Jack Baker · David A. Swanson
Jeff Tayman · Lucky M. Tedrow

Cohort Change Ratios and their Applications

Springer

Cohort Change Ratios and their Applications

Jack Baker • David A. Swanson • Jeff Tayman
Lucky M. Tedrow

Cohort Change Ratios and their Applications

Springer

Jack Baker
Health Fitness Corporation
Lake Forest, IL, USA

David A. Swanson
University of California Riverside
Riverside, CA, USA

Jeff Tayman
University of California San Diego
San Diego, CA, USA

Lucky M. Tedrow
Western Washington University
Bellingham, WA, USA

ISBN 978-3-319-53744-3 ISBN 978-3-319-53745-0 (eBook)
DOI 10.1007/978-3-319-53745-0

Library of Congress Control Number: 2017932371

Printed on acid-free paper

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

Foreword

What do these things have in common: school enrollment; persons with mobility limitations; the prevalence of obesity, diabetes, and cardiovascular disease; the size and composition of the civilian labor force; alcohol and cigarette consumption; and the number of households and housing units? Answer: they can all be analyzed, estimated, and projected using cohort change ratios.

A cohort is a group of people who experience the same demographic event during a given period of time, and a cohort change ratio (CCR) measures changes in cohort size over time. CCRs frequently focus on age cohorts but can be calculated for other types of cohorts as well, such as people starting college or getting married in the same year. They are often broken down by sex, race, ethnicity, or other demographic characteristics and can be used for a wide variety of purposes. Common uses include constructing population estimates by age, sex, and race; forecasting school enrollments by grade; and projecting the number and characteristics of people living in a particular city, county, or state. CCRs are conceptually simple but analytically powerful, and their minimal data requirements mean they can be applied at almost any level of geography.

This book presents an in-depth look at the construction and use of CCRs. It goes beyond previous treatments of this topic in several ways. It discusses modifications that can be made to a given set of CCRs, such as adjusting them to reflect the continuation of historical trends, calculating averages based on several sets of CCRs, and developing synthetic CCRs that incorporate information from other geographic areas. It describes techniques for splitting broader age groups into single years of age and for interpolating values between two points in time. It gives many step-by-step examples showing how CCRs can be used to construct different types of estimates and projections. It provides empirical evidence on the accuracy of estimates and projections made using different techniques or for places with differing characteristics. Its extensive list of references and websites makes it easy for readers to delve more deeply into specific aspects of the broader topic (the instructions for accessing data from American Factfinder are particularly helpful).

Applied demography is often defined as the use of demographic methods and materials for decision-making purposes. Its basic objective is to "get more bang for the buck" or to accomplish a given task in the shortest possible time and for the least possible cost. CCR estimation and projection models can play an important role in this regard because they have relatively low costs and small data requirements and are fairly simple to apply. In contrast, full-blown cohort-component models are more complex, costly, and data-intensive. When time is short and budgets are tight, this is an important advantage of CCR models. In terms of accuracy, CCR models have generally been found to perform as well as full-blown cohort-component models in most circumstances.

Cohort Change Ratios and Their Applications is a highly practical book, helping practitioners undertake a variety of projects and deal with the thorny issues that often complicate seemingly simple tasks. But it is more than a guidebook. It also investigates topics such as using CCRs to illustrate the findings of stable population theory, calculating life expectancy at birth, and examining the relationship between survivorship and net migration. It discusses the use of spatial weighting to adjust CCRs and describes a technique for constructing measures of uncertainty for CCR projections. This book's treatment of these topics takes CCR models well beyond their usual applications.

The authors are eminently qualified to write this book, given their broad academic training, deep knowledge of demographic data and methods, and many years working in academic, business, and government settings. They have extensive hands-on experience dealing with CCR models and their clear understanding of the issues is fully evident. This book promises to be a valuable addition to any demographer's or planner's library.

University of Florida, Stanley K. Smith
Gainesville, FL, USA
December 2016

Preface

Like many demographers our first exposure to the cohort change ratio (CCR) was as a census survival rate, which can be used as a measure of mortality in places lacking good vital statistics. In places with limited migration, census survival rates are a very good approximation to life table survival rates. We were also taught that the cohort-component method was the de facto standard for producing estimates and forecasts by age, sex, and other demographic characteristics. Once we finished our education and entered the field, it quickly became apparent that implementing the cohort-component method in places lacking good vital events and migration information was difficult, if not impossible. At the same time there was a rising and seemingly insatiable demand for small area (especially subcounty) estimates and forecasts of demographic characteristics. As a result and for many years, we have successfully used the Hamilton-Perry method (H-P) based on cohort change ratios to develop such estimates and forecasts for a wide range of geographic areas both inside and outside of the United States. The H-P method has gained acceptance as research has demonstrated its practical value and accuracy in estimating and forecasting population.

While estimation and forecasting has been the main use of the H-P method, over the past few years we have been investigating potential refinements to this method as well as other applications for the CCR including stable population analysis and estimating historical populations. The results of this research have mainly been presented at professional conferences and to date in only one publication. We decided it would be worthwhile to write a book that pulls together both published and unpublished research in one place to present a unified story of the CCR and to describe the various ways it can be used in both academic and applied demography. To our knowledge, this is the first book focused on the CCR. We had three goals in mind when writing this book: (1) enhancing the reputation and value of the CCR as being more than a second class citizen to its more widely valued cohort component method, (2) serving as a platform for future research into uses and applications of the CCR, and (3) providing a reference guide for those wanting to implement CCR applications.

In terms of cooperation, we want to thank Dr. Frank Trovato for permission to use a 2012 article published by Swanson and Tedrow in *Canadian Studies in Population* as the basis for Chapter 11. Similarly, the general idea found in Chapter 12 is based on a chapter by Swanson, Tedrow, and Baker in *Dynamic Demographic Analysis* (2016), a book edited by Robert Schoen. Materials in Chapter 6 are taken from a 2013 paper by Swanson and Tayman, which appears in *Proceedings of the 6th EUROSTAT/UNECE Work Session on Demographic Projections* (2014), edited by Marco Marsili and Giorgia Capacci. Materials in Chapters 5, 10, and 16 come from papers presented at various conferences, including those of the British Society for Population Studies, the Canadian Population Society, and the Southern Demographic Association. We are grateful for the comments and suggestions made at these conferences. Baker would also like to acknowledge the contribution of collaborators Dr. Adelamar Alcantara (Geospatial and Population Studies, University of New Mexico) and Mr. Xiaomin Ruan (Population Forecasts Program, Portland State University) for much of the material developed in Chapters 12 and 14. Importantly, Eddie Hunsinger deserves our thanks for his willingness to host the Excel files underlying tables in this book (see Chapter 3) on the "Applied Demographer's Toolbox" website.

We decided to undertake this project together because without our joint collaboration this book would likely never have been written. We thank Stan Smith for writing the foreword. Stan has made many significant contributions to the field of applied demography, and our collaboration with him over many years has made this a better book. We also thank Evelien Bakker and Bernadette Deelen-Manns for shepherding the book through Springer's production process; their assistance was invaluable. A project such as this requires a significant time commitment and required the patience and understanding of our families to complete. Dave, Jeff, and Lucky dedicate this book to their wives, Rita, Melinda, and Loretta, and Jack dedicates it to his parents (David and Betty Baker) and sons (Nate and Alex) who all helped in more ways than they can imagine.

Lake Forest, IL, USA — Jack Baker
Riverside, CA, USA — David A. Swanson
San Diego, CA, USA — Jeff Tayman
Bellingham, WA, USA — Lucky M. Tedrow

Contents

Chapter 1
Introduction

1.1 Why a Book on Cohort Change Ratios?

Why write a book about cohort change ratios (CCRs)? The answer is that CCRs have a wide range of uses and a high level of utility, features useful to applied demographers but which we believe have been largely overlooked. So, this book is aimed at showing how cohort change ratios can deliver a wide range of timely and cost-effective demographic information with a good level of precision. The book is primarily designed for use by applied demographers, but planners and others who generate and use demographic information to guide decision-making and policy in both the private and public sectors should find it both informative and accessible. The book also can be used in conjunction with a course on demographic methods or as a supplement from which chapters can be selected to fit into a number of courses, including applied demography, demographic methods, business forecasting, economic forecasting, and market research, among others. At the end of this chapter, we discuss its potential classroom use in more detail.

Although the general idea of a CCR has been around for at least 100 years (Hardy and Wyatt 1911) and it has been widely used to generate population forecasts since their "re-introduction" by Hamilton and Perry (1962), CCRs have largely remained a tool of applied demographers who generate population forecasts (Smith et al. 2013: 176–181). We started discovering (or more likely, re-discovering) more of their uses and features because many projects we have worked over the years called for techniques and data that were not generally found in the applied demographer's tool kit. Because we were familiar with them in the forecasting context, it did not take long to realize that they could be used more broadly (Swanson and Tayman 2012: 201–204). The more we used CCRs, the more we learned about their features, which revealed even more uses and features to us.

Before we jump into descriptions of these uses and features, we need to provide some background. Thus, in this chapter we first describe what cohorts are and give a brief introduction to their uses in sociology and demography. We then describe

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_1

CCRs and related measures and give an idea of their applications. Finally, we provide an outline of the book along with some suggestions about how it might be read and also how it might be used in the classroom.

1.2 Cohorts and Their Analyses

A cohort is a group of people who experience the same demographic event during a particular period of time such as their year of birth, marriage, or death (Swanson and Stephan 2004: 755). Cohorts typically are constructed using an "initiating" signal event, such as birth, but they also can be constructed using a "terminating" signal event, such as death (Swanson 1986). Cohorts also can be "synthetic." That is, they can be an analytic construct, rather than a set of empirical observations. The period life table, for example, can be viewed as a synthetic cohort (Kintner 2004: 306–307).

Norman Ryder (1951, 1965) is usually credited with establishing the use of a cohort as a unit of analysis and he did much to deserve this credit, especially in regard to the study of fertility (Quiñones 2010). However, others preceded him in using cohorts as a unit of analysis, notably in regard to the study of mortality (Dublin et al. 1949: 174–182).

Today, cohort analysis is widely used and not only in academic circles (Ahlburg 1986, Berger 1985, Carlson 1992, Easterlin 1987). It has found a home, for example, in the private sector, where it is used to study consumer behavior (Martins et al. 2012: 169–196), often in the form of defining first-time purchasers of a product or service as a cohort and following it in order to assess cumulative lifetime value.

Another use of a cohort as a unit of analysis is found in the construction of a "cohort change ratio," which was described as early as 1911 (Hardy and Wyatt 1911) and actually first specified and used by Hamilton and Perry (1962). This methodological and conceptual construct has gained traction in the field of population forecasting, especially for small areas (Smith et al. 2013: 176–181; Swanson and Tayman 2014; Swanson et al. 2010).

1.3 The Cohort Change Ratio

Cohort change ratios (CCRs) are found throughout this book, so it is appropriate to discuss this concept here. A cohort change ratio (CCR) is typically computed from age-related data in the two most recent censuses:

$$ {}_{n}CCR_{x,t} = {}_{n}P_{x,t} / {}_{n}P_{x-k,t-k}. \quad (1.1) $$

where,

${}_nP_{x,\,t}$ is the population aged x to $x + n$ at the most recent census (t),
${}_nP_{x-k,\,t-k}$ is the population aged $x-k$ to $x-k + n$ at the 2nd most recent census (t–k), and
k is the number of years between the most recent census at time t and the one preceding it at time (t–k).

As implied by Eq. 1.1, a cohort change ratio is not typically computed for a single cohort, but for all cohorts found in two successive census counts.

Given the nature of the CCR in Eq. 1.1, the youngest 5 year age group for which a CCRs numerator can be constructed is 10–14 if there are 10 years between censuses. That is, we can construct the denominator for this cohort aged 10–14 using age group 0–4 from the census taken 10 years earlier. By taking the ratio of those aged 10–14 in the most recent census to those aged 0–4 in the preceding census, we have a CCR. However, we cannot construct a CCR using either those aged 0–4 or those aged 5–9 in a given census as the numerator because the members of these two respective cohorts are not found in the preceding census, given 10 years between censuses. To analyze age groups younger than ten in a given application, a child-adult ratio (CAR) can be used. This ratio, computed separately for ages 0–4 and ages 5–9, relates young children to adults in the age groups most likely to be their parents (Smith et al. 2013: 178). Chapter 4 discusses other approaches for dealing with these age groups.

For the terminal, open-ended age group (e.g., ages 75 years and older), one uses the same approach found in life table construction (Smith et al. 2013: 178). As such, the CCR for a terminal, open-ended age group differs slightly from those for the closed age groups beyond age 10 preceding it. If, for example, the final closed age group is aged 70–74, with persons aged 75 years and older as the terminal open-ended age group, calculating the $CCR_{75+,t}$ requires the summation of the three oldest age groups (65–69, 70–74, and 75 years and older) to get the population age 65 years and older at time t–k:

$$ {}_{\infty}CCR_{75,t} = {}_{\infty}P_{75,t} / {}_{\infty}P_{65,t-k}. \tag{1.2} $$

Table 1.1 provides an example of a complete set of CCRs for the total population of Riverside County, California between 2000 and 2010.

In viewing Table 1.1, recall that the CCRs for those aged 0–4 and 5–9 are actually CARs computed from the 2010 census as follows: ${}_5CAR_{0,\,2010} = {}_5P_{0,2010}/{}_{15}P_{20,\,2010}$; and ${}_5CAR_{5,\,2010} = {}_5P_{5,\,2010}/{}_{15}P_{25,\,2010}$. That is, the CAR for those aged 0–4 in 2010 (0.37171) is found by dividing the number of persons aged 0–4 (162,438) by the number of adults aged 20–34 (154,572 + 143,992 + 138,437) and the CAR for those aged 5–9 in 2010 (0.39184) is found by dividing the number of persons aged 5–9 (167,065) by the number of adults aged 25–39 (143,992 + 138,437 + 143,992). We see the CCRs vary from a low of 0.20139 for the cohort of people aged 75 years and older in 2010 (found by dividing the number aged 75+ in 2010 by the number 65+ in 2000) to a high of 1.50517 for the cohort of people aged 35–39 in 2010 (found by dividing the

Table 1.1 Total population CCRs, Riverside County, California, 2000–2010

Age	Population 2000	2010	CCR[a]
0–4	121,629	162,438	0.37171
5–9	139,468	167,065	0.39184
10–14	133,886	177,644	1.46054
15–19	119,725	187,125	1.34171
20–24	96,374	154,572	1.15450
25–29	95,621	143,992	1.20269
30–34	108,602	138,437	1.43646
35–39	124,260	143,926	1.50517
40–44	117,910	149,379	1.37547
45–49	96,484	152,722	1.22905
50–54	79,538	140,016	1.18748
55–59	61,880	114,765	1.18947
60–64	54,046	98,974	1.24436
65–69	52,309	78,495	1.26850
70–74	50,845	62,103	1.14908
75–79	44,184	49,003	0.93680
80–84	27,542	36,793	0.72363
85–89	14,399	22,399	0.50695
90+	6,685	9,793	0.20139
Total	1,545,387	2,189,641	n/a

Source: U.S. Census Bureau (http://factfinder2.census.gov)
[a]Ages 0–4 = $P_{0-4,t}/{}_{15}P_{20,t}$
Ages 5–9 = $P_{5-9,t}/{}_{15}P_{25,t}$
Ages 10–89 = $P_{x+10,t}/P_{x,t-10}$
Ages 90+ = $P_{90+,t}/P_{80+,t-10}$

number aged 35–39 in 2010 by the number aged 25–29 in 2000). CCRs are never less than 0.00 and in principle can become very high.

A CCR in excess of 1.00 means that net in-migration occurred over the period between the two census counts used to construct it. For example, the CCR of 1.50517 for those aged 35–39 in 2010 means that there was net in-migration between 2000 and 2010 for those who were aged 25–29 in 2000. As you can see by perusing Table 1.1, all of the CCRs are higher than 1.00 for those aged 10–14 in 2010 to those aged 70–74 in 2.010. This indicates a substantial net in-migration occurred for these age groups. A CCR between 0.00 and 1.00 can imply net out-migration, mortality in excess of net in-migration or a combination of the two. CCRs that are less than 1.00 for younger age groups (i.e., those aged less than 55 in the most recent census) indicate net out-migration because the effect of mortality is low. CCRs less than 1.00 for older age groups (i.e., those aged 75 years and older in the most recent census) typically indicate the effects of mortality because migration is often low among these age groups. Given this, it is noteworthy that the age groups in Riverside County from 50 to 54 through 70 to 74 in 2010 all show CCRs in excess of 1.00 indicating volumes of net in-migration sufficient to offset the effects of higher mortality found at these ages.

As can be seen from Table 1.1, the data needed to assemble a set of CCRs is relatively easy to obtain and the calculations are easy to do. We have used census data to develop the example, but we could have used administrative records or survey data just as easily. The major requirement is that the width of the age groups for which CCRs are desired needs to be consistent with the length of time between the two points in time from which the input data are assembled. In the case of Table 1.1, we used 5 year age groups which are consistent (evenly divisible by) with the 10 years between the two points in time. We could have used 10 year age groups just as easily. With the exception of the terminal, open-ended age group (e.g., 85 years and over), the remaining age groups should all be of the same width even if they are consistent with the length of time between the two sets of data.

However, we could not have directly assembled CCRs if we used 5 year age groups and the data were taken from observations only 2 years apart. Situations where the width of the age group is not consistent with the length of time between the two sets of data can be accommodated, but they require "age splitting" (Judson and Popoff 2004) and re-assembly of the results into age groups that are consistent with the length of time between the two sets of data. For example, it is not uncommon to encounter 5 year age groups except for those between ages 15 and 24, where they may be tabulated, for example, as 15–18, 19–20, and 21–24. In such a case, age group 19–20 would have to be split such that those aged 19 are separated from those aged 20 so that the former could be added to those aged 15–18 and the latter to those aged 21–24, forming age groups 15–19 and 20–24, respectively. Fortunately, age related data provided by agencies in most countries are consistent with the length of time between two successive data sets, avoiding the need for age splitting and data re-assembly.

One feature of CCRs that is implicit in the preceding discussion is that they yield patterns representing demographic change. If, for example, the CCRs for those aged 10–14 are above 1.00 (indicating net in-migration), there will be one or more CCRS above 1.00 in the age groups that are likely to represent the parents of those aged 10–14. With the exception of special circumstances such as forced migrations, children move with their parents. CCRs also form broader patterns, usually associated with geography. For example, areas representing urban centers often have CCRs above 1.00 for those of college age and CCRs less than one for those of post-college age. In many respects, graphs of a full set of CCRs by age for urban, suburban, and rural areas can be seen to fit into net migration typologies similar to those developed by Pittenger (1974).

As suggested earlier, cohorts can be defined by criteria other than age. Sometimes the criteria are associated with age, such as K-12 school enrollment by grade or year of high school graduation. However, it is always the case that once a cohort has been defined at a given point in time, one can also determine age. This is a direct result of what is known as the "age-period-cohort" issue, whereby having defined two of the three, the third is determined (Bloom 1987). This can be illustrated by the following equation:

$$p = a + c \tag{1.3}$$

where,

p = period,
a = age, and
c = cohort.

As an example of Eq. 1.3, consider a person who is 50 years of age and who was part of the 1966 birth cohort. Knowing these two elements (a and c, respectively) we can determine that the period (p) in question is 2016. Similarly, if we know that the period is 2016 and that a person was part of the 2000 birth cohort, then we know the person is 16 years of age.

It is worthwhile to note that one could calculate a "Cohort Change Difference" as follows:

$${}_{n}CCD_{x,t} = {}_{n}P_{x,t} - {}_{n}P_{x-k,t-k}. \tag{1.4}$$

where,

${}_{n}P_{x,\,t}$ is the population aged x to $x + n$ at the most recent census (t),
${}_{n}P_{x-k,\,t-k}$ is the population aged $x-k$ to $x-k + n$ at the 2nd most recent census ($t-k$), and
k is the number of years between the most recent census at time t and the one preceding it at time $t-k$.

Unlike a CCR, a CCD has no lower boundary. That is, it can become negative, not bounded on the lower end by zero. In terms of comparisons, when a CCR is greater than 1.00, its corresponding CCD will be positive and when a given CCR is less than 1.00, its corresponding CCD will be negative (less than zero). Given this and other properties (e.g., neither division nor subtraction is associative and commutative in terms of their mathematical properties), there is no major advantage in using a CCD compared to its corresponding CCR. However, the conventional approach that can be directly traced to Hamilton and Perry (1962) is to use a CCR.

1.4 Reverse CCRs

As illustrated in Chapter 10, it is possible to construct CCRs that go backward in time. Not surprisingly, these are known as Reverse CCRs (RCCRs) and can generally be described as:

$${}_{n}RCCR_{x-k,t-k} = {}_{n}P_{x-k,t-k} / {}_{n}P_{x,t} \tag{1.5}$$

where,

${}_{n}P_{x,t}$ is the population aged x to $x + n$ at the most recent census (t),
${}_{n}P_{x-k,t-k}$ is the population aged $x-k$ to $x-k+n$ at the 2nd most recent census $(t-k)$, and
k is the number of years between the most recent census at time t and the one preceding it at time $t-k$.

As is the case with a CCR, there are special conditions that need to be taken into account when calculating an RCCR, which are discussed in Chapter 10. However, there is no major reason why a Reverse Cohort Change Difference could not be computed and used, other than following convention.

1.5 Census Survival Ratios

A Census Survival Ratio (CSR) is a special case of a CCR where $0 \leq {}_{n}\text{CSR}_{x,\,t} \leq 1.0$. With the upper limit of a CSR established as 1.0, we interpret that to mean that no migration is present. Although the lack of migration in a population is uncommon, non-migration does occur (or nearly so) in some actual populations; and it is an important idea in the analysis of mortality via a life table, whereby a CSR can be viewed as the probability of a member of a given cohort surviving into the future. These and related issues are covered in detail in Chapter 11.

1.6 Some Applications of Cohort Change Ratios

As you will see in this book, cohort change ratios (CCRs) have a wide range of applications. Many applications relate to the construction and evaluation of population forecasts and current and historical population estimates.[1] CCRs are very valuable when one wants to forecast or estimate the population of a small area, such as a census tract or school district, and other sub-county units (e.g., townships, fire districts, legislative districts), or statistical geographies (e.g., block group, block). Their value stems from the fact that there is a minimal amount of input data required to generate them and the forecasts and estimates they are used to make. Small area population projections and estimates are a major staple in both the private and public sectors (Swanson 2015, Swanson and Pol 2004, Swanson and Pol 2008, Swanson et al. 2010, Yusuf and Swanson 2010). Private sector uses include determining housing demand, business site location, market valuation, assessing profitability, and assembling consumer profiles. Public sector uses are often in

[1]In general, an estimate refers to information for a current or past date in the absence of a census, whereas a projection refers to information to a time beyond the current date. A forecast is a projection deemed or judged most likely to occur. Chapter 2 discusses these concepts in greater detail.

regard to transportation and strategic planning, land use zoning, and economic development.

CCRs are also useful in forecasting school enrollments and as the basis for forecasting a wide range of social, economic, and health outcome related characteristics. The utility of CCRs extends beyond general estimation and forecasting applications into other areas of demography such as the determination of life expectancy and stable population theory. CCRs can also be proxies for survival rates in places lacking vital statistic data of age-specific mortality as discussed in Chapter 11.

In summary, we believe that there are several theoretical and practical reasons to use cohort change ratios as measures of cohort change. One of them is that they are preferable to cohort change differences in terms of a measure of cohort change. Among other benefits, ratios will not fall below zero, which is not the case for differences. This issue is not a minor one, as has been pointed out by Swanson (2004) in regard to using the ratio-correlation method of population estimation, which yields a time-based regression model that meets the general condition of "stationarity," an important feature in constructing valid and reliable population models that incorporate temporal change. Another feature of cohort change ratios is they are non-linear whereas cohort change differences are essentially linear, and cohort change ratios have a natural affinity to probability and, by extension, to measures such as the odds ratio and relative risk, issues discussed in Chapter 2.

1.7 About This Book

Now that we know something about CCRs and have an overview of their uses, it is natural to ask what specifically can be done with them. In providing an answer to this question, this book is organized into 16 chapters (including this one) along with two appendices, an author index, and a subject index. Even if you are an experienced demographer, we suggest reading Chapters 2, 3, and 4 before moving on to the other chapters, which depending on your experience, can be read in any order. Chapter 2 covers basic demographic concepts and terms while Chapter 3 describes sources of data that can be used to develop CCRs. Chapter 4 shows how to forecast the size and composition of a population from two census counts. It is the foundation for the other chapters, which generally progress from basic to more advanced applications.

Chapter 5 shows how CCRs can be modified in order to more fully capture the dynamics of demographic change while Chapter 6 describes a method for generating formal measures of uncertainty for population forecasts made using CCRs. Using CCRs to develop both short-term and long-term school (K-12) enrollment forecasts is the subject of Chapter 7. Chapter 8 shows how CCRs can be used to generate forecasts of a wide range of characteristics of interest (health outcomes, labor force, and so forth), while Chapter 9 shows how CCRs can be used to generate population size, composition, and characteristics for a current point in time. In

Chapter 10, the use of CCRs for purposes of estimation is extended to developing population size, composition, and characteristics for the past. In Chapter 1, we show how CCRs can be used to develop a standard demographic measure, life expectancy.

An advanced application is found in Chapter 12, which applies the CCR approach to a major canon of formal demography, stable population theory. The theoretical aspects of CCRs are explored in Chapter 13, which deals with decomposing the factors making up differences in CCRs. Chapter 14 provides an overview of how CCRs can used in spatial applications while Chapters 15 and 16 offer summary remarks, with the former providing a discussion of the utility of the CCR approach and the latter providing concluding remarks, including some ideas about the future of the CCR approach.

A short, but important, proof is given in the Appendix whereby the CCR approach is shown to be consistent with the fundamental demographic theorem. Author and subject indices follow the appendix.

