (1880–1890 filled with 1900 data)
Number of metres from Wyandotte County, Kansas	Derived from county maps	Metres from Wyandotte Co, Kansas (adjacent to Kansas city, MO)
Longitude of county centroid (degrees)	Derived from county maps	1880–1940 boundaries
Latitude of county centroid (degrees)	Derived from county maps	1880–1940 boundaries
Depth of topsoil layer	USGS statsgo database	1990 observation year, 1890–1940 boundaries (1880 filled with 1890 data)
Percent of total land area irrigated	U.S. Agricultural Census (GP data)	1890, 1910–1940 observation years (1880 and 1900 filled with extrapolated data from 1890 and 1910)
5% or More of employed in railroad industry in 1930	U.S. Census of Population (GP data)	1930 Observation year (1880–1920 & 1940 filled with 1930 data)
5% or more of employed in mining industry in 1930	U.S. Census of Population (GP data)	1930 Observation year (1880–1920 and 1940 filled with 1930 data)
5% or more of employed in manufacturing industry in 1930	U.S. Census of Population (GP data)	1930 Observation year (1880–1920 and 1940 filled with 1930 data)

Methods

Our point of departure was to build a data file structure in which all Great Plains counties that were at risk of settlement between 1880 and 1940 contribute observations, with duration to settlement as the outcome (and censored observations are those that did not reach the settlement density by 1940). Each of these 'focal' counties (as defined by 1940 boundaries) is clustered with a 'precursor' county (or counties) if that focal county was founded after 1880. We then apply a member of the family of methods for the analysis of multivariate failure time data proposed by Wei et al. (1989). Multivariate failure time data arise when each study subject can potentially experience several events or when there exists some natural or artificial clustering of subjects that induces dependence among the failure times of members of the same cluster. Data in the former situation are referred to as multiple events data. The Wei, Lin and Weissfeld (WLW) method for multiple events data produces robust variance estimates that allow for dependence among multiple event times using the proportional hazards (Cox regression) model. With these variance estimates, we get efficient pooled estimates of the coefficients and their standard errors, allowing us to test a number of relevant hypotheses (Allison, 1995: p. 242). The technique is generally referred to as a marginal Cox, or population-averaged, method. Although there is now a sizeable class of Cox regression methods for multiple events and clustered data, the WLW method is particularly suited to our research question because there is no need to make assumptions about the nature or structure of the dependence among clustered observations.

Put simply, our method clusters newly founded and precursor counties into a hierarchy of strata, where the first stratum is composed of all counties at risk of settlement at a given moment in time, the second stratum contains all immediate or principal precursor counties at that same moment in time, the third stratum contains a second order of precursor counties, the fourth stratum holds a third order of precursor counties and so forth. Next, the effects of covariates on the risk of reaching the settlement threshold in a given stratum are simultaneously estimated with the risk of settlement in all higher order strata. This conditions the estimated effects of covariates for counties in the lower ordered strata of a cluster, thereby parsing their dependence on their precursor counties. Lastly, we form a single set of coefficients, their estimated standard errors and test statistics as a weighted average of their corresponding stratum-specific values, using the number of counties at risk of settlement in each stratum as weights.

We use the overall model fit to address the research questions delineated earlier. Model selection is based on the likelihood fit statistic. The likelihood (L) is the criterion function used by Maximum-Likelihood Estimation (MLE), which selects parameter values by maximizing the probability that the observed data were generated by the hypothesized model. We report the commonly applied transformation $-2^*\log(L_k)$. Two nested models can be compared by the ratio of their likelihoods, the likelihood ratio (LR) $= L_k/L_0$, where L_k is the value of the likelihood for the model with *k* predictors and an intercept and L_0 is the value of the likelihood for the intercept-only model (a model with no predictors); and $2^*\log(L_k/L_0) \sim X^2_k$, with *k* degrees of freedom (DF). The obvious problem with using the likelihood as the criterion for model selection is that it must increase with the number of predictors. All practical model-selection statistics therefore penalize for the number of predictors. Prominent among these is Akaike's Information Criterion (AIC), where $AIC = -2^*\log(L_k) + 2^*k$ (Burnham & Anderson, 2002; Raffalovich, Deane, Armstrong, & Tsao, 2008). We report our goodness of fit comparisons in Table 7.2. Because we are primarily interested in overall goodness of fit, we do not report individual regression coefficients for any but our final models. This strategy also mitigates any concern for multicollinearity among any repetitive covariates in our model specifications.

Two final points concern the proportional hazards assumption of the Cox regression model and an optimal method for handling tied occurrences. The appropriateness of the proportional hazards assumption is evidenced by the absence of time-varying coefficients. To this end, we investigated interactions between covariates and analysis time and found no evidence that the effects of our model covariates changed as a function of exposure to the risk of reaching the settlement threshold. It should also be clear that settlement densities change in continuous time. We can only observe that change in the discrete intervals given to us in the census records, but we find no logical reason to treat ties (the occurrence of two or more events with the same survival time) as discrete time phenomena. Accordingly, we use the exact method for continuous time to allocate ties when estimating the partial likelihoods of our Cox regression models.

The Baseline Model

If, as Turner, Webb and others suggest, the fundamental settlement process for western counties was one of spatial expansion, how do we measure it? The simplest way to see this is as a mechanism by which counties farther from the eastern fringe were settled later than those closer to that starting line. This view leads to a set of trend surface models that are readily specified as members of the family of spatial expansion models that have been applied to a wide range of problems in urban and regional analysis (Anselin, 1988: pp. 125–126).

Trend Surface Models

We begin with a set of marginal Cox regressions that explore the directionality of settlement using 'naïve' locational coordinates (longitude and latitude of county centroids). We explore both isotropic and anisotropic models, wherein the latter is a special case of the former. In isotropic specifications, directionality does not matter. If a pure east to west pattern holds, the risk of settlement should be a function only of longitude. We consider this model as isotropic because north-south adjustments should not improve the model fit. Alternatively, an isotropic specification may use only latitude. Here, we assess north to south patterns of settlement and whether adjustments should be made. If settlement was anisotropic, both longitude and latitude must be modelled. Because a linear trend surface model leaves no room for a loss of momentum in the spatial expansion, it is often an unrealistic model of spatial dependence from a point of origin. Typically the trend surface follows a nonlinear trajectory. Consequently we also specify the risk of settlement as a function of linear coordinates and their squared terms. Although it is reasonable to break the settlement process down into east-west and north-south paths, it is also clear that because Oklahoma was not open to settlement until 1892, it served as something of a barrier to a pure trend surface model of settlement, which assumes that settlement depends only on spatial location and ignores such location-specific socio-political factors. All our models therefore include a dummy variable for counties within Oklahoma, which greatly improves their fit (models without Oklahoma are not reported).

Our naïve spatial expansion models are extremely fine-grained, in that they include both longitude and latitude, along with squared terms and interactions. As such, they capture a multitude of underlying processes. At the same time they offer no clear theoretical underpinning, because they explain only the direction of movement and not reasons for movement in those directions. What makes our task more interesting – and certainly more complex – than it might otherwise be is that a number of natural factors that can enhance or impede settlement are analogous to spatial expansion, as they vary by longitude and latitude. For example, precipitation decreases and elevation increases from east to west, while temperature increases from north to south, but these variables also affect the desirability of a location for agricultural settlers. Our next task here is to explain those alternatives and evaluate them as starting points for the rest of our analysis.

A Single Starting Point

The easiest way to visualize the settlement process is to think of a single jumping-off point from which migrants who settled the region started and to estimate the likelihood of a county's reaching the settlement threshold based on distance from that location. In order to do this, we measured the distance (in kilometres) from the centroid of Wyandotte County, Kansas (near Kansas City) to the centroid of every other county and used that information in our model to predict settlement. We selected Wyandotte County because it is near the centre of the north-south dimension of the Great Plains and is therefore a reasonable place to consider as a single jumping-off point.

Temperature and Precipitation

In the Great Plains, temperature and precipitation are an effective alternative to latitude and longitude, because precipitation varies from east to west and temperature varies generally from south to north (Figs. 7.2 and 7.3). The advantage of temperature and precipitation in specifying the baseline model is that they also have an impact on agricultural success, which we believe is an important determinant of settlement timing. This approach is consistent with findings at the global scale, for example in Nordhaus (2006).

Elevation

The Great Plains region is a gently sloping plain that rises slowly from an elevation of about 600 m at its eastern edge to around 1,500 m as it approaches the Rocky Mountains. Elevation plays many roles in understanding the relationship between environment and population change, as documented in an increasing number of studies (Nordhaus, 2006; Pender, Jagger, Nkonya, & Sserunkuuma, 2004). Elevation has an impact on health, agricultural productivity and access to other resources. At the scale of the Great Plains in the later twentieth century, there is evidence that the desire to live in proximity to mountains gives elevation significance in studies of migration (Gutmann et al., 2005). Because of the gradual rise in elevation across the plains, elevation also serves as a proxy for longitude, but its impact is probably broader than that of a stand-in for the movement of population from east to west because the land itself and the ways it can be used change with elevation.

All of our analyses of the baseline models suggest that the conventional diffusion model, the movement of population from high to low densities, is a realistic way of describing the settlement of the Great Plains. This was both an east-west and a north-south process and we feel it is reflected better in a model that includes temperature, precipitation and elevation (Model 8) than in the more arbitrary latitude-longitude (Model 3) or latitude-longitude with squared terms and interactions (Model 4) models. This baseline model combines goodness of fit with a theoretical basis that includes elevation, basic suitability for agriculture (precipitation and temperature) and the political economy of settlement captured by the Oklahoma dummy. We will use that specification as our baseline model and see whether our other hypothesized influences (agricultural potential, access to transportation, natural resources and the emergence of an industrial economy) have a significant impact on the timing of settlement in the counties of the Great Plains.

Determinants of Settlement Timing: Baseline and More

We begin our discussion of ways to enhance our understanding of the timing of settlement in the Great Plains by asking whether enhanced suitability for crop-based agriculture, above and beyond the factors captured in our baseline model, accelerated settlement. Earlier research (Burke, Lauenroth, Parton, & Cole, 1994; Burke et al., 1997; Gutmann et al., 2004; Sala, Parton, Joyce, & Lauenroth, 1988; Cunfer,

2005) has shown that a relatively small number of factors – primarily precipitation, access to irrigation, temperature, soil texture, topsoil and slope – contribute to suitability for agriculture. Because our baseline already includes precipitation, temperature and elevation, in Model 9 of Table 7.3 we add only the depth of the topsoil layer and the per cent of land area irrigated. Prior to 1940 virtually all irrigation in the Great Plains was stream-fed, and not pumped from deep aquifers. The availability of irrigation water was therefore location-specific and would have been apparent to potential settlers.

The results of Model 9 show that adding soil and irrigation to our estimates improves the overall fit of the model. When we go beyond agriculture to ask whether other kinds of economic development contributed to the pace of settlement, we find additional confirmation. Here, we explored three concepts: the role of transportation, natural resources and manufacturing, as respectively measured by the amount of railroad, mining and manufacturing employment in each county. We recognize that all of these activities played a minor role when compared to agriculture in the

Table 7.3 Hazard models predicting settlement, population density of four or more persons (all models include dummy variable for state of Oklahoma)

					–2LL	AIC	DF
Baseline alternatives							
	Trend surface models						
1	Longitude				1,998.66	2,010.66	6
2	Latitude				2,279.21	2,291.21	6
3	Longitude	Latitude			1,934.00	1,952.00	9
4	Longitude	Latitude	Long-Lat squared	Interactions	1,783.46	1,825.46	21
5	Distance from origin (Wyandotte, KS)				2,134.24	2,140.24	3
	Geophysical models						
6	Temperature and precipitation				2,040.62	2,058.62	9
7	Elevation				2,160.84	2,172.84	6
8	Temperature and precipitation	Elevation			1,973.05	1,997.05	12
	Baseline models plus additions						
9	Temperature and precipitation	Elevation	Top soil	Irrigated land	1,902.18	1,938.18	18
	Potential for economic development models						
10	Temperature and precipitation	Elevation	Top soil Irrigated Land	Railroad Mining Manufacturing	1,780.18	1,834.18	27

N = 760.
–2LL = –2 * LogLikelihood.
AIC = Akaike's Information Criterion = 2DF – 2LL.
LR = Likelihood Ratio χ^2.
DF = Degrees of Freedom = # parameters in model.

generally rural Great Plains counties up through 1940. Hence, we create dichotomous variables that indicate whether at least 5% of those employed worked in each of those industries in 1930 (the first year for which data are available). We chose 5% as the cut-off because in most counties the employment in these activities was much less than 5%, making this a suitable level for identifying significant employment. The variable for rail employment measures transportation infrastructure differently from other studies (such as Gutmann and Sample (1995)) that use the length of track at different points in time and is probably less valuable than the indicators used in other research because it reflects employment rather than infrastructure in place. However, it is the only variable available at this time for a region the size of the Great Plains. The goodness of fit results are reported in Model 10 of Table 7.3, with the detailed regression results reported in Table 7.4.

The results for Model 10 show that adding these three variables improved the fit of the model and that two of the three variables (manufacturing and mining) had significant coefficients with signs in the hypothesized direction. In other words, counties with the potential for meaningful levels of mining or manufacturing were settled earlier than they would have been in the absence of these economic opportunities. Rail employment, on the other hand, does not predict settlement timing, but we believe that this is more likely to show that rail employment is a poor indicator of transportation infrastructure than that transportation is not significant, because rail employment (primarily for rail construction) in 1930 is itself a reflection of late settlement. Nonetheless, we find that adding agricultural and employment variables beyond those included in our baseline model improves our understanding of the pattern of Great Plains settlement.

Table 7.4 Full model predicting settlementpopulation density of 4 or more persons

Explanatory variables	log(risk ratio)	Pr>|z|
Baseline model		
Oklahoma	−1.481	0.004
Long-term average temperature	−0.0087	0.0016
Long-term average precipitation	0.011	< 0.0001
Minimum elevation	−0.0016	0.03
Suitability for agriculture		
Depth of topsoil layer	−0.0036	0.94
Percent of total land area irrigated	−0.135	0.98
Economic development		
5% or more of employed in railroad industry in 1930	−0.032	0.91
5% or more of employed in mining industry in 1930	0.491	0.11
5% or more of employed in manufacturing industry in 1930	0.123	0.002

N = 670.
–2LL = 1,780.18.
DF = 27.
LR = 493.49.

Discussion and Conclusions

This chapter has two purposes: to improve our understanding of the process of frontier settlement in the United States and to demonstrate a method of longitudinal data analysis that has wide potential for future research. It is apparent that the survival-analysis application brings great rewards, and we look forward to using it more ourselves and to seeing it used by others to explain processes occurring over time and space.

The challenge of areal unit boundary changes in panel data analyses is typically treated as a problem of rendering one boundary file onto another or combining the incompatible spatial units into spatial clusters. Earlier, we enumerated the limitations of creating areal clusters based on longitudinal templates, but we also urge analysts to think carefully about the nature of their research questions before adopting the other approach. For example, U.S. urban and community researchers have readily embraced Geolytics' Neighborhood Change Database (NCDB), which normalizes 1970–2000 summary files to 2000 census tract boundaries. However, producing comparability in this way is not purely benign. It is well-known that the U.S. Census Bureau redraws tract boundaries to encapsulate about 4,000 inhabitants in relatively homogeneous units with respect to population characteristics, economic status and living conditions. Therefore, the process by which tract boundaries change over time ensures increasing within-unit homogenization if those boundaries are defined at an endpoint and then held constant backwards through time. This artifactual generation of serial within-unit heterogeneity will be consequential in panel data analyses using the NCDB. In contrast, our approach imposes no artificial spatial structure on the panel data. The spatial units we analyse are the contemporary units of analysis at each moment in our analysis.

Our research objectives in this chapter were linked to sets of coefficients and their estimated standard errors, derived from weighted averages of their corresponding stratum-specific values. In other applications, attention to the consistency in the stratum-specific estimates may supersede interest in the global effects if the structure of dependence is of analytic value. Post-estimation tests for the equality of stratum-specific estimates easily attain this objective and give yet another dimension to the utility of this method.

Our exploration of the process of settlement has produced very interesting results. Widely accepted theory teaches that settlement is a diffusion process, with people moving away from a starting point in a steady progression. We started out by attempting to estimate a baseline model that measured diffusion, beginning with latitude and longitude, then using migration from a single point and finally evaluating precipitation, temperature and elevation, which are arrayed along east-west and north-south axes in the Great Plains. Our survival models show that each of these approaches has value, with reasonably good fit and individually significant variables. While the latitude-longitude model has advantages in terms of goodness of fit, we believe that the precipitation-temperature model, with its individually significant variables and strong fit, constituted a better starting point for further analysis because it ties settlement to specific characteristics of place. All these models include the important public policy determinant of settlement timing that we capture by including a dummy variable for counties in Oklahoma, which were opened to settlement later and more arbitrarily than other counties in the region. Our experience fitting the baseline model suggests that the process of diffusion does not occur evenly across space but is shaped by the political and environmental features of the landscape over which a population is diffusing.

Settlement is much more than just a diffusion process, however. Most of the early settlers to the Great Plains came to farm or to engage in commerce that supported farming and we hypothesized that suitability for agriculture would be a strong determinant of early settlement, all other things being equal. Previous research shows that the productivity of land is largely a function of availability of moisture and soil quality. In the period prior to 1940, when there was still relatively little irrigation in the Great Plains, this was largely a function of precipitation. In order to evaluate this hypothesis we added two variables, one that measures topsoil depth and one that measures land area irrigated. Neither was significant, although the model fit improved marginally. This is not surprising given the small amount of irrigation and the relatively weak effect of soil quality compared with precipitation in most other studies. What this shows is that the patterns of settlement were not very sensitive to small variations in suitability to agriculture, at least at the scale observed.

Other forms of economic development do matter for the settlement process. We tested three variables, transportation employment, natural resource endowments and manufacturing employment. The model that included these variables had an improved fit and two of the three variables – manufacturing employment and mining employment – were significant. We take this result as confirmation that economic development matters, but it also reveals that by the time we measure the presence of rail employment in the Great Plains, the region was thoroughly covered by rail connections and there was little benefit to being in one part of the region over another.

Given the strength of dispersion models in migration theory, not everything we conclude is surprising. It is important to emphasize that modern and methodologically sophisticated tests of the power and limitations of dispersion models are very important and that we have added significantly to what we know about how European origin populations move through the settlement process, especially in regions like the Great Plains during the late nineteenth and early twentieth centuries. There was a process of settlement diffusion in the Great Plains and variations in that diffusion process favoured areas well-suited to cropping, mining and manufacturing. This confirms theory that adds the environment and structural economic conditions as significant factors shaping settlement migration, in addition to short-term economic conditions that are widely known to operate. These new and strengthened findings make it possible to enhance knowledge in the future by studying local conditions where settlement was taking place and asking how those local conditions acted together with the broader structural factors that are now increasingly known to have an impact.

Acknowledgement This work has been supported by Grant no. R01HD33554 from the National Institute of Child Health and Human Development in addition to work supported by the National Science Foundation while Myron Gutmann worked at the Foundation. We are grateful to colleagues from the Great Plains population and environment project who gave advice about this paper and especially to Christie Wilson, who did important early work with us on this topic. Any opinion, finding, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Notes

1. Adams County exemplifies the difficulty of ascertaining when an area reached a density of four persons per square mile. We know that it had reached that threshold before 1910, because it had a density of 6.6 in that year. The procedure that we use in this paper assigns it to 1880, but it is likely that it reached four persons per square mile in 1900, or even in 1910.
2. We also used population density data for 1870 to ascertain counties at risk of settlement by 1880.
3. These data are available from ICPSR: http://www.icpsr.umich.edu/PLAINS as well as ICPSR Studies # 4254 and 4296 (Gutmann, 2005a, b).

References

Allison, P. D. (1995). *Survival analysis using the SAS system: A practical guide*. Cary, NC: SAS Institute, Inc.

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht: Kluwer Academic Publishers.

Burke, I. C., Lauenroth, W. K., & Parton, W. J. (1997). Regional and temporal variability in aboveground net primary productivity and net N mineralization in grasslands. *Ecology*, *78*, 1330–1340.