For someone who is interested in using this book in the classroom, it is organized so that the chapters proceed from basic ideas to applications to advanced applications, which fits the lecture format for a full term course. The writing style also fits this approach. In the event it is used as a supplemental textbook, both the organization and writing style should work for this purpose as well. As a textbook, *Cohort Change Ratios and Their Applications* is designed to accommodate the trend toward graded assignments as a form of student assessment rather than closed-book and other forms of examinations. The examples in this book can be assigned to students to replicate as tutorial (non-graded) assignments, with graded assignments being similar but using different data (e.g., a tutorial assignment would involve forecasting the multi-race population of California, while a graded assignment would involve forecasting the multi-race population of the U.S., for which the excel file for California can be used as a template for the U.S.). As already noted, Chapter 3 contains the URL for an online site where you can find excel templates and files containing the data and computational statements used to generate the book's examples.

If used in a class, learning outcomes can be derived from the book's general objectives, which are to provide (1) basic demographic and measurement concepts without requiring prior major demographic, mathematical or statistical skills; (2) useful analytical frameworks and tools; (3) a basic understanding of demographic factors that affect a wide range of applications, such as current or future market size and segments; (4) theoretical and conceptual foundations of the cohort perspective and its implementation in the form of cohort change ratios; and (5) examples that can be used as the basis for case studies (Swanson and Morrison 2010).

References

Ahlburg, D. (1986). A relative cohort size forecasting model of Canadian total live births. *Social Biology, 33*, 51–56.

Bloom, D. (1987). On the nature and estimation of age, period and cohort effects in demographic data. *Genus, 43*(1/2), 59–68.

Berger, M. (1985). The effect of cohort size on earnings growth: A re-examination of the evidence. *Journal of Political Economy, 93*, 561–573.

Carlson, E. (1992). Inverted Easterlin fertility cycles and Kornai's 'soft' budget constraint. *Population and Development Review, 18*(4), 669–688.

Dublin, L., Lotka, A. J., & Spiegelman, M. (1949). *Length of life*. New York: The Ronald Press Company.

Easterlin, R. (1987). *Birth and fortune* (2nd ed.). New York: Basic Books.

Hamilton, C. H., & Perry, J. (1962). A short method for projecting population by age from one decennial census to another. *Social Forces, 41*, 163–170.

Hardy, G., & Wyatt, F. (1911). Report of the actuaries in relation to the scheme of insurance against sickness, disablement, etc. *Journal of the Institute of Actuaries, XLV*, 406–443.

Judson, D., & Popoff, C. (2004). Selected general methods. In J. Siegel & D. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 677–732). San Diego: Elsevier Academic Press.

Kintner, H. (2004). The life table. In J. Siegel & D. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 301–340). San Diego: Elsevier Academic Press.

Martins, J., Yusuf, F., & Swanson, D. (2012). *Consumer demographics and behaviour: Markets are people*. Dordrecht: Springer.

Pittenger, D. (1974). A typology of age-specific net migration rate distributions. *Journal of the American Institute of Planners, 40*(4), 278–283.

Quiñones, E. (2010). Norman Ryder, renowned demographer and leader in fertility studies, dies. *News at Princeton* (12 July 2016). Retrieved from https://www.princeton.edu/main/news/archive/S27/89/47G33/index.xml?section=topstories.

Ryder, N. (1951). *The cohort approach*. Ph.D. dissertation, Princeton University (reprinted in 1980 by New York: Arno Press).

Ryder, N. (1965). The cohort as a concept in the study of social change. *American Sociological Review, 30*(6), 843–861.

Smith, S. K., Tayman, J., & Swanson, D. (2013). *A practitioner's guide to state and local population projection*. Dordrecht: Springer.

Swanson, D. (1986). Timing the second birth: Fecundability models for selected race and age groups in Hawai'i. *Janasamkhya, 4*(21), 82–113.

Swanson, D. (2004). Advancing methodological knowledge within state and local demography: A case study. *Population Research and Policy Review, 23*(4), 379–398.

Swanson, D. (2015). Applied demography. In J. Wright (Ed.), *International encyclopedia of the social and behavioral sciences* (2nd ed., pp. 839–844). Oxford: Elsevier.

Swanson, D., & Morrison, P. (2010). Teaching business demography using case studies. *Population Research and Policy Review, 29*(1), 93–104.

Swanson, D., & Pol, L. (2004). Contemporary developments in applied demography within the United States. *Journal of Applied Social Science, 21*(2), 26–56.

Swanson, D., & Pol, L. (2008). Applied demography: Its business and public sector components. In Y. Zeng (Ed.), *The encyclopedia of life support systems, demography volume*. Oxford: UNESCO-EOLSS Publishers.

Swanson, D., & Stephan, G. (2004). Glossary. In J. Siegel & D. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 751–778). San Diego: Elsevier Academic Press.

Swanson, D., & Tayman, J. (2012). *Subnational population estimates*. Dordrecht: Springer.

Swanson, D., & Tayman, J. (2014). Measuring uncertainty in population forecasts: A new approach. In M. Marsili & G. Capacci (Eds.), *Proceedings of the 6th EUROSTAT/UNECE work session on demographic projections* (pp. 203–215). Rome: National Institute of Statistics.

Swanson, D., Schmidt, R., & Schlottmann, A. (2010). Forecasting the population of census tracts by age and sex: An example of the Hamilton-Perry method in action. *Population Research and Policy Review, 29*(1), 47–63.

Yusuf, F., & Swanson, D. (2010). Business demography in the 21st century. *Population Research and Policy Review, 29*(1), 1–3.

Chapter 2
Basic Demographic Concepts

2.1 Introduction

We begin this chapter by discussing estimates, projections, and forecasts and continue with discussions of the size of the population; its distribution across geographic areas; its composition (e.g., age, sex, race, and other characteristics); and changes in population size, distribution, and composition over time. We also define a number of basic demographic concepts, define some commonly used terms, describe a number of statistical measures used in demography, and present the participation rate method for doing forecasts of population-related characteristics. This chapter is designed to give readers with little training or experience in demography a brief introduction to the field and sets the stage for the topics covered in the remainder of the book.

2.2 Estimates, Projections, and Forecasts

An important distinction is made between estimates, on the one hand, and on the other, projections and forecasts. The most fundamental difference is that estimates refer to the present or the past while projections and forecasts refer to the future. In addition, estimates are often based on data for corresponding points in time. For example, estimates for 2016 made in 2015 can be based on data (e.g., births, deaths, building permits, school enrollments, and Medicare enrollees) reflecting population growth through 2015. However, such data do not yet exist for forecasts for 2025 made in 2016.

The distinction between estimates and forecasts is not always clear-cut. Sometimes no data are available for constructing population estimates. For example, calculations of a city's age-sex composition in 2016 made in 2015 may have to be based on the extrapolation of 2000–2010 trends because data series reflecting post-

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_2

2010 changes in age-sex composition may not be available. Should these calculations be called estimates or forecasts? In this book, we refer to calculations extending beyond the date of the last observed data point as forecasts or projections and calculations for all prior dates as estimates.

A population projection is the numerical outcome of a particular set of assumptions regarding future population trends (Isserman 1985; Keyfitz 1972; Weeks 2014). Some projections refer to total population while others provide breakdowns by age, sex, race, and other characteristics. Some focus solely on changes in total population while others distinguish among the individual components of growth—births, deaths, and migration. Population projections are conditional statements about the future. They show what the population would be if particular assumptions were to hold true, but make no predictions as to whether those assumptions actually will hold true. Population projections are always "right," barring a mathematical error in their calculation, and can never be proven wrong by future events. A population forecast, on the other hand, is the projection the analyst (i.e., the person or agency making the projection) believes is most likely to occur in the future. Unlike projections, forecasts are explicitly judgmental. They are unconditional statements reflecting the analyst's views regarding the optimal combination of data sources, projection techniques, and methodological assumptions, leavened by personal judgment. Population forecasts can be proven right or wrong by future events and can be found to have relatively small or large errors.

Demographers have traditionally and typically use the term projection to describe calculations of the future population. There are several reasons for choosing this terminology. Projection is a more inclusive term than forecast. A forecast is a particular type of projection; namely, the projection the analyst believes is most likely to provide an accurate prediction of the future population. Given this distinction, all forecasts are projections but not all projections are forecasts. Also, demographers often intend their calculations of future population to be merely illustrative rather than predictive; projection fits more closely with this intention than forecast. In this book we use the term forecast when discussing calculations of future events, in part for simplicity, but also because most users view prognostication about the future as forecasts (rather than projections) regardless of the intent of the producer (Smith et al. 2013: 323)

2.3 Demographic Concepts

2.3.1 Size

Population estimates and forecasts start with the same basic consideration as a census: What is the size of a population? The concept of population size refers to the number of people residing in a specific area at a specific time (the de jure approach). According to January 1, 2014 population estimates, Loving County, Texas had a population of 82, whereas Harris County, Texas had a population of 4,365,601 (Texas State Data Center 2015). These were the largest and smallest counties in Texas in terms of population size. However, the concept of population size can also refer to the number of people actually present in a given area at a given time (the de facto approach). Under a de facto count, all tourists, business travelers, seasonal residents, and workers in downtown Boston, MA would be counted along with usual residents who are also in downtown that day (Swanson and Tayman 2011). Usual residents of downtown Boston who were out of town would not be counted. De facto population estimates have many uses including dealing with potential traffic congestion, long commuting times, or disaster and relief activities to know the number of people that may be affected in an emergency.

The de jure concept is more ambiguous in that it comprises all of the people who "belong" to a given area by virtue of legal residence, usual residence, or some similar criterion (Wilmoth 2004: 65). However, the de jure concept is used as the census definition of population in the United States, Canada, and most other developed countries and, as such, becomes the dominant concept in population estimation and forecasting. The methods covered in this book use the de jure concept. A discussion of methods for estimating de facto populations is found in (Swanson and Tayman 2012: 313–327).

2.3.2 Distribution

The distribution of a population refers to its geographic location; there are two major ways in which geographic areas have been identified. The first is the administrative approach, where areas are defined according to administrative or political criteria. Examples include states, counties, cities, U.S. congressional districts, and a wide variety of state and local administrative and political delineations (e.g., city council, water, and school districts). For many purposes these are the most important types of geographic areas that can be defined (Plane 2004). However, administrative areas also have several limitations. Their boundaries may not account for important economic, cultural, and social considerations. For example, Gary, Indiana is administratively distinct from the city of Chicago, Illinois, but it is economically, culturally, and socially linked to it. Another problem is that administrative boundaries may not remain constant over time–annexations by a city are a case in point–and changing boundaries make it difficult not only to make comparisons over time, but to produce consistent estimates and forecasts.

One way to avoid some of the limitations imposed by administrative definitions is to define geographic areas specifically for purposes of identifying areas that are economically, socially, and culturally linked. These so-called statistically defined areas are used in many countries, including the United States (Plane 2004).

In the United States, important statistical areas are based on geography used in the census—blocks, block groups, and census tracts. Blocks are small areas bounded on all sides by visible features such as streets or railroad tracks or by

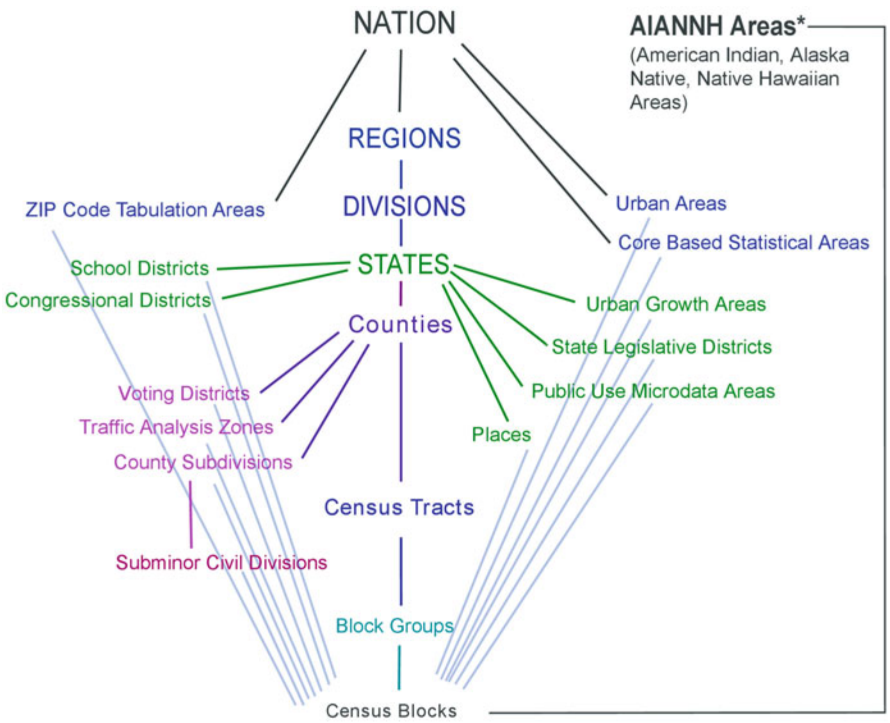

Fig. 2.1 Standard hierarchy of census geographic entities (Source: U.S. Census Bureau (http://www.census.gov/geo/www))

invisible boundaries such as city or township limits; they are the smallest geographic unit for which census data are tabulated. Block groups are clusters of blocks and generally contain 250–550 housing units; block groups do not cross census tract boundaries. Census tracts are small, relatively permanent areas that do not cross county boundaries. These areas generally contain between 2,500 and 8,000 persons and are designed to be relatively homogeneous with respect to population characteristics, living conditions, and economic status. Figure 2.1 shows a hierarchy of geographic areas built from the 2010 U.S. census geography. Geographic areas may work in a hierarchical fashion, with smaller areas nesting in larger ones (e.g., census tracts within counties, counties within states, and states within the U.S.), while others like Core Based Statistical Areas do not cover all of the U.S.

Geographic boundaries can also be defined according to other criteria. In the United States, for example, one can obtain census data for Postal ZIP code areas and data for market areas that are important for businesses. It is not uncommon to produce estimates and forecasts for a combination of administrative and statistical areas. Figure 2.2 shows an example of one such system used in San Diego County, California. Master Geographic Reference Areas combine census geography,

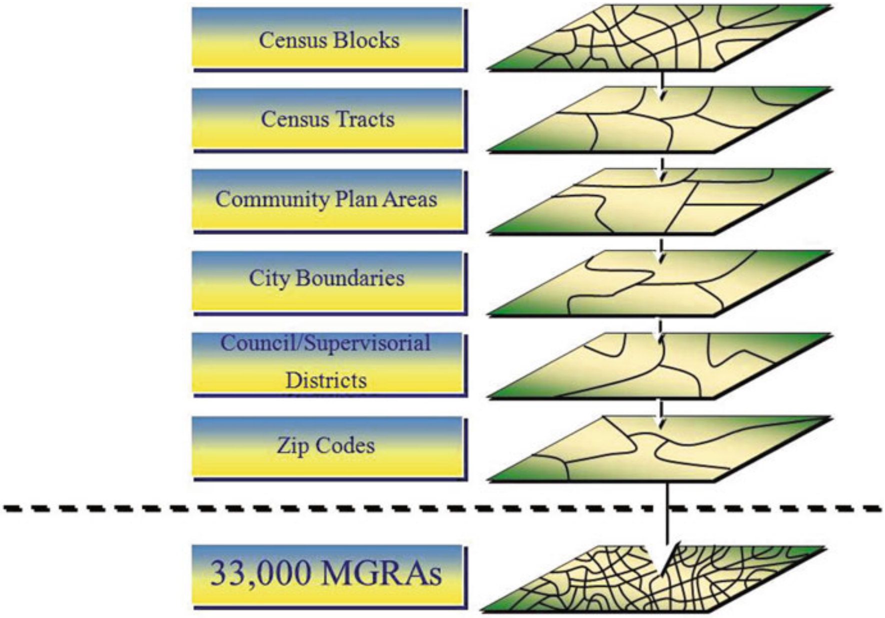

Fig. 2.2 Master Geographic Reference Areas in San Diego County, California (San Diego Association of Governments 2010)

political boundaries, and zip codes into a spatially detailed spatial system that supports a wide range of uses.

2.3.3 Composition

Composition refers to the characteristics of the population. For population estimates and forecasts the most commonly used characteristics are age, sex, race, and Hispanic origin. For many purposes, age is the most important demographic characteristic because it has such a large impact on so many aspects of life, for individuals as well as for society as a whole. The age structure of a population affects its birth, death, and migration rates, and the demand for public education, health care, and nursing home care. It also impacts the housing market, the labor market, and the marriage market. No other characteristic is more valuable for a wide variety of planning and analytical purposes than the age composition of the population (Smith et al. 2013: 23). Sex composition also is important for many purposes. It is often used in combination with age to show a population's age-sex structure.

The age-sex structure is often illustrated using population pyramids (Hobbs 2004: 161–166). Population pyramids are graphic representations showing the number (or proportion) of the population. The basic pyramid form consists of bars, representing age groups in ascending order from the lowest to the highest, pyramided horizontally on one another. The bars for males are given on the left of a central vertical axis and the bars for females on the right of the axis. The characteristics of pyramids (e.g., the length of a bar compared to others, the steepness and regularity of its slope) for different populations quickly reveal any differences in the proportion of the sexes, the proportion of the population in any particular age class or classes, and the general age structure of the population.

Figure 2.3 shows pyramids for four populations with different age–sex structures. The pyramid for Yemen has a very broad base and narrows very rapidly. This pyramid illustrates an age–sex structure with a very large proportion of children, a very small proportion of elderly persons, and a low median age. It reflects a "young" population with relatively high fertility rates. The pyramid for Japan is very different. It has a relatively narrow base and a somewhat larger middle section. It illustrates an age–sex structure with a very small proportion of children, a very large proportion of elderly persons, and a high median age. It reflects an "old" population and relatively low fertility rates. The pyramid for Singapore has a very narrow base indicative of its very low fertility rate (less than one child per woman on average), but it has relatively few elderly and a large number of young adults aged 20 to 34 due to the migration of workers into this country. The U.S. pyramid has a fairly uniform look for the population aged 0–59 with variations due to the large baby boom cohort (aged 50–59) that was followed by the much smaller baby bust cohort (aged 35–49), and then the larger baby boomlet cohort (aged 20–34).

One pattern is the same in every country shown in Fig. 2.3. Females outnumber males at the older ages due to the cumulative effect of higher male mortality rates at virtually every age (Wisser and Vaupel 2014). For the population aged 65 years and older in 2015, the relative number of males to females ranges from 77.1 males per 100 females in Japan to 86.6 males per 100 females in Yemen (U.S. Census Bureau 2015).

Race and ethnicity are two other widely used demographic characteristics. For example, the U.S. Census Bureau uses five broadly defined racial categories: African American, American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander, and White. Starting with the 2000 census, there was an important change to the collection of racial data. The Census Bureau for the first time allowed respondents to list themselves as belonging to more than one racial category; prior to that time, respondents could list only a single category (McKibben 2004). In addition to race, the census uses an ethnic dimension, with two general categories: Hispanic and non-Hispanic. It should be noted that "Hispanic" is not a racial category; that is, people are classified both by race and by Hispanic origin. Composition also can refer to other characteristics such as employment and marital status, income, education, and occupation (O'Hare et al. 2004).

As illustrated in Table 2.1, Hispanics and race groups often have different demographic characteristics and patterns of growth that influence population estimation and forecasting. Between 2000 and 2010 in Los Angeles County, California, the percentage change in the Hispanic population is more than triple that of the overall population. Consequently, the Hispanic share of the total population

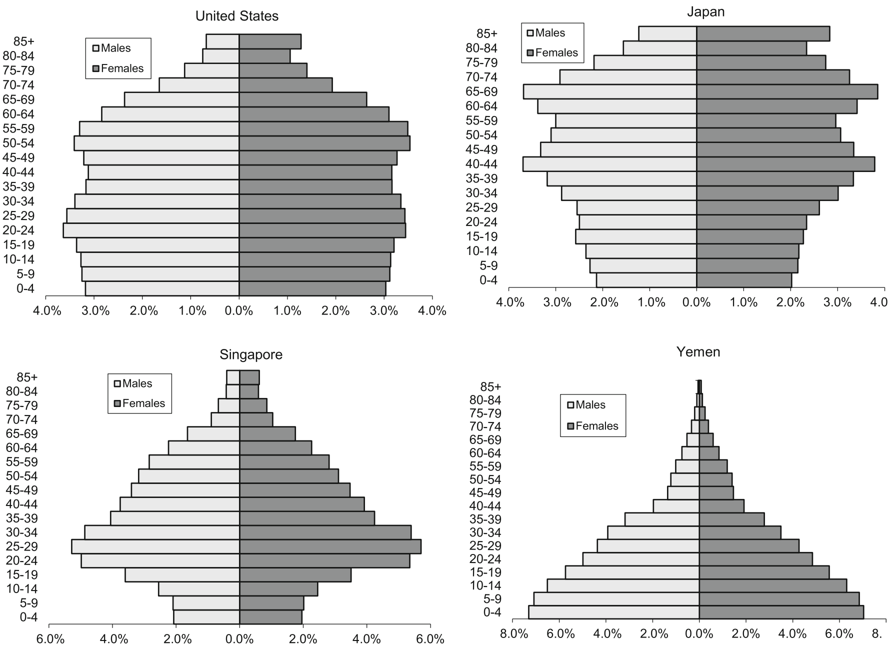

Fig. 2.3 Percent distribution by age and sex of the 2015 population of Japan, Singapore, United States, and Yemen (Source: U.S. Census Bureau International Data Base 2015)

increased from 44.5% in 2000 to 47.7% in 2010. Asians are the fastest growing race group, increasing by 18.4%, and their share of the total population rose from 11.9% in 2000 to 13.7% in 2010. Almost 2,141,000 people in Los Angeles County identified themselves as belonging to two or more race groups in 2010, almost 100,000 less than 10 years prior. Blacks, American Indians, and Native Hawaiians all lost population during the first decade of the twenty-first century in Los Angeles County. Non-Hispanic Whites, another widely used distinction, lost over 230,000 persons between 2000 and 2010, causing its share of the total population to drop from 31.3 to 27.8%.

Non-Hispanic Whites have the oldest age structure with a median age of 44.5 years in 2010, almost 16 years older than the Hispanic median age (see Fig. 2.4). Asians also have a relatively old population with a median age approaching 40 years. Other races and two or more races have the lowest median ages (27.4 and 25.2, respectively), indicative of their younger age structures.

2.3.4 Change

Population change is measured as the difference in population size between two points in time. A point in time can correspond to the date of a census or to the date

Table 2.1 Population by race and Hispanic origin, Los Angeles County, California, 2000 and 2010

	2000	2010	Change Number	Percent
All Races	9,519,338	9,818,605	299,267	3.1%
White	4,637,062	4,936,599	299,537	6.5%
Black or African American	930,957	856,874	−74,083	−8.0%
American Indian and Alaska Native	76,988	72,828	−4,160	−5.4%
Asian	1,137,500	1,346,865	209,365	18.4%
Native Hawaiian and Other Pac. Is.	27,053	26,094	−959	−3.5%
Other Races	2,239,997	2,140,632	−99,365	−4.4%
Two or More Races	469,781	438,713	−31,068	−6.6%
Hispanic or Latino	4,242,213	4,687,889	445,676	10.5%
Non-Hispanic White	2,959,614	2,728,321	−231,293	−7.8%

Source: U.S. Census Bureau (http://factfinder2.census.gov)

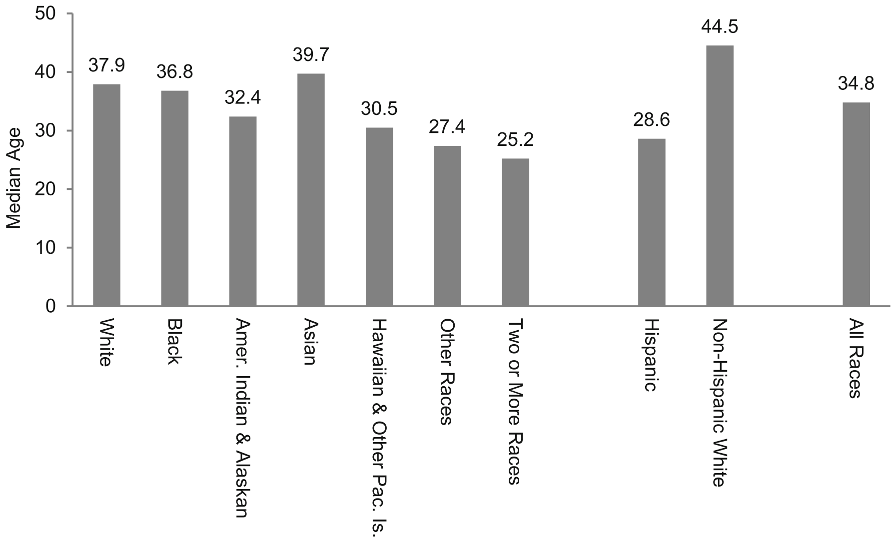

Fig. 2.4 Median age by race and Hispanic origin, Los Angeles County, California, 2010 (Source: U.S. Census Bureau, 2010 census (http://factfinder2.census.gov))

of a population estimate. Measures of population change always refer to a specific population and a specific period of time; in most instances, they refer to a specific geographic area as well. Population change can also be measured for various subgroups of the population (e.g., females, Asians, teenagers, etc.), different geographic areas (e.g., counties, cities), and different time periods (e.g., 2010–2015, 2000–2015). In other words, population change can refer to changes in size, distribution, or composition, or to any combination of the three.

Population change can be expressed in either numeric or percentage terms. Numeric change is computed by subtracting the population at the earlier date from the population of the later date. A percentage change is computed by dividing the numeric change by the population at the earlier date and multiplying by 100. Population change is often expressed in terms of an annual average. Average annual numeric change is computed by dividing the numeric difference by the number of years between the two endpoints. Average annual percentage changes or growth rates can be computed assuming discrete compounding (geometric) or assuming continuous compounding (exponential).

Measuring population change is simple and straightforward in many instances. However, changes in geographic boundaries, changes in the accuracy of base data, and changes in definitions makes measuring change difficult. Consistent measures of change are possible only if geographic boundaries are constant over time. This is generally the cases for states and counties, but may not be the case for many subcounty geographic areas. Changes in definition can also be problematic, such as comparing race in 1990 (where the respondents could choose only one racial category) with race in 2000 (where respondents could select multiple race categories). A more detailed discussion of population change is found in Smith et al. (2013: 25–27) and Perz (2004).