Burke, I. C., Lauenroth, W. K., Parton, W. J., & Cole, C. V. (1994). Interactions of land use and ecosystem function: A case study in the central Great Plains. In P. M. Groffman, G. E. Likens, (Eds.), *Integrated regional models: Interactions between humans and their environment* (pp. 79–95). New York: Chapman Hall.

Burnham, K. P. & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information theoretic approach* (2nd ed.). New York: Springer.

Cunfer, G. (2005). *On the Great Plains: Agriculture and environment*. College Station, TX: Texas A&M University Press.

Fite, G. C. (1966). *The farmers' frontier, 1865–1900*. New York: Holt, Rinehart and Winston.

Gutmann, M. P. (2005a). GREAT PLAINS POPULATION AND ENVIRONMENT DATA: AGRICULTURAL DATA [Computer file]. ICPSR version. Ann Arbor, MI: University of Michigan [producers], 2005. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005. Available online at http://dx.doi.org/10.3886/ICPSR04254.

Gutmann, M. P. (2005b). GREAT PLAINS POPULATION AND ENVIRONMENT DATA: SOCIAL AND DEMOGRAPHIC DATA [Computer file]. ICPSR version. Ann Arbor, MI: University of Michigan [producers], 2005. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005. Available online at http://dx.doi.org/10.3886/ICPSR04296.

Gutmann, M. P., Deane, G. D., Lauster, N., & Peri, A. (2005). Two population-environment regimes in the Great Plains of the United States, 1930–1990. *Population and Environment*, *27*, 191–225.

Gutmann, M. P., Parton, W. J., Cunfer, G. A., & Burke, I. C. (2005). Population and environment in the U.S. Great Plains. In B. Entwisle, P. Stern, (Eds.), *Population, land use, and environment: Research directions* (pp. 84–105). Washington, DC: National Academy Press.

Gutmann, M. P., Pullum-Piñon, S., Baker, S. G., & Burke, I. C. (2004). German-origin settlement and agricultural land use in the twentieth century Great Plains. In W. Kamphoefner, W. Helbich, (Eds.)., *German-American immigration and ethnicity in comparative perspective* (pp. 136–168). Madison: University of Wisconsin Press.

Gutmann, M. P., Pullum-Piñon, S., Cunfer, G. A., & Hagen, D. (1998). The Great Plains population and environment database: Sources and user's guide. Version 1.0. Texas Population Research Center Papers, Austin. Available electronically at http://www.icpsr.umich.edu/PLAINS/pdf/prc10.pdf

Gutmann, M. P. & Sample, C. G. (1995). Land, climate, and settlement on the Texas Frontier. *Southwestern Historical Quarterly*, *99*, 137–172.

Harley, C. K. (1978). Western settlement and the price of wheat, 1872–1913. *Journal of Economic History*, *38*, 865–878.

Horan, P. M. & Hargis, P. G. (1995). COUNTY LONGITUDINAL TEMPLATE 1840–1990 [Computer file]. ICPSR version. Athens, GA: Patrick M. Horan, University of Georgia, Dept. of Sociology/Statesboro, GA: Peggy G. Hargis, Georgia Southern University, Dept. of Sociology and Anthropology [producers], 1995. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1995. Available online at http://dx.doi.org/10.3886/ICPSR06576.

Kittel, T. G. F., Rosenbloom, N. A., Painter, T. H., Schimel, D. S., & VEMAP Modelling Participants. (1995). The VEMAP integrated database for modelling United States ecosystem/vegetation sensitivity to climate change. *Journal of Biogeography*, *22*, 857–862.

Kittel, T. G. F., Royle, J. A., Daly, C., Rosenbloom, N. A., Gibson, W. P., Fisher, H. H., et al. (1997). A gridded historical (1895–1993) bioclimate dataset for the conterminous United States. In *Proceedings of the 10th Conference on Applied Climatology, 20–24 October 1997, Reno, NV* (pp. 219–222). Boston: American Meteorological Society.

Nordhaus, W. D. (2006). Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences*, *103*, 3510–3517.

Pender, J., Jagger, P., Nkonya, E., & Sserunkuuma, D. (2004). Development pathways and land management in Uganda. *World Development*, *32*, 767–792.

Raffalovich, L. E., Deane, G. D., Armstrong, D., & Tsao, H. (2008). Model selection procedures in social research: Monte-Carlo simulation results. *Journal of Applied Statistics*, *35*, 1093–1114.

Sala, O. E., Parton, W. J., Joyce, L. A., & Lauenroth, W. K. (1988). Primary production of the central grassland region of the United States. *Ecology*, *69*, 40–45.

Turner, F. J. (1893). The significance of the frontier in American history. In *Report of the American Historical Association* (pp. 199–227). Washington, DC: Government Printing Office.

U.S. Department of Agriculture. (N. d.). *State soil geographic (Statsgo) data base*. Washington, DC: USDA Natural Resources Conservation Service.

Webb, W. P. (1931). *The Great Plains*. Dallas, TX: Ginn and Company (Reprint: University of Nebraska Press, 1981).

Webb, W. P. (1957). The American West. Perpetual mirage. *Harper's Magazine*, *104*, 25–31.

Wei, L. J., Lin, D. Y., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distribution. *Journal of the American Statistical Association*, *84*, 1065–1073.

Worster, D. (1987). New West, true West. *Western Historical Quarterly*, *18*, 141–156.

Chapter 8
Commonalities and Contrasts in the Development of Major United States Urban Areas: A Spatial and Temporal Analysis from 1910 to 2000

Andrew A. Beveridge

Introduction

How cities and urban agglomerations grow and develop is of intense interest to many social demographers and urban studies researchers. The question about whether such development follows any obvious pattern or law has animated speculation, research and discussion from the founding of social demography and urban studies in the United States to the present day. Despite much theoretical analysis, most empirical research in the area has relied upon data from one or a few locations, or made comparisons based upon impressionistic evidence. Many such studies in United States sociology and demography have used decennial materials created by the United States Census Bureau to compare and contrast urban patterns. However, since comparable information about different cities or urban regions was, until very recently, unavailable in a form that could be mapped, it has not been possible to rigorously track change at the small area level in multiple cities, nor has it been possible to compare their patterns of change.

With the advent of data from the National Historical Geographical Information System (NHGIS) in 2006, small area (census tract level) data on the development of about fifty urban areas now exist for the decennial censuses from 1940 to 2000. Even earlier data are available for nineteen cities from 1930 and ten cities from 1920, with some limited information going back to 1910 for another eight. Using these materials, it is possible to begin to address the patterns of change in a more rigorous manner:

Here three questions are examined, for the first time using comparable data:

1. What are the general patterns of growth and decline in United States urban settlements from early in the twentieth century through 2000?
2. How consistent are these patterns? Do they vary by urban area or by period?

A.A. Beveridge (✉)
Queens College and Graduate center, City University of New York, New York, NY, USA
e-mail: andy@socialexplorer.com

M.P. Gutmann et al. (eds.), *Navigating Time and Space in Population Studies*, International Studies in Population 9, DOI 10.1007/978-94-007-0068-0_8,

3. For three of the major metropolises in the United States, are the patterns of growth similar or different from 1910 to 2000, and in what ways?

Simple measures of small area population change will be presented to depict and compare growth and decline during three periods for all of the areas for which data exist. Maps will be used to visualize patterns and change in the three emblematic urban areas of New York, Los Angeles and Chicago. Finally, spatial analysis techniques will be used to compare the patterns of change from decade to decade in the three major metropolises. Suggestions for further work will be outlined.

Though this effort is largely one of first impressions since these data have never before been arrayed in this manner, it will lay out a series of hypotheses drawn from the literature of United States social demography, urban studies and sociology. A set of measures will be developed to address these hypotheses directly. Furthermore, even with the existence of the NHGIS data, many questions remain regarding the comparability of geographies and the measurement of growth from decade to decade.

The data used in these analyses only document total population size and not more refined measures, such as geographical features, race or ethnic composition or economic status. Nonetheless, the use of the NHGIS materials makes it possible to address rigorously some of the common ideas in social demography and urban sociology about the trajectory of urban change in the United States. Even with the relatively simple data provided here, it is possible to address one of the recent debates about the patterns of growth of urban populations, and to make the standards for that debate more rigorous. In short, this chapter should be seen as a starting point to put discussions of the trajectory of change in cities and urban areas in the United States and elsewhere on a more rigorous and comparative empirical footing.

Views of Urban Change

In the United States, urban studies, urban sociology and social demography have been dominated by the so-called ‘Chicago School’. Developed after the founding of the University of Chicago sociology department in the 1920s and 1930s and led by Robert Park and Ernest W. Burgess, it featured one of the earliest models designed to explain the spatial organization of urban areas. Indeed, the founding of both social demography and sociology in the United States academy dates from the development of the University of Chicago sociology department. From its beginning, the University’s demographers and sociologists were engaged with social and spatial changes in the city of Chicago.

The classic Chicago School’s model was Burgess’s concentric ring theory of how the city evolved. Figure 8.1 presents a diagram that displays the various rings. At the centre of this model was the Central Business District, which was also where the homeless of those days (‘hobos’) lived. This zone was surrounded by a zone

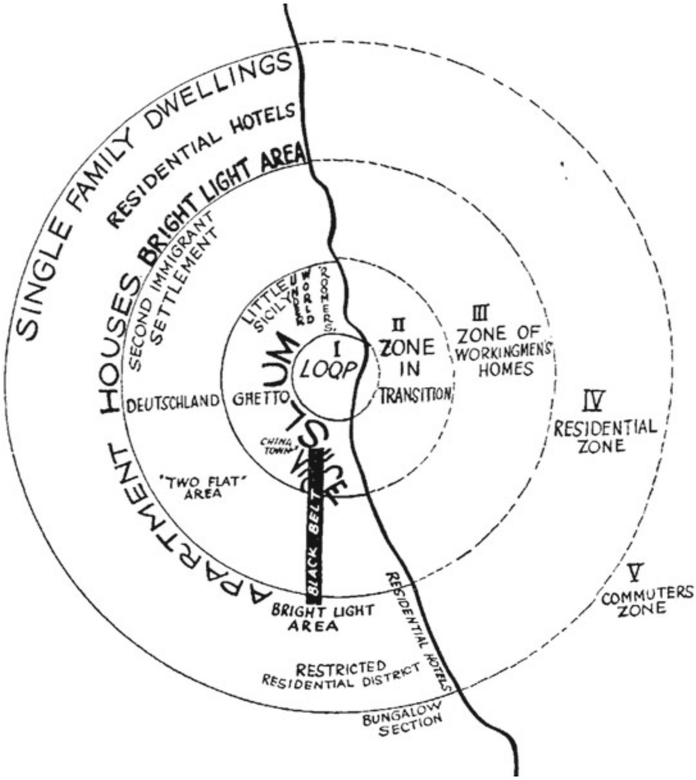


Fig. 8.1 Concentric zone map from *The City*

of transition into which business, light manufacturing and new immigrants were moving and where the city's slums and vice were concentrated. Moving outwards, the second zone was inhabited by workers, often second-generation immigrants, who wanted to live within easy access of work, albeit in modest homes. The inhabitants of this zone looked to move out to the promised land of the next two zones. The third zone included high-class apartment buildings and more exclusive areas of single-family housing for the middle class. The outermost zone contained the commuter suburbs (Burgess, 1925: pp. 47–62). The dynamic terms of the model were 'invasion', 'succession', and 'segregation', which, encapsulating the belief that moving outwards was the preferred goal, indicated the struggles through which social groups attempted to move from inner to outer rings.

Because of this emphasis on specific geographic areas, as well as the influence of both Amos Hawley and Donald Bogue, researchers at the University of Chicago working with the Census Bureau defined 75 community areas in the late 1920s. The Census Bureau collected and released data for these areas, which roughly encompassed large neighbourhoods. In this way, social demography in the United States developed in concert with urban sociology. The Chicago School included both, and it emphasized mapping population distribution. The typical map produced during this time was a dot or chloropleth map showing the rate or incidence of certain social characteristics throughout the city; researchers used these maps to identify gradients between areas that might indicate zone boundaries (Bulmer, 1984). Later work by Bogue (1950) and Taeuber and Taeuber (1965) used census data to map and measure segregation. Indeed, their work came to dominate the manner in which large scale urban change was viewed.

At the same time, the development of large cities was seen as part of a system of cities, as defined by Central Place Theory, which saw each city as a place within a hierarchical system of cities (for example, Berry (1961); Berry and Pred (1961); De Vries (1984)). This method owed a lot to developments in fractal geometry and was originally posited in 1933 by Christaller (1972). In a similar vein the concentric ring or zone theory, which was originally developed by Park and Burgess, was seen by human ecologists as being rooted in something analogous to what bioecologists had found 'on the structure and development of plant associations' (Hawley, 1981: p. 423). Those researchers showed that plant species adapt to their environment by distributing themselves over a localized area in a pattern, which enables them to engage in complementary uses of habitat resources (Hawley, 1981: p. 423).

Indeed, as late as 1981, Hawley, a member of the Chicago School and a leading proponent of applying this ecological approach to urban populations, affirmed the relevance of the classic model. He wrote the following on recent work that he found more rigorous:

> Although these several studies have called attention to minor deviations from and needed refinements in the pattern they have confirmed that the distribution of housing types and of the occupational, educational and family characteristics of city residents exhibit a gradient pattern of variation from low to higher values with distance from the central business district. That is a finding of some interest, for it suggests that the assumptions underlying the model are as pertinent in the third quarter as in the first quarter of the century (Hawley, 1981: p. 425).

It is interesting that few social demographers explicitly tested the propositions of the Chicago school. Among those who did were Schnore (1957), Duncan, Sabagh, and Van Arsdol (1962), and later Guest (1971, 1975, 1978 and 2002.) None of these researchers had access to the much more fine-grained data created by the NHGIS. Guest's most recent work is quite consistent with what is presented here. It should also be noted that Donald Bogue and Elizabeth Mullen Bogue, who created data that could be used for analysis from the 1940 and 1950 censuses by keying in tables, did not include New York City or Los Angeles in their 1950 file, and in 1940 the Census Bureau did not produce tract tabulations for New York City, but only health areas (See Documentation Elizabeth Mullen Bogue File, Census Tract Data, 1940–1970, National Archives and Records Administration, College Park, MD, May 28, 1999 NN3-CFS-96-999). Some of the published tables for health areas in New York City were published for census tracts in cooperation with the Census Bureau by the New York City Welfare Department.

Recently, the Chicago School has been explicitly challenged by the so-called Los Angeles School. The main focus of the Los Angeles School is the 'sprawling, polycentric character of the urban, built environment' (Scott & Soja, 1996: p. 6). Although the term 'Los Angeles School' came about in the 1980s, the school's founding text is arguably Fogelson's 1967 book *The Fragmented Metropolis: Los Angeles, 1850–1930*. Fogelson traces the evolution of Los Angeles from a 'nondescript agricultural village' in 1850 to a large city of 1.2 million in 1920 and then, already by 1930, to a dispersed and decentralized metropolis. He writes:

> More than any other American metropolis – and with remarkably few misgivings – Los Angeles succumbed to the disintegrative, though not altogether undesirable, forces of suburbanization and progressivism. And as a result it emerged by 1930 as the fragmented metropolis par excellence, the archetype for better or for worse of the contemporary American metropolis (Fogelson, 1967: p. 2).

Fogelson thus lays out the central idea of the current Los Angeles School, among whose major representatives are Michael Dear, Jennifer Wolch, Alan Scott, Edward Soja and Michael Davis. For example, Wolch and Dear (1993, xxiii) wrote that, 'the Los Angeles prototype – with its patterns of multicentred, dispersed low-density growth – is perhaps a new paradigm of metropolitan growth'. They contrasted this new urban paradigm with the formerly dominant 'Chicago model', which analysed urban development as a number of zones, concentric circles in their tidiest forms, radiating out from a perennially dominant core composed of the Central Business District (Wolch & Dear, 1993). In a similar vein, Scott (1996: p. 276) pointed out how, by the end of the twentieth century, Southern California, and especially the Los Angeles region, is made up of a 'multiplicity of discrete industrial districts (some of which are individually comparable to Silicon Valley in size and rapidity of growth) scattered over its entire geographic extent'. Following the Los Angeles School's tendency to argue that a new vocabulary is required to capture what it is uncovering, Scott proposed the term 'techno poles' for these dispersed, high technology, industrial districts.

To this is added the claim, also characteristic of the Los Angeles School, that developments in Los Angeles may be the new model that other urban areas will follow. This claim is typically hedged. Perhaps the Los Angeles region is the harbinger of the future, if not already the present, of many of America's other major urban areas. On the other hand, it may be unique. (Both claims are often contested by observers outside the Los Angeles region.) For example, Wolch and Dear (1993, xxiv) write of Los Angeles's strengths for study, 'It is already self-evident that southern California's emergent urbanism possesses more than local or regional significance. Given its implications as the prototype for future urban dynamics . . . Los Angeles is pre-eminently qualified to serve as a laboratory for our analysis'. They also concede that it may not be universally relevant, 'Whether Los Angeles represents the model twenty-first century urban development is a question we shall leave to others'.

Since the introduction of the Los Angeles model, the claim that the Los Angeles region represents the future of America's major metropolises has been disputed. For example, the geographer Jan Nieman (Miller, 2000) and the urban historian Robert Schneider (Schneider, 2000: p. 1672) argue that Los Angeles is too gargantuan to be typical of other American metropolises. Nieman suggests Miami exhibits many of the tendencies of Los Angeles but, being smaller, is more typical of the future of United States cities and urban areas and offers a more useful model. Schneider suggests Los Angeles, with its large population of homeless, may be more typical of contemporary Third World cities such as Mexico City, Cairo, Calcutta and Sao Paulo than of urban areas in the U.S.

Halle (2003) recently posited the existence of a ‘New York School’, characterized by an interest in the central city. What is going on there? What should the central city be like? Halle asserts that some of this new school’s key figures include Jane Jacobs, Sharon Zukin, Kenneth Jackson, Robert Stern, Richard Sennett, and William Whyte. According to Halle, these authors share a fascination with contemporary New York City, especially with Manhattan, and a belief, in some cases passionate, in the superiority of city life over suburban life. They also see New York City as a place for the middle class and the rich, not just the poor and working class.

The New York researchers have not, so far, been explicitly identified as a school, partly because they do not identify themselves in this way. Those in the Chicago School differed from the later New York researchers because they often agreed with the opinion of residents at the time that the outer suburban zones were the preferred places to live. They did not envision the central city as an especially desirable place for the residences of the middle class and even less for those of the rich. To the Chicago School, the inner city was a place for central business, new immigrants, vagrants, hobos and the working class.

Kenneth Jackson’s history of the suburbanization of America from 1814 to the present, *The Crabgrass Frontier* (Jackson, 1985), is a major statement about urban development. According to Halle, a central subtext of the book is a regret-filled analysis of how policies of the federal government undermined the economic and social health of America’s central cities and urban neighbourhoods in the twentieth century, especially from the 1920s to the 1970s. Such policies, for example, wrote public housing legislation that almost guaranteed it would be clustered in the central city rather than spread evenly through the region. Other initiatives encouraged mortgage policies that discriminated against central city neighbourhoods, especially those with large concentrations of minorities, with the effect of undermining the value of the housing there. Jackson also shows how twentieth century suburbanites, acting through their state legislatures, which set city boundaries, tried to inhibit the spread of older American cities by preventing them from doing what they had done throughout the nineteenth century – namely, expanding their borders geographically to capture for the city the new economic and residential developments on the periphery.

To justify their choice of which city to study, the New York adherents tend to use strategies similar to the members of the Los Angeles School. Some argue that what they uncover in New York is typical of many other big cities and helps us to understand them. Jacobs, for example, wrote: ‘In trying to explain the underlying order of cities, I use a preponderance of examples from New York, because that is where I live. But most of the basic ideas in this book come from things I first noticed or was told in other cities' (Jacobs, 1961: Chapter 1). Others justify what they do on the grounds that New York City is unique, or at least extremely unusual (Duneier, 1999).[1] What makes it unique – for example, its size and cultural importance together with the economic power of Wall Street – tends to be what makes it a crucial object of study.

In a sense then, each of the schools generalizes from its local city's experience. Those in Chicago experienced the movement of population away from the immigrant and African American ghettos out to suburbs, those in Los Angeles saw the sprawl, while those in New York City saw the advantages of city living, even if they could not afford it. However, the generalizations about the pattern of urban change are in part based upon incomplete and inaccurate data, and beg for answers to the following questions:

1. How different or similar were and are the patterns of urban growth in each of these cities and metropolitan regions?
2. How do they compare with the other metropolitan areas in the United States?
3. How did they change during the decades of the twentieth century?

Each school seeks to study and articulate developmental patterns, but unlike much of the work on Chicago and Los Angeles, which generally draws upon information (often impressionistic) at one point in time, this chapter explicitly addresses patterns over time and space. In short, it attempts to answer the question: How is the pattern of growth characterized in major cities and urban agglomerations during the twentieth century? The hypotheses emerging from the three general vantage points (New York, Los Angeles and Chicago) can be given explicit expression in terms of the location of population growth (and decline) in relation to the centre of the city, and of the spatial patterning of that growth.

1 *Chicago School Hypotheses:*

   a Population Growth and Decline: Major population growth should be away from the urban centre, while any population decline would be near the urban centre.
   b Spatial Patterning of Growth: Population growth and decline should be spatially patterned, that is growth areas in one decade should be near growth areas in the next decade, while decline areas in one decade should be near decline areas in the next decade.

2 *Los Angeles School Hypotheses:*

   a Population Growth and Decline: Major population growth should not follow a particular structure, so population growth and decline would be roughly the same distance from the urban centre.
   b Spatial Patterning of Growth: Population growth and decline should not be spatially patterned. Growth areas in one decade can be near decline areas in the next decade and vice versa.

3 *New York School Hypotheses (Which Would Mostly Apply to Recent Periods):*

a Population Growth and Decline: Major population growth should be nearer the centre of the city, and decline should be away from the urban centre.
b Spatial Patterning of Growth: Growth in the centre of the city will be patterned, but there is no reliable way to predict the patterning of growth elsewhere in the city.

Proponents of each school may object to the manner in which these hypotheses are laid out. However, to begin to clarify their empirical patterns, it is important to state explicitly their expectations. In this way, data may be arrayed and measures can be developed to directly address these questions.

Data on Urban Change and the NHGIS

To examine patterns of urban growth and change comprehensively requires data with spatial characteristics from all of the areas under study for the entire period of examination. The analysis presented here is based on census tracts, which are geographic areas developed by the Census Bureau in consultation with local officials. Currently, tracts have an average population of approximately 4,000 inhabitants. They were expressly designed to report on small areas in urban settings, and thus provide more detail for large places than using city or county level data. Indeed, the creation of census tracts began with a desire to make New York City more comprehensible.

The Census Tract Movement began in New York in 1906, when urban planner Walter Laidlaw suggested that the city be divided into units according to population for the 1910 Census. In 1910, for the most populous neighbourhoods, such as most of Manhattan and portions of Brooklyn and the Bronx, a measure of approximately forty acres was used, with each tract averaging about eight city blocks. The rest of the city was divided into larger geographic areas. Laidlaw also convinced the Census Bureau to create tracts and tabulate census tract data for other cities that had more than 500,000 inhabitants in 1910. These included Boston, Philadelphia, Baltimore, Pittsburgh, Cleveland, Chicago and St. Louis, as well as New York City. Data from Washington, D.C., as well as the original eight cities, became available in 1920. By 1930, the number of cities for which tracts were defined separately had increased to fifteen; that number reached nearly sixty by 1940. In addition, Detroit, Nashville, New Orleans and Milwaukee used the depression-era Works Project Administration to create census tracts retrospectively back to 1920. More areas were tracted between 1950 and 1980, and by 1990, the Census Bureau had divided the entire country into tracts.

As a result, since 1910 the United States has had a limited set of urban small area data based upon the census, which grew to include some sixty cities in forty-nine areas by 1940. For the year 1990, census data were available for every part of the United States. This pattern of data availability has several consequences: (1) Data are not provided for comparative areas for each decade; (2) Data are not provided for all areas for all decades; (3) Some of the variables may not be available or completely comparable from decade to decade; (4) The boundaries from decade to decade for those areas where small area data exist were changed from decade to decade.

The NHGIS has developed tract boundaries for all areas of the United States for which they existed, and also has experimented with using other methods of adding spatial data.[2] The NHGIS has created boundary files for each year of tracted data that match those reported by the Census Bureau. This means that there is no guarantee that the tract boundaries for one decade will match those in a succeeding decade. To make it possible to assess population change, for the purposes of this analysis all tract boundaries were normalized to those existing in 2000 by using areal interpolation.[3]

Two sets of comparisons were used for the analyses. One set included all cities for which tract-level data existed in 1910, 1940 and 1970. These were then compared to 1940, 1970 and 2000, as applicable. This means that the area that was New York City in 1910 would be compared to the same geographic areas in 1940, 1970 and 2000. Similarly, Chicago in 1940 would be compared based upon the area that was tracted in 1940. Using these sets of urban areas, the locations of growth and decline areas in relation to the urban centre were assessed.

A careful analysis of the changing tract patterns and the changing populations in the Chicago, New York City and Los Angeles metropolitan areas was performed using methods (described below) to understand spatial patterning of growth and decline. Los Angeles did not have any tracts until 1930, and the NHGIS had not yet completed its work with the 1930 Los Angeles tracts at the time of this writing. New York City was one of the first tracted cities, and in 1930 and 1940 other parts of the metropolitan area were added. To make the comparisons for the same areas where tract data did not exist, county data were used instead. This means that the units used for comparison are not strictly comparable over time, but the geographic areas are. Population densities were computed for all decades for each of the three cities.

The results of each analysis and the measures used are described below, along with an assessment of how the results either confirm or dispute the various hypotheses related to the three 'schools' posited earlier in this chapter.

Simple Measures of Population Growth and Decline

The question to be explored is the extent to which urban growth in Chicago, New York City, Los Angeles and other urban areas conforms to the hypotheses described as the Chicago, New York and Los Angeles schools. Figure 8.2 presents the pattern of population density in New York and Chicago for 1910 and Los Angeles for 1940 (the first year for which comprehensive tract data exist), for the areas tracted and the counties included. Chicago does seem to present something that approximates a

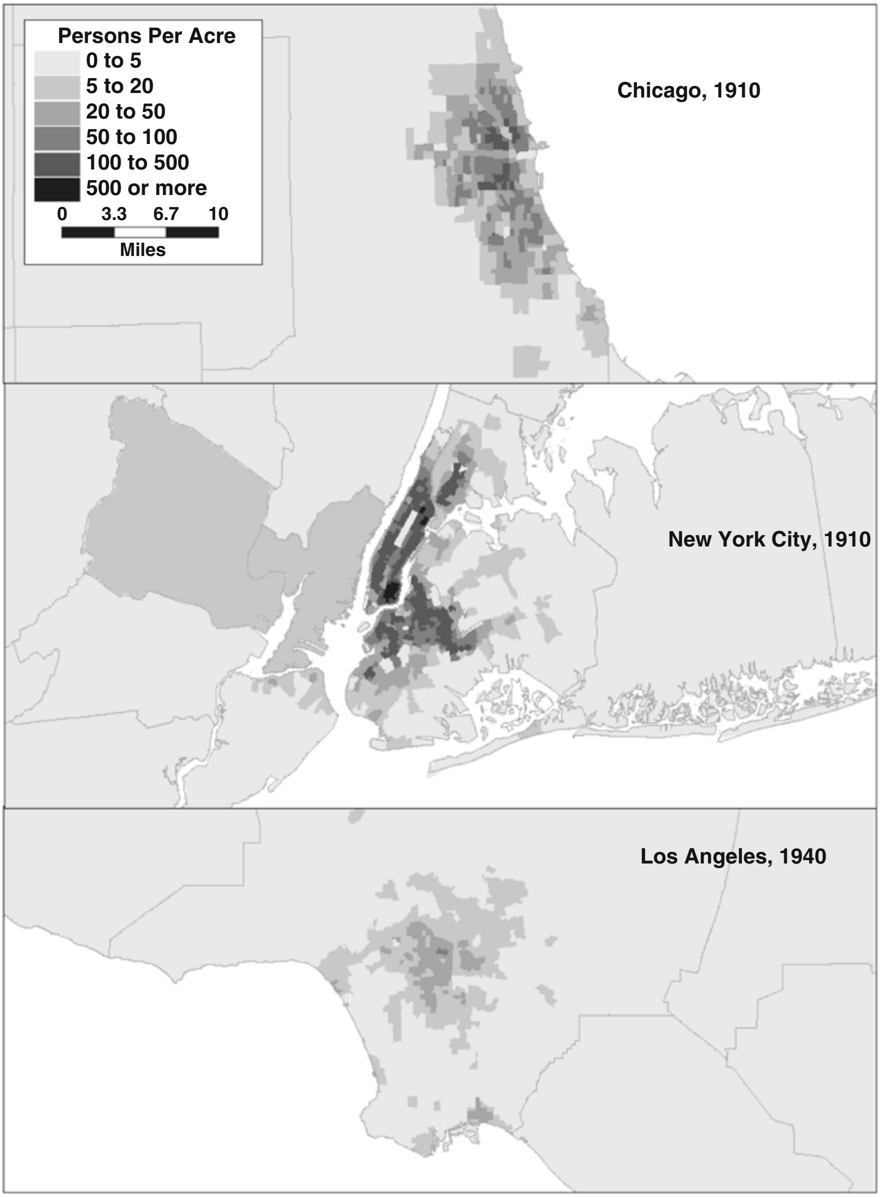


Fig. 8.2 Population density, tracted areas, New York city and Chicago, 1910, Los Angeles, 1940

concentric system of zones, as do New York City and Los Angeles. There is a core of high density areas in Chicago and New York City, and as one moves out from those areas, density declines. The same is true for Los Angeles. Figure 8.3 presents population density for the three large metropolises for 2000, and shows that the pattern of a uniform gradient of density from a central point out is diminished. As

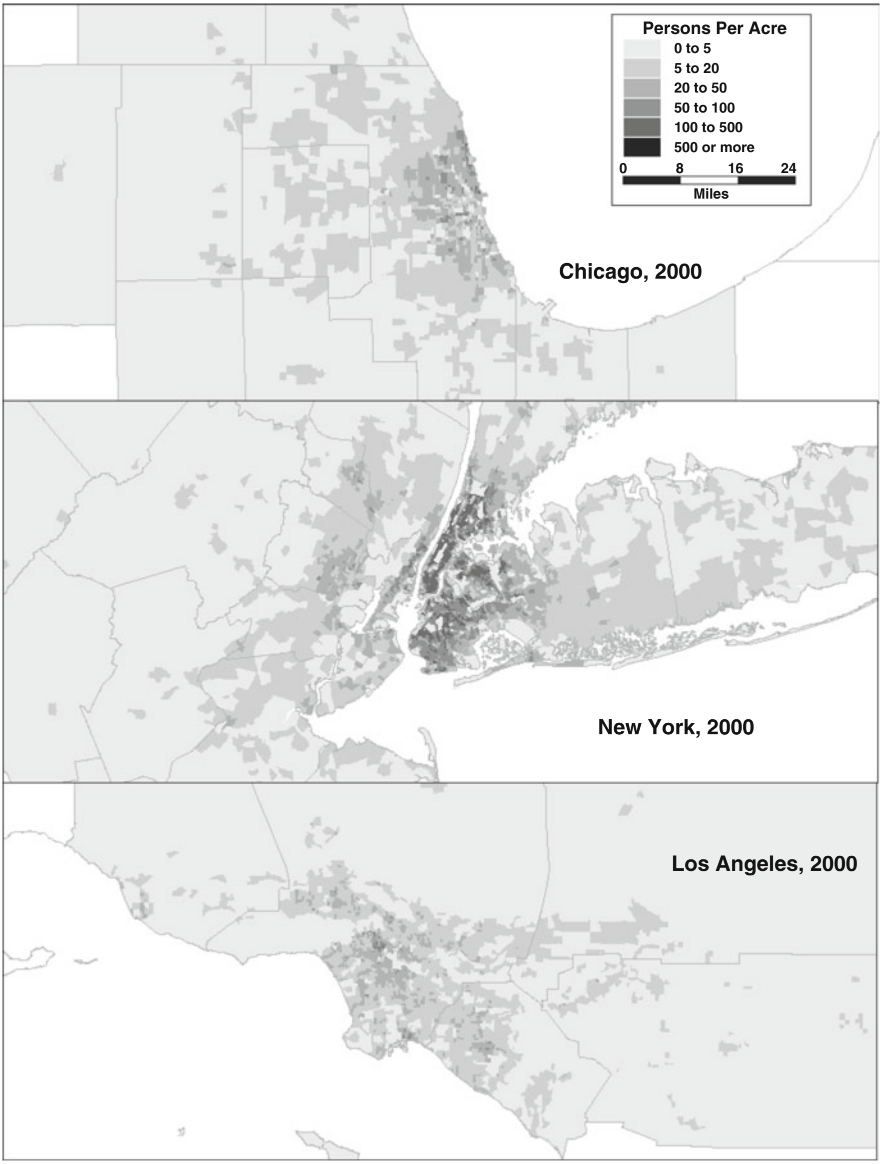


Fig. 8.3 Population density, tracted areas, New York City, Chicago and Los Angeles, 2000

shown in the map, all three cities have high-density locations away from the centre. Chicago and New York in 1910, and Los Angeles in 1940, are quite concentrated in the centre, but by 2000 the patterns are not as obvious.

Chicago in 1910 was plainly the centre of a largely undeveloped area, surrounded by farms. Chicago had tenements on the south side, very few African Americans, and a large number of immigrants. It did not have many natural barriers to growth, as do New York, where the coast and the river systems block growth, and Los Angeles, which has mountains as well. New York City had recently been consolidated. Brooklyn and New York City had added Staten Island and townships and villages from Westchester and Nassau county to form the city of five boroughs that we know today. Most population concentrations were in Manhattan, especially southern Manhattan, and in western Brooklyn. There were massive tenements on the Lower East Side and in some other areas, while much of Queens, Staten Island, the northern Bronx, eastern Brooklyn and even northern Manhattan remained undeveloped. Los Angeles in 1940 was on the brink of a major expansion. Though the post-war expansion of all three cities and urban areas was driven by the growth of highways and expressways, Los Angeles's expansion was especially based upon the rise of the automobile. In all three cities and urban areas in the early period, there are obvious population centres as shown already in Figure 8.2, but by 2000 population seems much less structured. Population density does not appear to follow a gradient from the centre to the periphery. In certain ways, the growth trajectory in all three cities and urban areas seems quite similar from this perspective. Population grows away from the centre, while the pattern is more mixed in the centre.

With these maps as a backdrop, measures were developed to track the location of growth and decline areas. Data are available for a total of six urban areas for 1910 (including Chicago and New York), some forty-nine for 1940 (including Los Angeles) and forty-four for 1970. By 1970 some urban areas had joined with others, while new areas had been added. There were about 3,600 tracts in 1910, 12,000 in 1940 and 24,000 in 1970, all using 2000 census tract boundaries. Two simple measures were developed and applied to these tracted areas for this period of time. First, tract data were used to measure population growth or decline from the first period to the second period and to report the total population decline in areas of population decline and total population growth in areas of population growth. Then, in each case, a population centre for the tracted area in the earlier period was computed to serve as a comparison.

The population centre was based upon the centroids (the geometric centre of a region) of the census tracts, weighted by the population in each tract. Population centres have a time-honoured tradition in the tracking of population change in the United States. In every atlas produced by the U.S. Census Bureau and after every census, the U.S. Census Bureau computes a 'geographic centre of population'. Simply put, a population centre is where an area would 'balance' if all population weighted the same was distributed uniformly over each region (U.S. Bureau of the Census 2007: p. 11). Using these centres and the centroids weighted by the population, one is able to compute the average distance from the population centre of the areas of growth or decline. These two measures are computed for the six cities for which information exists from 1910. Table 8.1 presents figures using two measures of change for New York, Chicago and four other cities by census tract.

From 1910 to 1940, there was a much greater growth in population in areas where population was growing than there was decline in population in those areas that were in decline. Also, these areas of growth were all farther from the population centre of

Table 8.1 Patterns of growth and decline in the six largest city areas from 1910–1940, 1940–1970 and 1970–2000. Based upon areas tracted in 1910

	Total increase in growth areas 1910–1940	Total increase in decline areas 1910–1940	Ratio of population growth areas to decline areas	Distance from population centre of growth areas 1910–1940	Distance from population centre of decline areas 1910–1940	Ratio of distance to population centres of growth areas to decline areas
A. 1910–1940						
Boston	218, 330	61, 418	3.6	4.3	1.8	2.4
Chicago	1, 569, 914	424, 957	3.7	7.8	2.2	3.5
Cleveland	242, 483	13, 718	17.7	4.6	2.5	1.8
New York	3, 990, 485	2, 379, 156	1.7	7.3	2.9	2.5
Pittsburgh	162, 924	49, 231	3.3	3.9	1.3	2.9
St. Louis	237, 172	25, 902	9.2	4.0	2.4	1.7
Means			6.5	5.3	2.2	2.5

	Total increase in growth areas 1940–1970	Total increase in decline areas 1940–1970	Ratio of population growth areas to decline areas	Distance from population centre of growth areas 1940–1970	Distance from population centre of decline areas 1940–1970	Ratio of distance to population centres of growth areas to decline areas
B. 1940–1970						
Boston	54, 200	126, 402	0.4	4.6	2.1	2.2
Chicago	238, 492	414, 350	0.6	9.7	5.0	1.9
Cleveland	9, 971	95, 671	0.1	4.8	2.3	2.1
New York	1, 011, 727	1, 314, 206	0.8	9.2	6.2	1.5
Pittsburgh	3, 371	56, 685	0.1	4.5	2.3	2.0
St. Louis	120	118, 548	0.0	5.7	2.2	2.6
Means			0.3	6.4	3.3	2.1

Table 8.1 (continued)

	Total increase in growth areas 1970–2000	Total increase in decline areas 1970–2000	Ratio of population growth areas to decline areas	Distance from population centre of growth areas 1970–2000	Distance from population centre of decline areas 1970–2000	Ratio of distance to population centres of growth areas to decline areas
C. 1970–2000						
Boston	194, 106	103, 779	1.9	3.0	3.4	0.9
Chicago	593, 317	688, 325	0.9	5.5	5.3	1.0
Cleveland	189, 464	224, 544	0.8	3.3	3.6	0.9
New York	1, 898, 966	937, 355	2.0	10.9	6.6	1.7
Pittsburgh	201, 244	176, 144	1.1	1.8	3.0	0.6
St. Louis	219, 753	274, 130	0.8	6.0	2.7	2.2
Means			1.3	5.1	4.1	1.2

1910 than were the areas of decline. Simply put, population growth was away from the centre of the cities while decline was much more likely to be nearer the centre, and in every city, growth outpaced decline in those areas originally tracted in 1910. So, from 1910 to 1940, one can say that this measure followed the hypothesis that would be predicted by the Chicago School: growth was away from the city centres, while decline was near the city centres.

The pattern from 1940 to 1970 is strikingly different. In every one of the six cities, there was greater decline in population in those areas in decline than there was growth in the growth areas. At the same time, the areas of growth were farther from the population centres of 1910 than were the areas of decline. This finding also supports the hypothesis of the location of growth of the Chicago School. Since these only reflect the areas tracted in 1910, by 1940 they are mostly the centres of larger urban regions. The pattern shown from 1970 to 2000 is more mixed. In three cities, Boston, New York and Pittsburgh, the growth areas grew more than the decline areas declined. However, in only two cities were the growth areas farther from the 1910 population centres than were the decline areas (on average). In three of the cities, Boston, Cleveland and Pittsburgh, the growth areas were closer to the population centres than were the decline areas. In sum, there is a consistent pattern of growth or decline in the six cities for which we have data for 1910–1940 and 1940–1970, and it supports the Chicago School. However, the patterns found since 1970 are less consistent than the earlier patterns. In three of those cities, growth would be that predicted by the Chicago School, but in the other three there is support for the hypotheses associated with the New York School, implying a resurgence of population in the centres.