2.3.4.1 Components of Population Change

There are three components of population change: births, deaths, and migration. A population grows through the addition of births and migrants moving in, and declines through the subtraction of deaths and migrants moving out. Understanding these three demographic processes is essential to understanding the nature and causes of population change. Fertility is the reproductive performance of a woman, man, couple, or group; it also is a general term for the incidence of births in a population or group (Swanson and Stephan 2004: 760). Although fertility rates are generally low in the United States and other developed countries, they can vary substantially from place to place and from one race, ethnic or socioeconomic group within a given country. In 2013, the total fertility rate (average number of children a cohort of women will have during their lifetime) for states ranged from 1.6 in Vermont to 2.3 in Utah (Martin et al. 2015: Table 12). Mortality is a general term for the incidence of deaths in a population or group (Swanson and Stephan 2004: 767). While mortality rates do not vary greatly within high income countries, there are differences between race, ethnic and socioeconomic groups. In 2011, there was a 17.3 year difference in life expectancy at birth (average number of years of remaining life) between Black Males (69.5 years) and Asian Females (86.8 years) in Los Angeles County, CA (Los Angeles County Department of Public Health 2015).

Migration is a general term for the incidence of movement by individuals, groups, or populations seeking to make permanent changes of residence (Swanson and Stephan 2004: 766). It refers to changes in usual place of reference and excludes short-term temporary movements such as commuting, visiting friends or relatives, or taking a business trip. Migration levels and rates can vary considerably from place to place and from country to country and can undergo large sudden changes, making migration often the most difficult component of change to estimate and forecast (Smith et al. 2013: 103–104).

Migration can be viewed from several perspectives (Smith et al. 2013: 106–109). Gross migration refers to the total number of migrants into or out of an area (e.g. 500 migrants; 200 in-migrants plus 300 out-migrants). Net migration is the difference between the two (e.g., a net outflow of 100 persons); it shows the net effect of migration on the change in population. It is often useful to make the distinction between migration that occurs within a country and migration that occurs between countries. Internal or domestic migration refers to changes of residence within a country, while foreign or international migration refers to changes of residence from one country to another. The terms in-migrant and out-migrant refer to domestic migration. People leaving a country are known as emigrants and those entering a country are known as immigrants.

2.3.4.2 Fundamental Demographic Equation

The overall change in a population is formalized in the fundamental demographic equation:

$$\mathrm{P_l} - \mathrm{P_b} = \mathrm{B} - \mathrm{D} + \mathrm{IM} - \mathrm{OM} \tag{2.1}$$

where,

P_l is the population at the end of the time period,
P_b is the population at the beginning of the time period, and
B, D, IM, OM are the number of births, deaths, in-migrants, and out-migrants during the time period.[1]

The difference between births and deaths ($B - D$) is called natural change coming from the population itself. It may be either positive (natural increase) or negative (natural decrease) depending on whether births exceed deaths or deaths exceed births. The difference between *IM* and *OM* reflects the change in population due to migration and can be either positive or negative depending on whether in-migrants exceed out-migrants or out-migrants exceed in-migrants. The demographic balancing equation is a basic formula in demography and has other uses including deriving estimates of population and net migration (Smith et al. 2013: 30).

To illustrate the fundamental demographic equation, Table 2.2 shows births, deaths, natural change, and estimates of net domestic and net international migration from 2010 to 2014 for the 15 counties in Arizona. Natural increase accounts for

[1]The *IM* and *OM* terms include both domestic and foreign migrants. If information is only available on net migration the *IM* and *OM* terms would be replaced by $\pm NM$ (net migration).

Table 2.2 Components of population change, Arizona counties, 2010–2014

	Population[a] Change	Natural Increase	Births	Deaths	Net Migration Total	Inter-national.[b]	Domestic
Apache	310	1,974	4,371	2,397	−1,608	49	−1,657
Cochise	−3,909	2,040	7,194	5,154	−5,955	1,463	−7,418
Coconino	3,245	4,218	7,353	3,135	−1,134	503	−1,637
Gila	−478	−385	2,650	3,035	−120	119	−239
Graham	737	1,269	2,413	1,144	−543	1	−544
Greenlee	909	263	501	238	644	71	573
La Paz	−258	−66	841	907	−166	11	−177
Maricopa	269,834	118,394	232,032	113,638	146,372	38,922	107,450
Mohave	3,175	−2,949	8,005	10,954	5,464	2	5,462
Navajo	607	3,070	7,011	3,941	−2,465	148	−2,613
Pima	24,253	13,771	51,183	37,412	9,905	9,141	764
Pinal	26,148	9,435	20,023	10,588	14,917	3,088	11,829
Santa Cruz	−725	1,693	2,876	1,183	−2,485	339	−2,824
Yavapai	7,829	−3,417	7,763	11,180	10,862	637	10,225
Yuma County	7,497	7,713	13,509	5,796	−840	1,999	−2,839
Arizona	339,174	157,023	367,725	210,702	172,848	56,493	116,355

Source: U.S. Census Bureau, Population Division, March 2015. (http://www.census.gov/popest/data/counties/totals/2014/CO-EST2014-02.html)

[a]Total population change includes a residual. This residual represents the change in population that cannot be attributed to any specific demographic component

[b]Net international migration for the United States includes the international migration of both native and foreign-born populations. Specifically, it includes: (a) the net international migration of the foreign born, (b) the net migration between the United States and Puerto Rico, (c) the net migration of natives to and from the United States, and (d) the net movement of the Armed Forces population between the United States and overseas.

48% of the population change in Arizona, followed by domestic net migration (35%) and international net migration (17%). There is substantial variability in the demographic reasons for population growth among these counties. Eleven counties show natural increase, while deaths exceed births in Gila, La Paz, Mohave, and Yavapai Counties. All counties show positive growth due to international migration, but nine counties (60%) lose population as the result of domestic migration. Cochise County has the largest loss from domestic migration as well as the largest loss in total population of any county. In Pima County, the second largest county and home to Tucson, growth due to international migration is 11 times greater than the positive growth due to domestic migration. Growth due to domestic migration is substantially larger than growth due to international migration in Maricopa County, the largest county and home to Phoenix, and the adjacent Pinal County. The same pattern is seen in Mohave and Yavapai Counties. In the five counties with positive natural increase and positive total migration, the share of growth due to natural increase ranges from 29 to 59%.

2.4 Statistical Measures

Absolute measures focus on single numbers such as shown in Table 2.2. Relative measures emphasize the relationship between two numbers; they are typically expressed as ratios, proportions, percentages, rates, or probabilities. All the relative measures are similar to each other, but each has a distinct meaning.

2.4.1 Ratios

A ratio is simply one number divided by another. These could be any two numbers and do not need to have any particular relationship to each other. For example, one could calculate the ratio of storks to babies, the ratio of sunspots to gross domestic product, or the ratio of public transportation riders to automobile passengers in a major traffic corridor. To be useful, a ratio should provide some type of meaningful information.[2]

A commonly used ratio in demography is the sex ratio, which is the number of males divided by the number of females and is usually multiplied by 100. A sex ratio below 100 indicates an excess of females, while a sex ratio above 100 indicates an excess of males. Table 2.3 shows that sex ratios by age vary between the United States, Japan, Singapore, and Yemen in 2015. Sex ratios for ages 0–4 reflect the fact more males are born than females. Sex ratios at birth typically range from 103 to 106. However, China, India, and some other Asian countries have shown abnormally high sex ratios at birth since the 1980s (Poston and Bouvier 2010: 252). For example, in 2015, China's and India's sex ratios at age 0 are 114.9and 112.1, respectively (U.S. Census Bureau 2015). Sex ratios under 100 occur in every country for the population 60 years and older, with the ratios declining consistently to ages 85 years and older. The greatest excess of males in the U.S. is in ages 15–24; in Japan in ages 0–9; in Singapore in ages 15–24; and in Yemen in ages 35–39. In the U.S. and Yemen, there are more males in ages 25–34 compared to females than in Japan and Singapore. For ages 50–59 the pattern reverses, Japan and Singapore have more males and the U.S. and Yemen have more females. In all countries except Yemen, the total population contains more females than males.

Dependency ratios also are widely used measures in demography (Siegel 2002:12). They measure the pressure of those typically not in the labor force on the productive (or working-age) population, and are usually split into youth and

[2] A proportion is a type of ratio where the numerator is a subset of the denominator, such as the portion of the population aged 65 years and older, males, employed, or married. Proportions have a range from zero to one. A percentage is a proportion multiplied by 100.

Table 2.3 Sex ratio by age, Japan, Singapore, United States, and Yemen, 2015

Age	U.S.	Japan	Singapore	Yemen
0–4	104.7	106.0	105.5	103.9
5–9	104.3	104.1	105.6	103.4
10–14	104.3	104.0	108.6	103.1
15–19	104.8	102.8	113.4	103.0
20–24	105.6	93.2	106.8	103.0
25–29	103.8	92.7	97.6	102.3
30–34	101.4	90.4	95.5	112.1
35–39	100.1	95.7	95.4	114.4
40–44	98.7	96.0	97.6	102.9
45–49	98.4	98.1	99.4	92.7
50–54	96.5	102.4	101.3	87.1
55–59	94.6	101.2	101.5	84.7
60–64	91.7	98.9	99.2	88.6
65–69	89.7	93.9	95.7	90.4
70–74	85.7	86.1	89.5	87.0
75–79	80.1	78.9	79.7	83.7
80–84	71.3	72.0	66.8	80.8
85+	52.9	66.2	43.6	73.6
Total	97.1	96.0	94.3	102.5

Source: U.S. Census Bureau International Data Base (2015)

elderly dependency ratios. The youth dependency ratio (YDR) is the population aged 18 and younger divided by the population aged 18–64, and the elderly dependency ratio (EDR) is the population aged 65 and older divided by the population aged 18–64. There is not universal agreement on the age groups that defines youth, workers, and retirees. For example, Poston and Bouvier (2010: 245) use ages under 15 and ages 15–64 to define youths and workers; and Meyers (2007: 46) uses ages 25–64 to define workers. For a discussion of the strengths and weaknesses of dependency ratios based solely on age, see Donoghue (2003), Ervik (2009), and Siegel (2002: 595–598).

The last ratio we cover in this chapter is the child-woman ratio (CWR), a surrogate way to examine the level of fertility, computed by dividing the population in young ages by the female population in childbearing years. The CWR is usually computed for ages 0–4 and ages 5–9 by dividing the population in these age groups by the female population aged 15–44 and 20–49, respectively (Smith et al. 2013: 178). The CWR is influenced by past mortality and migration, as well as by past fertility behavior. However, the CWR does not require any information on births making it useful in areas lacking vital statistics information and when making forecasts with the Hamilton-Perry (H-P) method (see Chapter 4).[3]

[3]Another important ratio used in the H-P method is the cohort change ratio (CCR), which is the population aged x at time t divided by the population aged x-n at time t-n, where n is the number of years between the two time points of the population data (e.g., n = 10 if the CCR is based on the previous two decennial censuses). Chapters 1 and 4 discuss CCRs in detail.

Table 2.4 Selected demographic ratios, Japan, Singapore, United States,and Yemen, 2015

	U.S.	Japan	Singapore	Yemen
Youth Dependency Ratio[a]	39.8	25.6	30.1	106.8
Elderly Dependency Ratio[b]	24.4	47.1	12.2	5.7
Child-Woman Ratio, Females[c]	0.1537	0.0696	0.1165	0.3077
Child-Woman Ratio, Males[d]	0.1609	0.0738	0.1230	0.3197

Source: U.S. Census Bureau International Data Base (2015)
[a]Pop < 18/Pop 18–64
[b]Pop 65+/Pop 18–64
[c]Fpop 0–4/Fpop 15–44
[d]Mpop 0–4/Fpop 15–44

Table 2.4 shows considerable variation in the YDR, EDR, and CWR ages 0–4 for United States, Japan, Singapore, and Yemen in 2015. Japan and Singapore have the lowest YDRs, consistent with their low fertility rates. Strikingly, Yemen has more children aged 18 and under compared to its working-age population. Not surprising, the EDR is highest in Japan and lowest in Yemen, reflecting the effects of low and high fertility rates, respectively. However, Singapore's EDR is much smaller than the EDRs of the U.S. and Japan. Despite Singapore's low fertility rate, the impact of a high level of immigration of the working-age population is evident in its EDR. The CWRs correspond with the fertility rates in each country; Yemen has by far the highest CWRs and the Japan and Singapore have has the lowest CWRs. In each country, the male CWR is higher than the female CWR, reflecting the slightly higher likelihood of having a male child.

2.4.2 Rates and Probabilities

A rate is the number of events occurring during a given time period divided by the population at risk of the occurrence of those events. For example, the death rate is the number of deaths divided by the population exposed to the risk of dying and the birth rate is the number of births divided by the population exposed to the risk of giving birth. Strictly speaking, the population at risk to the occurrence of an event is the number of person-years of exposure experienced by the population during the period under consideration (typically one-year) (Newell 1988: 7). It is very difficult to develop an exact measure of the population at risk to the occurrence of an event and the mid-period population is often used as an approximation of the population at risk. This approximation assumes that births, deaths, and migration occur evenly throughout the year.

Demographers make a distinction between crude rates and refined rates. For example, the crude birth rate (CBR) is calculated by dividing the number of births during the year by the mid-year total population. It is often multiplied by 1,000 to express the CBR as the number of births per 1,000 persons. The crude death rate (CDR) is similarly defined replacing births with deaths in the numerator. These rates are called crude because their denominators are only a rough approximation of the population at risk. For the CBR, the denominator includes males and females outside the childbearing years; and for the CDR not everyone in the denominator has the same chance of dying (i.e., males have higher death rates than females and older people have higher death rates than younger people). To overcome the problem with crude rates, rates are refined to reflect specific age-sex groups (racial and ethnic groups can be used as well). For age groups, the general formula for an age specific rate is (ASR):

$$_{n}ASR_{x} = {}_{n}E_{x} / {}_{n}P_{x} \tag{2.2}$$

where,

x is the youngest age in the age interval,
n is the number of years in the age interval,
$_{n}E_{x}$ is the number of events, and
$_{n}P_{x}$ is the mid-year population.

For example, if $x = 35$ and $n = 5$, the ASR would be based on data for the population aged 35–39, or if $x = 35$ and $n = 1$, the ASR would refer to the population aged 35.

Figure 2.5 contains age-specific fertility (ASFR) and death rates (ASDR) for the United States in 1990 and 2013. The age pattern of ASFR has changed from 1990 to 2013. In the earlier year, the highest rates occur in ages 20–29 with a substantial drop after the age of 30. By 2013, the effects of a drop in the overall fertility level and of delayed childbearing are evident. The fertility rate for ages 30–34 is now higher that the rate for ages 20–24 and much closer to the rate for ages 25–29. The rates for all ages below 30 have declined and the rates for all ages above 30 have increased over the 23 year period. These changes are reflected in the total fertility rate, which declined from 2.08 in 1990 to 1.86 in 2013 (Martin et al. 2015).

ASDRs show a J-shaped pattern that reflects relatively high death rates for newborns, considerably lower rates for young children, slowly increasing rates for the middle ages, and rapidly increasing rates for the older population. The pattern of the ASDRs is very similar for the 2 years and is found for virtually all population and population subgroups throughout the world (Smith et al. 2013: 54). Death rates in 2013 are lower in every age group compared to the 1990 rates. The percentage declines range from −11% for ages 85 years and older to −49% for ages 15–19. As a result, the life expectancy at birth for both sexes increased from 75.4 years in 1990 to 78.8 years in 2013 (National Center for Health Statistics 1990, 2015).

In addition to the distinction between crude and age-specific rates, a distinction is also made between probabilities and central rates. A probability is a special type of rate that measures the chance or likelihood that a population will experience a given event over a given time period (Siegel 2002: 11). In a central rate, the denominator is an area's population at the midpoint of a time period (typically, the middle of a year) and the numerator is the number of events occurring in the

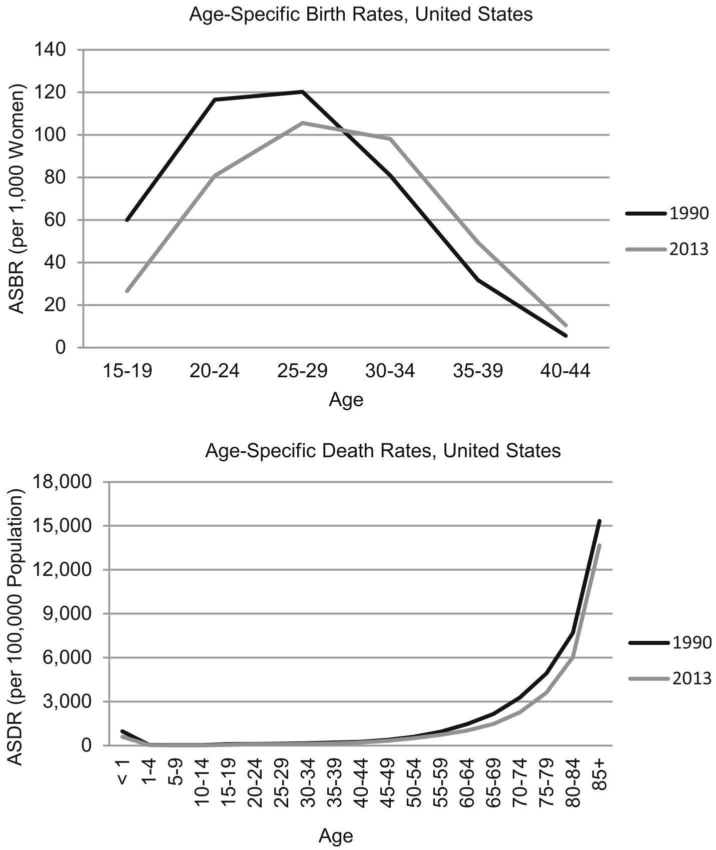

Fig. 2.5 Age-specific birth and death rates, United States, 1990 and 2013 (Sources: National Center for Health Statistics 1990, 2015)

area during the time period. The denominator is meant to represent the average population during the time period, or the total number of person-years of exposure to the risk of an event. The CBR, CDR, and ASR defined above are all central rates. In a probability, the denominator is the population at the beginning of the time period and the numerator is the number of events occurring to that population during the time period (Rowland 2011: 32).

For example, a single-year probability for a death rate can be computed by dividing the deaths occurring over the year by the population at the beginning of the year. However, some deaths will be missed for people who leave the area and then die and deaths will be improperly included for people who died after moving into the area. Consequently, it is next to impossible to construct true probabilities for demographic measures and central rates are often used to approximate their probabilities (Smith et al. 2013: 35).

2.4.3 The Odds Ratio

Given that a Census Survival Ratio (discussed briefly in Chapter 1 and in detail in Chapter 11) can be defined as the probability of survival, it is natural to ask if it is related to other measures that are based on probability. One such notable measure is the "odds ratio," which provides the odds of an event occurring among those exposed to a condition that is related to the event in question divided by the odds of the event in question occurring among those not exposed to the condition. One example would be the odds of those being diagnosed with asbestos who were exposed to asbestos relative to those diagnosed with asbestos who were not exposed to asbestos. The Odds Ratio is defined as:

$$\text{Odds ratio} = (\mathrm{P_e}/(1-\mathrm{P_e}))/(\mathrm{P_n}/(1-\mathrm{P_n})) \tag{2.3}$$

where,

P_e is the probability (condition | exposed), and
P_n is the probability (condition | not exposed).

If we substitute the concept of "dying" for the concept of "condition" and the concept "cohort" for the concept of "exposed", we can see that the idea of an odds ratio can be used to measure the "odds of dying" among members of a given cohort during a given period of time divided by the "odds of dying" for the same period of time among those who are members of a different cohort:

$$\text{Odds Ratio} = (\mathrm{P_{dx,t}}/(1-\mathrm{P_{dx,t}}))/\left(\mathrm{P_{dy,t}}/\left(1-\mathrm{P_{dy,t}}\right)\right) \tag{2.4}$$

where,

$P_{dx,t}$ is the probability of dying among those in cohort x during time t, and
$P_{dy,t}$ is the probability of dying among those in cohort y during time t.

Note that the probability of dying during the time period defined by *t* by members of cohort *x* is given by $P_{dx,t}$ and that the probability of surviving during the time period defined by *t* for those in cohort *x* is given by $(1 - P_{dx,t})$. These respective values have the same interpretation when applied to cohort *y*.

As an example of using the odds ratio in this way, we use a 2010 complete USA life table (both sexes combined) taken from the Human Mortality Database (2009) and compare the odds of a person dying in the next year who has reached his or her 10th birthday relative to a person dying in the next year who has reached his or her 65th birthday. In the life table, there are: (1) 99,221 persons in cohort "*y*" who reached their 10th birthday of whom 10 died before reaching their 11th birthday; and 84,492 persons in cohort "*x*" who reached their 65th birthday of whom 1079 died before reaching their 66th birthday. Inserting these numbers into Eq. 2.4, we have the odds ratio for those who reached their 65th birthday (cohort *x*) and dying before reaching their 66th birthday relative to those who reached their 10th birthday (cohort *y*) and dying before reaching their 11th birthday:

$$\begin{aligned}
\mathrm{P_{d65,t}} &= 0.01277\ \ (1.079/84,492),\\
\mathrm{P_{d10,t}} &= 0.00010\ \ (10/99,221),\text{and}\\
\text{odds ratio} &= (0.01277/(1-0.1227))/(0.00010/(1-0.00010))\\
&= (0.01277/0.98723)/(0.00010/0.99990)\\
&= 0.01294/0.00010\\
&= 129.4.
\end{aligned}$$

Thus, the odds of a member of the cohort of people reaching his or her 65th birthday and dying before reaching the 66th birthday are about 129 times higher than that of a member of the cohort of people reaching the 10th birthday and dying before reaching his or her 11th birthday. Alternatively, we can state that the risk of dying in the next year is about 129 times higher for those who reached their 65th birthday than it is for those who reached their 10th birthday.

2.5 Participation-Rate Method

2.5.1 Logic and Formulas

In the participation-rate method, current and historical data are used to construct rates reflecting the proportion or prevalence of the population having the attribute of interest (e.g., in the labor force). Rates are typically stratified by demographic characteristics rather than being defined using the total population. They are constructed separately for each age group and can be further stratified by sex and racial/ethnic groups as well. Rates can be forecast into the future by holding them constant, extrapolating recent trends, tying them to forecast changes in other places, using structural models, or relying on expert judgment. The forecast rates are then applied to forecasts with the matching demographic characteristics to obtain forecasts of the attribute of interest:

$$\text{Launch year participation rate}\ \ \mathrm{PR_{a,d,t}} = \mathrm{P_{a,d,t}}/\mathrm{P_{d,t}} \tag{2.5}$$

$$\text{Forecasted participation rate}\ \ \mathrm{PR_{a,d,t+i}} \tag{2.6}$$

$$\text{Forecasted characteristic}\ \ \mathrm{P_{a,d,t+i}} = \mathrm{PR_{a,d,t+i}} \times \mathrm{P_{d,t+i}} \tag{2.7}$$

where,

PR is the participation rate,
P is the population,
a is the attribute of interest (e.g., in the labor force),
d is the demographic characteristic (e.g., an age cohort),

t is the launch year; and
i is the length of forecast interval.

These computations are followed for each demographic group and for each interval over the forecast horizon.

2.5.2 Implementation Issues

What issues must be addressed when preparing population-related forecasts? Perhaps the most fundamental is obtaining the necessary data. The participation-rate method requires age-specific data on the variable of interest, and perhaps sex- and race/ethnicity-specific data as well. These data are often available from administrative records (e.g., labor force status) or surveys (e.g., the ACS and health surveys). Clearly, the availability of reliable data is essential for the production of reasonable forecasts.

The participation-rate method requires population data for constructing rates and a set of population forecasts to which the projected rates can be applied. Population data from the decennial census or post-censal estimates can generally be used as denominators in the rates. If reliable data for either the numerator or denominator are not available for a particular area, rates from similar areas can be used as proxies (e.g., county rate forecasts used for census tract rate forecasts). If independently produced population forecasts are not available, they can be constructed using the H-P method described in Chapter 4 or other techniques (Smith et al. 2013).

The participation-rate method requires that rates be forecast into the future. Making reasonable choices regarding future rates is crucial to the reliability of the forecasts but is largely a subjective process. Thorough knowledge of historical trends and the factors affecting the variable of interest are essential. In some circumstances, it may be advisable to consult an expert in the field before making these choices and to apply several alternative assumptions in order to provide a range of forecasts. Reasonable forecasts of population-related variables can be made only if the analyst makes reasonable choices regarding future participation rates. Thorough knowledge of the population-related variables—and how they are related to the stages of the life cycle—are essential (Martins et al. 2012: 83–938; Modigliani 1970; O'Rand and Krecker 1990).

References

Donoghue, C. (2003, August). Deconstructing aged dependency: An assessment of the dependency ratio as an indicator of population aging. Paper presented at the annual meeting of the American Sociological Association, Atlanta, GA, 16–19.