These patterns closely mirror what locals know about each of these cities. For instance, in Chicago from 1910 to 1940, population grew rapidly outside the area of settlement, while high density areas lost population. From 1940 to 1970, most of the downtown area lost population, while there were some spots of downtown growth. From 1970 to 2000, there was a decline in the South Side, while growth continued in some other areas. The development in New York City is not all that patterned or stable. From 1910 to 1940, growth was away from downtown Manhattan, with decline near the original area of population growth. There was development throughout the city's five boroughs, especially growth in Brooklyn, the Bronx and Queens, while Manhattan became less dense. From 1940 to 1970, there was growth outwards, coupled with growth and decline in areas previously settled. From 1970 to 2000, there was continued growth in Staten Island, with growth and decline areas all over the rest of city.

Table 8.2 presents the results for the forty-nine areas for the periods 1940–1970 and 1970–2000 based upon the areas tracted in 1940. The nine areas of Atlantic City, NJ, Buffalo, NY, Boston, MA, Kansas City, MO, Nashville, TN, Trenton, NJ, Providence, RI, Milwaukee, WI and Richmond, VA had more decline in areas of decline than growth in areas of growth. In all cities, the areas of growth were farther from the population centres in 1940 than were the areas of decline. Once again, this supports the hypothesis of the Chicago School. For the same areas from 1970 to 2000, the following pattern is found: thirty-seven of the forty-nine city areas

Table 8.2 Patterns of growth and decline in the forty-nine urban areas from 1940–1970 and 1970–2000. Based upon areas tracted in 1940

	Total increase in growth areas 1940–1970	Total increase in decline areas 1940–1970	Ratio of population growth areas to decline areas	Distance from population centre of growth areas 1940–1970	Distance from population centre of decline areas 1940–1970	Ratio of distance to population centres of growth areas to decline areas
Akron, OH	167, 865	50, 677	3.3	5.5	1.4	3.9
Atlanta, GA	507, 171	90, 893	5.6	7.4	1.2	6.1
Atlantic City, NJ	6, 577	21, 285	0.3	1.9	0.5	3.6
Augusta, GA	43, 579	23, 246	1.9	3.9	1.1	3.5
Austin, TX	86, 501	11, 584	7.5	2.8	0.9	3.0
Baltimore, MD	248, 253	202, 161	1.2	4.7	1.7	2.7
Bergen-Passaic, NJ	17, 673	12, 505	1.4	1.7	0.8	2.2
Birmingham, AL	63, 978	56, 997	1.1	5.6	1.7	3.4
Boston, MA	79, 487	212, 272	0.4	4.5	2.5	1.8
Buffalo, NY	42, 031	134, 863	0.3	5.5	2.1	2.6
Chicago, IL	564, 891	593, 372	1.0	10.0	5.2	1.9
Cincinnati, OH	422, 862	125, 638	3.4	8.2	3.2	2.5
Cleveland, OH	794, 928	223, 501	3.6	12.0	2.7	4.5
Columbus, OH	152, 302	57, 991	2.6	4.2	1.0	4.2
Dallas, TX	167, 338	63, 149	2.6	5.5	1.3	4.2
Dayton, OH	86, 146	42, 317	2.0	4.0	0.8	5.1
Denver, CO	137, 248	59, 123	2.3	4.2	1.4	2.9
Detroit, MI	1, 183, 959	492, 347	2.4	12.2	3.2	3.8
Duluth–Superior, MN–WI	36, 582	22, 846	1.6	6.8	1.9	3.5
Flint, MI	206, 814	22, 621	9.1	5.8	1.2	4.8
Hartford, CT	171, 679	32, 978	5.2	5.5	1.4	3.9
Houston, TX	162, 130	73, 152	2.2	5.5	0.7	8.2
Indianapolis, IN	370, 454	86, 853	4.3	7.2	1.5	4.9
Kansas City, MO	43, 939	105, 630	0.4	4.8	2.0	2.3
Los Angeles, CA	4, 436, 739	198, 867	22.3	16.9	3.3	5.2

Table 8.2 (continued)

	Total increase in growth areas 1940–1970	Total increase in decline areas 1940–1970	Ratio of population growth areas to decline areas	Distance from population centre of growth areas 1940–1970	Distance from population centre of decline areas 1940–1970	Ratio of distance to population centres of growth areas to decline areas
Louisville, KY	139, 293	85, 338	1.6	5.5	1.6	3.3
Macon, GA	73, 645	20, 872	3.5	3.3	2.0	1.6
Memphis, TN	265, 954	66, 950	4.0	6.2	2.0	3.0
Milwaukee, WI	108, 478	141, 269	0.8	5.3	1.6	3.3
Minneapolis–St. Paul, MN	421, 161	161, 017	2.6	9.8	5.5	1.8
Nashville, TN	22, 980	44, 054	0.5	2.7	1.1	2.5
New Haven, CT	46, 647	46, 702	1.0	3.1	1.3	2.5
New Orleans, LA	217, 491	112, 799	1.9	4.8	1.2	4.1
New York, NY–NJ	2, 663, 429	2, 133, 797	1.2	9.8	6.4	1.5
Oklahoma City, OK	230, 134	69, 569	3.3	5.6	1.5	3.8
Philadelphia, PA	450, 384	447, 207	1.0	9.1	3.0	3.1
Pittsburgh, PA	521, 866	328, 405	1.6	9.5	4.6	2.1
Portland, OR	125, 019	39, 548	3.2	5.4	2.1	2.5
Providence, RI	47, 386	79, 386	0.6	3.5	1.2	3.0
Richmond, VA	45, 649	48, 737	0.9	3.1	1.1	2.8
Rochester, NY	87, 824	50, 403	1.7	4.7	1.0	4.8
San Francisco–Oakland, CA	400, 727	111, 277	3.6	9.4	5.3	1.8
Savannah, GA	99, 339	29, 650	3.4	4.2	1.1	3.7
Seattle, WA	121, 304	51, 515	2.4	6.3	1.5	4.2
St. Louis, MO–IL	805, 615	249, 436	3.2	11.2	3.3	3.4
Syracuse, NY	45, 244	36, 906	1.2	3.6	0.9	3.9
Toledo, OH	56, 401	47, 077	1.2	5.5	1.6	3.5
Trenton, NJ	13, 260	25, 376	0.5	3.3	1.0	3.4
Washington, DC–MD–VA	255, 430	162, 162	1.6	5.1	1.8	2.7
Means			2.7	6.0	2.0	3.4

Table 8.2 (continued)

	Total increase in growth areas 1970–2000	Total increase in decline areas 1970–2000	Ratio of population growth areas to decline areas	Distance from population centre of growth areas 1970–2000	Distance from population centre of decline areas 1970–2000	Ratio distance to population centres of growth areas to decline areas
Akron, OH	12, 748	84, 858	0.2	7.0	3.2	2.2
Atlanta, GA	177, 470	163, 714	1.1	10.2	4.8	2.1
Atlantic City, NJ	8, 048	11, 139	0.7	2.6	0.9	2.8
Augusta, GA	9, 128	27, 985	0.3	3.9	2.9	1.3
Austin, TX	75, 162	22, 713	3.3	5.9	2.4	2.4
Baltimore, MD	22, 302	276, 344	0.1	2.8	3.1	0.9
Bergen-Passaic, NJ	15, 736	11, 338	1.4	1.5	0.8	1.9
Birmingham, AL	6, 854	90, 662	0.1	3.9	3.1	1.2
Boston, MA	63, 548	115, 896	0.5	2.7	3.4	0.8
Buffalo, NY	1, 574	177, 156	0.0	2.2	1.9	1.2
Chicago, IL	240, 978	717, 383	0.3	5.3	5.5	1.0
Cincinnati, OH	130, 382	204, 287	0.6	13.0	5.7	2.3
Cleveland, OH	167, 185	494, 422	0.3	15.4	7.1	2.2
Columbus, OH	2, 524	121, 250	0.0	5.8	2.8	2.0
Dallas, TX	73, 061	80, 013	0.9	6.2	3.1	2.0
Dayton, OH	557	88, 717	0.0	0.1	2.4	0.1
Denver, CO	25, 316	68, 550	0.4	5.5	3.2	1.7
Detroit, MI	40, 561	891, 058	0.0	12.8	6.2	2.1
Duluth–Superior, MN–WI	4, 786	16, 289	0.3	4.7	4.9	0.9
Flint, MI	54, 267	91, 285	0.6	8.8	2.6	3.3
Hartford, CT	27, 116	65, 867	0.4	6.0	2.4	2.5
Houston, TX	61, 231	120, 883	0.5	4.6	3.1	1.5
Indianapolis, IN	110, 022	177, 621	0.6	9.5	3.4	2.8
Kansas City, MO	1, 675	121, 869	0.0	4.6	2.0	2.3
Los Angeles, CA	2, 863, 253	365, 347	7.8	22.7	17.3	1.3
Louisville, KY	5, 609	118, 429	0.0	5.5	4.4	1.2

Table 8.2 (continued)

	Total increase in growth areas 1970–2000	Total increase in decline areas 1970–2000	Ratio of population growth areas to decline areas	Distance from population centre of growth areas 1970–2000	Distance from population centre of decline areas 1970–2000	Ratio distance to population centres of growth areas to decline areas
Macon, GA	43, 511	35, 775	1.2	8.8	2.3	3.8
Memphis, TN	17, 832	171, 782	0.1	10.0	3.8	2.7
Milwaukee, WI	10, 355	123, 358	0.1	2.7	2.6	1.0
Minneapolis–St. Paul, MN	64, 710	175, 598	0.4	10.7	7.6	1.4
Nashville, TN	3, 023	41, 524	0.1	1.6	2.2	0.7
New Haven, CT	8, 635	21, 259	0.4	2.7	1.4	1.9
New Orleans, LA	78, 701	189, 669	0.4	9.0	3.2	2.9
New York, NY NJ	1, 123, 593	1, 159, 975	1.0	10.9	8.2	1.3
Oklahoma City, OK	38, 616	77, 121	0.5	8.2	4.0	2.1
Philadelphia, PA	60, 989	514, 502	0.1	9.1	3.4	2.7
Pittsburgh, PA	93, 570	416, 906	0.2	13.6	6.3	2.2
Portland, OR	24, 424	29, 760	0.8	6.3	2.7	2.3
Providence, RI	23, 086	21, 683	1.1	2.4	1.7	1.4
Richmond, VA	1, 564	56, 166	0.0	3.1	2.1	1.5
Rochester, NY	5, 455	95, 471	0.1	2.8	2.8	1.0
San Francisco–Oakland, CA	210, 024	99, 579	2.1	7.5	6.1	1.2
Savannah, GA	81, 812	37, 420	2.2	9.1	2.1	4.4
Seattle, WA	53, 392	28, 941	1.8	3.2	3.7	0.9
St. Louis, MO–IL	217, 889	525, 862	0.4	18.5	5.5	3.4
Syracuse, NY	5, 630	55, 949	0.1	3.5	1.6	2.1
Toledo, OH	4, 989	88, 758	0.1	4.9	2.1	2.3
Trenton, NJ	1, 522	21, 956	0.1	1.5	1.3	1.1
Washington, DC–MD–VA	35, 968	220, 271	0.2	3.4	3.4	1.0
Means			0.7	6.7	3.7	1.9

had more population decline in areas of decline than they had growth in areas of growth. In six cases, the areas of growth were closer to the population centre of 1940 than were the areas of decline. So the most urban parts of most areas follow the hypothesis derived from the Chicago School, but six of the forty-nine follow that posited by the New York School.

Table 8.3 presents similar figures based upon the 1970 tract data for the change from 1970 to 2000. Pittsburgh, PA, Buffalo–Niagara Falls, NY, Milwaukee–Racine, WI, Duluth–Superior, MN–WI, Dayton–Springfield, OH, Cleveland–Akron, OH, Saginaw–Bay City–Midland, MI, Chicago–Gary–Kenosha, IL–IN–WI, Syracuse, NY and Detroit–Ann Arbor–Flint, MI have larger population declines in the decline areas than they do growth in the growth areas. Yet, in only two areas, Bakersfield, CA and Duluth–Superior, MN–WI were the growth areas nearer to the population centre of 1970 than were the decline areas.

Though some large urban areas have experienced declines, one of the main findings is that growth continues far from the centre in most cases (especially Los Angeles), while decline is more likely to occur in the centre. However, there are definitely variant patterns. In short, using measures of the location of growth near traditional urban centres, many areas followed the pattern posited by the Chicago School for most of the twentieth century, but there are some variations in the centres of large urban areas.

Spatial Analysis of Growth Patterns

This first comparison of all the tracted areas yields some interesting results based upon very simple measures. Since the area tracted generally shifts from year to year, to a get a fuller picture it is important to use all the data available to understand changing patterns. The changing tract patterns and the changing populations in the Chicago, New York City and Los Angeles metropolitan areas were analysed. Los Angeles did not have any tracts until 1930, and at this writing the NHGIS has not completed its work with the Los Angeles tracts for that decade. New York City was one of the first tracted cities, and other areas in the metropolitan area vicinity were added in 1930 and 1940. To make the comparisons for the same areas where tract data did not exist, county data were used instead. This means that the units used are not strictly comparable over time, but the geographic areas are. Population densities were computed for all decades since 1910 for each of the three cities.

As shown in Figures 8.2 and 8.3, the areas with tract data changed over time. From this, one can see that for each metropolitan area population density did not just expand outwards; there were also spots of increase not connected with other densities. Figures 8.4 and 8.5 examine the per cent change from the earliest year tracted to the next decade and for 1990–2000. It is plain in the maps of later years that there are areas of decline near areas of growth. Put another way, the small area data reveal patterns of growth and decline that are much less obvious, especially from 1990 to 2000.

Table 8.3 Patterns of growth and decline in the 44 Urban areas from 1970–2000. Based upon areas tracted in 1970

	Total population increase in growth areas 1970–2000	Total population decrease in decline areas 1970–2000	Ratio of population growth areas to decline areas	Distance from population center of growth areas 1970–2000	Distance from population center of decline areas 1970–2000	Ratio of distance to population centers of growth areas to decline areas
Atlanta, GA	1,281,839	177,904	7.2	18.8	5.7	3.3
Augusta–Aiken, GA–SC	131,338	42,469	3.1	12.4	10.5	1.2
Austin–San Marcos, TX	550,051	33,290	16.5	9.2	3.4	2.8
Bakersfield, CA	353,347	20,870	16.9	27.0	80.4	0.3
Birmingham, AL	159,895	142,843	1.1	13.2	5.1	2.6
Boston–Worcester–Lawrence, MA	319,400	270,769	1.2	18.5	7.7	2.4
Buffalo–Niagara Falls, NY	114,936	278,006	0.4	12.0	4.2	2.9
Chicago–Gary–Kenosha, IL–IN–WI	939,383	1,116,664	0.8	22.9	11.2	2.0
Cincinnati–Hamilton, OH–KY	552,906	270,045	2.0	16.8	7.8	2.2
Cleveland–Akron, OH	484,789	634,646	0.8	21.4	12.8	1.7
Columbus, OH	411,889	176,372	2.3	10.7	4.0	2.7
Dallas–Fort Worth, TX	1,053,638	162,061	6.5	12.7	5.6	2.3

Table 8.3 (continued)

	Total population increase in growth areas 1970–2000	Total population decrease in decline areas 1970–2000	Ratio of population growth areas to decline areas	Distance from population center of growth areas 1970–2000	Distance from population center of decline areas 1970–2000	Ratio of distance to population centers of growth areas to decline areas
Dayton–Springfield, OH	94, 924	142, 015	0.7	9.3	3.4	2.7
Denver–Boulder–Greeley, CO	959, 541	122, 062	7.9	13.2	5.7	2.3
Detroit–Ann Arbor–Flint, MI	1, 088, 614	1, 189, 826	0.9	22.3	16.1	1.4
Duluth–Superior, MN–WI	18, 275	39, 073	0.5	27.6	35.0	0.8
Hartford, CT	143, 332	101, 582	1.4	12.3	5.8	2.1
Houston–Galveston, TX	1, 922, 317	263, 325	7.3	18.1	5.7	3.2
Indianapolis, IN	395, 070	178, 412	2.2	13.2	3.8	3.4
Kansas City, MO–KS	519, 455	253, 985	2.0	17.4	5.9	3.0
Los Angeles–Riverside, CA	5, 800, 702	474, 329	12.2	39.0	32.9	1.2
Louisville, KY–IN	209, 824	175, 473	1.2	11.4	5.2	2.2
Macon, GA	131, 554	42, 773	3.1	12.7	5.7	2.2
Memphis, TN–AR–MS	383, 410	205, 173	1.9	12.8	7.0	1.8

Table 8.3 (continued)

	Total population increase in growth areas 1970–2000	Total population decrease in decline areas 1970–2000	Ratio of population growth areas to decline areas	Distance from population center of growth areas 1970–2000	Distance from population center of decline areas 1970–2000	Ratio of distance to population centers of growth areas to decline areas
Milwaukee–Racine, WI	86, 460	200, 158	0.4	9.9	4.4	2.3
Minneapolis–St. Paul, MN	886, 992	218, 298	4.1	17.3	8.3	2.1
Nashville, TN	188, 981	67, 095	2.8	10.2	4.1	2.5
New Orleans, LA	284, 225	258, 849	1.1	9.4	6.7	1.4
New York–Northern, NJ–CT	1, 605, 778	1, 638, 156	1.0	21.9	11.7	1.9
Oklahoma city, OK	415, 025	99, 752	4.2	13.8	6.4	2.2
Philadelphia–Wilmington, PA–DE	830, 301	826, 758	1.0	26.9	9.5	2.8
Pittsburgh, PA	181, 348	546, 617	0.3	19.0	10.2	1.9
Portland–Salem, OR–WA	310, 223	33, 874	9.2	12.4	4.7	2.6
Providence–Fall River, RI–MA	82, 140	40, 682	2.0	7.6	4.6	1.6
Richmond–Petersburg, VA	129, 407	73, 304	1.8	10.4	3.1	3.4

Table 8.3 (continued)

	Total population increase in growth areas 1970–2000	Total population decrease in decline areas 1970–2000	Ratio of population growth areas to decline areas	Distance from population center of growth areas 1970–2000	Distance from population center of decline areas 1970–2000	Ratio of distance to population centers of growth areas to decline areas
Rochester, NY	145,368	121,944	1.2	11.7	3.5	3.4
Saginaw–Bay City–Midlan, MI	32,477	42,181	0.8	9.0	4.7	1.9
San Francisco–Oakland, CA	1,157,831	175,743	6.6	24.4	14.6	1.7
Savannah, GA	81,812	37,420	2.2	8.6	2.7	3.2
Seattle–Tacoma–Bremerto, WA	636,653	55,588	11.5	15.5	8.2	1.9
St. Louis, MO–IL	658,998	596,573	1.1	22.9	8.0	2.8
Syracuse, NY	80,488	94,897	0.8	9.3	4.4	2.1
Toledo, OH	115,048	112,314	1.0	9.3	5.4	1.7
Washington–Baltimore, DC	1,157,309	763,641	1.5	20.9	18.2	1.1
Means			3.5	15.8	9.9	2.2

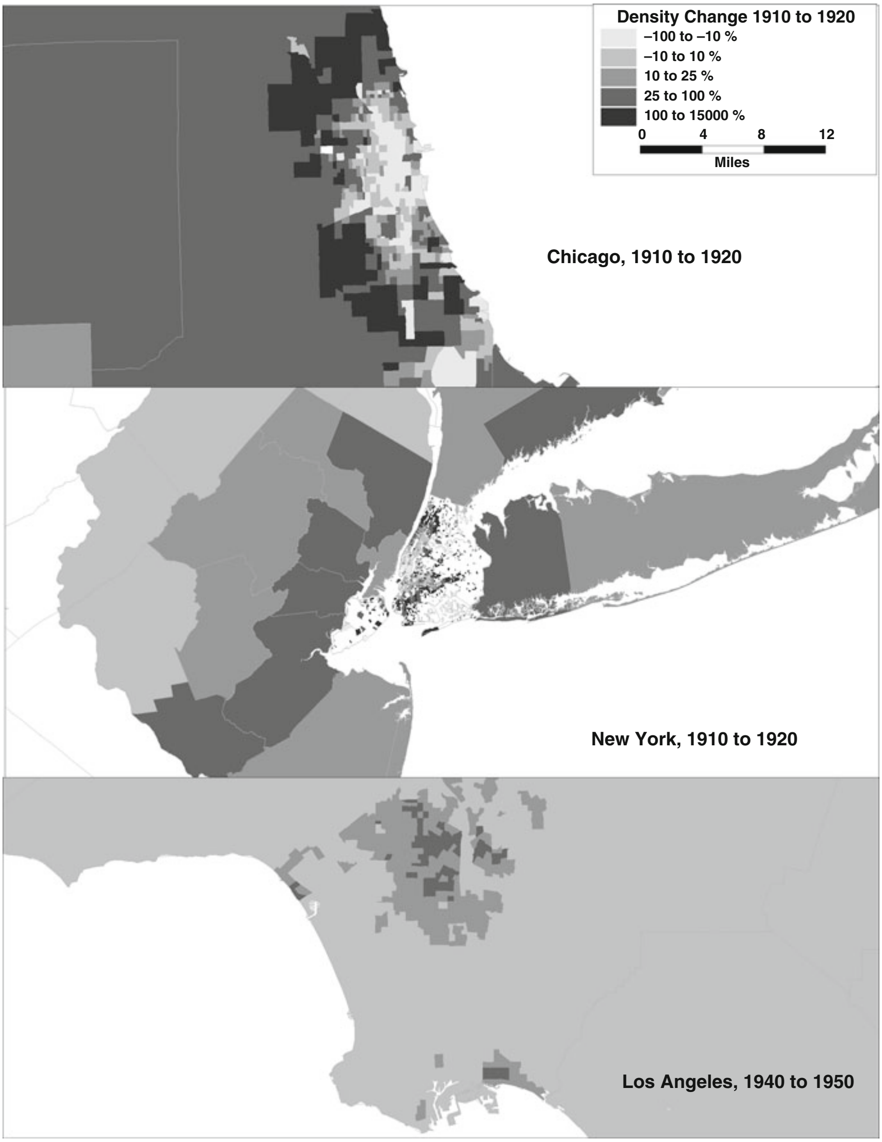


Fig. 8.4 Density change, tracted areas, Chicago and New York, 1910–1920, Los Angeles, 1940–1950

To examine the patterns of population change more rigorously, a Moran's *I* was computed for the percent population change by tract from year to year. (Per cent change was used because it effectively controls for differences in the actual size of the population in different areas.) Moran's *I* is a measure of the extent to which a variable is spatially autocorrelated. Here it would measure whether increases and

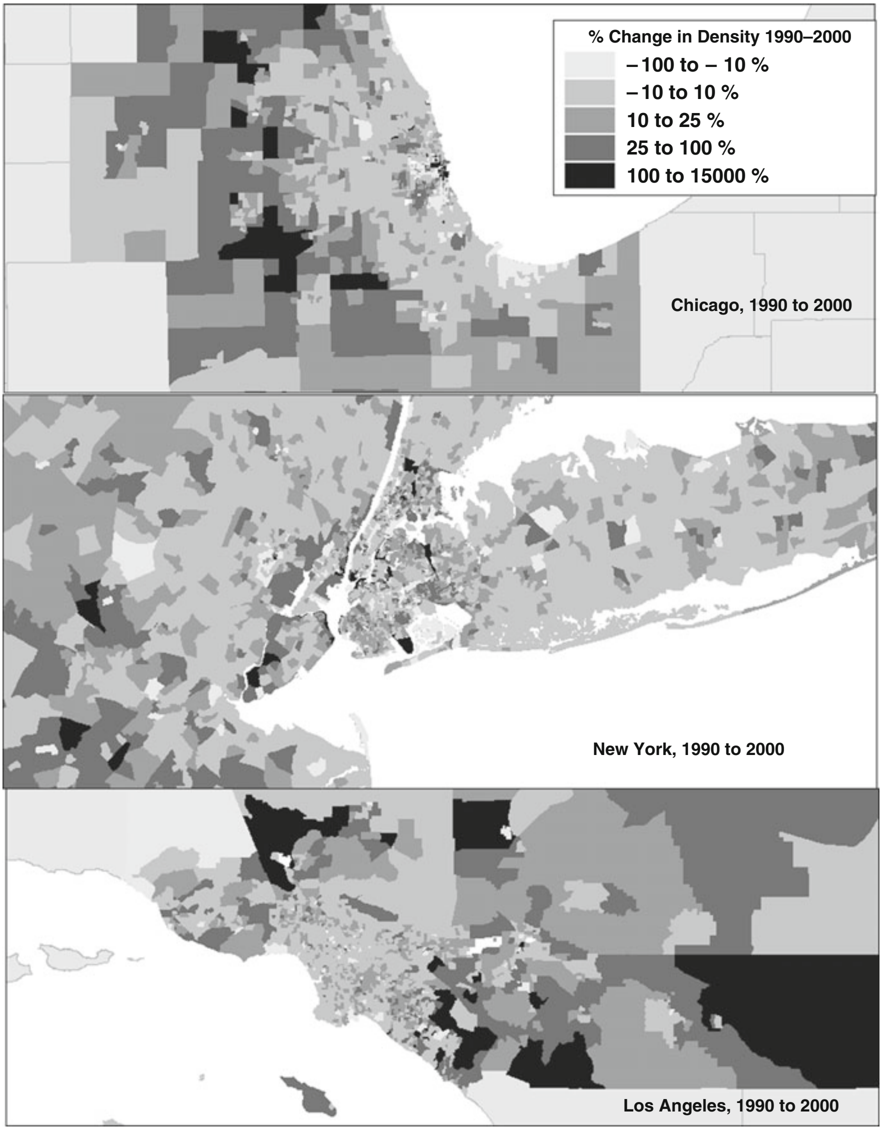


Fig. 8.5 Density change, tracted areas, Chicago, New York, and Los Angeles, 1990–2000

decreases in population would occur together (a positive Moran's *I*) or whether increases were near decreases (a negative Moran's *I*). Moran's *I* is a correlation coefficient and has the same properties, so a positive correlation means that growth is related to growth, a negative correlation that growth was related to decline, and something approaching zero means that overall there was little or no patterning of growth and decline (Moran, 1950).

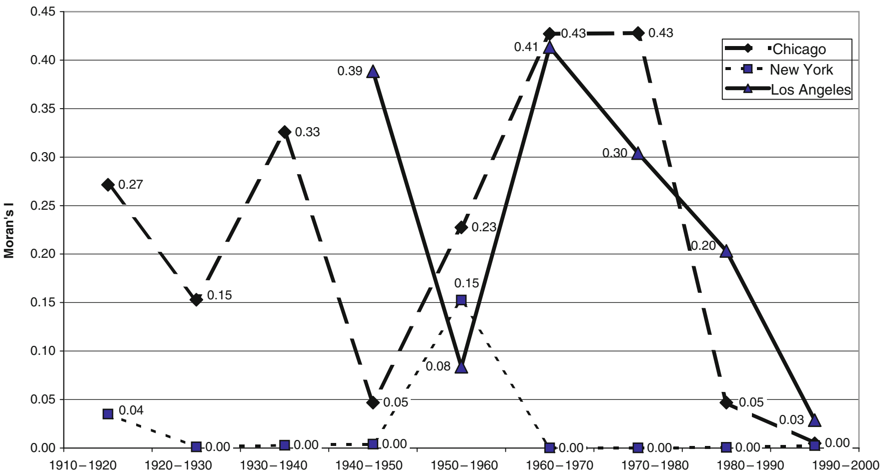


Fig. 8.6 Moran's *I* for per cent change in population for Chicago, New York and Los Angeles, 1910–1920 to 1990–2000

Figure 8.6 presents the results of this analysis, which show that the growth in the New York City metro area was less spatially autocorrelated while the region was building out than was growth in either Chicago or Los Angeles. Indeed, one could say that the pattern of growth in Los Angeles and Chicago followed the pattern posited by the Chicago School until very recently (1980–1990 for Chicago, and 1990–2000 for Los Angeles). New York, however, seems to have followed the pattern posited by the Los Angeles School more closely during its growth from 1910 through 2000, while Los Angeles and Chicago more closely followed the pattern of the Chicago School. By 1990 and 2000, however, all three cities and urban areas have virtually no spatial autocorrelation in their growth patterns and therefore are all behaving according to the Los Angeles School model.

These larger patterns can be refined using Local Indicators of Spatial Autocorrelation (LISA). This measure allows one to look at the structure of growth in a local area rather than globally, as does Moran's *I* (Anselin, 1995; Ord & Getis, 1995). The results are presented as maps, which indicate where there is no structuring of growth or decline and where positive and negative values are clustered (growth or decline in this examination), and where decline is near growth and where growth is near decline.

Figure 8.7 presents the results for Chicago for 1910–1920 and for 1990–2000. Between 1910 and 1920, it appears that decline was spatially autocorrelated in the downtown area, while growth was correlated with growth in some of the adjacent suburban counties and some areas away from the city centre. Low growth in the rest of Cook County was related to high growth in the surrounding counties and in a couple of other locations in Chicago. This is consistent with the overall pattern as depicted by the earlier Moran's *I* analysis. The patterns found for 1990–2000 indicate spatially autocorrelated growth in the downtown area near the

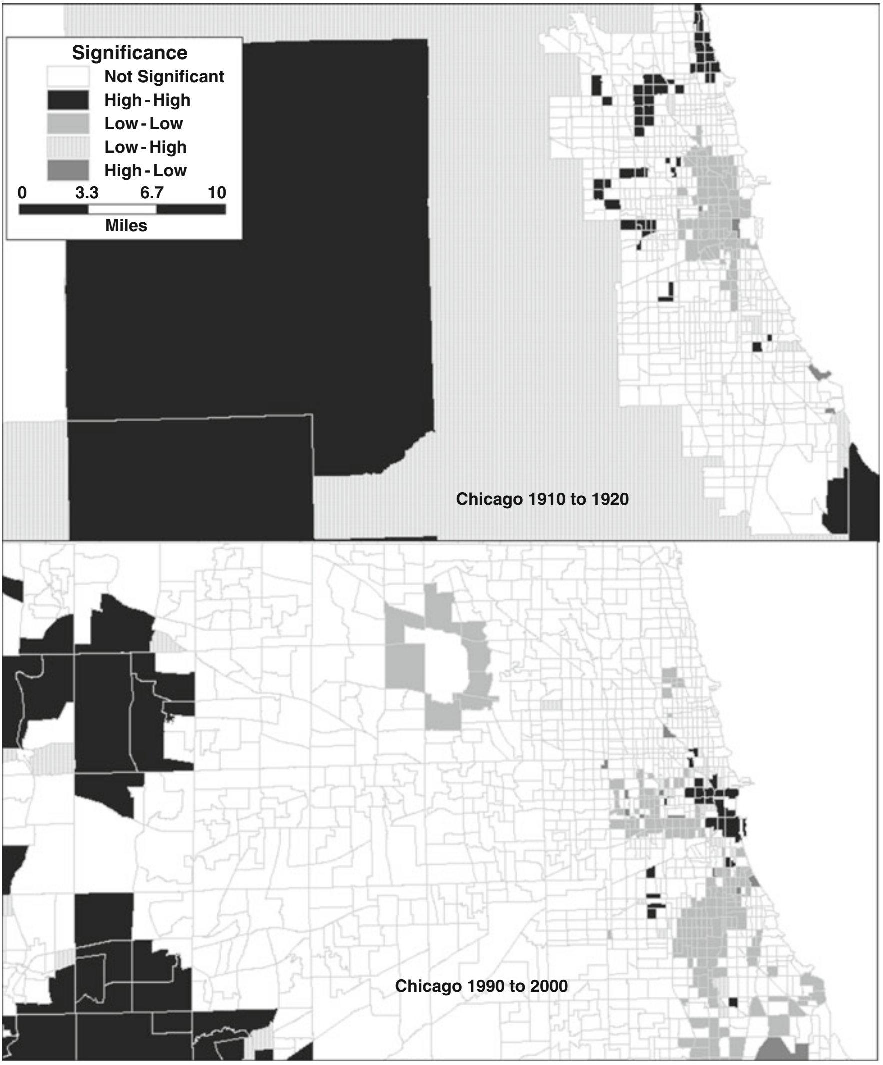


Fig. 8.7 Local indicators of spatial autocorrelation, Chicago, percent population growth, 1910–1920 and 1990–2000

lake and a decline in some other parts of urban Chicago (with spots of growth nearby), but also highly clustered growth outside the city. In short, the underlying pattern of little global or overall autocorrelation includes areas of clustered high growth (both within the city and in the suburbs), as well as decline near growth, growth near decline, and decline near decline. In the last decade or two, it appears that while some growth continued to be outside the city, there was also growth in the downtown areas. The pattern found in downtown Chicago between 1990 and 2000, which is mirrored by similar patterns found in New York City and Los

Angeles, gives some support to both the New York and Chicago schools and their models.

Conclusion

One of the hypotheses examined in this chapter suggests that in urban areas, patterns of growth take place away from urban centres, and the preceding analysis does seem to support that hypothesis. It is particularly true from 1970 to 2000 for the tracted areas in 1970. This prediction is also broadly true for those areas tracted from 1910 to 1940; it is always true for the period from 1940 to 1970 and mainly true from 1970 to 2000. If one defines the Chicago School model as the simplistic notion that growth radiates toward areas that have yet to be built up, one can conclude that this analysis generally supports the Chicago School pattern.

On the other hand, when one looks at the actual spatial patterning of growth, one can find evidence that supports exponents of the Chicago, Los Angeles and New York Schools of urban studies in various ways. The univariate Moran's *I* results indicate that earlier growth was highly structured in Los Angeles and Chicago, while in New York City there was no obvious pattern. More recently, it appears that all three regions have become like New York City (but following the image developed by the Los Angeles School), and that their recent growth has little of the overall spatial pattern that would be expected by those adhering to the Chicago School model.

At the same time, the Local Indicator of Spatial Autocorrelation (LISA) indicates specific growth patterns in specific areas, but the areas simply do not follow some master plan for growth. Downtowns are growing, just like outlying areas. This implies that to understand the patterns of growth and decline in cities and urban areas, one may need more than a simple explanation from a master pattern or model, whether that of the Chicago, Los Angeles or New York schools. Instead, factors related to overall population change and current land use patterns, the characteristics of transportation facilities and other factors all need to be considered. This is especially true after the initial settlement of large urban areas has occurred.

Much remains to be done to understand the patterns uncovered in the analysis presented here. More spatial information will enrich the analysis, particularly by making use of all the available tracted areas, rather than just the three large regions of New York, Los Angeles and Chicago. The addition of tract-level census data about race and ethnic status, social class (including education, occupation and income) and land use will permit researchers to estimate spatial regression models that predict growth from decade to decade. Other measures of dispersion should be developed and used. Data at the minor civil division or even the enumeration district level could make it possible to use uniform units from decade to decade.

Despite the need for additional research, this first-ever look at the growth patterns of major United States urban areas, which relies upon population data and geo-referenced tract boundaries, provides a glimpse into the power and possibility of analysis using newly emerging sources of data with spatial techniques. One can give real meaning to notions of dispersion or spatially structured growth far beyond what older models could articulate. With these techniques, true comparison can replace more impressionistic work, and patterns of growth and decline can become the subject of rigorous analysis. From this analysis it is obvious that the loose formulation of the Los Angeles and Chicago Schools should give way to more nuanced statements about change over time. Growth can be both patterned and multi-focal. Decline and growth can go hand in hand, and have done so in the major metropolises. Whatever patterns are found must take into account both a local and global dimension.

Even with the relatively simple analysis presented here two things stand out:

(1) During early periods of growth, growing 'out' is the norm in virtually all areas. This is true even for Los Angeles, which seems to have followed a quite common model. New York, on the other hand, was much more complex.
(2) In later periods, however, growth patterns become much less obvious. All three of the major areas examined here now have globally quite unstructured patterns of growth, but this may not be true of all areas.

Plainly, there is not one simple model that encapsulates growth in all areas for all periods. Rather, various factors affect growth at various times. Now with the NHGIS, along with developing temporal-spatial techniques, it will be possible to address the real complexity that underlies urban growth and change.

Acknowledgement I acknowledge the assistance of Jordan Segall, Ahmed Lacevic and Susan Weber. Comments from Myron Gutmann and anonymous reviews were very helpful. This work was in part supported by grants from the National Science Foundation (0226279, 0618456, and 0647902)

Notes

1. In his path-breaking study of sidewalk vendors on Sixth Avenue in Manhattan, Duneier writes: 'New York City and Greenwich Village are unique in many ways . . . I cannot hope to show how the sidewalk works in low-income neighbourhoods where the majority of dense sidewalk interactions occur among members of the same class or social group . . . I must leave it to my readers to test my observations against their own, and hope that the concepts I have developed to make sense of this neighbourhood will prove useful in other venues' (Duneier, 1999: Introduction).
2. At this writing, there are some areas where tract data exist but are not yet available from the NHGIS. These will be included in the final NHGIS release.
3. Areal interpolation was used to estimate the population in earlier decades in the areas that followed the 2000 tract boundaries. In this method, population is assumed to be uniformly distributed throughout a given area. If part of a tract is moved to another tract, the population that is moved is proportional to the

area that this moved. This is a standard method, but does have some limitations. However, given the available information, it was the only method that could be used.

References

Anselin, L. (1995). Local Indicators of Spatial Association – LISA. *Geographical Analysis*, *27*, 93–115.

Berry, B. (1961). City size distribution and economic development. *Economic Development and Cultural Change*, *9*, 573–587.

Berry, B. & Pred, A. (1961). *Central place studies: A bibliography of theory and applications*. Philadelphia: Regional Service Institute.

Bogue, D. J. (1950). *The structure of the metropolitan community: A study of dominance and subdominance*. Ann Arbor, MI: Horace H. Rackham School of Graduate Studies, University of Michigan.

Bulmer, M. (1984). *The Chicago School of Sociology: Institutionalization, diversity, and the rise of sociological research*. Chicago: University of Chicago Press.

Burgess, E. W. (1925). The growth of the city. In R. E. Park, E. W. Burgess, R. D. McKenzie, (Eds.), *In The City* (reprinted 1967). Chicago: University of Chicago Press.

Christaller, W. (1972). How I discovered the theory of central places: A report about the origin of central places. In P. W. English, R. C. Mayfield, (Eds.), *Man, space and environment* (pp. 601–610). Oxford: Oxford University Press.

De Vries, J. (1984). *European urbanization, 1500–1800*. Cambridge: Harvard University Press.

Duncan, B., Sabagh, G., & Van Arsdol, M., Jr. (1962). Patterns of city growth. *American Journal of Sociology*, *67*, 418–429.

Duneier, M. (1999). *Sidewalk*. New York: Farrar, Strauss and Giroux.

Fogelson, R. (1967). *The fragmented metropolis: Los Angeles, 1850–1930*. Berkeley, CA; Los Angeles: University of California Press.

Guest, A. M. (1971). Urban growth and population densities. *Demography*, *10*, 51–69.

Guest, A. M. (1975). Population suburbanization in American metropolitan areas, 1940–1970. *Geographical Analysis*, *7*, 267–283.

Guest, A. M. (1978). Suburban territorial differentiation. *Sociology and Social Research*, *62*, 523–536.

Guest, A. M. (2002). Suburban population growth and decline: The emergence of new patterns in the 1990s. Paper presented at the Annual Meeting of the Population Association of America, May 8–11, 2005, Atlanta.

Halle, D. (2003). Introduction. In D. Halle, (Ed.), *New York and Los Angeles: Politics, society and culture: A comparative view*. Chicago: University of Chicago Press.

Hawley, A. H. (1981). Human ecology: Persistence and change. *American Behavioral Scientist*, *24*, 423–444.

Jackson, K. T. (1985). *Crabgrass frontier: The Suburbanization of the United States*. New York: Oxford University Press.

Jacobs, J. (1961). *The death and life of great American cities*. New York: Random House.

Miller, D. W. (2000). The new urban studies. *Chronicle of Higher Education*, August 18: A15.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*, 17–23.

Ord, J. K. & Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis*, *27*, 286–306.

Schneider, R. (2000). The postmodern city from an early modern perspective. *American Historical Review*, *105*(5), 1668–1675.

Schnore, L. F. (1957). Metropolitan growth and decentralization. *American Journal of Sociology*, *63*, 171–180.

A. Scott, E. Soja, (Eds.). (1996). *The City: Los Angeles and urban theory at the end of the twentieth century*. Berkeley, CA: University of California Press.