Ervik, R. (2009, August). *A missing leg of aging policy ideas: Dependency ratios, technology, and international organizations*. Paper presented at the annual conference of the International

Sociological Association's Research Committee on Poverty, Social Welfare, and Social Policy, Montreal, Canada, 20–22.
Hobbs, F. (2004). Age and sex composition. In J. Siegel & D. Swanson (Eds.), *Methods and materials of demography* (2nd ed., pp. 125–173). New York: Elsevier Academic Press.
Human Mortality Database (2009). University of California, Berkeley and Max Planck Institute for Demographic Research. Retrieved from www.mortality.org.
Isserman, A. (1985). Forecasting regional population change with endogenously determined birth and migration rates. *Environment and Planning A, 17*, 25–45.
Keyfitz, N. (1972). On future population. *Journal of the American Statistical Association, 67*, 347–362.
Los Angeles County Department of Public Health. (2015). *LA Health Data Now*. Retrieved from https://dqs.publichealth.lacounty.gov/default.aspx.
Martin, J., Hamilton, B., Osterman, M., Curtin, S., & Mathews, T. (2015). Births: Final data for 2013. National Center for Vital Statistics Reports, Vol 64, No. 1. Hyattsville: National Center for Health Statistics.
Martins, J., Yusuf, F., & Swanson, D. (2012). *Consumer demographics and behaviour: Markets are people*. Dordrecht: Springer.
McKibben, J. (2004). Racial and ethnic composition. In J. Siegel & D. Swanson (Eds.), *Methods and materials of demography* (2nd ed., pp. 175–189). New York: Elsevier Academic Press.
Meyers, D. (2007). *Immigrants and boomers: Forging a new social contract for the future of America*. New York: Russell Sage Foundation.
Modigliani, F. (1970). The life-cycle hypothesis and intercountry differences in the saving rate. In W. Eltis, M. Scott, & J. Wolfe (Eds.), *Induction, growth, and trade: Essays in honour of Sir Roy Harrod* (pp. 197–225). Oxford: Oxford University Press.
National Center for Health Statistics. (1990). *Advanced report of final mortality statistics 1990*. Monthly Vital Statistics Report, Vol 41 No 7, suppl. Hyattsville, MD: Public Health Service, 1993.
National Center for Health Statistics. (2015). *Deaths: Final Data for 2013*. National Vital Statistics Report. Retrieved from http://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_02.pdf.
Newell, C. (1988). *Methods and models in demography*. New York: The Guilford Press.
O'Rand, A., & Krecker, M. (1990). Concepts of the life cycle: Their history, meanings, and uses in the social sciences. *Annual Review of Sociology, 16*, 241–262.
O'Hare, W., Pollard, K., & Ritualo, A. (2004). Educational and economic characteristics. In J. Siegel & D. Swanson (Eds.), *Methods and materials of demography* (2nd ed., pp. 211–215). New York: Elsevier Academic Press.
Perz, S. (2004). Population change. In J. Siegel & D. Swanson (Eds.), *Methods and materials of demography* (2nd ed., pp. 253–263). New York: Elsevier Academic Press.
Plane, D. (2004). Population distribution-geographic areas. In J. Siegel & D. Swanson (Eds.), *Methods and materials of demography* (2nd ed., pp. 81–104). New York: Elsevier Academic Press.
Poston, D., & Bouvier, L. (2010). *Population and society: An introduction to demography*. New Cambridge: Cambridge University Press.
Rowland, D. (2011). *Demographic methods and concepts*. New York: Oxford University Press.
San Diego Association of Governments. (2010). 2050 Regional Growth Forecast process and model documentation. San Diego, CA.
Siegel, J. (2002). *Applied demography: Applications to business, government, law, and public policy*. San Diego: Academic Press.
Smith, S., Tayman, J., & Swanson, D. (2013). *A practitioner's guide to state and local population projections*. Dordrecht: Springer.
Swanson, D., & Stephan, G. (2004). Glossary. In J. Siegel & D. Swanson (Eds.), *Methods and materials of demography* (2nd ed., pp. 571–778). New York: Elsevier Academic Press.
Swanson, D., & Tayman, J. (2011). On estimating a de facto population and its components. *Review of Economics and Finance, 5*, 17–31.

Swanson, D., & Tayman, J. (2012). *Subnational population estimates*. Dordrecht: Springer.
Texas State Data Center. (2015). Total population estimates for Texas counties. Retrieved from http://osd.texas.gov/Data/TPEPP/Estimates/.
U.S. Census Bureau. (2015). International Database. Retrieved from http://www.census.gov/population/international/data/idb/informationGateway.php
Weeks, J. (2014). *Population: An introduction to concepts and issues* (12th ed.). Boston, MA: Cengage Learning.
Wilmoth, J. (2004). Population size. In J. Siegel & D. Swanson (Eds.), *Methods and materials of demography* (2nd ed., pp. 65–80). New York: Elsevier Academic Press.
Wisser, O., & Vaupel, J. (2014). The sex differential in mortality: A historical comparison of the adult-age pattern of the ratio and difference. In *MPIDR Working Paper WP 2014–005*. Rostock: Max-Planck Institute for Demographic Research.

Chapter 3
Sources of Demographic Information

3.1 Introduction

Several sources of data to make cohort change ratios (CCRs) are available. We cover four major sources for data that can be used in conjunction with CCR methods: (1) demographic data for the United States; (2) administrative records data for the United States (2) demographic data for other countries and (4) other data sources that can be used in conjunction with CCR methods.

In terms of U.S. demographic data, the Census Bureau is the primary supplier, with the decennial census being the most important. However, the Census Bureau also produces population estimates and conducts regular surveys, so we also provide an overview of data from these sources that can be used in conjunction with CCR methods. These data cover a wide range of geographic areas, including: (1) administrative areas such as states, counties, cities, townships, legislative districts; and (2) statistical areas, such as census tracts and block groups. Our primary need is the data by age (and sex, ethnicity, and race if desired) for the U.S. and its geographic subdivisions. The most common application of CCR methods uses 5-year age groups.

In terms of administrative records, the most widely used in terms of CCR methods are school enrollment and related data, as shown in Chapter 7, but also as is shown in Chapters 12 and 13 vital statistics can be useful. Hence, we briefly cover these types of administrative data.

For demographic data in other countries, we first cover the U.S. Census Bureau's International Data Base (IDB) that provides the basic data for CCR construction for countries of the world. We describe this resource in some detail, but also give links to data available from the national statistical offices of several countries.

The fourth source we consider is largely made up of the collections assembled under the auspices of the University of Minnesota. We also describe data found elsewhere that can be useful for CCR methods, such as the University of Michigan (ICPSR program) and the University of California Berkeley/Max Planck Institute

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_3

for Demographic Research (The Human Mortality data Base). Useful data can also be obtained from provincial and state demographic agencies (e.g., **B**ritish **C**olumbia Stats (BC Stats), California Department of Finance, and the Washington State Office of Financial Management) and the private sector (e.g., ESRI).

3.2 United States Census Bureau

The U.S. Census Bureau is the "go to" source of data for making CCRs. Data by age (and sex, ethnicity, and race) are readily available for no cost to anyone with an internet connection using *American FactFinder,* which is discussed below. Keep in mind that the Census Bureau produces a wide range of data beyond demographics and generates these data through different programs, including the decennial census, estimates, and surveys.

In addition to programs used to generate and disseminate data, the Census Bureau sponsors the State Data Center (SDC) Program; one of its longest and most successful partnerships. Started in 1978, the SDC is a partnership between the 50 states, the District of Columbia, Puerto Rico, the island areas and the Census Bureau. This partnership has made data readily available locally to the public through a network of state agencies, universities, libraries, and regional and local governments. The SDC lead organization is appointed by the Governor of each state/commonwealth, Puerto Rico, island area (American Samoa, Guam, The Commonwealth of the Northern Mariana Islands, Virgin Islands), and the Mayor of the District of Columbia. Since the beginning, the SDC has provided access and education on Census Bureau data and products as well as other statistical resources to millions of data users. For access to the data to make CCRs, the SDC in each state is a first stop source. Current contact information for the SDC program lead and affiliate agencies is available at: http://www.census.gov/about/partners/sdc/member-network.html.

3.3 Decennial Census

Decennial censuses are by far the most important source of demographic data produced by the Census Bureau. They form the basis for the CCR population estimates and forecasts described throughout this book. No other source of data is as comprehensive or used for as many purposes.

Decennial Census data were used to compute the CCRs found in Chapter 1 for Riverside County, California. The CCRs found in Table 1.1, for example, were computed using data drawn from the quick Tables (QT-P1) for the 2000 and 2010 censuses, a feature of the "Advanced Search" facility of the *American Factfinder* data extraction system. *American Factfinder* can be found at http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml. The Riverside County data were found by entering the preceding site, opening the “Advanced Search” tab, and then clicking on the “show me all” tab, which led to two selection features, “topic or table name” and “state, county or place (optional).” Under the “topic or table name,” we typed “QT-P1” and under “State, County or place (optional),” we typed “Riverside County, California,” and then clicked on the “GO” tab. This opened up a dropdown list that showed six sources of data, each from a decennial census. We opened up “QT-P1, Age Groups and Sex 2010, 2010 SF1 100% data” by clicking on “Age Groups and Sex 2010.” This led to the display of a table from the 2010 census that showed, along with other information the 2010 population of Riverside County, California by 5 year age groups. We then clicked on the download tab and selected the option to download the table as an Excel file.

Once the 2010 data were downloaded, we returned to the original display of the table, where a box can be seen that states “versions of this table are available for the following years, 2010 and 2000.” Because we had the 2010 data, we clicked on the “2000” tab, which opened the 2000 census version of this same table for Riverside County, California, which we downloaded as an Excel file. From the two downloaded files, we copied and pasted the 2000 and 2010 counts by age (both sexes combined) into an excel template set up to construct CCRs for any two decennial censuses; the 2000–2010 CCRs were automatically calculated. We could have found these same two tables via several routes available from *American Factfinder*, including the “Guided Search” feature and the “Download” feature.

The historical U.S. Census data (1910, 1920, and 1930) used to generate the estimate of the Native Hawaiian population in 1778 described in Chapter 10 were not found using *American Factfinder*, which does not include decennial census data prior to the year 2000. Instead, the data were taken from online copies of census reports (.pdf files) that the Census Bureau makes available online at https://www.census.gov/prod/www/decennial.html. Once at this site scroll down and you will see tabs for each of the decennial census counts starting with the most recent (2010) and ending with the first (1790). If you click on the 1910 census, for example, a window will open that provides several choices, including: (1) information about the 1910 census; (2) abstracts; (3) bulletins; (4) final reports; and (5) other 1910 census documents. If you click on the fourth choice, “final reports,” another window will open showing each of the 11 volumes of the 1910 census reports. The 1910 data found in Chapter 10 were taken from Volume 3. If you click on this choice, you can click either on “title page” or a “full document (a condensed file).” If you open the “title page” you will see where the data for Hawai’i are located in Volume 3, which is “clickable” and will lead you to the data for Hawai’i. Unfortunately like all of the pre-2000 historical decennial data from the Census Bureau, the source files are all in .pdf format, which means data extraction is time-consuming and more error prone. Fortunately, there is an alternative data source, which allows for direct downloads into a file that can be analyzed by Excel and statistical packages such as SPSS and STATA. We discuss this source in Sect. 3.6.

Like any method requiring data from two time points, the areas used in CCR methods must be based on constant geographic boundaries. Boundary changes, while uncommon at the county level and higher geographies, are common in many subcounty areas such as cities and census tracts. Since the CCR method is often used for forecasting subcounty populations by age, sex, and other demographic characteristics, the data for such application must be adjusted for changes in geographic boundaries. Although we discuss information on tracking these changes over time and assembling data sets with common geographical definitions in Chapter 14, here we give an overview of the available information from the Census Bureau for making such adjustments.

The Census Bureau's Geography Division tracks boundary changes for all levels of geography in several resources. One is "Geographic Change Notes (GCN)," and can be found at the following website: http://www.census.gov/geo/reference/boundary-changes.html. The GCN lists selected changes to incorporated places (cities and towns), census designated places, county subdivisions, counties and equivalent areas, and American Indian, Alaska Native, and Native Hawaiian areas, as recognized by the Census Bureau, within the 50 States, the District of Columbia, Puerto Rico, and the Island Areas (American Samoa, Guam, the Commonwealth of the Northern Mariana Islands, and the U.S. Virgin Islands). Once at this website, you use a selection tab that starts with the state (or its equivalent) in which the geography of interest is located. Once a state is selected, a popup appears that shows a description of changes subsequent to the 2000 census up to a current point in time (which, as of the writing of this book was 2013) by area and effective data.

The preceding website is useful for determining if a boundary change took place for a city or town of interest, but it does not provide detailed information on the land area (in acres) affected. To obtain this detail, you need to use the Geography Division's "Boundary and Annexation Survey" data, which can be found at: https://www.census.gov/geo/partnerships/bas.html. Once at this site, open the tab marked "Legal Boundary Change/Annexation Data," which will take you to a site that shows the years for which data are available and a tab identified by "Download Legal and Boundary Change Files," with the tab itself identified by state. This will take you to a set of files identified by year. Here, you have a choice of downloading text files, .pdf files, or Excel files.

Like the "Geographic Change Notes," the "Boundary and Annexation Survey" are limited in that the latter only provides the land area affected. When a more exact description of boundary changes over time is needed, we can turn to a third resource available from the Geography Division, namely, the relationship files. These files (for block, census tracts, places, counties, and urban areas) show precise relationships between 2000 boundary definitions and 2010 boundary definitions and can be found at: https://www.census.gov/geo/maps-data/data/relationship.html. Relationship files can be downloaded as compressed Excel files, decompressed, and read by following the file record layout guidelines available at this website. At this same website, one can identify the geographic relationships between types of geography (e.g. a city and one or more census tracts) at the same point in time. Also useful in this regard are the **T**opologically **I**ntegrated **G**eographic **E**ncoding **R**eferencing (TIGER) tools, including shapefiles and geodatabases. Again, Chapter 14 provides more detail on these resources and the procedures needed to utilize them for assembling data with the same geographical definitions over time.

3.4 Population Estimates

A second source of demographic data is found in the Census Bureau's population estimates program (PEP), which produces annual population estimates by age and sex for states and counties (http://www.census.gov/popest/). Population estimates are not primary data in the same sense as the decennial census data just discussed; rather, they are derived from (or based on) decennial census data. They play an important role in supplementing and updating data from the decennial census in that they are more current and cover years other than those ending in zero.

Population estimates are also produced by a variety of state and local government agencies. Many state agencies participate with the Census Bureau in the Federal-State Cooperative Program for Population Estimates (FSCPPE); this program serves as a clearinghouse for demographic data and as a forum for discussing methods and exchanging ideas related to population estimation (https://www.census.gov/popest/fscpe/). Some states produce independent population estimates at the state, county, and/or city level (e.g., The Washington State Office of Financial Management (http://www.ofm.wa.gov/pop/estimates.asp) and The California Department of Finance, (http://www.dof.ca.gov/Forecasting/Demographics/Estimates). Some city and county governments—and Councils of Governments for large metropolitan areas—also produce population estimates, often for small areas such as census tracts and traffic analysis zones (e.g., The New York City Department of City Planning (http://www1.nyc.gov/site/planning/data-maps/nyc-population/current-future-populations.page) and The San Diego Association of Governments (http://www.sandag.org/resources/demographics_and_other_data/demographics/estimates/index.asp).

One issue to keep in mind in using population estimates is the point in time to which they are referenced. The decennial census data are referenced to April 1st, as, for example, are the annual estimates produced by the State of Washington, but the Bureau's PEP estimates are referenced to July 1st. These different reference points could affect the integrity of CCRs constructed from a combination of decennial and PEP data.

3.5 Surveys

The U. S. Census Bureau conducts a wide range of sample surveys, about 130 each year (http://www.census.gov/programs-surveys/are-you-in-a-survey/survey-list/household-survey-list.html). Notable for purposes of this book is the American Community Survey (ACS). The ACS is a nationwide, continuous survey designed to provide communities with demographic, housing, social, and economic data every year (U.S. Census Bureau 2008). The ACS samples nearly 3 million addresses each year, resulting in nearly 2 million final interviews. For each area with an estimated population of 65,000 or more, the ACS provides single year estimates, for each area with an estimated population of 20,000 or more, the ACS provides estimates based on an aggregation of three years of sample data, and for all areas, the ACS provides annual estimates based on an aggregation of five years of sample data. The 3-year and 5-year aggregations are problematic in that the data refer to a period of time rather than a point in time, which makes them inconsistent with the 1-Year ACS data as well as the decennial and PEP data, which all are referenced to a single year (Swanson 2010). Another issue is that the definition of a resident is different than the definition used in the decennial census, which makes the ACS data inconsistent with decennial data in locations that experience seasonal variations in the de-facto population (Swanson 2010, Swanson and Tayman 2011). This inconsistency can affect the integrity of CRRs constructed from a combination of decennial and ACS data.

ACS data can also be a bit confounding in that one can obtain differences in the 1-year, 3-year, and 5-year estimates for the same period of time for an area with an estimated population of 65,000 or more. For example, the 2013 total population of Spokane, Washington is estimated at 210,722; 209,876; and 209,478, respectively, by the 1-year (2013), 3-year (2011–2013), and 5-year (2009–2013) ACS samples. Margins of error (90% confidence interval) are available from the ACS but the choice of which to use as a 2013 estimate is up to the user. Another issue, even for areas with populations estimated at 65,000 and over, is because of sampling error the 1-year ACS can yield erratic annual values for some variables (e.g., average number of persons per household) that is generally not subject to sudden change (Swanson and Hough 2012). The ACS has been providing data for areas since 2010 and its data can be easily retrieved using *American FactFinder*.

Other surveys conducted by the Census Bureau that have potential relevance for this book include the Current Population Survey (http://www.census.gov/programs-surveys/cps.html), The American Housing Survey (http://www.census.gov/programs-surveys/ahs.html), the National Health Interview Survey (http://www.cdc.gov/nchs/nhis/index.htm), and the Consumer Expenditure Survey (http://www.bls.gov/respondents/cex/).

One annual survey funded by the federal government, but specific to each state is the Behavior Risk Factor Surveillance System (BRFSS), which was established in 1984 in 15 states (http://www.cdc.gov/brfss/). Funded by the Centers for Disease Control (CDC), it now collects data in all 50 states as well as the District of Columbia and three U.S. territories and completes more than 400,000 adult interviews each year on a wide range of health-related topics. CDC has set up a set of interactive tools (http://www.cdc.gov/brfss/data_tools.htm) that can be used to extract data that can be used to construct measures similar to the examples discussed in Chapter 8.

There are other entities that also conduct regular surveys containing data that may be useful. For example, the Survey Research Center at the University of Michigan conducts the National Survey of Family Growth, the Survey of Consumer Attitudes, and the Panel Study of Income Dynamics, among others (http://www.src.isr.umich.edu/about/). The National Opinion Research Center (NORC) at the

University of Chicago also conducts a wide range of surveys (http://www.norc.org/Research/Topics/Pages/default.aspx).

3.6 Administrative Records

Administrative records are records kept by agencies of federal, state, and local governments for purposes of registration, licensing, and program administration. Although not always designed explicitly to do so, these records provide valuable information on specific demographic events or subgroups of the population. In terms of data that can be used either directly or in conjunction with the CCR method, these sources include school enrollment and vital statistics. Other types of administrative data include employment, voter registration, and property tax records.

Administrative data can be used for various types of demographic analyses, including the production of population estimates and projections. In Chapter 7, for example, data from the California Department of Education are used in conjunction with the CCR method to generate a short-term enrollment forecast for Riverside County, California. These data were taken directly from the Department's website (http://www.cde.ca.gov). Once at the website, we clicked on the "Data and Statistics" tab, which led to (http://www.cde.ca.gov/ds). At this site, one of the choices is "*DataQuest*," an interactive data query system used to obtain the (fall) enrollment data by grade for Riverside County in 2013, 2014, and 2015. Because the site is interactive it is rather tedious to describe the exact steps used to extract the data. Fortunately, the *DataQuest* site is very well designed and easy to navigate, so we invite you to use it to replicate the data we use in Chapter 10.

Data on events such as births, deaths, marriages, and divorces are called vital statistics. In the United States, the collection of these data is the responsibility of individual states and not the federal government. As early as 1639, the Massachusetts Bay Colony began reporting births, deaths, and marriages as part of its administrative/legal system. Other states gradually began doing the same thing, and today all states have complete (albeit imperfect) records of births, deaths, and other vital events. The federal government sets standards for the collection and reporting of the data, compiles summaries from data collected by each state, and publishes a variety of reports based on these data. The quality of vital statistics data is generally very good in the United States and other high-income countries.

Before 1945, vital statistics reports were published by the U.S. Census Bureau. Beginning in 1945, this task was taken over by the U.S. Public Health Service, National Office of Vital Statistics. In 1960, this office was reorganized and became part of the National Center for Health Statistics (NCHS), which today is a branch of the Centers for Disease Control (CDC). Annual and monthly reports on births, deaths, marriages, and divorces are available from the NCHS. It should be noted that some of the concepts and definitions used by the NCHS do not precisely match those used by the Census Bureau. Consequently, adjustments may have to be made when combining population data from the Census Bureau with vital statistics data from the NCHS.

Data from the NCHS are available only at the national and state levels; vital statistics data for local areas must be obtained elsewhere. Most states tabulate data at the county (or county-equivalent) level, but few go beyond that to develop regular data series for subcounty areas. Although individual records generally contain the information needed to allocate them to different types of subcounty areas (e.g., cities, census tracts), actually doing so requires a substantial effort. In addition, there are often errors in geocoding birth and death records at the subcounty level (Flotow and Burson 1996). Analysts needing vital statistics data for subcounty areas may have to develop those data themselves.

3.7 International Data

For the international data, the Census Bureau has an International Data Base (IDB) program that contains data for many countries for a series of years. Located at (http://www.census.gov/population/international/data/idb/informationGateway.php) it is possible to download age data for two time periods for a host of countries permitting CCR computations. Additional data if required for other geographies within a particular country may be available at the respective statistical agencies within each country. Many individual countries also have high quality data available online. Often, these population data are taken from a population registration system, which is a common method for collecting demographic data in the Scandinavian region. Statistics Finland has such data and has produced an English-language report that documents the methods and quality found in its population register (Statistics Finland 2004). The U.S. Census Bureau maintains a list of the URLs for international statistical agencies, which can be found at (https://www.census.gov/population/international/links/stat_int.html).

3.8 Other Data Sources

Another data source to make CCRs is compiled by Minnesota Population Center (MPC) at the University of Minnesota. There are several collections assembled and held by the MPC, including Integrated Public Use Microdata (IPUMS), which is extensively used by researchers, policymakers, students, and faculty. The MPC is a leading developer of demographic data resources. All data obtained through the MPC are available over the internet at no cost (https://www.ipums.org/). The MPC has both an IPUMS collection for the United States and the IPUMS-International collection. According to the MPC website, the IPUMS-USA collection has harmonized data on people in the U.S. census and American Community Survey from 1850 to the present. The IPUMS-International collection contains harmonized data for 1960 forward covering 560 million people in 258 censuses from around the world.

Historical population data used in a study by Swanson and Verdugo (2016) on the demographic effects of the Civil war on the former states of the Confederacy (described in Chapter 10) were obtained from IPUMS files. Specifically, the IPUMS data were taken from samples of the original census records from the 1850, 1860, and 1870 census counts, with a sampling ratio of approximately 1 in every 100 records. There are several advantages to using these data in lieu of hardcopy census reports. First, they are machine-readable and can be easily imported into Excel or a statistical system, such as SPSS, SAS, or some other analytical software package. Second, the IPUMS files have been cleaned, edited, and assembled using high levels of quality control. Third, because they are individual level data, the extracted sample data can be aggregated in different ways to suit a given analysis and automatically weighted in order to reproduce the census counts. Finally, MPC provides an online assembly and tabulation feature so that aggregated data, properly weighted can be extracted from its IPUMS collections.

In the initial extraction, Swanson and Verdugo (2016) selected, non-Hispanic white males by age and state for each of the three census counts, 1850, 1860, and 1870 along with their weights. They then used the recode, frequency, and filter procedures provided by MPC to generate output that could be import directly into Excel. The result was aggregated census counts for each of the 11 Confederate states that contained the selected five-year age groups appropriately weighted: (1) from 10–14 to 40–44 for the 1850 and 1860 census years; and (2) from 20–24 to 50–54 in the 1860 and 1870 census years.

MPC also manages The National Historical Geographic Information System (NHGIS), which provides, free of charge, aggregate census data and GIS-compatible boundary files for the United States between 1790 and 2015 (https://www.nhgis.org/). Data of this type can be found in Chapter 14. Another useful data set is the Human Mortality Database (University of California Berkeley). Data from this site are used in Chapter 12. These and other mortality data can be accessed by starting at (http://www.mortality.org).

There also are commercial databases, For example, ESRI provides market segmentation data for the U.S. and other countries using its “Tapestry” market segmentation system. Current data typically can only be obtained on a cost basis, but older data are often made available at no cost. These data can be found at (http://www.esri.com/landing-pages/tapestry/), which has an interactive query system. These data can be used in conjunction with the methods described in Chapter 8.

3.9 On-line Location of Excel Files

All tables and most figures in this book have a corresponding Excel file, which can be found at the “Applied Demographer’s Toolbox,” a website created and maintained by Eddie Hunsinger. The toolbox is a collection of applied demography programs, scripts, spreadsheets, databases, and texts. As the title suggests, this website contains far more than the excel files used in this book, but navigating to the folder (zipped) containing our Excel files is straightforward. The starting point is (http://u.demog.berkeley.edu/~eddieh/toolbox.html). Once there, scroll down until you see "Excel files for the book, Cohort Change Ratios and Their Applications" which contains the zipped folder.

3.10 Conclusions

There are, of course, many sources of data available that are relevant to this book. The list is obviously too large to provide in a single chapter, so we have given an illustrative sample of data resources. Virtually all of the data we have described herein are reliable and of good quality. The sample data resources we have described, such as the American Community Survey, have in some cases reliability issues, but the ACS variables have estimated margins of error, which are very useful in gauging in reliability.

References

Flotow, M., & Burson, R. (1996). *Allocation errors of births and deaths to subcounty geography*. Paper presented at the meeting of the Population Association of America, New Orleans, LA, 9–11, May.

Statistics Finland. (2004). Use of register and administrative data sources for statistical purposes: Best practices of Statistics Finland. Helsinki, Finland. Statistics Finland. Available online at http://www.stat.fi/tup/julkaisut/kasikirjoja_45_en.html

Swanson, D. (2010). *Perspectives on the American Community Survey*. Presented at the annual Conference of the Latin American Association for Population Studies, Havana, Cuba, 15–19, November.

Swanson, D., & Hough, G. (2012). An evaluation of persons per household (PPH) data generated by the American Community Survey: A demographic perspective. *Population Research and Policy Review, 31*, 235–266.

Swanson, D., & Tayman, J. (2011). On estimating a de facto population and its components. *Review of Economics and Finance, 5*, 17–31.