Taeuber, K. E. & Taeuber, A. F. (1965). *Negroes in cities; Residential segregation and neighborhood change*. Chicago: Aldine Publishing Co.

U.S. Bureau of the Census. (1942a). Census Tract Data Files, 1940–1970 (Elizabeth Mullen Bogue Files) Electronic and Special Media Records Services Division Reference Report. Census Tract Data Files *Archival Research Catalog (ARC) identifier:* 617523 *Collection:* Records of the Community and Family Study center (CFS), University of Chicago. National Archives and Records Administrion, 1999.

U.S. Bureau of the Census. (1942b). Census Tract Data on Population and Housing, New York City: 1940. Welfare Council Committee on 1940 Census Tract Tabulations for New York City. New York, NY, 1942.

U.S. Bureau of the Census. (2007). *Census atlas of the United States*. Washington, DC: Government Printing Office.

Wolch, J. & Dear, M. (1993). *Malign neglect: Homelessness in an American city*. San Francisco: Jossey Bass.

Chapter 9
Economic Transition and Social Inequality in Early Twentieth Century Puerto Rico

Katherine J. Curtis

Introduction

Puerto Rico provides a persuasive case for the study of economic, demographic and stratification transitions. The island was a Spanish colony for nearly 400 years until possession was transferred to the U.S. in 1898 as a result of the Spanish-American War. This political transfer created fundamental shifts in the relationship between the population, the land and the state that had significant demographic and social consequences for the racially diverse population. Under U.S. control, Puerto Rico's economic infrastructure was entirely revamped with substantial effects on the system of land use, the regional distribution of the population and the social, and largely racialized, class structure (Dietz, 1986). Previous research describes a process of uneven economic and social development that varies spatially and temporally according to shifts in geographically constrained crop production, prompted by the change in economic interests that accompanied the transfer of political rule. This characterization motivates the current analysis of one socioeconomic consequence of the economic shift: racial inequality in literacy. I explicitly investigate whether the type and extent of crop production is associated with the degree of racial inequality in literacy rates among Puerto Rican *municipios*.

It has been a tenet of scholarly literature that the structures of social and economic life within the American tropics were fundamentally determined by the social-environmental linkages associated with the 'after-dinner crops', especially crops dominant in Puerto Rico: sugar, coffee and tobacco (Wagley, 1960; Steward, 1956). The social organization in sugar-producing areas is distinct from the organization in coffee- and tobacco-producing areas; sugar production is based on a plantation model characterized by a relatively large and highly stratified labour structure, whereas tobacco and coffee production is based on a family-farm model that is smaller and less stratified. Importantly, sugar production in Puerto Rico, like many plantation crops, has an historical dependence on African slave labour. The

K.J. Curtis (✉)
Department of Community and Environmental Sociology, University of Wisconsin-Madison, Madison, WI 53706, USA
e-mail: kcurtis@ssc.wisc.edu

M.P. Gutmann et al. (eds.), *Navigating Time and Space in Population Studies*, International Studies in Population 9, DOI 10.1007/978-94-007-0068-0_9,

different organizational structures characteristic of the different crops might produce varying degrees of racial inequality. According to what I call the *stratifying* thesis, higher racial inequality is presumably found in areas shaped by a plantation model of production that depends on an often non-white subordinate labouring class.

Alternatively, in what I call the *equalizing* thesis, improvements in infrastructure were made in the years immediately following U.S. governance, including greater access to education. In general, better and more equal access to education is one way of reducing inequality in a society. Within the Puerto Rican context, while advances were made across the island, more capital was flowing into sugar-producing areas. Therefore, greater improvements in access to education and a narrowing in racial inequality in literacy may have resulted in sugar-producing areas relative to areas dependent on other crops.

No statistical analysis of racial inequality, to my knowledge, has been conducted for the early decades of U.S.-governed Puerto Rico. Historical accounts suggest that a racial hierarchy was in effect and had meaningful consequences even though official policies were silent on the issue of race. Racist policies prevailed throughout much of the nineteenth century. Further, analyses of more contemporary periods suggest that skin colour had implications for education, occupation and income (see, for example, Tumin (1961)). Scholarly accounts of earlier and later periods and of the racial climate in mainland U.S. during the early twentieth century, give little reason to suspect that the racial hierarchy would have weakened in the early years of U.S. governance.

The question remains whether the extent of racial inequality was larger in sugar-producing areas relative to areas dependent on other crops; it has heretofore not been systematically examined in the Puerto Rican context. Importantly, a spatio-temporal process is implicit in both of the propositions – one suggesting greater racial inequality and one suggesting less racial inequality – through the central focus on crop production. A review of Puerto Rico's economic and population history suggests that the process of societal development is largely influenced by the environment – the mountainous regions did not permit sugar production and, therefore, did not invite the larger populations that accompanied industry expansion in the early decades of the twentieth century. Drawing from this work, space is an inherent factor in the relationship between crop production and social development (one component of which is racial inequality); space acts as a variant in the large-scale temporal distribution of socio-economic status and a conditioning factor in the relationship between economic production and inequality. That is, the process of economic change and its social consequences do not unfold evenly across the geographic landscape, but vary according to the type of production. Likewise, time is an inherent factor in the relationship between economic production and inequality It acts as a variant in the spatial distribution of economic production and population distribution, such that shifts in the economic base and, in turn, population distribution and inequality, unfold over time. That is, change is not immediate but occurs over time.

This chapter is fundamentally descriptive and is part of a larger project examining ecological- and individual-level data to explore spatial and temporal processes of change in literacy inequalities at a unique moment in Puerto Rico's history – the transfer from Spanish to U.S. rule. I begin this analysis by exploring the spatial and temporal distribution of literacy across the island in 1910 and 1920, and, given limitations on detailed crop data, use 1899 as a baseline to contextualize general shifts according to pre- and post-U.S. governance. I then turn to an analysis of racial literacy ratios over time and across space during the early years of U.S. governance in Puerto Rico. Within this focus, I analyse the relationship between racial differences in literacy ratios and crop production as well as population composition using exploratory spatial data analysis (ESDA) techniques available in GeoDa (Anselin, 2003) that incorporate a temporal component and regression techniques that simultaneously examine spatial and temporal dynamics.

Background and Theory

Economic Restructuring

The acquisition of Puerto Rico by the U.S. in 1899 occurred in a climate of 'national destiny' steeped in Frederick Jackson Turner's frontier thesis, which argued that conquering new frontiers was essential to the country's progress (Dietz, 1986: p. 80). Conditions on the island generally improved even though the U.S.'s primary interest in Puerto Rico was military and economic. For example, health care, communications, transportation and civil services were established and enhanced. Educational institutions also expanded; there were 529 public (primary) schools in 1898 and 733 in 1901 and by 1910 there were 1,025 elementary schools and 19 high schools (Dietz, 1986: p. 84; see also Osuna 1975).

Civic infrastructure was not the central concern of the new government. Indeed, development on the island had a distinct and narrow economic focus. Corporate capital from the mainland flooded the island's sugar industry and trade policies further supported development of the sugar industry by crippling the export of other goods (Dietz, 1986; Steward, 1956). Tariffs on coffee and tobacco to Puerto Rico's main markets in Spain and the Spanish colonies increased immediately after Puerto Rico became foreign to the crown. For example, tariffs increased for tobacco imported to Cuba from 15–20 cents per pound to 5 dollars per pound (Dietz, 1986). Despite the setback, the tobacco industry eventually recovered when the U.S. found some economic value in supporting its production in Puerto Rico.

The impact was more devastating for Puerto Rico's coffee industry. The U.S. imported coffee from Brazil and was interested in maintaining the already established trade relationship because Brazil was safely outside of the hurricane belt and was a much larger producer than Puerto Rico; therefore, a greatly cheaper product could be imported from Brazil than Puerto Rico (Wolf, 1956: p. 197). Consequently, Puerto Rican coffee received no economic support from the U.S. and, like tobacco, was subject to high tariffs once the island was no longer part of the Spanish crown. The doors to key markets were virtually closed to Puerto Rican coffee producers. The decline in coffee production led to a redistribution of credit flow into other areas of the island and undermined the existing credit system (Wolf, 1956). In response, workers looked for other sources of income and, no longer tied to the land, engaged in regional out-migration.

The shift in economic interest altered socio-economic patterns across the Puerto Rican landscape. Because the new socio-economic patterns were rooted in the changing crop production that advanced the sugar industry, they also took on a distinct geographical pattern. The restructured economy produced new and racialized occupational groups and economic classes and modified old ones; in general, whites were among the landholders and operators while non-whites composed the labouring class. The economic transition also spurred a labour-motivated internal migration and urbanization that the island had not before experienced.[1] Existing regional distinctions sharpened and reflected differences in production and systems of racial exploitation (Wolf, 1956: p. 64). For example, population growth between 1899 and 1910 was greater in the sugar-producing regions; the island's population grew by 17%, yet Guánica, the largest sugar *municipio*, grew by 121% (Dietz, 1986: p. 125). As sugar productivity intensified, it required more land and more wage labourers. The sugar industry created and attracted landless wage labourers who were largely of African descent. The industry's labour requirement not only affected the island's population distribution, but also reshaped the relative and racialized socio-economic distribution of the island's residents.

By the time the U.S. arrived on the island, the sugar industry had virtually halted – many plantations were no longer under cultivation and interest rates were at an all time high – but U.S. capital made quick work of building a larger, more efficient and more productive enterprise. The new system gave rise to a new class of competitors called *colonos*. Independent peasant sugar cane farmers could not compete with the economically privileged *colonos* in a structure of production that required a close coordination between crop production and factory processing. The complex status hierarchy of the sugar industry made it distinct from other crop industries. The sugar industry was organized around foreign capital, absentee ownership, local managers and foremen and wage labourers ranging from skilled processors to unskilled field workers.

The sugar industry had a sophisticated stratified system defined. The question of central concern in the current analysis is the extent to which it was stratified along racial lines. The industry's roots in slavery suggests that it was (Scarano, 1984), and the island's historical race policies and American ideas about race could have further promoted racial stratification in Puerto Rico. The overall level of literacy may have increased as a result of educational expansion, but racial inequality in literacy might also have increased. Not all Puerto Ricans may have had equal access to the expanding educational institutions and access might have varied according to race. Moreover, the extent of racial differences in access and eventual literacy may have varied further by the local racial stratification system promoted by the specific system of crop production.

Economic Restructuring and Inequality Dynamics

In general terms, economic restructuring can either reduce or increase inequality in the area undergoing change. The relative standing of potential labourers can be reorganized in the new economy depending on industry demands and requirements. For example, in the context of gender wage inequality in the contemporary U.S., scholars suggest that the wage gap decreased because of reduced male earnings (Armstrong, 1996) and increased female income (DiPrete & Forrestal, 1995; Gittleman & Howell, 1995) that accompanied the manufacturing-to-service industry shift since, it is argued, the service industry valued more traditionally female skills and experience. Others have argued that wage inequality has actually increased because, in a sense, the gendered hierarchy has been reapplied to the new economy; males occupy higher status and higher paying positions in the service economy while women fill low-wage positions (Jensen, 1989; McDowell, 1991; Reskin & Roos, 1990). In this example, women remain more vulnerable members of the new economy (McCall, 1998: p. 383).

In her spatially informed analysis, McCall (1998, see also McCall (2000)) finds that the U.S.'s new service economy increased and decreased the wage gap depending on the spatial context and depending on the measure of industry change; inequality increased in regions where manufacturing was dominant (the 'old' economy), but decreased in areas where the new service industry was dominant. Yet, there was a difference in the inequality dynamic when industry was measured in terms of growth rather than dominance. Areas that experienced growth in the new economy did not witness a reduction in inequality but, instead, saw an increase in inequality through benefits to men's wages.

I extend this framework to the study of racial inequality in historical Puerto Rico. Economic restructuring had two important consequences for inequality. First, it required producers to have significant capital to compete successfully with the U.S.-owned, high-production sugar plantations. Second, the new sugar economy also created a large demand for wage labour since sugar production required a substantial force of field workers and created a large labour supply of former producers of economically obsolete crops or capital-poor operations. These changes had potential implications for an increase or decrease in racial inequality given the racially dichotomized owner–labourer class structure.

In the Puerto Rican context, the black or 'darker' members of the population were among the most vulnerable in the new economy. Although these members had skills to compete in the new economy, they lacked capital since this race group historically occupied socially and economically subordinate positions. The economic restructuring might have magnified this subordination by eliminating alternative markets that had greater autonomy (such as coffee) and demanding wage labour. Further, the U.S. immediately expanded the education system, yet access might have been uneven. Children in urban areas had greater access than rural children and, importantly, darker – phenotypically expressed and socially perceived – Puerto Ricans were overrepresented in rural, crop-producing areas, especially sugar-producing areas. Additionally, some landowners opposed increased education 'out of fear that no one would then wish to do unskilled, low-paid agricultural work' (Dietz, 1986: p. 129; see also Bobonis and Toro 2007). Together, these factors stimulated by the U.S. occupation might have increased racial inequality in literacy between white and non-white Puerto Ricans.

Yet, there are reasons to suspect that racial inequality might have diminished in the years immediately following U.S. governance. One could argue that the new economy could have reduced racial inequality by providing sugar labourers with opportunities for social mobility. Literacy rates are used as a proxy for social and economic standing. The new economy increased the infrastructure across the island, including public schools and the U.S. sugar industry was superior in production and profit relative to earlier versions of the educational and economic structures under Spanish rule and, perhaps, gave residents of sugar regions an advantage, including non-white sugar labourers. The generally improved conditions might have reduced racial inequality in literacy by increasing the literacy of the children of non-white labourers. In the current analysis, I empirically assess whether the US-led economic restructuring, indicated by crop production, increased or decreased racial inequality in literacy in Puerto Rico.

Physiographic Characteristics

The spatial dimension of the distribution of inequality is rooted in the physical location of the various industries. Puerto Rico's new economy favoured the sugar industry, yet sugar was not and could not be produced everywhere on the island. Sugar cane needs large amounts of water and at least a 10-month growing season (Picó, 1950). The strict environmental requirements were met in the lowlands and adjacent lower hills of Puerto Rico, although some attempts were made to produce the cash crop in less perfect conditions (for example, in Vieques, 'in spite of the scarcity and poor distribution of rainfall, the backwardness of production techniques and the continuously diminishing yields' (Picó, 1950: p. 209)). Coffee and tobacco production were similarly concentrated according to environmental factors, namely topography and climate. Coffee was found in the rugged western central mountains, whereas tobacco was produced in the gentle eastern central mountains (see also the analyses by Griffith (1979, 1992) of spatial autocorrelation in crop production for late twentieth-century Puerto Rico).

Environmental features heavily shaped the distribution of crop production in Puerto Rico and, in turn, the social life that was structured around the crop. Sugar production was organized around a plantation system with a complex hierarchy and significant wage labour demands, whereas coffee and tobacco were largely produced by small, independent farmers and their families (Steward et al., 1956). As a social institution, slavery is one historical aspect of the plantation system and sugar industry that had important implications for the racial distribution of the island's population. Geographically, the distribution of the black population corresponded with sugar production (Scarano & White, 2007). In addition to the racially stratified system characteristic of sugar production, racial inequality has a greater potential in the sugar producing areas due to the disproportionate presence of a non-white population relative to other areas in Puerto Rico.

Race Before U.S. Governance

Race and class structure were the strongest forms of subordination under Spanish colonialism. The 'purity of blood' criteria were applied in Puerto Rico; whites, Catholic Spaniards and American natives were among the 'pure' and advantaged, while people of African slave decent and the immigrant Jewish and Islamic populations were among the disadvantaged (Cubano Iguina, 2006). During the period of slave importation (the eighteenth and first half of the nineteenth centuries), the coloured population was legally excluded from certain occupations, government offices and racial intermarriage. Slavery was abolished in Puerto Rico in 1873, segregationist laws were eliminated from the books in the 1880s and by 1893 the title of *Don* (meaning a person of some importance) could be applied to black people with the recognized economic status. But both progressive racial views and racism persisted before and after the abolition of formal policies – intermarriage occurred before the 1880s and a 'private' racism persisted afterwards (Cubano Iguina, 2006). The Mortgage Law of 1880 reflected an intermediate point in the movement from the public (formal) to the private (informal) type of racism; the law was intended to keep former slaves and their descendants available as labourers by making it difficult for them to own or access land (Carrasquillo, 2006: p. 60).

Race, it seems, was largely embedded in issues of class in early twentieth century Puerto Rico. Intermarriage was most common among the lower class; blacks were wooed by politicians as 'urban labourers' and elites used scientific racism to affirm superiority over the peasant class regardless of skin colour. Indeed, peasants who looked white were not considered white since they were 'polluted' by their living conditions and interactions with blacks (Carrasquillo, 2006: p. 70). Racial harmony may have occurred among the common people through intermarriages and based in shared poverty, but racial tensions persisted in the larger society through the maintenance of class structure by elites.

Race After U.S. Governance

Race had been excluded from Spanish censuses after the abolition of slavery, but was reintroduced with the first U.S. enumeration. The extent to which the racial classification reflected race rather than class, however, is debatable. 'Passing' as white became desirable for all Puerto Ricans after the U.S. occupation given U.S. race politics, but it was more feasible for some, typically for members of the upper class. Debates quickly ensued about the whiteness of various groups of people and the mountain people and rural folk in general were at the centre of the discussion (Carrasquillo, 2006). Not being able to 'whiten' was equated with not being able to improve culturally and, perhaps, intellectually and economically (see Loveman and Muniz (2007) for a thorough discussion and analysis of the whitening of Puerto Rico between 1910 and 1920).

The objective of the current analysis is to measure whether the U.S.-led economic transition impacted racial inequality in Puerto Rico's levels of literacy. The present analysis takes advantage of historical data to test the extent to which land use shaped differences in socio-economic status. I assess whether and how the distribution of inequality corresponded with crop production. In general, greater literacy is anticipated in areas with an expanding economic sector (sugar) relative to areas with a contracting or non-expanding economic sector (coffee) due to higher capital investments in infrastructure (including educational institutions). Yet, according to the stratifying thesis, greater inequality in literacy is anticipated in areas with higher sugar production because they are organized around a system of production dependent on racial subordination; the system reinforced inequality by restricting the subordinate (non-white) population's access to education. In contrast, the equalizing thesis suggests that racial inequality in literacy will be lower in areas with higher sugar production since these are the areas that disproportionately benefited from higher capital in-flows and resultant infrastructure development, equally accessible to all area residents.

Analytical Approach

I use *municipio*-level data from the 1899, 1910 and 1920 published census volumes (Department of Commerce & Bureau of the Census, 1913, 1922; War Department 1900). The 1899 census was conducted by the War Department rather than the Census Bureau. Despite this operational difference, statistics for comparable population categories are available and will serve as a reference point for the more pointed analyses to be conducted for 1910 and 1920. *Municipios* are comparable to U.S. counties.[2] Unfortunately, detailed data on crop production are not available until 1910 and, thus, limit the crop-specific analysis to 1910 and 1920.

Measures

The three variables central to the analysis are inequality in literacy, crop production and population composition. Primary focus is on the relationship between crop production and literacy, whereas population composition is of concern because it is entangled with the temporal and spatial distribution of crop production and literacy ratios. Inequality in literacy is measured by the non-white (black and mulatto) literacy as a percentage of white literacy for people aged 10 years or more. Black and mulatto were collapsed into a single non-white category because both are 'darker' than the white population, in colour and in relative social status. The ratio is calculated by dividing the proportion literate for non-whites by the proportion literate for whites and multiplying the value by 100. A value of 100% indicates that white and non-white literacy rates are the same, whereas a value less than 100% indicates that

Table 9.1 Descriptive statistics for racial literacy ratio and its correlates (N = 68)

	Mean	SD	Min	Max
1910				
Racial literacy ratio	66.79	17.53	35.54	124.70
Sugar acreage	0.08	0.07	0	0.25
Coffee acreage	0.07	0.10	0	0.44
Black	0.04	0.04	0	0.23
Mulatto	0.31	0.13	0.05	0.59
White	0.65	0.15	0.32	0.93
1920				
Racial literacy ratio	81.89	15.26	53.21	137.75
Sugar acreage	0.11	0.10	0	0.39
Coffee acreage	0.08	0.11	0	0.48
Black	0.04	0.04	0	0.20
Mulatto	0.25	0.13	0.06	0.67
White	0.72	0.15	0.27	0.94

Note: All variables are proportions except racial literacy ratio.

non-whites have lower literacy rates than whites. A value greater than 100% implies non-whites have higher literacy rates. Descriptive statistics on the literacy ratios and the other variables under analysis are reported in Table 9.1. Non-whites have lower literacy rates in both decades, although the disadvantage decreased between the two decades; the non-white literacy rate was 67% of the white literacy rate in 1910, but increased to 82% in 1920. The measure of literacy is consistent across observation points and reflects the population that can both read and write.

Crop production is measured as the percentage of farmland devoted to the production of a specific crop and is reported in acres. Sugar and coffee production are among the after-dinner crops central to the economic functioning of the island. Previous research argues that sugar production rapidly expanded after U.S. governance, whereas coffee production declined. In 1910, nearly equal proportions of farm acreage were devoted to sugar and coffee production (7%) and both increased over the next ten years. Sugar production, however, out-paced coffee production by 3% on average (11% of the island's total farmland was devoted to sugar production whereas 8% was in coffee production).

Sugar, coffee and tobacco production were unevenly distributed across the island. The dominant crop produced in each *municipio* is illustrated in Figure 9.1 – dominance reflects a greater proportion of the indicated crop relative to the other two after-dinner crops in 1910. Also reported in the figure is the extent of crop expansion. Sugar-dominant *municipios* expanded production by nearly 50% between 1910 and 1920 and tobacco grew by 114% (due, in large part, to the smaller number of initial acres devoted to the crop). In contrast, coffee production increased by only 14% among the coffee dominant *municipios*. That is, while production increased for all crops, expansion was least pronounced among coffee and, given the small

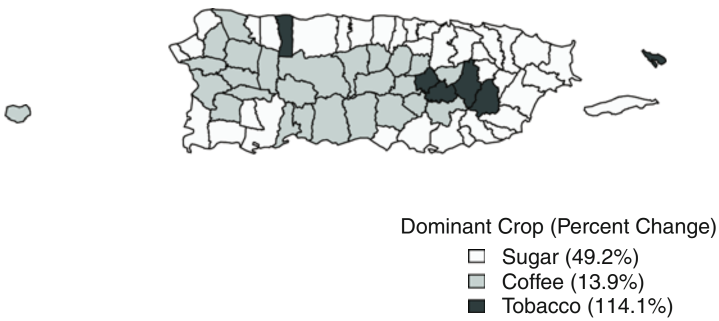


Fig. 9.1 Crop production in 1910 and change in crop production between 1910 and 1920

initial acreage investment in tobacco, sugar expansion was the greatest. The remainder of the analysis focuses on sugar and coffee since tobacco was produced on a considerably smaller scale than the other two crops.

Finally, population composition represents the proportion of the *municipio* population that is black, mulatto and white. The expansion of sugar production presumably was accompanied by population growth because of the labour demands and a topography that is more conducive to larger populations (flat coasts as opposed to mountainous interiors). In contrast, coffee producing *municipios* tended to have smaller populations and less growth. The growth was not only uneven in terms of spatial distribution, but also in terms of population composition.

Previous research on Puerto Rico suggests that the racial composition of the population varied according to crop production (Mintz, 1956; Scarano, 1984; Steward et al., 1956). In 1910, only 4% of the island's population was black while 31% was mulatto and the remaining 65% was white. The proportion of the population that was black remained unchanged in 1920, although the percentage mulatto decreased to 25% and the white population increased to 72% (Loveman & Muniz, 2007). The population composition is anticipated to vary according to crop production, where a higher proportion black and mulatto is expected to accompany greater sugar production.

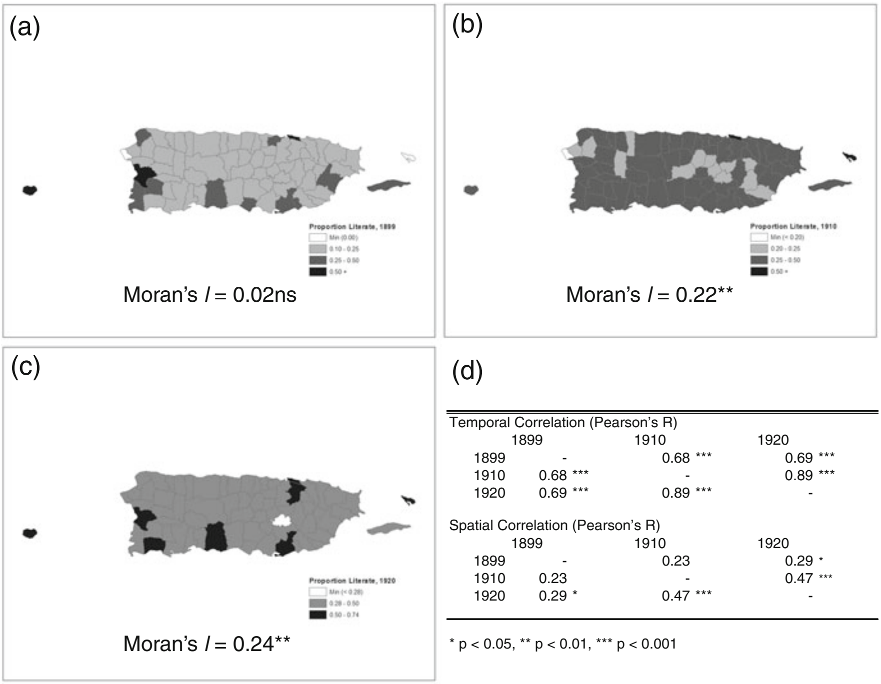
Findings

Overall Literacy

A general introduction to the distribution of literacy in Puerto Rico at the beginning of the twentieth century is useful to set the stage. Literacy rates increased by 10% or more with the passing of each decade, although they remained substantially lower than U.S. literacy rates in the same period. In 1899, 20% of the Puerto Rican population was literate. Yet, 31% was literate by 1910 and approximately 42% was by 1920. Even though the rates were improving during this period, they are considerably lower than those in the U.S. For example, 92% of the mainland U.S. population was literate in 1910 and 94% was literate in 1920 (Carter et al., 2006). So, while gains in literacy were made in Puerto Rico during the early decades of the twentieth century, there was room for much improvement. Moreover, literacy was not uniformly distributed across the island.

The distribution of literacy in Puerto Rico is presented for 1899, 1910 and 1920 and is accompanied by the estimated correlation over time in Figure 9.2. These estimates show that there is a weaker correspondence in the distribution of literacy rates between pre- and post-U.S. occupation ($r = 0.68$, $p < 0.001$ and $r = 0.69$, $p < 0.001$) compared to the correlation between the decades following U.S. rule ($r = 0.89$, $p < 0.001$). Results suggest a significant redistribution of literacy immediately following U.S. occupation and greater stability after 1910.

Similarly, there is little correspondence in the spatial distribution of literacy in the periods before and after U.S. occupation estimated by correlating the distribution of literacy at time 2 (e.g. 1920) with the spatially lagged distribution at time 1 (e.g. 1910). The spatial distribution of literacy was not the same in 1899 and 1910 ($r = 0.23$) and weak spatial correspondence is evidenced for 1910 and 1920 ($r = 0.47$, $p < 0.001$). Results show that the spatial patterning of literacy changed



Temporal Correlation (Pearson's R)

	1899	1910	1920
1899	-	0.68 ***	0.69 ***
1910	0.68 ***	-	0.89 ***
1920	0.69 ***	0.89 ***	-

Spatial Correlation (Pearson's R)

	1899	1910	1920
1899	-	0.23	0.29 *
1910	0.23	-	0.47 ***
1920	0.29 *	0.47 ***	-

* p < 0.05, ** p < 0.01, *** p < 0.001

Fig. 9.2 Distribution of the proportion literate in (a) 1889, (b) 1919, and (c) 1920 with (d) temporal and spatial correlations. Moran's *I* statistic 0.02 ($p > 0.1$) in 1899, 0.22 ($p < 0.01$) in 1910, and 0.24 ($p < 0.01$) in 1920 dramatically between the pre- and post-U.S. occupation periods and changed to a lesser degree under U.S. rule.

The extent of spatial clustering in literacy within each decade is measured by the Moran's I statistic (Moran, 1950). A first-order queen weights matrix was used throughout the analysis and includes information on the value of the variable of interest for all contiguous neighbours of each *municipio* – neighbours to each side and on the diagonal, as a queen moves on a chess board. The matrix was manually adjusted to insure that islands were neighboured by the nearest *municipios* since no *municipio* is socially isolated. A significant finding is that spatial clustering was present in the post-U.S. occupation decades (Moran's $I = 0.22$, $p < 0.01$ and Moran's $I = 0.24$, $p < 0.01$ in 1910 and 1920, respectively) but not in the pre-occupation decade (Moran's $I = 0.02$ in 1899). In 1899, literacy in one *municipio* was not associated with literacy in a neighbouring *municipio*; literacy was randomly distributed across the island. In 1910 and 1920, however, there was a positive and statistically significant spatial relationship in the distribution of literacy; *municipios* tended to neighbour other *municipios* with similar literacy rates and the distribution was not random.

Literacy Inequality

The discussion about the patterns of general literacy provides the backdrop to the analysis of racial literacy ratios. The *municipio* racial literacy ratio is the measure of inequality and is the focus of the remaining analyses. The ratios represent the non-white literacy rate as a percentage of the white literacy rate, where 100% indicates the same literacy rate between the two groups, values less than 100% suggest a lower literacy rate among non-whites and values greater than 100% indicate a higher literacy rate among non-whites.

The distribution of racial inequality in literacy in Puerto Rico is illustrated in Figure 9.3. In 1910, the majority of *municipios* had lower literacy rates for non-whites compared to whites (64 of the 68 *municipios*), although one *municipio*

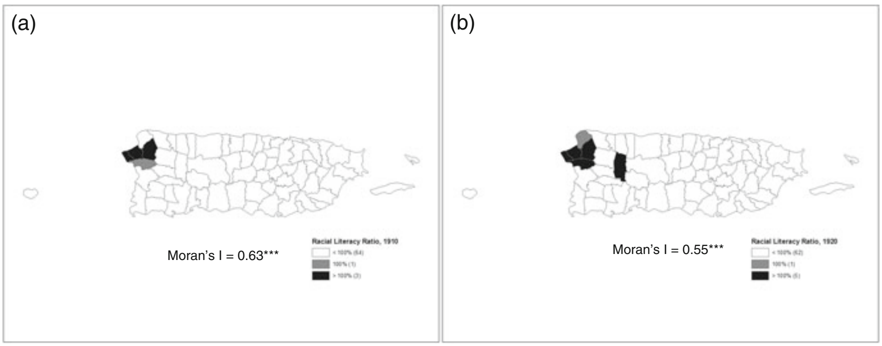


Fig. 9.3 Distribution of racial literacy ratios in (a) 1910 and (b) 1920. Temporal correlation (Pearson's R) = 0.77 ($p < 0.001$), Spatial Correlation (Pearson's R) = 0.74 ($p < 0.001$)

had equal literacy rates (Añasco) and three had higher literacy rates among non-whites than whites (Rincón, Aguada and Moca). These *municipios* are among, or neighbour, sugar-producing *municipios*. The distribution changed only slightly in 1920; the majority of *municipios* still had lower literacy rates for non-whites (n = 62), one *municipio* had equal literacy rates (Aguadilla, this time) and five had higher literacy rates among non-whites (Lares and Añasco joined Rincón, Aguada and Moca). All, with the exception of Lares, were or neighboured sugar-producing municipios; Lares is a coffee-producing *municipio* and a southern neighbour to sugar-producing municipios.

Local indicators of spatial association (LISA) clusters, presented in Table 9.2 (which corresponds with Figure 9.4), re-confirm stability in the spatial distribution of 'hot' and 'cold' spots between the two post-occupation decades. LISA is a method of identifying local patterns of association and assessing local instability in the spatial association (Anselin, 1995: p. 94). In the univariate context, LISA statistics are used to identify areas where racial inequality in literacy is especially high ('hot') and especially low ('cold'). In the bivariate context, the technique permits the analyst to assess whether the bivariate association between racial inequality in literacy and crop production varies across space. In the spatio-temporal context, the local estimates are used to examine spatial stability in the distribution of the variables of interest over time.

Returning to the univariate context, there is evidence of high and low clusters of racial inequality in literacy. 'Hot spots' indicate high-high non-white literacy (high non-white literacy *municipios* neighbour high non-white literacy *municipios*). 'Cold spots', in contrast, represent low-low non-white literacy rates (low non-white literacy *municipios* neighbour low non-white literacy *municipios*). Both hot and cold spots indicate clusters of racial inequality; the high-high pertains to a non-white advantage whereas low-low reflects a white advantage in literacy rates. Evidence of spatial clusters in the presence of global spatial autocorrelation (illustrated in Fig. 9.3) suggests that the *municipios* contribute more than their expected share to the global statistic (Anselin, 1995).

One large hot spot (non-white advantage) is found in the north-western *municipios* of Puerto Rico in 1910 and 1920 and corresponds with the highlighted area in the preceding figure of racial literacy ratios. Here, non-whites have a literacy advantage over whites. One large cold spot (white advantage) is located in the eastern *municipios* in 1910. In 1920, however, three cold spots are distributed in the central and eastern *municipios*. The cold spots have non-white literacy disadvantages ranging from 35% to 55% in 1910 and 53% to 80% in 1920. The spatial clustering of literacy ratios remains relatively stable with the exception of the emergence of two cold spots in 1920.

Crop Production and Literacy

Next, I examine whether the distribution of inequality corresponds with crop production in the temporal context. Drawing from previous research, literacy

Table 9.2 Univariate and bivariate LISA associations for racial literacy ratio and covariates, 1910 and 1920 ($p < 0.05$)

	Racial literacy ratio		Racial literacy ratio and sugar production		Racial literacy ratio and coffee production		Racial literacy ratio and population composition (white)	
	1910	*1920*	*1910*	*1920*	*1910*	*1920*	*1910*	*1920*
High-High	Aguada (2)	Aguada (2)	Aguada (2)	Yauco (68)	Adjuntas (1)	Adjuntas (1)	Adjuntas (1)	Aguada (2)
	Aguadillo (3)	Aguadillo (3)	Rincón (50)		Añasco (6)	Añasco (6)	Aguada (2)	Aguadillo (3)
	Añasco (6)	Añasco (6)			Lares (35)	Lares (35)	Aguadillo (3)	Añasco (6)
	Isabela (31)	Isabela (31)			Las Marías (36)	Las Marías (36)	Añasco (6)	Camuy (15)
	Las Marías (36)	Moca (42)			Maricao (39)	Maricao (39)	Camuy (15)	Hatillo (29)
	Moca (42)	Rincón (50)			Mayagüez (41)	Mayagüez (41)	Hatillo (29)	Isabela (31)
	Rincón (50)	San Sebastian (58)			San Sebastian (58)	Ponce (48)	Isabela (31)	Juncos (33)
	San Sebastian (58)				Utuado (63)	San Sebastian (58)	Lares (35)	Lares (35)
						Utuado (63)	Moca (42)	Las Marías (36)
						Yauco (68)	Quebradillas (49)	Moca (42)
							San Sebastian (58)	Quebradillas (49)
							Utuado (63)	San Sebastian (58)
								Utuado (63)
								Yauco (68)

Table 9.2 (continued)

	Racial literacy ratio		Racial literacy ratio and sugar production		Racial literacy ratio and coffee production		Racial literacy ratio and population composition (white)	
	1910	*1920*	*1910*	*1920*	*1910*	*1920*	*1910*	*1920*
Low-Low	Caguas (14)	Aguas Buenas (4)	Aguas Buenas (4)	Barranquitas (10)	Carolina (16)	Carolina (16)	Bayamón (12)	Carolina (16)
	Carolina (16)	Barranquitas (10)	Aibonito (5)	Barros (11)	Culebra (23)	Culebra (23)	Carolina (16)	Dorado (24)
	Gurabo (28)	Barros (11)	Barranquitas (10)	Cidra (19)	Dorado (24)	Dorado (24)	Dorado (24)	Fajardo (25)
	San Lorenzo (57)	Bayamón (12)	Barros (11)	Comerio (21)	Fajardo (25)	Fajardo (25)	Gurabo (28)	Río Piedras (52)
	Trujillo Alto (62)	Caguas (14)	Caguas (14)	Corozal (22)	Gurabo (28)	Gurabo (28)	Humacao (30)	San Juan (56)
	Yabucoa (67)	Cidra (19)	Cidra (19)	Naranjito (45)	Humacao (30)	Humacao (30)	Ponce (48)	Toa Alta (60)
		Comerio (21)	Comerio (21)		Loíza (37)	Loíza (37)	Río Piedras (52)	Toa Baja (61)
		Corozal (22)	Corozal (22)		Naguabo (44)	Naguabo (44)	San Juan (56)	Trujillo Alto (62)
		Fajardo (25)	Culebra (23)		Vieques (66)	Trujillo Alto (62)	Toa Alta (60)	
		Morovis (43)	Naranjito (45)			Vieques (66)	Toa Baja (61)	
		Naranjito (45)					Trujillo Alto (62)	
		Vega Alta (65)						

Fig. 9.4 *Municipio* (ID) to correspond with Table 9.2

inequality may vary across space and according to economic restructuring; higher racial inequality in literacy may be present in areas with high sugar production due to the more highly and racially stratified system of production relative to other crop industries. An alternative thesis, the equalizing thesis, suggests that racial inequality could be lower in areas with a strong sugar presence due to infrastructure improvements, including education, which may have disproportionately accompanied the new economic expansion.

Change in *municipio* literacy rates and racial literacy ratios are examined relative to changes in sugar and coffee production. Results are reported in Table 9.3. The findings suggest that there is no relationship between changes in general literacy rates and changes in crop production, either sugar or coffee. There is, however, evidence that racial inequality in literacy is associated with crop production. Results show that racial inequality is reduced (the non-white advantage in literacy increases) as sugar production increases ($r = 0.266$, $p < 0.05$). In contrast, a negative association is found between change in the literacy ratio and the spatially lagged values of change in coffee acreage; increases in coffee production in the larger spatial context are associated with white literacy advantages. Results are consistent with the equalizing thesis that racial inequality is lower in areas with higher sugar production since these areas experienced greater infrastructure development, including the expansion of the educational system, relative to other crop-producing areas that had less capital in-flows, like coffee-producing areas.

Results show that change in sugar acreage is associated with change in racial inequality in literacy. Within-decade patterns are examined since it is reasonable

Table 9.3 Bivariate associations (Pearson's R) between change in crop production and literacy (N = 68)

	Δ Sugar acreage	*Δ Coffee acreage*
Δ Literacy rate	–0.014	0.150
Δ Racial literacy ratio	0.266*	0.006
	Spatially lagged Δ Sugar acreage	*Spatially lagged Δ Coffee acreage*
Δ Racial literacy ratio	0.011	–0.277*

*$p < 0.05$.

to expect that where a *municipio* begins in terms of literacy ratios is part of the larger process of change in racial literacy ratios. Bivariate LISAs were estimated to assess the spatial distribution of literacy relative to the spatial distribution of sugar production and are reported for 1910 and 1920 separately (Table 9.2).

The bivariate LISA clusters show local patterns of *municipio* spatial correlation between literacy ratios and the average sugar production for its neighbours. Results suggest that *municipios* with a non-white literacy advantage tend to neighbour *municipios* with high sugar production. In contrast, *municipios* with a white literacy advantage tend to neighbour areas with low sugar production. The results demonstrate that where a relationship exists, racial inequality in literacy tends to be lower in higher sugar-producing areas. Comparing the two cluster distributions suggests stability in the concentration of local area relationships over time.

Review of the bivariate LISA clusters that related the distribution of coffee production to the distribution of literacy, also in Table 9.2, shows results similar to those reported for sugar production (i.e. higher crop production is associated with a non-white literacy advantage and the distribution appears stable over time). One difference is in the pattern of the local area relationships: two sizeable clusters emerge and the hot spot directly corresponds with the distribution of coffee production (results available from author at request). The results of the LISA cluster analysis characterize the relationship between crop production and literacy ratios within a decade and not changes in the values over time. Relating where a *municipio* begins the process of change to where the *municipio* ends is revisited in the multivariate analysis.