Swanson, D., & Verdugo, R. (2016). *The civil war's demographic impact on non-Hispanic white males in the 11 Confederate states: An analysis by state and selected age groups*. Paper presented at the annual conference of the Southern Demographic Association, Athens, GA, 12–14, October.

U.S. Census Bureau. (2008). *A compass for understanding and using American Community Survey Data: What general data users need to know*. Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/content/dam/Census/library/publications/2008/acs/ACSGeneralHandbook.pdf

Chapter 4
Forecasting Population Size and Composition

4.1 Introduction

In a seminal paper, Hamilton and Perry (1962) proposed cohort-change ratios (CCRs) as a variant of the cohort-component method for purposes of short-term population projections. The Hamilton-Perry (H-P) method has much smaller data requirements than its more data-intensive cousin while still providing a forecast of population by age (as well as sex, race, ethnicity, if so desired), which is the hallmark of the cohort-component method. Instead of specific rates for the components of population change, forecasts from the H-P method are based on cohort change ratios and child-woman ratios (CWR) or more generally child-adult ratios (CARs) as previously discussed in Chapter 1. CCRs and CWRs are most often obtained from the two most recent censuses, but can be based on age distributions from any two points in time.

Consequently, the H-P method requires much less time and resources to implement than the full cohort-component model. Not surprisingly, it has mainly been used for small geographic areas in which mortality, fertility, and migration data are non-existent, unreliable, or very difficult to obtain (Baker et al. 2014; Smith et al. 2013:176; Swanson et al. 2010). Although the H-P method has primarily been used for small geographic areas, its minimal data input requirements combined with its capability for forecasting age and other characteristics make it attractive for use at higher levels of geography such as states and counties when detailed information on the components of population change is not needed.

The H-P method has gained acceptance as research has demonstrated its practical value and accuracy in forecasting populations (Kodiko 2014; Smith and Shahidullah 1995; Swanson and Tayman 2017; Swanson et al. 2010). Smith and Tayman (2003) found for U.S. states and counties in Florida that the H-P and cohort- component methods produced similar projections of the age-sex structure of the population; neither approach consistently produced more accurate projections

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_4

than the other for 10- and 20-year forecast horizons. Wilson (2016), however, found that the H-P method had modestly larger errors compared to some variants of the cohort-component method for local government areas in New South Wales, Australia.

In this chapter, we present two step-by-step examples each based on commonly used procedures to develop forecasts using CCRs and CWRs. These examples illustrate forecasts by age and sex and by age only for the city of Bellingham, Washington. We also show a one-year forecast of major league pitchers by consecutive years in the league developed from the H-P method. We next investigate the impact of adjusting H-P forecasts to independent total population controls because in rapidly growing (declining) areas CCRs applied to a beginning population can lead to large forecast errors and a strong upward (downward) biases (Smith et al. 2013: 180). Controlling also may be useful in other instances such a reaching a capacity constraint in small areas or adhering to a previously established population control. We end this chapter by discussing the strengths and weaknesses of the H-P method.

Our strategy is to describe the simplest, straightforward, and most used application of the H-P method; namely, holding CCRs and the CWRs constant over the forecast horizon. In Chapter 5, we evaluate the accuracy of the H-P method using modified ratios.

4.2 Hamilton-Perry Forecast

The H-P method is illustrated by producing 2020 forecasts for the city of Bellingham in Washington State based on data from the 2000 and 2010 censuses. Bellingham's 2016 population is estimated at 84,850 (State of Washington 2016). Bellingham provides an interesting example because it is home to a large special population of college students. Western Washington University, Whatcom Community College, and Bellingham Technical College enroll around 22,100 students, or 26.2% of the 2016 total population (National Center for Educational Statistics 2016). We also show the adaptability of the H-P method by producing a one-year forecast pertaining to pitchers according to their number of consecutive years in the major leagues in this position.

4.2.1 Forecast by Age and Gender

The first example forecasts the 2020 population of Bellingham by age and gender. CCRs are calculated by dividing the population aged x in year t by the population aged $x{-}10$ in year $t{-}10$ calculated separately for males and females, where t is 2010 and $t-10$ is 2000. These CCRs are applied to each age, gender group in year t to provide forecasts by age and gender in the year $t + 10$ (i.e., 2020). Given the nature of the CCRs, 10–14 is the youngest five-year age group for which forecasts can be made if there are 10 years between censuses. Children younger than age 10 are forecast using CWRs from the launch year (i.e., males or females younger than 5/females aged 15–44 and males or females aged 5–9/females aged 20–45). Eqs. 4.1 through 4.3 represent the usual application of the H-P framework (i.e., holding CCRs and CWRs from the most recent 10-year period constant over the horizon):

$$_{n}P_{x+10,g,t+10} = {}_{n}CCR_{x,g,t} \times {}_{n}P_{x,g,t} \quad \text{(Ages 10+)}, \tag{4.1}$$

$$_{4}P_{0,g,t+10} = {}_{4}CWR_{0,g,t} \times {}_{44}FP_{15,t+10} \quad \text{(Ages0}-\text{4)}, \tag{4.2}$$

$$_{9}P_{5,g,t+10} = {}_{9}CWR_{5,g,t} \times {}_{49}FP_{20,t+10} \quad \text{(Ages5}-\text{9)}, \tag{4.3}$$

where,

n is the width of the age group;
x is the beginning of the age group,
g is gender,
t is the launch year,
P is the population,
CCR is the cohort change ratio,
CWR is the child-woman ratio, and.
FP is the female population.

Table 4.1 shows the 2020 forecast for Bellingham by age and gender using eqs. 4.1–4.3. As the table shows, the H-P method requires only a limited set of calculations. For example, the male populations aged 0–4 and 40–44 in 2020 are calculated as:

0.09219 × (3,993 + 7,459 + 3,683 + 2,907 + 2,976 + 2,601) = 2,177 Ages (0–4),
0.94521 × 2,644 = 2,499 Ages (40–44).

Forecasts of the oldest age group differ slightly from the forecasts of the other age groups. The population aged 75 years and older in the launch year (2010) forms the basis of the forecast for the population aged 85 years and older in 2020. For example, females aged 85 years and older in 2020 are calculated as:

0.43982 × (998 + 1,028 + 1,374) = 1,495.

Remember, the CCRs for ages 10 years and older combine the effect of mortality and migration. The large CCRs for ages 15–24 reflect the in-migration of college students and the dramatic decrease in CCRs for ages 25–39, especially ages 30–34, reflect the out-migration of college students. The CCRs for ages 40–69 suggest in-migration in these age groups, except for females aged 40–44. At the oldest ages, the female CCRs are uniformly larger than the male CCRs reflecting, in part, the higher survivorship of females. Unlike the cohort-component model, the H-P method does not require adjustments for special population such as college students (Smith et al. 2013: 251–258). As the table shows, the college age groups have not

Table 4.1 Population forecast by age and sex, Bellingham, Washington, 2020

Female Age	2000	2010	CCR[a]	2020[b]	Male Age	2000	2010	CCR[a]	2020[b]
0–4	1,738	1,808	0.08645	2,041	0–4	1,764	1,928	0.09219	2,177
5–9	1,603	1,667	0.08639	1,895	5–9	1,625	1,809	0.09375	2,056
10–14	1,547	1,652	0.95052	1,719	10–14	1,604	1,686	0.95578	1,843
15–19	3,544	3,840	2.39551	3,993	15–19	2,829	3,339	2.05477	3,717
20–24	5,964	6,985	4.51519	7,459	20–24	5,639	6,950	4.33292	7,305
25–29	2,408	3,399	0.95909	3,683	25–29	2,935	3,868	1.36727	4,565
30–34	2,010	2,482	0.41616	2,907	30–34	2,263	2,644	0.46888	3,259
35–39	2,019	2,102	0.87292	2,967	35–39	2,077	2,274	0.77479	2,997
40–44	2,142	2,106	1.04776	2,601	40–44	1,948	2,139	0.94521	2,499
45–49	2,465	2,222	1.10054	2,313	45–49	2,210	2,179	1.04911	2,386
50–54	2,057	2,291	1.06956	2,252	50–54	1,958	2,073	1.06417	2,276
55–59	1,323	2,547	1.03327	2,296	55–59	1,261	2,236	1.01176	2,205
60–64	986	2,222	1.08021	2,475	60–64	895	2,053	1.04852	2,174
65–69	918	1,558	1.17763	2,999	65–69	805	1,356	1.07534	2,404
70–74	1,032	1,120	1.13590	2,524	70–74	782	899	1.00447	2,062
75–79	1,133	998	1.08715	1,694	75–79	744	741	0.92050	1,248
80–84	922	1,028	0.99612	1,116	80–84	493	637	0.81458	732
85+	1,069	1,374	0.43982	1,495	85+	459	673	0.39682	814
Total	34,880	41,401		48,429	Total	32,291	39,484		46,719

Source: 2000 and 2010, U.S. Census Bureau (http://factfinder2.census.gov)

[a] ${}_4P_{0,g,t}$ / ${}_{30}P_{15,g=f,t}$ Ages 0–4 (Child-Woman Ratio)
${}_9P_{5,g,t}$ / ${}_{30}P_{20,g=f,t}$ Ages 5–9 (Child-Woman Ratio)
$P_{x,g,t}$ / $P_{x-10,g,t-10}$ Ages 10–84
$P_{85+,g,t}$ / $P_{75+,g,t-10}$ Ages 85+

[b] ${}_4CCR_{0,g,t} \times {}_{30}P_{15,g=f,t+10}$ Ages 0–4
${}_9CCR_{5,g,t} \times {}_{30}P_{20,g=f,t+10}$ Ages 5–9
$CCR_{x,g,t} \times P_{x,g,t}$ Ages 10–84
$CCR_{75+g,t}$ / $P_{75+g,t}$ Ages 85+

aged between 2010 and 2020, a desirable feature for this population and something that the cohort-component method would usually require special adjustments to accomplish.

4.2.2 *Forecast by Age*

The H-P method can also be implemented without regard to gender by modifying Eq. 4.1:

$$ {}_nP_{x+10,t+10} = {}_nCCR_{x,t} \times {}_nP_{x,t} \quad \text{(Ages 10+)}. \tag{4.4}$$

The CCR and forecast calculations are now based to the population for both genders. Instead of child-woman ratios, child-adult ratios (CARs) are used to forecast the two youngest age groups. These ratios, computed separately for ages 0–4 and ages 5–9, relate young children to adults in the age groups most likely to be their parents. In this chapter and other examples in the book, the population aged 0–4 is related to the population aged 20–35 and the population aged 5–9 is related to the population aged 25–39. Of course other adult age groups could be used to calculate CARs, such as the populations aged 15–44 and 20–49 following the convention used in the CWR. The choice of the adult population age groups is not critical as long as the same age groups used to compute the CARs are used in the forecasting computations.[1]

Table 4.2 shows a 2020 forecast for Bellingham by age using Eq. 4.4 and CARs for ages 0–4 and 5–9. The CCRs for all genders shows the in- and out-migration of college students and indicate net in-migration up to age 79, with CCRs above 1.0 even in ages 70–79, where mortality rates are relatively high. For comparison, the table includes the forecast of males plus females from Table 4.1 (Bottom-Up column). The two forecasts are quite similar in ages 10 years and older, but show more variation in the youngest age groups. This variation is due to the different age ranges used in the CWRs and CARs. We ran a forecast using CARs for ages 0–4 and 5–9 based on ages 15–44 and 20–49, respectively and found the difference between the two forecasts drops to 28 persons (0.7%) for ages 0–4 and 11 persons (0.3%) for ages 5–9.

4.2.3 Forecast of Major League Pitchers

The H-P method is not only applicable to human and other populations stratified by age cohorts, but can be used to forecast other attributes that change in predictable ways over time. In this example, we forecast major league pitchers according to the number of consecutive years they have been in the major leagues. Years in the majors are measured by consecutive integers ranging from 0–10+ years, making them analogous to single-years of age. As such, we can construct CCRs using two consecutive annual time points for each amount of time in the majors. In this simulated forecast example, we use historical information for 1980 and 1981 to develop a one-year (1982) forecast of major league pitchers by consecutive years in the majors (see Table 4.3). These years were not affected by changes in the number

[1]Another approach for forecasting the youngest age groups is to take their ratios at two points in time and apply that ratio to the launch year age group. This approach is used in Chapter 6 where regression models are used to measure uncertainty in H-P forecasts. We prefer using CARs and CWRs in point forecast applications of the H-P method because they can account for changes in the at-risk population over the forecast horizon.

Table 4.2 Population forecast by age, Bellingham, Washington, 2020

				2020		2010–2020 Difference	
Age	2000	2010	CCR[a]	No Gender[b]	Bottom-Up[c]	Number	Percent (%)
0–4	3,502	3,736	0.14190	4,130	4,218	−88	−2.1
5–9	3,228	3,476	0.20729	4,207	3,951	256	6.1
10–14	3,151	3,338	0.95317	3,561	3,562	−1	0.0
15–19	6,373	7,179	2.22398	7,731	7,710	21	0.3
20–24	11,603	13,935	4.42241	14,762	14,764	−2	0.0
25–29	5,343	7,267	1.14028	8,186	8,248	−62	−0.8
30–34	4,273	5,126	0.44178	6,156	6,166	−10	−0.2
35–39	4,096	4,376	0.81902	5,952	5,964	−12	−0.2
40–44	4,090	4,245	0.99345	5,092	5,100	−8	−0.2
45–49	4,675	4,401	1.07446	4,702	4,699	3	0.1
50–54	4,015	4,364	1.06699	4,529	4,528	1	0.0
55–59	2,584	4,783	1.02310	4,503	4,501	2	0.0
60–64	1,881	4,275	1.06476	4,647	4,649	−2	0.0
65–69	1,723	2,914	1.12771	5,394	5,403	−9	−0.2
70–74	1,814	2,019	1.07337	4,589	4,586	3	0.1
75–79	1,877	1,739	1.00929	2,941	2,942	−1	0.0
80–84	1,415	1,665	0.91786	1,853	1,848	5	0.3
85+	1,528	2,047	0.42469	2,315	2,309	6	0.3
Total	67,171	80,885		95,250	95,148	102	0.1

Source: 2000 and 2010, U.S. Census Bureau (http://factfinder2.census.gov)
[a] ${}_4P_{0,t}\,/\,{}_{15}P_{20,t}$ Ages 0–4 (Child-Adult Ratio)
${}_9P_{5,t}\,/\,{}_{15}P_{25,t}$ Ages 5–9 (Child-Adult Ratio)
$P_{x,t}\,/\,P_{x\text{-}10,t\text{-}10}$ Ages 10–84
$P_{85+,t}\,/\,P_{75+,t\text{-}10}$ Ages 85+
[b] ${}_4CCR_{0,t} \times {}_{15}P_{20,t+10}$ Ages 0–4
${}_9CCR_{5,t} \times {}_{15}P_{25,t+10}$ Ages 5–9
$CCR_{x,t} \times P_{x,t}$ Ages 10–84
$CCR_{75+,t} \times P_{75+,t}$ Ages 85+
[c] Male + female forecasts from Table 4.1

of major league teams (there were 26 in each of the three years) and the three-year set allows us to compare the 1982 forecast to the recorded 1982 numbers to get an idea of the accuracy of using the method for this purpose.

A traditional CCR cannot be computed for zero years in the majors, so we averaged the ratio of pitchers with zero consecutive years in the league to pitchers with one and two consecutive years in the league for 1980 and 1981. CCRs for the other years spent in the league are analogous to traditional CCRs. For example, the CCR for 4 consecutive years in the majors is computed by dividing number of pitchers in the league for 4 consecutive years in 1981 by the number of pitchers in the majors for 3 consecutive years in 1980 (0.93939 = 31/33). The CCR for the open-ended category (10+ years) is the number of pitchers with 10 or more consecutive years in the majors in 1981 divided by the number of pitchers with 9 or more consecutive years in 1980 (0.78571 = 44/(12 + 44)).

Table 4.3 Forecast of pitchers by the number of consecutive years in the major leagues, 1982

Years in League	1980	1981	CCR[a]	1982 Forecast[b]	Actual	Numeric Error	Percent Error (%)
0	48	44	0.61884	40	48	−8	−16.7
1	41	40	0.83333	37	39	−2	−5.1
2	39	29	0.70732	28	33	−5	−15.2
3	33	30	0.76923	22	21	1	4.8
4	21	31	0.93939	28	28	0	0.0
5	21	18	0.85714	27	24	3	12.5
6	20	18	0.85714	15	15	0	0.0
7	18	18	0.90000	16	17	−1	−5.9
8	11	15	0.83333	15	12	3	25.0
9	12	9	0.81818	12	13	−1	−7.7
10+	44	44	0.78571	42	43	−1	−2.3
Total	308	296		282	293	−11	−3.8
						MAPE	8.7%
						MALPE	1.0%

Source: Thorn (2004)

Note, the number of teams (26) was constant between 1980 and 1982

[a]$(P_{0,t} / (P_{1,t} + P_{2,t}) + P_{0,t-1} / (P_{1,t-1} + P_{2,t-1})) \times 0.5$ Year 0

$P_{x,t} / P_{x-1,t-1}$ Years 1–9

$P_{10+,t} / P_{9+,t-1}$ Years 10+

[b]$CCR_{0,t} \times (P_{1,t+1} + P_{2,t+1})$ Year 0

$CCR_{x,t} \times P_{x,t}$ Years 1–9

$CCR_{9+,t} \times P_{9+,t}$ Years 10+

The 1982 forecast for the all pitchers is 11 lower than the actual count for an error of-3.8%. There is a wide range of errors for the individual consecutive years in the league, ranging in absolute terms from 0.0% to 25.0%. Errors for 7 categories are less than 8.0%, with the other 4 categories showing double digit percentage errors. On average, the forecast has a slight downward bias with a MALPE of −1.0% and a lower degree of accuracy with a MAPE of 8.7%.

While we are not experts in baseball statistics, the data shown in Table 4.3 indicate that there is a high level of volatility in the major league career of a pitcher, especially in the initial years. The factors likely include: (1) injuries that lead to one or more missed seasons; (2) being sent down to the minors to gain more experience (one common example is that in the initial season a pitcher is "called up" for a couple of games at the end of the season to "have a cup of coffee," followed by a return to the minors); and (3) outright release. Given this volatility, this example can be viewed as a rather strenuous test of how well the H-P method can perform in subject areas where there is less stability year to year than found in large populations of people. These areas would include any highly competitive activity such as professional sports. In this regard, we note that while there is variation in the accuracy of the 1982 forecast by number of consecutive years in the majors by pitchers, the overall forecast is reasonably accurate. Given this, other potential applications might include forecasting the number of professional football players by position (e.g., quarterback), as well as forecasting the numbers by position in basketball, hockey, and soccer. The same idea (forgetting positions) might be applied to NASCAR drivers and those in other racing circuits.

To assemble the data found in this illustration, we used a hardcopy of the eighth edition of *Total Baseball* (Thorn 2004), which means that identifying the correct data and transcribing it (into Excel) was subject to a high level of potential error, given what we wanted to record. To minimize transcription error we developed and used the following protocol. First, we identified all of the pitchers who played in any one of the years of interest, 1980, 1981, and 1982 in the hardcopy edition. We next weeded out those who played in those years but did not play consecutively in prior years (In his 27 year career, Nolan Ryan, for example, was in the majors in each of the three years but because he initially is listed in the majors in 1966 but then is not listed again until 1968, as such he is not included in data found in Table 4.3. Following these two steps we then did three separate counts for each of the three years of interest. Between the first and second count we weeded out additional errors (e.g., in the first pass we erroneously included Joey McLaughlin), but by the second pass we realized that he first was listed in 1977 and then in 1979, 1980, 1981, and 1982, but he was not listed as being in the majors in 1978. As such he was deleted because he did not have consecutive years in the majors between the year he was first listed and any of the three years of interest. By the third count, the number matched up with well with the second count and we were satisfied that the data were of sufficient accuracy to use. As an additional test, we counted the number of pitchers by year found in a third count that was separate from the count by number of consecutive years played and then compared these counts with the sum of the number by year and consecutive years played as found in the third set of counts. For 1980, the counts matched at 308, as did the count for 1981 (296) and the count for 1982 (293).

4.3 Controlling a Hamilton-Perry Forecast

The H-P method is essentially a set of cohort growth rates applied to a launch year population. It is well known that a set of constant growth rates can lead to large forecast errors in places with rapid changes. We believe it is advisable to control H-P forecasts to independent forecasts of total population in such instances (Smith and Shahidullah 1995; Swanson et al. 2010). There may be other more general reasons for controlling that are specific to any forecasting situation (Smith et al. 2013: 259). One is to make composition forecasts consistent with an "official" forecast adopted or sanctioned by a government body or some decision making unit. Another is to tie the demographic composition from an earlier forecast to an updated forecast of total population. Finally, controlling will ensure forecasts are consistent across demographic subgroups and geographic areas. For example, a forecast by age for census tracts will sum to the total population of each census tract and the age distribution for the county. Controlling an H-P forecast will require a single racking factor procedure if an age forecast is adjusted to the total population, but will require an iterative proportional fitting procedure if an age forecast is adjusted to total population and to an age distribution for a larger geographic area (Smith et al. 2013: 260–261 and 266–272).

In this section, we examine the controlling issue relative to H-P forecasts in four instances 1) Pinal County, a rapidly growing county in Arizona; 2) Gila County, a slow growing county in Arizona; 3) Pacific Beach, a slow growing community in San Diego, California and 4) Mission Valley, a fast growing community in San Diego, California. Forecasts in Pacific Beach and Mission Valley are constrained by a capacity limit. In all four examples, we prepared H-P forecasts by age for 2020 and 2030 based on the procedures discussed in section 4.2.2. This example uses 2000 and 2010 as the base period. We computed total population forecasts by summing the age groups, compared them to the controls, and computed adjustment factors needed to match the controls.

Table 4.4 shows the forecasts, independent population controls, and adjustment factors for the two Arizona counties. The total population in Pinal County more than doubled between 2000 and 2010. That rapid growth is reflected in CCRs for ages 10–69 that range from 1.65 (ages 20–24) to 2.64 (ages 30–34). Growth is also indicated is ages 70–79, where the CCRs are sizable despite of the relatively high mortality in these age groups. As a result, the total population forecasts from the H-P method are much higher than the controls taken from the "official" medium series forecast produced by the State of Arizona (2015). Consequently, the adjustment factors are substantially below 1.0 and decrease as the horizon length increases from 10- to 20-years. The H-P forecasts are too high by 41% in 2020 and are too high by 64% in 2030 relative to the controls.

In contrast, Gila County's total population grew slowly by about 0.4% per year on average between 2000 and 2010. This slow growth is reflected in CCRs for ages 10–69 that range from 0.655 (ages 20–24) to 1.241 (ages 65–69) and imply out-migration of young adults (ages 20–29) and in-migration of ages 30–69, with the most rapid in-migration occurring in the retirement ages. As a result, the total population forecasts from the H-P method align closely with the controls. The 2020 and 2030 total population forecasts are about 1% higher and about 1%, lower than the controls, respectively.

Table 4.5 shows the forecasts, independent population controls, and adjustment factors for the two communities in San Diego, California. Pacific Beach (PB) is a beach community 10 miles north of downtown San Diego. It has grown very slowly over the past 25 years and showed no change in the total population between 2000 and 2010. Historically, new houses in PB have come from small in-fill and redevelopment projects. The age composition in PB has been very stable overtime as illustrated in the 2000 and 2010 age distributions. PB is home to large number of young adults (aged 20–29) that tend to leave the area once they reach their thirties. This pattern is suggested in the large CCRs for ages 20–29 that drop dramatically to below 1.00 for the ages (35–44).

Table 4.4 Population forecast by age, Pinal and Gila counties, Arizona, 2020 and 2030

Pinal County						Gila County					
Age	2000	2010	CCR	2020	2030	Age	2000	2010	CCR	2020	2030
0–4	12,066	30,182	0.38736	58,534	132,199	0–4	3,116	3,059	0.43176	3,012	2,637
5–9	12,775	29,021	0.34566	59,076	121,379	5–9	3,578	3,120	0.45822	3,409	3,004
10–14	12,823	26,374	2.18581	65,972	127,944	10–14	3,946	3,135	1.00610	3,078	3,030
15–19	12,078	22,650	1.77299	51,454	104,741	15–19	3,413	3,342	0.93404	2,914	3,184
20–24	11,057	21,171	1.65102	43,544	108,921	20–24	2,117	2,584	0.65484	2,053	2,016
25–29	11,858	27,603	2.28539	51,764	117,592	25–29	2,165	2,360	0.69147	2,311	2,015
30–34	11,736	29,143	2.63571	55,801	114,769	30–34	2,561	2,141	1.01134	2,613	2,076
35–39	13,034	27,212	2.29482	63,344	118,789	35–39	3,165	2,308	1.06605	2,516	2,464
40–44	12,350	23,147	1.97231	57,479	110,057	40–44	3,539	2,848	1.11207	2,381	2,906
45–49	11,109	22,659	1.73845	47,307	110,120	45–49	3,463	3,489	1.10237	2,544	2,774
50–54	10,296	21,729	1.75943	40,726	101,130	50–54	3,588	4,028	1.13817	3,242	2,710
55–59	9,687	20,754	1.86821	42,332	88,379	55–59	3,351	4,222	1.21917	4,254	3,102
60–64	9,687	22,054	2.14200	46,544	87,235	60–64	3,174	4,511	1.25725	5,064	4,076
65–69	9,581	19,092	1.97089	40,904	83,432	65–69	3,121	4,159	1.24112	5,240	5,280
70–74	8,409	14,163	1.46206	32,244	68,050	70–74	2,627	3,096	0.97543	4,400	4,940
75–79	5,978	9,356	0.97652	18,644	39,944	75–79	2,117	2,449	0.78468	3,263	4,112
80–84	3,195	5,733	0.68177	9,656	21,983	80–84	1,309	1,555	0.59193	1,833	2,604
85+	2,008	3,727	0.33333	6,272	11,524	85+	985	1,191	0.27001	1,403	1,755
Total	179,727	375,770		791,597	1,668,188	Total	51,335	53,597		55,530	54,685
Total population control[a]				463,500	604,800	Total population control[a]				55,100	55,300
Adjustment factor[b]				0.5855	0.3625	Adjustment factor[b]				0.9923	1.0112

Source: 2000 and 2010, U.S. Census Bureau (http://factfinder2.census.gov)
[a]State of Arizona (2015)
[b]Population control/total population forecast

Table 4.5 Population forecast by age, Pacific Beach and Mission Valley communities, San Diego, California, 2020 and 2030

Pacific beach						Mission valley					
Age	2000	2010	CCR	2020	2030	Age	2000	2010	CCR	2020	2030
0–4	1,428	1,362	0.06749	1,259	1,162	0–4	377	864	0.09603	1,634	2,876
5–9	1,184	1,064	0.06249	1,014	911	5–9	260	434	0.05799	872	1,485
10–14	993	918	0.64286	876	809	10–14	188	267	0.70822	612	1,157
15–19	1,064	909	0.76774	817	778	15–19	256	608	2.33846	1,015	2,039
20–24	6,003	6,254	6.29809	5,782	5,517	20–24	1,685	2,973	15.81383	4,222	9,678
25–29	8,071	8,742	8.21617	7,468	6,713	25–29	2,299	3,545	13.84766	8,419	14,055
30–34	5,341	5,184	0.86357	5,401	4,993	30–34	1,521	2,479	1.47122	4,374	6,211
35–39	3,458	3,101	0.38422	3,359	2,869	35–39	1,170	1,460	0.63506	2,251	5,347
40–44	2,379	2,195	0.41097	2,130	2,220	40–44	837	1,086	0.71400	1,770	3,123
45–49	2,002	1,891	0.54685	1,696	1,837	45–49	740	1,024	0.87521	1,278	1,970
50–54	1,834	1,782	0.74905	1,644	1,595	50–54	636	967	1.15532	1,255	2,045
55–59	1,370	1,495	0.74675	1,412	1,266	55–59	491	854	1.15405	1,182	1,475
60–64	1,001	1,476	0.80480	1,434	1,323	60–64	341	618	0.97170	940	1,219
65–69	846	1,056	0.77080	1,152	1,088	65–69	318	477	0.97149	830	1,148
70–74	977	790	0.78921	1,165	1,132	70–74	318	324	0.95015	587	893
75–79	937	652	0.77069	814	888	75–79	236	278	0.87421	417	726
80–84	782	669	0.68475	541	798	80–84	182	241	0.75786	246	445
85+	630	762	0.32439	676	659	85+	162	350	0.60345	524	716
Total	40,300	40,302		38,640	36,558	Total	12,017	18,849		32,428	56,608
Total population control[a]				43,247	46,150	Total population control[a]				24,894	34,588
Adjustment factor[b]				1.1192	1.2624	Adjustment factor[b]				0.7677	0.6110

Source: 2000 and 2010, U.S. Census Bureau (http://factfinder2.census.gov)
[a]San Diego Association of Governments (2013)
[b]Population control/total population forecast

H-P forecasts for PB imply a decline in the total population after 2010, which is at odds with the controls developed by the San Diego Association of Governments (SANDAG) (2013). H-P forecasts are too low by 12% in 2020 and too low by 26% in 2030 relative to the controls. SANDAG's forecast assumes that substantial redevelopment will occur in PB as envisioned in the City of San Diego's City of Villages Plan (2016).

Mission Valley (MV) is six miles northeast of downtown San Diego and is home to Qualcomm stadium. For many years MV was largely developed with retail, commercial, and hospitality services with relatively little residential activity. That pattern started to change in the early 1990s, as MV began to attract large multi-family residential developments and the City of San Diego updated the MV Community Plan. MV grew substantially between 2000 and 2010, increasing by 57%. MV's age structure is similar to PBs, with relatively few people either under the age of 20 or over 64 and a concentration in the young adult ages (20–34). Young adults tend to leave between the ages of (35–49), but MV seems to attract persons in their 50s. This pattern is suggested in the very large CCRs for ages 20–29, a CCR of 1.47 for 30–34, CCRs below 1.00 for 35–49, and CCRs of 1.15 for those aged 50–59.

H-P forecasts for MV that assume continuation of the rapid growth seen from 2000–2010 are substantially higher than the controls. It is estimated that the MV Community Plan has a capacity for around 36,000 people (San Diego Association of Governments 2013); this capacity constraint is reflected in the controls. The H-P age forecasts are too high by 23% in 2020 and too high by 39% in 2030 relative to the controls.

4.4 Conclusions

The major advantage of the H-P method compared to the cohort-component method is that it has much smaller data requirements. Consequently is it far less expensive, much quicker to implement, and particularly useful for subcounty forecasts where data on the components of population change are very limited, if they exist at all. The 2000 U.S. census was the first to allow respondents to list themselves as belonging to one or more racial categories, and as a result racial data are inconsistent with racial data prior to 2000. In addition, racial classifications from the decennial census and America Community Survey are not completely consistent with the classification system used for vital statistics data, making it difficult to develop reliable estimates of the components of change for racial groups. Because it is based solely on data for two age distributions, the H-P method avoids these complications and provides a viable alternative to the full cohort-component method for forecasts of race, especially for forecasts of the multi-racial population (Swanson 2013).

Another attractive feature of the H-P method is that the CCRs, which embody both changes in mortality and migration, can handle special populations or unique age structures, like Pacific Beach and Mission Valley, without any adjustments to the basic model. The same is not true for the cohort-component model, where in these situations adjustments might be needed to the base population and/or to fertility, mortality, and migration rates.

The H-P method does not provide information on the components of change and will not be useful if this information is needed. (In Chapter 13, we present a decomposition method of the CCR that yields forecasts of migration and deaths). In rapidly changing areas, the H-P method can lead to large forecast errors and a strong upward or downward bias depending on whether the change in population is increasing or decreasing. We saw the strong upward bias in the rapidly increasing areas of Pinal County, Arizona and the Mission Valley community in the City of San Diego, California. To illustrate the impact of a large decline on H-P forecasts, we chose census tract 9.0 in Curry County, New Mexico; home to Cannon AFB. The population in that census tract dropped from 4,307 to 2,193 between 2000 and 2010, and H-P forecasts for 2020 and 2030 were 949 and 436, respectively. These forecasts likely have unreasonably large downward biases.

Census enumerations are generally high quality, but are not perfect. Some people are missed, others counted twice, and others counted in the wrong place. Coverage rates that differ from one subgroup to another and change over time may introduce error into the CCRs. While H-P calculations can be characterized as "quick and easy," application of this method can be anything but, especially dealing with subcounty areas. As Swanson et al. (2010) point out in their census tract forecasts for Clark County, Nevada, the H-P method required calibrations, many adjustments, and knowledge of the growth patterns for specific areas to generate plausible forecasts. One major effort, especially in subcounty areas, is creating data for geographic areas with constant boundaries at both points in time.

Finally, the H-P method has been shown to produce accurate and reasonable forecasts for 10- and 20-year forecast horizons relative to other forecasting methods. We know of no studies that have evaluated the H-P method for longer forecasting horizons.

References

Baker, J., Alcantara, A., Ruan, X., Watkins, K., & Vasan, S. (2014).Spatial weighting improves accuracy in small area demographic forecasts of urban census tract populations. *Journal of Population Research*. DOI 10.1007/s12546-014-9137-1

City of San Diego. (2016). *General Plan*. Retrieved from https://www.sandiego.gov/planning/genplan

Hamilton, C. H., & Perry, J. (1962). A short method for projecting population by age from one decennial census to another. *Social Forces, 41*, 163–170.

Kodiko, H. (2014). *Subnational projection methods: Application to the counties in the former Nyanza Province, Kenya*. Master's thesis. Nairobi, Kenya: Population Studies and Research Institute, University of Nairobi. Retrieved from http://psri.uonbi.ac.ke/sites/default/files/chss/psri/psri/HERBERT%20ODHIAMBO%20KODIKO%20FINAL%20PROJECT_0.pdf

National Center for Educational Statistics. (2016). College navigator. Retrieved from http://nces.ed.gov/collegenavigator

San Diego Association of Governments. (2013). Series 13 regional growth forecast. Retrieved from http://datasurfer.sandag.org/download/sandag_forecast_13_cpa_mission-valley.pdf

Smith, S. K., & Shahidullah, M. (1995). An evaluation of population projections errors by census tract. *Journal of the American Statistical Association, 90*, 64–71.

Smith, S. K., & Tayman, J. (2003). An evaluation of population projections by age. *Demography, 40*(4), 741–757.

Smith, S. K., Tayman, J., & Swanson, D. (2013). *A practitioner's guide to state and local population projections*. Dordrecht: Springer.

State of Arizona. (2015). 2015–2050 state and county population projections. Retrieved from https://population.az.gov/population-projections

State of Washington. (2016, April 1). *Population of cities, towns and counties used for allocation of selected state revenues State of Washington*. Retrieved from http://www.ofm.wa.gov/pop/april1/ofm_april1_population_final.pdf

Swanson, D. (2013). People of the Inland Empire: Changes in age, ethnicity, and race. In *Paper presented at the center of sustainable suburban development*. Riverside: University of California Riverside.

Swanson, D., & Tayman, J. (2017). A long term test of the accuracy of the Hamilton-Perry method for forecasting state populations by age. In D. Swanson (Ed.), *The frontiers of applied demography* (pp. 491–514). Dordrecht: Springer.

Swanson, D., Schlottmann, A., & Schmidt, R. (2010). Forecasting the population of census tracts by age and sex: An example of the Hamilton–Perry method in action. *Population Research and Policy Review, 29*, 47–63.

Thorn, J. (2004). *Total baseball: The ultimate baseball encyclopedia* (8th ed.). Sport Media Publishing.

Wilson, T. (2016). Evaluation of alternative cohort-component models for local area population forecasts. *Population Research and Policy Review, 35*, 241–261.

Chapter 5
Forecasting Using Modified Cohort Change Ratios

5.1 Introduction

Assessments of the H-P method have been based on the usual assumption that cohort change ratios (CCRs) developed over the base period (e.g., 2000 and 2010) and child-woman ratios (CWRs) or, more generally, child-adult ratios, developed for the launch year (e.g., 2010) are held constant over the forecast horizon (horizon). This basic or constant model was presented in Chapter 4. Smith et al. (2013: 179) discuss the possibility of relaxing this assumption by averaging together CCRs and CWRs from several recent censuses, by extrapolating historical trends, or using a synthetic approach based on CCR and CWR forecasts from a population in a larger geographic area. We are not aware of any studies that have evaluated the accuracy of the H-P method using modified CCRs and CWRs.

In this chapter, we evaluate several approaches for modifying CCRs and CWRs over the horizon. These approaches include: (1) averaging; (2) trending; (3) and a synthetic method that generates changes in local CCRs and CWRs by linking them to state-level changes. We evaluate three dimensions of error (accuracy, bias, and allocation error) and compare forecasts using modified CCRs and CWRs against forecasts holding them constant (the basic H-P framework). Our focus here is not on the utility of the H-P method for population forecasting (something we discuss in Chapter 15), but on the size of the errors of the modification alternatives relative to basic H-P framework. Errors for two 10-year horizons and one 20-year horizon are examined for counties in Washington State and a 10-year horizon is examined for census tracts in New Mexico. This evaluation suggests that averaging or trending CCRs and CWRs are not worthwhile strategies. However, we find that the synthetic method has lower errors in comparison to the errors found using the basic H-P framework.

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_5

5.2 Modifying Cohort Change and Child-Woman Ratios

As described in Chapter 4, the H-P method uses CCRs, which are calculated by dividing the population aged *x* in year *t* by the population aged *x–10* in year *t–10*. For the analysis conducted in this chapter, CCRs are calculated separately for males and females. These CCRs are applied to each age, gender group in year *t* to provide forecasts by age and gender in the year *t + 10*. Given the nature of the CCRs, 10–14 is the youngest 5-year age group for which forecasts can be made if there are 10 years between censuses. Children by gender younger than age 10 are forecast using CWRs from the launch year (i.e., males or females younger than age 5/females aged 15–44 and males or females aged 5–9/females aged 20 to 45). Eqs. 5.1, 5.2 and 5.3 represent the usual application of the H-P framework (i.e., holding CCRs and CWRs from the most recent 10-year period constant over the horizon):

$$_{\mathrm{n}}\mathrm{P}_{\mathrm{x+10,g,t+10}} = {}_{n}\mathrm{CCR}_{\mathrm{x,g,t}} \times {}_{\mathrm{n}}\mathrm{P}_{\mathrm{x,g,t}} \quad (\text{Ages } 10+), \tag{5.1}$$

$$_{4}\mathrm{P}_{0,\mathrm{g,t+10}} = {}_{4}\mathrm{CWR}_{0,\mathrm{g,t}} \times {}_{44}\mathrm{FP}_{15,\mathrm{t+10}} \quad (\text{Ages } 0-4), \tag{5.2}$$

$$_{9}\mathrm{P}_{5,\mathrm{g,t+10}} = {}_{9}\mathrm{CWR}_{5,\mathrm{g,t}} \times {}_{49}\mathrm{FP}_{20,\mathrm{t+10}} \quad (\text{Ages } 5-9), \tag{5.3}$$

where,

n is the width of the age group,
x is the beginning of the age group,
g is gender,
t is the launch year,
P is the population,
CCR is the cohort change ratio,
CWR is the child-woman ratio, and
FP is the female population.

What are the effects on the errors from H-P forecasts if the assumption of constancy is relaxed? To study this question, we evaluate three approaches for modifying CCRs and CWRs over the horizon. Following the suggestions the Smith et al. (2013: 179), we create forecasts using averages and then produce them using trends in the CCRs and CWRs. Our third approach uses a synthetic method that links changes in the CCRs and CWRs for one area to the changes forecast for a different area, which is usually a higher level of geography (e.g., county ratios modified by forecast changes in state ratios). The synthetic method is frequently used in state and local forecasting (Smith and Rayer 2012; Smith et al. 2013: 65) and has a long history of use in population estimation (Swanson and Tayman 2012: 209–213).

The average alternative (AVG) combines two CCRs over the most recent 20-year period (e.g. 1990–2010) and two CWRs over the most recent 10-year period (e.g. 2000 and 2010):

$$_{n}CCR_{x,g,t} = {}_{n}P_{x,g,t}/{}_{n}P_{x-k,g,t-10}, \tag{5.4}$$

$$_{n}CCR_{x,g,t-10} = {}_{n}P_{x,g,t-10}/{}_{n}P_{x-k,g,t-20}, \tag{5.5}$$

$$AVG_{n}CCR_{x,g} = \left({}_{n}CCR_{x,g,t} + {}_{n}CCR_{x,g,t-10}\right)/2, \tag{5.6}$$

$$AVG_{4}CWR_{0,g} = \left({}_{4}CWR_{0,g,t} + {}_{4}CWR_{0,g,t-10}\right)/2, \tag{5.7}$$

$$AVG_{9}CWR_{5,g} = \left({}_{9}CWR_{5,g,t} + {}_{9}CWR_{5,g,t-10}\right)/2, \tag{5.8}$$

where,

AVGCCR is the average CCR over a 20-year period, and
AVGCWR is the average CWRs over10-year period.

Forecasts for AVG assume the averaged CCRs and CWRs stay constant over the horizon. These forecasts are made by substituting $AVG_nCCR_{x,g}$, $AVG_4CWR_{0,g}$, and $AVG_9CWR_{5,g}$ for the CCRs and CWRs into Eqs. 5.1, 5.2 and 5.3, respectively.

The trend alternative (TREND) is based on the proportionate change (or ratio) in the CCRs over the most recent 20-year period and the proportionate change in the CWRs over the most recent 10-year period as follows:

$$RATIO_{n}CCR_{x,g} = \left({}_{n}CCR_{x,g,t}/{}_{n}CCR_{x,g,t-10}\right), \tag{5.9}$$

$$RATIO_{4}CWR_{0,g} = \left({}_{4}CWR_{0,g,t}/{}_{4}CWR_{0,g,t-10}\right), \tag{5.10}$$

$$RATIO_{9}CWR_{5,g} = \left({}_{9}CWR_{5,g,t}/{}_{9}CWR_{5,g,t-10}\right), \tag{5.11}$$

where,

RATIOCCR is the ratio of the CCRs over a 20-year period, and
RATIOCWR is the ratio of CWRs over a 10-year period.

Forecasts based on TREND are computed by:

$$_{n}P_{x+10,g,t+10} = \left({}_{n}CCR_{x,g,t} \times RATIO_{n}CCR_{x,g}\right) \times {}_{n}P_{x,g,t} \quad \text{(Ages 10+)}, \tag{5.12}$$

$$_{4}P_{0,g,t+10} = \left({}_{4}CWR_{0,g,t} \times RATIO_{4}CWR_{0,g}\right) \times {}_{44}FP_{15,t+10} \quad \text{(Ages 0–4), and} \tag{5.13}$$

$$_{9}P_{5,g,t+10} = \left({}_{9}CWR_{5,g,t} \times RATIO_{9}CWR_{5,g}\right) \times {}_{49}FP_{20,t+10} \quad \text{(Ages 5–9).} \tag{5.14}$$

AVG and TREND make individual adjustments to the CCRs and CWRs. That is, the average and proportionate adjustments are specific to each area being forecast. The synthetic method (SYN) does not make area-specific adjustments, but applies the same proportionate change to each area based on a forecast for a larger geographic area (i.e., state changes applied to each county). Including SYN allows examination of the efficacy of using area-specific modifications to the CCRs and CWRs as compared to a more global approach to modification. Unlike AVG and TREND, SYN requires an independent forecast of CCRs and CWRs for the larger area. The synthetic method globally incorporates information pertinent to the horizon, which may be an advantage over the average and trend alternatives, which rely solely on historical patterns.

For the CCRs, the proportionate adjustments for the larger area are based on CCRs from times *t and t-10* and CCRs from times *t* + *10 and t*; that is, from the decades prior and subsequent to the launch year. For the CWRs, the adjustments represent CWRs from times *t and t* + *10*; that is, from beginning and end of the decade after the launch year. Forecasts using SYN are computed by (bold indicates the larger area):

$$\mathbf{SYN_nCCR_{x,g}} = \left(\mathbf{{}_nCCR_{x,g,t+10}}/\mathbf{{}_nCCR_{x,g,t}}\right), \tag{5.15}$$

$$\mathbf{SYN_4CWR_{0,g}} = \left(\mathbf{{}_4CWR_{0,g,t+10}}/\mathbf{{}_4CWR_{0,g,t}}\right), \tag{5.16}$$

$$\mathbf{SYN_9CWR_{5,g}} = \left(\mathbf{{}_9CWR_{5,g,t+10}}/\mathbf{{}_9CWR_{5,g,t}}\right), \tag{5.17}$$

$${}_n\mathrm{P}_{x+10,g,t+10} = \left({}_n\mathrm{CCR}_{x,g,t} \times \mathbf{SYN_nCCR_{x,g}}\right) \times {}_n\mathrm{P}x,g,t \ \ (\text{Ages } 10+), \tag{5.18}$$

$${}_4\mathrm{P}_{0,g,t+10} = \left({}_4\mathrm{CWR}_{0,g,t} \times \mathbf{SYN_4CWR_{0,g}}\right) \times {}_{44}\mathrm{FP}_{15,t+10} \ \ (\text{Ages } 0\text{–}4), \tag{5.19}$$

$${}_9\mathrm{P}_{5,g,t+10} = \left({}_9\mathrm{CWR}_{5,g,t} \times \mathbf{SYN_9CWR_{5,g}}\right) \times {}_{49}\mathrm{FP}_{20,t+10} \ \ (\text{Ages } 5\text{–}9), \tag{5.20}$$

where,

SYNCCR is the CCR adjustments for the larger area, and
SYNCWR is the CWR adjustments for the larger area.

5.3 Measures of Forecast Error

We employ several commonly used measures that capture three dimensions of forecast error—accuracy, bias, and allocation error (Swanson 2015; Swanson et al. 2011). Error is defined as the difference between the simulated forecast and a census count. The mean algebraic percent error (MALPE) measures bias in which positive and negative values offset each other. A positive MALPE reflects the tendency for the forecasts to be too high on average and a negative MALPE reflect the tendency for the forecasts to be too low on average. The mean absolute percent error (MAPE) measures forecast accuracy in which positive and negative errors do not offset each other. It shows the average percentage difference between the forecast and observed population, ignoring the sign of the error.

The error distribution underlying the MAPE is often asymmetrical and right-skewed, causing the MAPE to overstate the error represented by most of the observations. MAPE-R and the median absolute percent error (MEDAPE) are two measures of accuracy that can be used when the error distribution underlying the MAPE is highly asymmetrical (Swanson et al. 2011; Swanson et al. 2012). Because forecast errors are generally stable across a variety of error measures (Rayer 2007), we forego using MEDAPE and MAPE-R in this study.

MAPE and MALPE are based on forecast errors for a particular geographic area. Another perspective views the misallocation of the forecast across geographic space or across a given variable such as age. Our focus here is not on geographic misallocation, but on the accuracy of the age distribution forecast. A number of measures can be used to measure allocation error (Duncan et al. 1961; Massey and Denton 1988). We use the Index of Dissimilarity (IOD) that compares the percentage distribution of the forecast population by age group with the corresponding percentage distribution in the census. The IOD calculates the percentage that the forecast distribution would have to change to match the census distribution. The IOD ranges from 0 to 100, with 0 indicating identical percentage distributions and 100 indicating complete disparity between the forecast and census distributions.

5.4 Empirical Data

Our samples consist of: (1) the 39 counties in Washington State; and (2) census tracts in New Mexico. Data were collected for 18 age groups (0–4, 5–9,. . ., 80–84, and 85+) for males and females. For the counties, we assembled census data for each decade from 1970 to 2010. Boundary changes are an issue when using longitudinal data for census tracts and limited how far back in time we could get data with consistent boundary definitions. As a result, we used census data from 1990, 2000, and 2010 for 471 of the 499 census tracts in New Mexico. Census tract boundaries for 2010 formed the basis of this data set. Census 1990 and 2000 data were extracted at the block level and then re-aggregated to census 2010 census tract boundaries.

Aside from boundary changes, implementing the H-P method in census tracts is effected by zero and small non-zero population counts. A CCR is undefined if the earlier census count (the denominator) is zero. The variability inherent in small population counts can also lead to abnormally large or abnormally small CCRs when percentage changes increase or decrease by large amounts. For this study, we made some general adjustments to the census tract data to deal with these issues. We excluded 28 census tracts that contained zero population in any age and sex group in 1990 and 2000. These excluded census tracts accounted for around 5% of the total males and females in the state, with a range from 2.4% for males and females aged 85 and older to 8.1% for males aged 20–24.

CCRs were also set at a maximum value of 3.0 and a minimum value of 0.4. The minimum and maximum values were applied to 5.6% and 7.0% of the 15,072 1990–2000 CCRs, respectively. Remember, the objective here is to compare the errors from the basic H-P model to modification alternatives and not to produce the most accurate H-P forecast. In the latter case, different adjustments strategies and a much closer inspection of the census tract CCRs would be necessary. Such adjustment strategies could include using county-level CCRs, using CCRs from an aggregation of census tracts around the census tract in question, and/or setting upper and lower limits on the total population size (Swanson et al. 2010).

For Washington State counties, we constructed H-P models for the four alternatives: (1) CONST (holding CCRs and CWRs constant), (2) AVG (averaging CCRs and CWRs and holding the averages constant; (3) TREND (proportionate changes applied to CCRs and CWRs; and (4) SYN (county CCRs and CWRs adjusted by a forecast of state trends in CCRs and CWRs). We prepared forecasts for three launch year and target year combinations, which yielded two 10-year horizons and one 20-year horizon: (1) 1990 launch year and 2000 horizon year; (2) 2000 launch year and 2010 horizon year; and (3) 1990 launch year and 2010 horizon year.

For the 10-year forecast with the 1990 launch year, CONST used CCRs from the 1980–1990 decade and CWRs from 1990; AVG and TREND used CCRs from the 1970–1980 and 1980–1990 decades and CWRs from 1980 and 1990; and SYN used CCRs from the 1980–1990 decade, CWRs from 1990, state-level CCRs from the 1980–1990 and 1990–2000 decades, and state-level CWRs from 1990 and 2000. For the 20-year horizon, SYN used state-level CCRs from the 1990–2000 and 2000–2010 decades and state-level CWRs for 2000 and 2010.

For the 2000 launch year, CONST used CCRs from the 1990–2000 decade and CWRs from 2000; AVG and TREND used CCRs from the 1980–1990 and 1990–2000 decades and CWRs from 1990 and 2000; and SYN used CCRs based on the 1990–2000 decade, CWRs from 2000, state-level CCRs from the 1990–2000 and 2000–2010 decades, and state-level CWRs for 2000 and 2010.

Because the census tract data started in 1990, fewer alternatives and only one 10-year horizon was analyzed. We prepared forecasts for CONST and SYN using the 2000 launch year and the 2010 horizon year (1980 data would have been needed for the average and trend alternatives). CONST used CCRs from the 1990–2000 decade and CWRs from 2000; and the SYN used CCRs from the 1990–2000 decade, CWRs from 2000, state-level CCRs from the 1990–2000 and 2000–2010 decades, and state-level CWRs for 2000 and 2010.

To test the synthetic alternatives for counties, we would like to have 2000 and 2010 forecasts by age and sex for Washington State based on a 1990 launch year and a 2010 forecast by age and sex based on a 2000 launch year. For the census tract synthetic alternative, we would like to have a 2010 forecast by age and sex for New Mexico based on a 2000 launch year. We were only able to obtain a 2010 forecast for Washington State based on a 2000 launch year (Office of Financial Management 2002). So for this analysis, we used state-level census data for all synthetic alternatives.

We do not believe the use of state-level census data, in place of state-level forecasts contemporaneous with the launch and target years examined, has a significant impact on the forecasts from SYN. To examine this claim, we compared 2010 forecasts (from a 2000 launch year) based on the SYN alternative using 2010 census data and the 2010 forecast for Washington State. The patterns and levels of bias and accuracy for age groups and total population were quite similar for both forecasts. Differences in MALPEs were less than 2% points and the differences in MAPEs were less than 1% point. The errors using state-level census data were not uniformly lower than errors using the state-level forecast, and misallocation errors were virtually identical in both forecasts.