Crop Production and Population Composition

Previous research on Puerto Rico suggests that the racial composition of the population varies by crop region. For example, Scarano (1984) demonstrates that large increases in the black population accompanied sugar expansion during the 1800s. Remnants of the uneven racial distribution reinforced by the slave trade during the nineteenth century persisted in the twentieth century (Mintz, 1956; Steward, 1956). The correlation between *municipio* crop production and racial composition in 1910 and 1920 is presented in Table 9.4. Sugar production is positively associated with the proportion black and mulatto and is negatively correlated with the proportion white during the decades immediately following U.S. governance. In contrast, coffee production is negatively correlated with the proportion black and mulatto, but is positively associated with the proportion white. This finding corroborates discourse about the ‘whiteness’ and ‘whitening’ of the mountain (coffee) region (Carrasquillo, 2006).

Recall from Table 9.3 that an increase in area coffee production is correlated with an increase in literacy inequality (white advantage); the uneven distribution of coffee expansion is related to the uneven distribution of literacy ratios. Analysis of population composition suggests that the uneven distribution of racial literacy ratios may be related to the uneven racial distribution of the population. This proposition

Table 9.4 Bivariate association (Pearson's R) between crop production and population composition (N = 68)

	Sugar acreage	Coffee acreage
1910		
Black	0.313**	–0.337**
Mulatto	0.196	–0.336**
White	–0.260*	0.388**
1920		
Black	0.327**	–0.353**
Mulatto	0.299*	–0.397***
White	–0.328**	0.424***

Note: All variables are proportions.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

is pursued further by analysing the spatial distribution of population composition and literacy ratios.

Bivariate LISA clusters plotting the spatial distribution of literacy relative to the spatial distribution of the white population are reported for 1910 and 1920 in the final column of Table 9.2. The bivariate LISA clusters show local patterns of municipio spatial correlation between literacy ratios and the average white proportion of the population for its neighbours. A striking relationship emerges. *Municipios* with a non-white literacy advantage tend to neighbour *municipios* with a high white proportion of the population. In contrast, *municipios* with a white literacy advantage tend to neighbour areas with a low white proportion of the population.

The reverse is observed when plotting the spatial distribution of literacy relative to the distribution of the non-white population (maps are available from the author at request). Combined, the bivariate LISA clusters, where a spatial relationship exists, show the minority race has the literacy advantage. Where there are fewer whites, there is a white literacy advantage. Where there are fewer non-whites, there is a non-white literacy advantage (the cluster of the non-white literacy advantage in the univariate LISA distribution corresponds with a low mulatto population in 1910 and a low black population in 1920). Moreover, there is stability in this relationship over time.

An increase in coffee production is related to an increase in literacy inequality, and *municipios* in areas with higher coffee production tend to have a higher proportion of whites relative to other crop-producing areas. In contrast, *municipios* with higher sugar production tend to have a lower proportion of whites. Importantly, *municipios* with lower proportions of whites tend to neighbour *municipios* with a white literacy advantage. Sugar-producing *municipios* tend to house higher proportions of non-whites and there is a relationship between *changes* in sugar production and changes in literacy ratios. Analysis of the *within-decade* association between sugar production and literacy ratios also shows a statistically significant positive relationship for 1910. Where a *municipio* began in its sugar production appears to have mattered and whether production expanded and the level of production in 1920, further informed changes in, or levels of, literacy ratios. This temporal distinction will re-emerge in the multivariate analysis of literacy ratios.

Multivariate and Simultaneous Space-Time Analysis of Literacy

The preceding descriptive analyses informed a simultaneous space-time analysis of the influence of crop production and racial population on inequalities in literacy. I employ regression techniques to examine temporal and spatial dynamics in the distribution of literacy ratios. Specifically, I regress racial literacy ratios in 1920 on the change in crop and race conditions in the respective *municipio* and in the neighbouring *municipios* while accounting for conditions in 1910; the temporal dynamics are addressed by earlier conditions and changes in conditions while the spatial dynamics are accounted for by spatially lagged conditions. I initially employed a cross-regressive model denoted as

$$y = X\beta + \rho Wx + \varepsilon$$

where y is the racial literacy ratio in 1920, $X\beta$ represents the crop and race conditions, ρWx represents the crop and race conditions in the neighbouring *municipios*, and ϵ is the error term. Spatial autocorrelation was evidenced in the cross-regressive model residuals (i.e. $MI = 0.22$, $p < 0.01$) and therefore, I used a spatial autoregressive approach to correct for the residual spatial autocorrelation that would otherwise threaten to bias parameter estimates. Model diagnostics indicated a spatial lag model, denoted as

$$y = X\beta + \rho Wy + \varepsilon$$

where ρWy represents the racial literacy ratio in 1920 in the neighbouring *municipios*. In general, use of the spatial lag model suggests the value of y in neighbouring spatial units affects the value of y in the spatial unit of interest. In the current analysis, there is no definitive conceptual argument to support this interpretation and modelling strategies were not taken to disentangle the interactive process (ρWy) from the reactive process (ρWx) (see Beggs, Wayne Villemez and Arnold (1997) for an example of one approach to statistically separating the spatially interactive and reactive processes). Instead, the parameter (ρ) is interpreted as a statistical correction for the autocorrelation not accounted for by the covariates ($X\beta$ and ρWx).

The final model used to analyse the distribution of racial literacy ratios in 1920 includes the spatially lagged covariates as well as the spatial correction and is denoted as

$$y = X\beta + \rho Wx + \rho Wy + \varepsilon.$$

One intention of the modelling strategy is to estimate the observed temporal and spatial dynamics shaping inequality; earlier conditions and change in conditions capture the temporal dynamic and spatially lagged conditions approximate the spatial dynamic. A second intention of the model is to purge the estimates of spatial autocorrelation, which may compromise model estimates; spatial autocorrelation is introduced by analysing areal units.

Three significant findings are gleaned from the multivariate analysis. First, regression results yield support for the equalizing thesis; higher sugar production – and crop expansion, more generally – is largely associated with a non-white literacy advantage. Results from the regression analysis, reported in Table 9.5, show that initial sugar acreage is strongly associated with lower inequality in 1920. The association between inequality and change in *municipio* sugar acreage, observed in the bivariate analysis reported in Table 9.3, is attenuated in the multivariate context. Yet, change in sugar production within the larger spatial context is directly associated with racial inequality in literacy, net of the influence of all other covariates.

Second, the spatio-temporal dynamic of the relationship between crop production and racial inequality, highlighted in model 2, is complex. *Municipios* neighbouring places that experienced increases in sugar acreage had greater racial inequality. At the same time, *municipios* neighbouring places that had higher sugar acreage in 1910 had lower racial inequality in literacy in 1920. It is important to emphasize that the *municipio*-specific associations are positive; higher sugar production and increases in sugar production are associated with lower racial inequality within the *municipio*. The nature of the relationship, however, is less straightforward in the larger spatial context. That is, *municipios* neighbouring other *municipios* with increased sugar production had greater racial inequality. All sugar effects combined, *municipios* with a one-standard deviation higher value in sugar acreage – in 1910, change between 1910 and 1920, and in the spatially lagged values of each – had 6% lower racial inequality than *municipios* with the average values in sugar acreage and

Table 9.5 Spatially informed regression analysis of racial inequality in municipio literacy rates for 1920 (N = 68)

	Model 1		Model 2	
	β	*SE*	*β*	*SE*
Constant	10.94	8.82	–3.36	13.50
Δ Sugar acreage	17.89	21.16	10.76	20.60
Δ Coffee acreage	146.45*	58.33	157.77**	56.14
Δ White	41.78**	14.48	37.69**	14.45
Sugar acreage, 1910	73.68***	19.83	67.78**	21.69
Coffee acreage, 1910	9.99	14.33	14.53	21.29
White, 1910	13.86	9.04	–2.88	12.44
Spatially lagged Δ sugar acreage			–73.41*	33.63
Spatially lagged Δ coffee acreage			–25.23	104.48
Spatially lagged Δ white			44.97	39.39
Spatially lagged sugar acreage, 1910			102.83*	44.94
Spatially lagged coffee acreage, 1910			–38.81	35.65
Spatially lagged white, 1910			58.66*	24.52
Spatial correction (ρ)	0.617***		0.408**	
AIC	520.28		517.66	
Moran's *I* (error)	3.46		2.80	

Note: All change variables are differences in proportions except the outcome.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

12% lower inequality than *municipios* with a one-standard deviation lower than the average values.

In contrast to the spatially extensive effects of sugar production, the influence of coffee production is concentrated within the *municipio*; change in coffee acreage is associated with reductions in racial inequality, yet coffee production in neighbouring places has no association. With regard to the *municipio*-specific influence, places with a one-standard deviation higher than average change in coffee acreage had 4% lower racial inequality than *municipios* with average change in coffee acreage and 7% lower inequality than places with a one-standard deviation lower than average change in acreage. Although less dramatic than the influence of sugar production, expansion in coffee production was also associated with reductions in racial inequality in literacy and suggests that crop expansion, in general, influenced racial inequality.

Finally, results demonstrate that changes in the racial composition of the *municipio* matter for inequality, net of the influence of crop production. *Municipios* that 'whitened' between 1910 and 1920 – either through population changes like migration or socially malleable categorization – or that neighboured *municipios* with a higher proportion white in 1910 had lower racial inequality in literacy in 1920. For example, places that experienced a one-standard deviation higher than average increase in the white population had 3% and 6% lower racial inequality than *municipios* with the average change and a one-standard deviation lower than average change in the white population, respectively.

Discussion

The analysis aimed to explore the extent to which the U.S.-led economic restructuring of early twentieth century Puerto Rico influenced racial inequality by examining the spatially and temporally dynamic association between crop production and racial inequality in *municipio* literacy rates. Results show that while overall literacy rates and racial inequality in literacy improved between 1910 and 1920, literacy was unevenly distributed across the *municipios* in a pattern that corresponded with crop production.

A significant finding is that crop expansion, in general, was associated with lower racial inequality. The historiography argues that conditions were generally improving where economic expansion was underway in Puerto Rico during the first decades of U.S. governance. That is, crop expansion was enabled by new capital in-flows from the U.S. that extended to other forms of infrastructure, including educational institutions. Results show that this expansion was not limited to a particular segment of the population the way that proponents of the stratifying thesis suggest; racial inequality in literacy (a non-white disadvantage) was lower among *municipios* that increased crop production between 1910 and 1920.

Support for the equalizing thesis is found in the general pattern of less racial inequality among even the sugar-producing *municipios* organized around racial subordination inherent to the plantation structure. Indeed, although crop expansion was generally related to lower racial inequality in literacy, sugar had the greatest impact in magnitude and in reach; *municipios* with a one-standard deviation higher value in sugar acreage had 6% lower inequality than *municipios* with average values, and sugar production within the *municipio* influenced racial literacy ratios within the corresponding *municipio* as well as racial literacy ratios in neighbouring *municipios*. The spatial context of sugar production matters for inequality in literacy.

Moreover, the spatio-temporal dynamic of the relationship between sugar production and racial inequality is complex. Sugar production was associated with lower racial inequality within the *municipio* whereas it was associated with lower *and* higher racial inequality among neighbouring *municipios*; higher sugar production in the initial period was related to lower racial inequality among neighbouring *municipios* and an increase in sugar production over the decade was associated with higher racial inequality among neighbours. This pattern suggests that the initial conditions and the change in conditions within the larger regional context impact racial inequality in the *municipios* within the region.

Results also demonstrate the importance of the racial composition of the *municipio* population in terms of the direct and indirect association with racial inequality in literacy. Population 'whitening' was associated with lower racial inequality and while the white population tended to be in the coffee producing areas, *municipios* that experienced growth in the white population either through demographic change or social re-categorization had lower racial inequality than places with less growth in the white population, net of the influence of coffee production. Like crop expansion, growth of the white population is arguably associated with capital in-flows and development. The theoretical implications of this finding are not resolved here, but results demonstrate inequality decreased in areas with greater 'whitening' (see Carrasquillo, 2006; Loveman & Muniz, 2007).

Demographic processes occur within the spatial and temporal contexts. The methodological approaches used here have added a new understanding to the relationship between economic transitions and social inequality within the early twentieth-century Puerto Rican case by explicitly investigating the contextual dimensions of the transition, which can be applied to other cases. Indeed, this research demonstrates that initial conditions and changes in conditions matter. Moreover, conditions within a place and conditions within the larger spatial context in which a place is embedded also matter. A richer understanding of complex demographic processes can be gained by engaging the conceptual and analytical complexity of that process.

Acknowledgement This research was supported by Research Grant F32 HD052345-01 from the National Institute for Child Health and Human Development, National Institutes of Health. Technical support was given by Bill Buckingham at the Applied Population Laboratory through the University of Wisconsin Center for Demography and Ecology GIA Core. The study has benefited from the research support from Adam Slez and the helpful comments given by Glenn Deane, Myron Gutmann, Rick Nordheim, Kenneth Sylvester, Paul Voss, the anonymous reviewers, and the participants of the International Conference on Space and Time in Historical Demographic Research – New Methods and Models sponsored by the IUSSP Scientific Committee on Historical Demography.

Notes

1. Migration from Puerto Rico was modest during the period under analysis and became sizeable in the years following World War II (Duany, 2009). A cursory review of individual micro-data (PUMS) suggests that the number of Puerto Ricans living in mainland U.S. was only 1,513 in 1910 and 11,811 in 1920 (Ruggles et al., 2003). By contrast, the number grew to 52,774 in 1930, 69,967 in 1940 and increased dramatically to 226,110 in 1950.
2. The 1899 *municipio* boundaries are preserved throughout the analysis; they are consistent with the 1910 boundaries although they differ from the 1920 boundaries – five *municipios* were developed from the original 68 *municipios* by 1920. *Municipio* information on the five 'new' *municipios* in 1920 was included in the 'original' *municipios*.

References

Anselin, L. (1995). Local Indicators of Spatial Association – LISA. *Geographical Analysis*, *27*, 93–115.

Anselin, L. (2003). *GeoDa 0.9 user's guide*. Urbana, IL: Spatial Analysis Laboratory, University of Illinois.

Armstrong, P. (1996). The feminization of the labor force: Harmonizing down in a global economy. In I. Bakker, (Ed.), *Rethinking restructuring: Gender and change in Canada* (pp. 29–54). Buffalo, NY: University of Toronto Press.

Beggs, J. J., Wayne Villemez, J., & Arnold, R. (1997). Black population concentration and black-white inequality: Expanding the consideration of place and space effects. *Social Forces*, *76*(1), 65–91.

Bobonis, G. J., & Toro, H.J. (2007). Modern colonization and its consequences: The effects of U.S. educational policy on Puerto Rico's educational stratification, 1899–1910. *Caribbean Studies, 35*, 31–76.

Carrasquillo, R. E. (2006). *Our landless patria: Marginal citizenship and race in Caguas, Puerto Rico, 1880–1910*. Lincoln, NE: University of Nebraska Press.

Carter, S. B., Gartner, S. S., Haines, M. R., Olmstead, A. L., Sutch, R., & Wright, G. (2006). *Historical statistics of the United States: Earliest time to the present* (Millennial Edition): Vol. 2. Work and Welfare. New York: Cambridge University Press.

Cubano Iguina, A. (2006). *Rituals of violence in 19th-Century Puerto Rico: Individual conflict, gender, and the law*. Gainesville, FL: University Press of Florida.

Department of Commerce, Bureau of the Census. (1913). *Thirteenth census of the United States taken in the year 1910: Statistics for Porto Rico*. Washington, DC: Government Printing Office.

Department of Commerce, Bureau of the Census. (1922). *Fourteenth census of the United States taken in the year 1920: Statistics for Porto Rico*. Washington, DC: Government Printing Office.

Dietz, J. L. (1986). *Economic history of Puerto Rico: Institutional change and capitalist development*. Princeton, IL: Princeton University Press.

DiPrete, T. & Forrestal, J. (1995). Socioeconomic change and occupational location for successive cohorts of American male and female workers. *Social Science Research*, *24*, 390–438.

Duany, J. (2009). The nation in the diaspora: The multiple repercussions of Puerto Rican emigration to the United States. In P. Clegg & E. Pantojas-Garcia, (Eds.), *Governance in the non-Independent Caribbean: Challenges and opportunities in the twenty-first century*. Kingston: Ian Randle.

Gittleman, M. B. & Howell, D. R. (1995). Changes in the structure and quality of jobs in the United States: Effects by race and gender, 1973–1990. *Industrial and Labor Relations Review*, *48*, 420–440.

Griffith, D. A. (1979). Urban dominance, spatial structure, and spatial dynamics: Some theoretical conjectures and empirical implications. *Economic Geography*, *55*, 95–113.

Griffith, D. A. (1992). A spatially adjusted N-way ANOVA model. *Regional Science and Urban Economics*, *22*, 347–369.

Jensen, J. (1989). The talents of Women, the skills of men: Flexible specialization and Women. In S. Wood, (Ed.), *The transformation of work* (pp. 141–155). Boston: Unwin Hyman.

Loveman, M. & Muniz, J. (2007). How Puerto Rico became white: Boundary dynamics and intercensus racial reclassification. *American Sociological Review*, *72*(6), 915–939.

McCall, L. (1998). Spatial routes to gender wage (in)equality: Regional restructuring and wage differentials by gender and education. *Economic Geography*, *74*, 379–404.

McCall, L. (2000). Explaining levels of within-group wage inequality in U.S. labor markets. *Demography*, *37*, 415–430.

McDowell, L. (1991). Life without father and Ford: The new gender order of post-Fordism. *Transactions of the Institute of British Geographers*, *16*, 400–419.

Mintz, S. W. (1956). Cañamelar: The subculture of a rural sugar plantation proletariat. In H. J. Steward, et al. (Eds.), *The people of Puerto Rico: A study in social anthropology* (pp. 314–417). Urbana, IL: University of Illinois Press.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*, 17–23.

Osuna, J. J. (1975). *A history of education in Puerto Rico*. (2nd ed.). New York: Arno Press. (Reprint, Editorial Universitaria, Universidad de Puerto Rico).

Picó, R. (1950). *The geographic regions of Puerto Rico*. Río Piedras: University of Puerto Rico Press.

Reskin, B. F. & Roos, P. A. (1990). *Job queues, gender queues: Explaining women's inroads into male occupations*. Philadelphia: Temple University Press.

Ruggles, S., et al. (2003). *Integrated public use microdata series: Version 3.0*. Minneapolis, MN: Historical Census Projects, University of Minnesota.

Scarano, F. A. (1984). *Sugar and slavery in Puerto Rico: The plantation economy of Ponce, 1800–1850*. Madison, WI: University of Wisconsin Press.

Scarano, F. A. & White, K. J. C. (2007). A window into the past: Household composition and distribution in Puerto Rico, 1910 and 1920. *Caribbean Studies*, *35*, 115–154.

Steward, J. H. (1956). Introduction. In J. H. Steward, R. Manners, E. R. Wolf, E. Padilla Seda, S. W. Mintz, R. L. Scheele, (Eds.), *The People of Puerto Rico: A Study in Social Anthropology* (pp. 1–27). Urbana, IL: University of Illinois Press.

Steward, J. H., Manners, R. A., Wolf, E. R., Padilla Seda, E., Mintz, S. W., & Scheefe, R. L. (1956). *The people of Puerto Rico: A study in social anthropology*. Urbana, IL: University of Illinois Press.

Tumin, M. M. (1961). *Social class and social change in Puerto Rico*. Princeton, IL: Princeton University Press.

Wagley, C. (1960). Plantation America: A culture sphere. In V. Rubin, (Ed.), *Caribbean studies: A symposium* (pp. 3–13). Seattle, WA: University of Washington Press.

War Department. (1900). Office Director Census of Porto Rico. *Report on the census of Porto Rico, 1899*. Washington, DC: Government Printing Office.

Wolf, E. R. (1956). San José: Subcultures of a "traditional" coffee municipality. In J. H. Steward, R. A. Manners, E. R. Wolf, E. Padilla Seda, S. W. Mintz, R. L. Scheefe, (Eds.), *The people of Puerto Rico: A study in social anthropology* (pp. 171–264). Urbana, IL: University of Illinois Press.

Index

M.P. Gutmann et al. (eds.), *Navigating Time and Space in Population Studies*, International Studies in Population 9, DOI 10.1007/978-94-007-0068-0,