Some applications control H-P forecasts to an independent forecast of total population (Smith et al. 2013: 180–181; Swanson et al. 2010).[1] We opted not to control the Hamilton-Perry forecasts by age because we wanted to evaluate the total population forecast error derived directly from the method itself for each alternative. Total population forecasts are derived by summing the forecast over all age groups.

5.5 Empirical Results[2]

5.5.1 Total Population Forecast Error

5.5.1.1 Washington State Counties

We begin the analysis by examining forecast error for the total population (sum of the age group forecasts) for Washington State counties. Table 5.1 contains the average bias and accuracy for the four alternatives and the three launch and horizon year combinations. In terms of bias, TREND clearly performs the worst in all launch year and horizon years. Its MALPEs range from −41.8% to 24.5%, compared to the largest and smallest MALPEs (ignoring signs) of the other alternatives (0.8% and 20.8%). For 10-year horizons, AVG performs the best, followed by the SYN, and then CONST. For the 20-year horizon, SYN has the lowest bias (−8.7%), followed by CONST (−14.4%). The biases in AVG and TREND, especially TREND, are considerably larger than the biases in CONST and SYN for the 20-year horizon.

In terms of accuracy, SYN has the least error of any alternative for all but one launch and horizon year combination; SYN's MAPE (5.7%) is slightly larger than

[1]Smith and Tayman (2003) found that while uncontrolled H-P forecasts generally had larger errors than the controlled forecasts for all states and counties in Florida, the patterns of errors by age groups was generally very similar for both the controlled and uncontrolled forecasts.

[2]We analyzed separate projections for males and females, but do not report the results here because of space limitations. Forecast errors for males and females were similar for most age groups and the total population. For ages 65 years and older, females generally had greater accuracy and lower bias.

Table 5.1 Forecast bias and accuracy for total population by alternative, Washington State counties

		MALPE			
Launch year	Horizon year	Constant	Average	Trend	Synthetic
1990	2000	−10.4%	4.5%	−24.5%	−5.9%
2000	2010	6.3%	0.8%	20.7%	2.4%
1990	2000	−14.4%	20.8%	−41.8%	−8.7%
		MAPE			
Launch year	Horizon year	Constant	Average	Trend	Synthetic
1990	2000	11.2%	12.3%	25.2%	8.3%
2000	2010	7.6%	5.6%	20.7%	5.7%
1990	2000	17.7%	31.3%	47.3%	15.3%

AVG's MAPE (5.6%) for the 2010 horizon using the 2000 launch year. TREND universally has the lowest accuracy of any alternative, with MAPEs ranging from 20.7% to 47.3%. For 10-year horizons, CONST has greater accuracy (11.2%) than AVG (12.3) using the 1990 launch year, but AVG (5.6%) has greater accuracy than CONST (7.6%) using the 2000 launch year. For the 20-year horizon, the AVG MAPE (31.3%) is now considerably larger than the CONST MAPE (17.7%), which is somewhat higher than the SYN MAPE (15.3%).

5.5.1.2 New Mexico Census Tracts

Turning to the forecast errors for the total populations of New Mexico census tracts, Fig. 5.1 shows bias is lower in SYN (29.2%) compared to CONST (34.7%); a difference of almost 19%. Accuracy is greater in SYN (48.3% vs 51.3%), but the difference of just over 6% is less than the improvement in bias. The distribution of absolute percent errors (APEs) is similar for SYN and CONST (see Fig. 5.2). SYN does have more relatively small APEs under 25% (52.9% vs 50.7%) and fewer extreme APEs greater than or equal to 100% (13.8% vs. 15.5%).

To provide a more detailed geographic perspective on the total population forecast error in New Mexico's census tracts, Table 5.2 shows the MALPE and MAPE for census tracts in each of New Mexico's 34 counties. SYN has lower bias in 29 counties (85.3%). The percentage improvement in bias in these counties ranges from 8.5% in Sierra County to 91.4% in Chaves County. In the other counties, excluding Curry, the bias in SYN is between 29% and 41% larger than the bias in CONST. In Curry County the MALPEs for CONST and SYN are 0.1% and −4.0%, respectively, leading to a large percentage increase in bias in SYN (3,900%). SYN has greater accuracy in 28 counties (82.3%). The percentage improvement in accuracy in these counties ranges from 0.7% in Cibola County to 46.3% in Debaca County. The percentage loss in accuracy in SYN compared to CONST ranges from 1.6% in Roosevelt County to 41.5% in Union County.

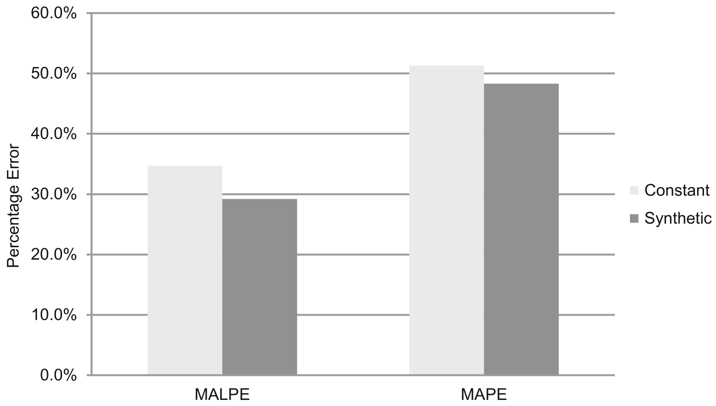

Fig. 5.1 Forecast bias and accuracy for total population by alternative, New Mexico census tracts, 2010

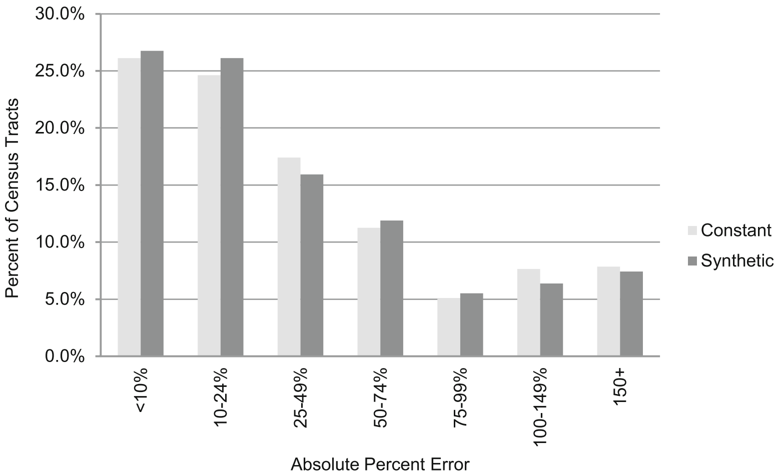

Fig. 5.2 Distribution of absolute percentage errors for total population by alternative, New Mexico census tracts, 2010

5.5.2 Forecast Error by Age Group

5.5.2.1 Washington State Counties

We continue the analysis by examining forecast error by age group for Washington State counties. We constructed the forecasts using 5-year age groups and a terminal age group of 85 years and older, but evaluate forecast errors using a reduced set of seven categories that cover the full age spectrum, adequately capture the age-specific performance of the alternatives, and make the analysis easier to follow.

Table 5.2 Forecast accuracy and bias for total population by alternative, census tracts within New Mexico counties, 2010

	MALPE			MAPE		
County	Constant	Synthetic	% Diff.[a,b]	Constant	Synthetic	% Diff.[a]
Bernalillo	20.3%	15.4%	24.1%	40.0%	37.9%	5.3%
Catron	35.6%	31.9%	10.4%	35.6%	31.9%	10.4%
Chaves	3.9%	−0.3%	92.3%	10.4%	10.1%	2.9%
Cibola	39.5%	33.9%	14.2%	64.3%	63.8%	0.8%
Colfax	14.3%	10.2%	28.7%	14.5%	15.1%	−4.1%
Curry	0.1%	−4.0%	−3900.0%	12.9%	13.3%	−3.1%
Debaca	8.2%	4.4%	46.3%	8.2%	4.4%	46.3%
Dona Ana	69.5%	62.2%	10.5%	106.7%	101.7%	4.7%
Eddy	41.3%	35.6%	13.8%	76.7%	72.9%	5.0%
Grant	48.2%	43.1%	10.6%	51.3%	47.0%	8.4%
Guadalupe	12.8%	8.3%	35.2%	12.8%	8.3%	35.2%
Harding	−9.6%	−12.4%	−29.2%	9.6%	12.4%	−29.2%
Hidalgo	18.2%	13.6%	25.3%	18.2%	16.7%	8.2%
Lea	−11.4%	−15.2%	−33.3%	13.7%	15.4%	−12.4%
Lincoln	44.0%	39.3%	10.7%	44.0%	39.3%	10.7%
Los Alamos	1.7%	−2.3%	−35.3%	7.8%	7.6%	2.6%
Luna	40.7%	35.0%	14.0%	45.0%	40.4%	10.2%
Mckinley	119.4%	108.9%	8.8%	130.1%	121.3%	6.8%
Mora	26.2%	21.6%	17.6%	26.2%	21.6%	17.6%
Otero	53.6%	47.4%	11.6%	76.2%	72.1%	5.4%
Quay	58.5%	53.3%	8.9%	62.0%	59.5%	4.0%
Rio Arriba	31.2%	25.9%	17.0%	44.8%	40.8%	8.9%
Roosevelt	11.2%	6.4%	42.9%	41.6%	42.3%	−1.7%
Sandoval	48.9%	42.7%	12.7%	68.9%	64.8%	6.0%
San Juan	54.9%	48.2%	12.2%	62.0%	56.6%	8.7%
San Miguel	87.2%	79.6%	8.7%	101.5%	97.6%	3.8%
Santa Fe	22.2%	17.4%	21.6%	31.9%	29.6%	7.2%
Sierra	56.0%	51.2%	8.6%	56.0%	51.2%	8.6%
Socorro	57.9%	51.4%	11.2%	78.2%	73.1%	6.5%
Taos	17.2%	12.8%	25.6%	20.2%	17.0%	15.8%
Torrance	60.6%	53.6%	11.6%	60.6%	54.7%	9.7%
Union	−8.2%	−11.6%	−41.5%	8.2%	11.6%	−41.5%
Valencia	48.0%	41.6%	13.3%	48.3%	42.7%	11.6%
New Mexico	34.7%	29.2%	15.9%	51.5%	48.3%	6.2%

[a](Constant – synthetic) / constant × 100
[b]Calculated using the absolute value of the MALPE

The seven age groups are: younger than age 10, 10–19, 20–34, 35–54, 55–64, 65–74, and 75 years and older.

Table 5.3 contains the level of bias by age group along with the average across age groups for the four alternatives and the three launch and horizon year

Table 5.3 Forecast bias (MALPE) by age group and alternative, Washington State counties

	Launch Year 1990 and Horizon Year 2000			
Age Group	Constant	Average	Trend	Synthetic
<10	−6.4%	16.8%	−28.3%	−3.3%
10–19	−13.3%	−10.0%	−13.4%	−8.4%
20–34	−10.5%	25.9%	−44.6%	−1.7%
35–54	−12.3%	−3.2%	−20.3%	−8.3%
55–64	−8.7%	0.1%	−19.3%	−6.9%
65–74	−7.0%	8.0%	−26.6%	−5.5%
75+	−5.9%	2.6%	−15.1%	−0.4%
Average	−9.2%	5.7%	−23.9%	−4.9%
	Launch Year 2000 and Horizon Year 2010			
Age Group	Constant	Average	Trend	Synthetic
<10	8.9%	6.0%	18.9%	3.0%
10–19	9.3%	1.9%	28.9%	3.6%
20–34	10.9%	5.0%	27.3%	1.7%
35–54	7.2%	0.7%	24.1%	2.4%
55–64	4.6%	0.0%	16.2%	4.0%
65–74	−2.4%	−5.9%	6.1%	1.0%
75+	−3.3%	−6.3%	4.2%	0.7%
Average	5.0%	0.2%	18.0%	2.4%
	Launch Year 1990 and Horizon Year 2010			
Age Group	Constant	Average	Trend	Synthetic
<10	−9.5%	46.8%	−53.1%	−4.9%
10–19	−10.8%	17.5%	−20.3%	−7.9%
20–34	−12.3%	40.2%	−60.9%	−6.6%
35–54	−18.1%	24.0%	−44.0%	−11.2%
55–64	−13.2%	−1.9%	−24.6%	−9.2%
65–74	−16.7%	8.3%	−50.4%	−10.6%
75+	−15.1%	7.9%	−38.0%	−3.6%
Average	−13.7%	20.4%	−41.6%	−7.7%

combinations. SYN has lower bias than TREND and CONST in every age group for all launch and target year combinations. SYN's bias over the age groups is roughly 50% lower than CONST and four to five times lower than TREND. SYN has lower bias than AVE in 14 of 21 comparisons (7 age groups and 3 launch and target years). SYN's average bias across age groups is substantially smaller than AVE in the 20-year horizon, slightly lower than AVE in the 10-year horizon using the 1990 launch year, and low (2.4%), but larger than AVE (0.2%) in the 10-year horizon using the 2000 launch year.

TREND shows the largest bias of any alternative in all but one comparison. In only one instance (ages 75 years and older, launch year 2000 and horizon year 2010) does TREND have a smaller (in absolute value) MALPE compared to AVE (4.2% vs −6.3%). The performance related to bias is less clear comparing CONST and AVE. AVE has lower bias than in CONST in 12 of 21 comparisons and a lower average bias across age groups in the 10-year horizons, but CONST has a decidedly lower average bias across age groups in the 20-year horizon (−13.7 vs 20.4). Also,

Table 5.4 Forecast accuracy (MAPE) by age group and alternative, Washington State counties

		Launch Year 1990 and Horizon Year 2000		
Age Group	Constant	Average	Trend	Synthetic
<10	10.3%	21.2%	29.2%	9.3%
10–19	14.1%	15.6%	21.6%	11.1%
20–34	13.9%	28.5%	45.5%	10.2%
35–54	12.9%	12.4%	23.2%	10.1%
55–64	9.4%	12.8%	20.4%	8.3%
65–74	8.0%	13.0%	26.6%	7.0%
75+	6.4%	9.0%	17.8%	4.1%
Average	10.7%	16.1%	26.3%	8.8%
		Launch Year 2000 and Horizon Year 2010		
Age Group	Constant	Average	Trend	Synthetic
<10	10.6%	9.5%	20.2%	7.6%
10–19	10.6%	6.9%	29.7%	6.9%
20–34	12.7%	9.4%	28.2%	8.2%
35–54	8.6%	5.2%	24.3%	5.2%
55–64	6.9%	5.8%	16.8%	6.7%
65–74	5.3%	7.3%	9.4%	4.5%
75+	5.6%	7.2%	8.5%	4.7%
Average	8.6%	7.3%	19.6%	6.3%
		Launch Year 1990 and Horizon Year 2010		
Age Group	Constant	Average	Trend	Synthetic
<10	18.6%	53.1%	57.8%	17.3%
10–19	18.1%	31.6%	50.2%	17.4%
20–34	19.6%	47.5%	64.9%	18.1%
35–54	21.0%	34.7%	53.7%	17.7%
55–64	16.0%	25.8%	37.2%	14.3%
65–74	18.5%	26.4%	50.4%	15.0%
75+	15.3%	19.9%	41.6%	8.3%
Average	18.1%	34.1%	50.8%	15.4%

AVE has more large outlying MALPEs than CONST; 25.9% in the 10-year horizon using the 1990 launch year and 46.8% and 40.2% in the 20-year horizon.

Table 5.4 contains the level of accuracy by age group along with the average across age groups for the four alternatives and the three launch and horizon year combinations. SYN has greater accuracy than TREND and CONST in every age group for all launch and horizon year combinations. SYN's average accuracy across age groups is roughly 20% lower than CONST and three times lower than TREND. SYN has a greater accuracy than AVE in 19 of 21 comparisons and its average accuracy across age groups is substantially smaller than AVE in the 20-year horizon and in the 10-year horizon using the 1990 launch year, and somewhat lower than AVE in the 10-year horizon using the 2000 launch year (6.3% vs 7.3%).

TREND shows the lowest accuracy of any alternative in all 21 comparisons. TREND's average across age groups range from 150% higher to 372% higher compared to the corresponding figures for the other alternatives. In general, CONST performs better than AVE regarding accuracy as compared to bias.

CONST has a greater accuracy than AVE in 15 of 21 comparisons, a lower average across age groups in the 10-year horizon using the 1990 launch year, and a decidedly lower average across age groups in the 20-year horizon (18.1% vs 34.1). AVE does show greater accuracy than CONST in the 10-year horizon using the 2000 launch year. AVG has smaller MAPEs in 6 of 7 age groups, (the sole exception is for the age group 65–74), and a smaller average across age groups (7.3% vs 8.6%).

We now turn to the last forecast error evaluation criterion, misallocation error across age groups. Table 5.5 contains the average of the IODs across counties for the four alternatives and the three launch and horizon year combinations. Under this criterion, SYN once again shows the smallest allocation errors of any alternative; although the differences with CONST are relatively small. CONST generally outperforms AVE; its average IODs are roughly half the size of those for AVE in two of three launch and horizon year combinations and its average IOD is only 0.1% larger than the AVE value for 2010 forecast using the 2000 launch year. TREND by far has the largest allocation errors of any alternative.

We conclude this section by looking at the relative performance of the alternatives in individual counties. We use a non-parametric approach that measures the percentage of counties where one alternative has a smaller absolute percent error (APE) or IOD than another, regardless of the magnitude of the difference. We compare SYN against CONST stratified by launch and horizon year combinations for age groups and the total population.[3]

SYN has a smaller APE than CONST in at least 61% of counties in all age groups and for the total population for each launch year and horizon year combination (See Fig. 5.3). SYN generally performs the best, according to this criterion, in the 2000 forecast using the 1990 launch year. The percentage of counties with smaller APEs for SYN ranges from 74.4% to 89.7% and 84.6% for the total population. For the 2010 forecast using the 1990 launch year, the percentages for age groups range from 69.2% to 82.1%, and 76.9% for the total population. For the 2010 forecast using the 2000 launch year, the percentages for age groups range from 61.5% to 82.1%, and 74.6% for the total population.

IODs for SYN are smaller than CONST in all launch and horizon year combinations, with percentages ranging from 59.0% of counties for the 2010 forecast

[3]To conserve space we do not present comparisons of SYN with AVE and SYN with TREND. To summarize these results, SYN had smaller APEs than AVE in more counties in 20 of the 21 combinations, with percentages in the age groups that ranged from 51.3% to 84.6%. For ages 10 to 19 in the 2010 forecast using the 2000 launch year, SYN had a lower MAPE in 48.7% of the counties. SYN also had smaller IODs than AVE in more counties, with percentages that ranged from 64.1% for the 2010 forecast using the 2000 launch year to 97.4% of counties for the 2000 forecast using the 1990 launch. SYN had smaller APEs than TREND in more counties in all age groups and launch year and horizon year combinations, with percentages that ranged from 69.2% to 100%. In terms of allocation error, SYN had a lower IOD than TREND ranged from 92.3% of the counties in both 10-year horizons and 100% in the 20-year horizon.

Table 5.5 Forecast allocation error across age groups by alternative, Washington State counties

		Mean index of dissimilarity			
Launch year	Horizon year	Constant	Average	Trend	Synthetic
1990	2000	2.6%	5.6%	7.2%	2.5%
2000	2010	2.3%	2.2%	4.0%	1.8%
1990	2010	3.7%	6.5%	15.5%	3.6%

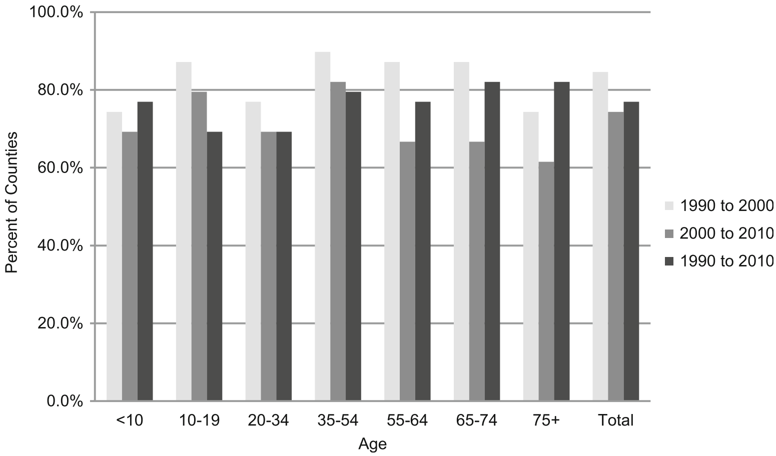

Fig. 5.3 Synthetic APE lower than constant APE by age group and total population by launch and horizon years, Washington State counties

using the 1990 launch year to 76.9% for the 2010 forecast using the 2000 launch year.

5.5.2.2 New Mexico Census Tracts

Turning to the analysis by age group for New Mexico census tracts, Table 5.6 contains the level of bias and accuracy by age group along with the average across age groups for the 2010 forecasts from SYN and CONST. Across all age groups, SYN has 13.1% less bias than CONST. SYN has lower bias than CONST in all age groups, but especially in age groups under 65 years of age. Bias is between 8.5% and 21.1% lower in SYN in these age groups, compared to around 2% lower for ages 65 and older.

A similar pattern is seen for forecast accuracy. Across all age groups, SYN has 4.7% greater accuracy than CONST. SYN has greater accuracy than CONST in all age groups, but especially in age groups less than 65 years of age. Accuracy is between 2.1% and 8.3% greater in these age groups, compared to less than 1% for ages 65 years and older. While SYN does lower bias and raise accuracy compared to CONST, SYN has a greater impact on lowering bias than on increasing accuracy.

Table 5.6 Forecast accuracy and bias by age group and alternative, New Mexico census tracts, 2010

	MALPE			MAPE		
Age group	Constant	Synthetic	Pct.[a,b] Diff.	Constant	Synthetic	Pct.[a,b] Diff.
<10	43.1%	37.4%	13.3%	65.8%	62.6%	4.8%
10–19	42.7%	33.7%	21.1%	63.4%	58.2%	8.3%
20–34	51.4%	41.7%	18.9%	73.0%	67.1%	8.1%
35–54	28.1%	22.8%	19.0%	46.3%	43.6%	5.9%
55–64	22.8%	20.9%	8.5%	42.2%	41.3%	2.1%
65–74	21.5%	21.1%	1.8%	40.9%	40.7%	0.5%
75+	42.6%	41.8%	1.9%	62.6%	62.0%	0.9%
Average	36.1%	31.3%	13.1%	56.3%	53.6%	4.7%

[a](Constant – synthetic) / constant × 100
[b]Calculated using the absolute value of the MALPE

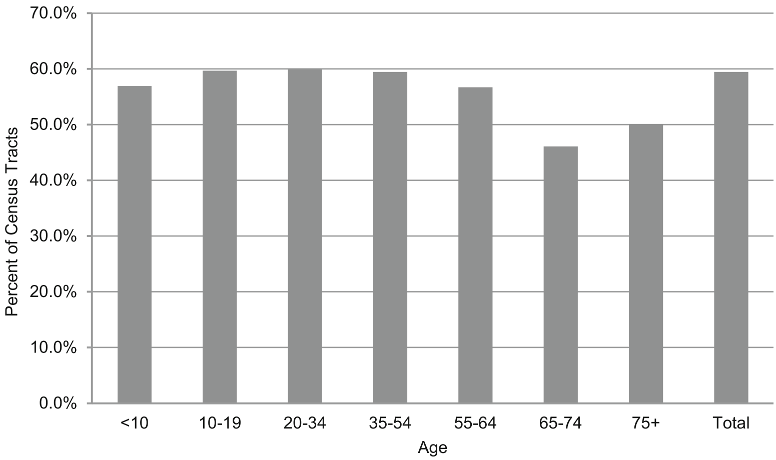

Fig. 5.4 Synthetic APE lower than constant APE by age group and total population, New Mexico census tracts, 2010

There is no difference in allocation error across age groups between SYN and CONST. The average IOD is 9.0% for both alternatives. The distribution of IODs by IOD size was virtually identical as well (data not shown).

We conclude this section by looking at the relative performance of SYN and CONST in individual census tracts using the same non-parametric technique used for Washington State counties. Figure 5.4 shows the percentage of census tracts where SYN has a smaller APE than CONST by age group and for the total population. The difference in forecast error between SYN and CONST is more muted using this criterion. In all age groups less than 65 years of age and for the total population, SYN has a smaller APE in more than 50% of the census tracts. For these age groups, the percentages are quite similar ranging from 56.7% to 59.9% and 59.4% for the total population. In age range of 65–74, SYN has a smaller APE in 46.1% of the census tracts. The percentage is 50.1% for ages 75 years and older showing almost no advantage for either SYN or CONST. Regarding allocation error; SYN has a smaller IOD than CONST in 54.1% of the census tracts.

5.5.3 *Total Population Forecast Error by Population Size and Growth Rate*[4]

5.5.3.1 Washington State Counties

In this section we analyze county total population forecast errors by county population size and growth rate for CONST and SYN. Size represents the population at the launch year and growth rate represents the percentage change over the decade prior to the launch year. We present detailed tables for only launch year 2000 and horizon year 2010, but discuss any major differences with the other two launch year and target year combinations. We begin by taking an aggregate look at the MALPE, MAPE, and IOD by four size categories and three growth rate categories (see Table 5.7). We use three categories for growth rate because it had much less variation compared to size. These results should be viewed with caution given the relatively small samples sizes in each category (9 or 10 for size and 10 to 15 for growth rate).

In terms of bias, the MALPE for SYN is lower than the MALPE for CONST in every size and growth rate category by an appreciable amount. For size, SYN has between 37.5% and 81.8% less bias and for growth rate SYN has between 44.0% and 97.1% less bias. These results are generally similar to those for the other launch and horizon year combinations. However in the forecast with a 20-year horizon, the MALPE for SYN is larger than the MALPE for CONST in the largest counties (4.2% vs −2.1%) and in the fastest growing counties (8.9% vs 2.1%).

A similar pattern by size and growth rate is seen for the MAPE, except the percentage differences are considerably smaller than those for the MALPE. In fact, for the slowest growing counties, the MAPEs for SYN and CONST are almost identical. These results are generally similar to those for the other launch and horizon year combinations with a few exceptions. In the forecast with a 20-year horizon, the MAPE for SYN is larger than the MAPE for CONST in the largest counties (13.0% vs 11.0%) and in the fastest growing counties (13.4% vs 10.6%). Similarly, in the 10-year forecast using the 1990 launch year, the SYN MAPE (6.5%) is a larger than the CONST MAPE (5.5%) in the largest counties and slightly larger in the fastest growing counties (6.9% vs 6.8%).

In terms of allocation error, the IODs for all size and growth rate categories are very low for both SYN and CONST. Except for counties with between 20,000 and

[4]We analyzed forecast errors by size and growth rate for each age group, but do not present these results to save space. The results for all groups were very similar to those for total population.

Table 5.7 Total population forecast bias and accuracy by size, growth rate, and alternative (launch year 2000 and horizon year 2010), Washington State counties

	MALPE			MAPE			Average IOD		
Size[a]	Constant	Synthetic	Pct. Diff.[b,c]	Constant	Synthetic	Pct. Diff.[b]	Constant	Synthetic	Pct. Diff[b]
< 20,000	8.8%	5.5%	37.5%	9.0%	6.6%	26.7%	3.0%	2.3%	32.4%
20,000–49,999	4.4%	0.8%	81.8%	8.4%	7.5%	10.7%	2.2%	2.3%	−4.5%
50,000–141,999	6.2%	2.4%	61.3%	7.4%	5.6%	24.3%	2.1%	1.6%	23.8%
142,000+	5.7%	1.3%	77.2%	5.7%	3.7%	35.1%	2.0%	1.3%	35.0%
Total	6.3%	2.4%	61.9%	7.6%	5.7%	25.0%	2.3%	1.8%	21.7%
	MALPE			MAPE			Average IOD		
Growth[a] rate	Constant	Synthetic	Pct. Diff.[b,c]	Constant	Synthetic	Pct. Diff[b]	Constant	Synthetic	Pct. Diff[b]
< 15%	3.4%	0.1%	97.1%	5.3%	5.2%	1.9%	2.9%	2.5%	10.7%
15%–24.9%	5.6%	1.6%	71.4%	5.6%	3.4%	39.3%	1.9%	1.4%	26.3%
25 + %	9.1%	5.1%	44.0%	11.5%	8.9%	22.6%	2.5%	1.9%	25.0%
Total	6.3%	2.4%	61.9%	7.6%	5.7%	25.0%	2.3%	1.8%	21.7%

[a]Size is the 2000 population and growth rate is the 1990–2000 percent change
[b](Constant – synthetic) / constant × 100
[c]Calculated using the absolute value of the MALPE

Table 5.8 Absolute percent errors for total population by size and growth rate, synthetic alternative (launch year 2000 and horizon year 2010), Washington State counties

	2000 Population				
Size	< 20,000	20,000–49,999	50,000–141,999	142,000+	Total
Synthetic lower[a]	90.0%	70.0%	66.7%	70.0%	74.4%
Odds ratio[b]	3.857	1.000	0.857	n/a	
Sample size	10	10	9	10	39
Chi square	1.762	p = 0.623			
Kendal's Tau-c	−0.150				

	1990–2000 Growth rate			
Growth rate	< 15%	15% to 24.9%	25 + %	Total
Synthetic lower[a]	50.0%	80.0%	85.7%	74.4%
Odds ratio[c]	0.167	0.667	n/a	
Sample size	10	15	14	39
Chi square	4.309	p = 0.116		
Kendal's Tau-c	0.281			

[a]Compared to the absolute percent error from the constant model
[b]142,000+ is the reference group
[c]25 + % is the reference group

50,000 people, the IOD for SYN is lower than the IOD for CONST in all other size and growth rate categories. These percentage differences range between 23.8% and 35.0% for size and between 10.7% and 25.0% for growth rate. In counties with between 20,000 and 50,000 people, the IOD for SYN is 0.1 of a percentage point larger than the IOD for CONST (2.6% vs 2.5%). These results are nearly identical to those for the other launch and horizon year combinations.

We now look at individual counties where the APEs for SYN are lower than the APEs for CONST by the size and growth rate of the counties (see Table 5.8). There is no discernable relationship between population size and a lower SYN APE. In all size groups, the APE for SYN is lower in over 60% of the counties. These figures are very close for populations above 20,000 (66.7% to 70.0%), and SYN most outperforms CONST in the smallest counties (90.0%). The relationship between a lower SYN APE with size is not statistically significant and weak (Kendal's Tau-c = −0.150).

The relationship between population size and a lower SYN APE is similar in the other combinations using the 1990 launch year, and quite different from these results. The relationship with size is now statistically significant, moderate in strength (Kendal's Tau-c of −0.379 and −0.396), and negative in direction. As growth rates increase, the percent of counties where SYN has greater accuracy decreases. For example, in the forecast with a 20-year horizon, SYN has a lower APE than CONST in 91.7% of the smallest counties. That percentage decreases to 37.5% in the largest counties.

There is a stronger and positive relationship between growth rate and a lower SYN APE. For counties in the lowest growth rate category, SYN is lower than

CONST just as often as CONST is lower than SYN. As growth rates increase, SYN has a lower APE in a larger percent of the counties. While this relationship is still not significant, the Kendal's Tau-c = 0.281 is almost double the figure for size. The relationship between population growth rate and a lower SYN is similar in the combinations using the 1990 launch year, and quite different from these results. The relationship with growth rate is now statistically significant, considerably larger in strength (Kendal's Tau-c's of −0.365 and −0.570), and negative in direction. As growth rates increase the percent of counties where SYN has a lower APE than CONST decreases. For example, in the forecast with a 20-year horizon, SYN has a lower APE in 100% of the counties with the lowest growth rate. That percentage decreases to 30.0% in the fastest growing counties.

5.5.3.2 New Mexico Census Tracts

We now turn to the analysis of total population forecast errors for SYN and CONST by population size and population growth rate for New Mexico census tracts. The number of census tracts enabled a more detailed breakdown of size and growth rate into five groupings for these characteristics. Table 5.9 provides an aggregate look at the MALPE, MAPE, and IOD by size and growth rate.

In terms of bias, the MALPE for SYN is between 9.5% and 31.2% lower than the MALPE for CONST in every size category. A different pattern related to bias occurs for growth rate. For census tracts that grow by more than 10%, SYN has a lower MALPE than CONST by between 7.2% and 24.8%. The MALPEs for SYN and CONST are −34.0% and −31.4%, respectively, in declining census tracts. In the most stable census tracts (growth rate between −10% to 9.9%), the MALPEs for SYN and CONST are −5.8% vs −1.9%, respectively.

Similar patterns by size and growth rate are seen for the MAPE, except the percentage differences are generally smaller than those for the MALPE. For size, the SYN MAPEs are in a tight range and lower than the CONST MAPEs by between 4.2% and 6.8%. For growth rate, CONST has greater accuracy than SYN for the declining and relatively stable census tracts, with the MAPEs for SYN around 6.0% higher. For census tracts that grow by 10% or more, SYN has greater accuracy than CONST, with MAPEs between 6.9% and 12.7% lower.

In terms of allocation error, there is not much difference between SYN and CONST across size and growth rate categories. For size, the IODs for SYN and CONST are identical for census tracts with 3,000 or more people. For smaller census tracts, IODs for SYN are smaller by only trivial amounts. IODs for census tracts with between 2,000 and 2,999 people differ by 0.2 of a percentage point and for census tracts with less than 2000 people they differ by 0.1 of a percentage point. For growth rate, IODs for CONST are smaller than IODs for SYN in declining and stable census tracts. For the faster growing census tracts (growth rates of 10% or more), SYN has slightly less allocation error. But the largest difference in IOD

Table 5.9 Total population forecast bias and accuracy by size, growth rate, and alternative, New Mexico census tracts[a], 2010

	MALPE			MAPE			Average IOD		
Size	Constant	Synthetic	Pct. Diff.[b,c]	Constant	Synthetic	Pct. Diff[b]	Constant	Synthetic	Pct. Diff[b]
<2,000	16.6%	11.7%	29.5%	49.5%	46.8%	5.5%	11.3%	11.1%	1.8%
2,000–2,999	20.2%	15.3%	23.9%	40.1%	38.0%	5.3%	10.1%	10.0%	1.0%
3,000–3,999	15.2%	10.4%	31.2%	34.4%	33.0%	4.2%	8.3%	8.3%	0.0%
4,000–4,999	38.3%	32.7%	14.7%	48.9%	45.7%	6.5%	7.5%	7.5%	0.0%
5,000+	78.9%	71.4%	9.5%	84.6%	78.9%	6.8%	8.8%	8.8%	0.0%
Total	34.7%	29.2%	16.0%	51.3%	48.3%	5.9%	9.0%	9.0%	0.0%

	MALPE			MAPE			Average IOD		
Growth rate	Constant	Synthetic	Pct. Diff[b,c]	Constant	Synthetic	Pct. Diff[b]	Constant	Synthetic	Pct. Diff[b]
< −10%	−31.4%	−34.0%	−8.1%	34.1%	36.3%	−6.3%	12.0%	12.2%	−1.7%
−10% to 9.9%	−1.9%	−5.8%	−201.5%	10.2%	10.9%	−6.6%	5.9%	6.0%	−1.7%
10% to 49.9%	19.0%	14.3%	24.8%	25.8%	22.5%	12.7%	7.6%	7.5%	1.3%
50% to 99.9%	71.1%	63.8%	10.2%	76.2%	69.8%	8.4%	10.8%	10.6%	1.9%
100.0%+	157.1%	145.9%	7.2%	157.8%	146.9%	6.9%	11.9%	11.7%	1.7%
Total	34.7%	29.2%	16.0%	51.3%	48.3%	5.9%	9.0%	9.0%	0.0%

[a]Size is the 2000 population and growth rate is the 1990–2000 percent change
[b](Constant – synthetic) / constant × 100
[c]Calculated using the absolute value of the MALPE

Table 5.10 Absolute percent errors for total population by size and growth rate, synthetic alternative, New Mexico census tracts, 2010

	2000 Population					
Size	< 2,000	2,000–2,999	3,000–3,999	4,000–4,999	5,000+	Total
Synthetic lower[a]	55.9%	52.9%	49.6%	64.1%	75.7%	59.4%
Odds ratio[b]	0.407	0.361	0.315	0.573	n/a	
Sample size	59	102	125	78	107	471
Chi square	19.548	$p = 0.001$				
Kendal's Tau-c	0.177					

	1990–2000 Growth rate					
Growth rate	< −10%	−10% to 9.9%	10% to 49.9%	50% to 99.9%	100.0%+	Total
Synthetic lower[a]	5.2%	34.7%	79.0%	89.0%	95.9%	59.4%
Odds ratio[c]	0.002	0.023	0.162	0.348	n/a	
Sample size	77	124	124	73	73	471
Chi square	212.036	$P < 0.001$				
Kendal's Tau-c	0.717					

[a]Compared to the absolute percent error from the constant model
[b]5000+ is the reference group
[c]100.0% + is the reference group

between SYN and CONST across growth rate categories is only 0.2 of a percentage point.

In looking at individual census tracts, we find that the APEs for SYN are lower than the APEs for CONST by the size and growth rate (see Table 5.10). There is a weak positive and statistically significant relationship between population size and a lower SYN APE (Kendal's Tau-c of 0.177). The percent of census tracts with lower SYN APEs is relatively close for census tracts with less than 4,000 people, ranging from 49.6% to 55.9%. SYN does outperform CONST in census tracts with 4,000 or more people. SYN has a lower APE in 64.1% of census tracts with between 4,000 and 5,000 persons and 75.7% of the census tracts with more than 5,000 persons.

Compared to size, there is a much stronger and statistically significant positive relationship between growth rate and a lower SYN APE (Kendal's Tau-c of 0.717). For declining areas, SYN is lower than CONST in only 5.2% of census tracts, and for stable areas SYN is lower than CONST in 34.7% of census tracts. As growth rates increase, SYN has a lower APE in a larger percentage of the census tracts. For census tracts that grow from between 10% and 50%, SYN has a lower APE in 79% of them. The percentage steadily increases, reaching 95.9% in census tracts that more than doubled in size.

5.6 Conclusions

Assessments of the H-P method have been based on the basic H-P framework in which CCRs developed over the base period and CWRs developed for the launch year are held constant over the forecast horizon. In this chapter, we evaluated several alternatives to modifying CCRs and CWRs and compared the errors from those forecasts to errors from forecasts using the basic H-P framework. These alternatives included: (1) averaging; (2) trending; and (3) and a synthetic method that based local CCRs and CWRs changes on state-level changes in the corresponding CCRs and CWRs. We evaluated three dimensions of forecast error (accuracy, bias, and allocation error) from forecasts created for counties in Washington State (10-year and 20-year horizons) and census tracts in New Mexico (10-year horizon). Forecast errors were computed by comparing the simulated forecasts to results from the 2000 and 2010 censuses.

How did the basic H-P framework (CONST) stack up against the alternatives (AVE, TREND, and SYN)? The short answer is very well against AVE and TREND in Washington State counties. Forecasts from CONST were almost universally better (lower error) than forecasts (total population and population by age) from TREND for all forecast launch year and target year combinations, and generally better than forecasts from AVE. AVE forecasts had lower forecast errors than CONST in the one of the forecasts with a 10-year horizon (launch year 2000 and horizon year 2010), but for the 20-year horizon AVE had much larger forecast errors than CONST. Incorporating historical information for 20-years for CCRs and 10-years for the CWRs did not lower forecast errors, but in fact increased them compared to the basic H-P framework.

Incorporating forecast information from a larger geographic area outperformed the basic H-P framework for Washington State counties. County forecasts from SYN had less bias, greater accuracy, and less allocation error than forecasts than CONST. This finding was very pervasive. It held for total population and population by age group for all launch and horizon year combinations based on an aggregate analysis of MALPEs, MAPEs, and IODs and analysis of the relative sizes of the APEs and IODs in individual counties. The advantage of SYN over CONST was greater in forecasts with a 10-year horizon than it was in the forecast with a 20-year horizon. Total population forecasts from SYN were also better than forecasts from CONST across virtually all size and growth rate categories. In counties with 20,000–50,000 persons, however, CONST's average IOD (2.2%) was marginally lower than SYN's average IOD (2.3%). These results by size and growth rate were based on rather small samples and should be viewed with caution.

Because of limited historical data with comparable boundaries, we were able to evaluate only SYN and CONST for census tracts. Like counties, census tract forecasts based on SYN had less bias, greater accuracy, and less allocation error than forecasts based on CONST, but the advantages of SYN were less dramatic and not quite as universal for census tracts. For example in the 2010 county total population forecast using a 2000 launch year, SYN had a lower MALPE and

MAPE than CONST by 62% and 25%, respectively. For the total population in census tracts, SYN had 18.8% less bias and 6.2% greater accuracy. While SYN had less bias and greater accuracy than CONST in all age groups, SYN's advantage was least in age groups 65 years and older. Forecasts from SYN were also better than forecasts from CONST across size and growth rate categories with a few exceptions. For census tracts with 3,000 or more people, there was no difference in averages IOD between SYN and CONST. The most noticeable exception was in declining or stable census tracts where CONST has less bias and allocation error and greater accuracy than SYN.

This analysis has shown that a synthetic approach is a viable alternative to the basic H-P framework in Washington State counties and New Mexico census tracts. We have shown that applying methods that altered CCRs and CWRs based on their history was not an effective strategy in Washington State counties. There is more to be gained by applying a global (the same) adjustment covering the horizon being forecast rather than basing adjustments on county-specific historical changes. Even though we did not test averaging or trending of CCRs and CRWs in the census tract forecasts, we believe these alternatives would not be useful at this level of geography. Assembling the necessary data would be daunting and costly, and historical variations in CCRs and CWRs for individual census tracts would be much more volatile and inconsistent than the county-level ratios analyzed here. Perhaps, averaging and trending CCRs and CWRs might be more viable strategies for states and other geographic areas larger than counties.

Although not as apparent in the county forecasts, the census tract forecasts suggest the basic H-P framework may produce forecasts with less error than a synthetic method in declining or stable census tracts. More study is needed to determine the generality of this finding. Perhaps this finding was not as apparent in Washington State counties because of a lack of variation in growth rates, which were almost always positive. It could also be an artifact of the way we applied the synthetic adjustment. We used state-level forecasts for both the counties and census tracts. This is how a synthetic forecast for counties would likely be implemented, but application of state-level changes to census tracts throughout the state might ignore important substate variation in CCRs and CWRs over the forecast horizon, affecting the performance of the synthetic approach.

Our results are conclusive enough to recommend using the synthetic approach when implementing the H-P forecasting method. In most forecasting situations a forecast for the higher level of geography would be available, so the synthetic approach would be a low cost alternative to the basic H-P framework. These findings are also conclusive enough warrant additional research into the efficacy of the synthetic alternative. Evaluating counties with more varied size and growth rate characteristics might shed additional light on the performance of the synthetic alternative, especially in declining or stable counties or in counties with small populations. It would be useful to examine whether more geographically-specific CCR and CWR forecasts as global adjustments is a better synthetic strategy for census tracts. Perhaps, the synthetic approach may have an even greater advantage over the basic H-P framework using such a strategy. Finally, we only examined uncontrolled forecasts. Would forecast errors and their patterns change for the synthetic alternative and the basic H-P framework, if the forecasts by age were adjusted to the total population for each geographic area?

References

Duncan, O. D., Cuzzort, R. P., & Duncan, B. (1961). *Statistical geography: Problems in analyzing area data*. Glencoe: Free Press.

Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces, 67*, 281–315.

Office of Financial Management. (2002). *Washington State county population projections for growth management by age and sex 2000–2025*. Olympia: OFM Population Division.

Rayer, S. (2007). Population forecast accuracy: Does the choice of summary measure of error matter? *Population Research and Policy Review, 26*, 163–184.

Smith, S. K., & Rayer, S. (2012). *Projections of Florida population by county, 2011–2040*, *Florida Population Studies, Bulletin 162*. Gainesville: Bureau of Economic and Business Research, University of Florida.

Smith, S. K., & Tayman, J. (2003). An evaluation of population projections by age. *Demography, 40*(4), 741–757.

Smith, S. K., Tayman, J., & Swanson, D. (2013). *A practitioner's guide to state and local population projections*. Dordrecht: Springer.

Swanson, D. (2015). On the relationship among values of the same summary measure of error when it is used across multiple characteristics at the same point in time: An examination of MALPE and MAPE. *Review of Economics and Finance, 5*(1), 1–14.

Swanson, D., & Tayman, J. (2012). *Subnational population estimates*. Dordrecht: Springer.

Swanson, D., Schlottmann, A., & Schmidt, R. (2010). Forecasting the population of census tracts by age and sex: An example of the Hamilton–Perry method in action. *Population Research and Policy Review, 29*, 47–63.

Swanson, D., Tayman, J., & Bryan, T. (2011). MAPE-R: A rescaled measure of accuracy for cross-sectional, sub-national forecasts. *Journal of Population Research, 28*, 225–243.

Swanson, D., Tayman, J., McKibben, J., & Cropper, M. (2012). *A "blind" ex post facto evaluation of total population and total household forecast for small areas made by five vendors for 2010: Results by geography and error criteria*. Paper presented at the 2012 Conference of the Canadian Population Society, Waterloo, Ontario, Canada, 30 May – 1, June.

Chapter 6
Forecasting Uncertainty

6.1 Introduction

This chapter explores the idea of creating statistical intervals for population forecasts based on stochastic forecasts of the cohort-change ratios (CCR). We provide an overview of the three approaches that have been used to assess population forecast uncertainty, judgment and personal opinion, a range of forecasts based on alternative scenarios, and statistical forecast intervals. This chapter focuses on the latter approach. We describe and evaluate a method for developing statistical intervals around population forecasts by age and for the total population. The method combines regression modeling of the cohort change ratios used in the Hamilton-Perry (H-P) method. The evaluation of state-level forecasts shows the intervals are neither so wide to be meaningless nor too narrow to be overly restrictive and that, overall, the percent of the forecasts contained within the intervals is consistent with the uncertainly level of the intervals. We make some observations regarding the limitations of this approach to measuring forecast uncertainty, and conclude with suggestions for further work.

6.2 Forecast Uncertainty

Although they are widely used, population forecasts entail an amount of uncertainty, especially for long time horizons and for places with small or rapidly changing populations (Alho 1984; Alho and Spencer 1985, 1990, 1997, 2005; Lutz et al. 1999; Smith et al. 2013: 365; Tayman et al. 2007, 2011; Wilson 2012). As such, virtually every forecast is wrong, making the task of an accurate forecast impossible, but the task is unavoidable (Keyfitz 1987: 236). It is impossible in that the forecasted numbers turn out to be different from what actually occurs, but

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_6

unavoidable in that forecasts must be done in the modern world. Swanson and Tayman (1995) describe this irony as the "rock" and the "hard place."

Demographers have developed several strategies for dealing with the "irony" of forecasting. They include the use of the term "projection" rather than "forecast," (Keyfitz 1972; Pittenger 1978; Smith and Bayya 1992; Smith et al. 2013: 323), "normative" forecasting (Moen 1984), and providing measures of forecast uncertainty. One way to assess forecast uncertainty is to use judgment and methods based on judgment (Linstone and Turoff 1975, Ševčíková et al. 2013). A second way is to produce several alternative forecasts or scenarios based on different sets of assumptions (Campbell 1996; Cheeseman-Day 1992; Spencer 1989; Tayman 2011; Thompson and Whelpton 1933). A third approach is to develop statistical forecast intervals (Alho and Spencer 2005; Rayer et al. 2009; Stoto 1983; Swanson and Beck 1994).

6.3 Statistical Forecast Intervals

Forecast intervals based on statistical theory and data on error distributions provide an explicit estimate of the probability that a given range will contain the future population. These intervals are sometimes called prediction intervals, probability intervals, confidence intervals, or confidence limits. We call them forecast intervals to distinguish them from traditional confidence intervals, which—strictly speaking—apply only to sample data.

Two approaches have been used to develop statistical forecast intervals. The first is based on the development of statistical (or stochastic) models of population growth, and the second is based on empirical analyses of errors from past population forecasts. Both rely on the assumption that historical or simulated error distributions can be used to predict future error distributions. To a large extent, the two approaches complement one another, but neither is fully satisfactory. On the one hand, model-based intervals tend to exploit the theories and underlying inferential statistics, but can fall short in utilizing the information available in historical data. On the other hand, empirically-based intervals must utilize information from historical data and forecasts, but fall short in exploiting the theories underlying inferential statistics. Our method for developing statistical intervals around population forecasts uses a model-based approach enhanced with information in historical data, a feature found in the empirically-based approach.

6.3.1 Model-Based Intervals

Model-based forecast intervals capitalize on the stochastic (or random) nature of population processes and offer one important benefit: they provide explicit probability statements to accompany point forecasts and provide consistency among demographic trajectories and parameters. These intervals often exceed the low and high projections produced using alternative scenarios (McNees 1992). Given that many data users (and producers) tend to overestimate the accuracy of population forecasts, model-based probability intervals can provide an important reality check. However, model-based forecast intervals are valid only to the extent that the assumptions underlying the models are valid. In spite of their objective appearance, they are strongly influenced by the analyst's judgment. The models themselves are often complex and require a substantial amount of base data. They are subject to errors in the base data, errors in specifying the model, errors in estimating the model's parameters, and future structural changes invalidating the model's parameter estimates (Lee 1992). In addition, it is the case that many alternative forecasting models can be specified, each providing different (perhaps dramatically different) forecast intervals (Cohen 1986; Lee 1974; Tayman et al. 2007).

Model-based intervals can be developed in a number of ways. Past applications have included maximum likelihood estimators of population growth rates (Cohen 1986), Monte Carlo simulations of fertility and migration rates (Pflaumer 1988), simulations incorporating uncertainty from other methods (Wilson and Terblanche 2017); regression-based forecasting models (Swanson and Beck 1994), Bayesian forecasting models (Alkema et al. 2011; Raftery et al. 2013), models based on the opinions of a group of experts (Lutz et al. 1999; San Diego County Water Authority 2002), and time series models covering mortality rates (Lee and Carter 1992), life expectancy (Torri and Vaupel 2012), fertility rates (Lee 1993), net migration (De Beer 1993), and total population size (Alders et al. 2007; Hyndman and Booth 2008). Although much of the research on model-based intervals has focused on national or regional forecasts, research has extended the analysis to subnational forecasts as well (Cameron and Poot 2011; Tayman et al. 2007; Wilson and Bell 2004). Providing a detailed description of model-based forecast intervals is beyond the scope of this chapter, but we can give several examples of the intervals produced by these models and compare them to the high and low projection series produced using alternative scenarios.

Lee and Tuljapurkar (1994) forecast a population of 398 million for the United States in 2065, with a 95% forecast interval of 259–609 million. This range is considerably wider than the spread between the low and high projections produced by the Census Bureau at about the same time; those projections ranged from 276–507 million in 2050, with a medium projection of 383 million (Cheeseman-Day 1992). The previous set of Census Bureau projections reported much lower numbers and a slightly smaller range, with a medium projection of 300 million and a range of 230–414 million for 2050 (Spencer 1989).

Pflaumer (1992) made two time series forecasts of the U.S. population: one based on population size and the other based on the natural logarithm of population size. The first model produced a medium forecast of 402 million in 2050, with a 95% forecast interval of 277–527 million. These numbers are similar to the Census Bureau's projections from the same time. The second model produced a medium forecast of 557 million, with a 95% forecast interval of 465–666 million. These numbers are much higher and provide a narrower range than the Census Bureau's projections.

McNown et al. (1995) made time series forecasts of the components of growth for the U.S. population, as well as total population size. For 2050, they forecasted a total population of 373 million; with a 95% forecast interval ranging from 243 million to 736 million. The total fertility rate was forecasted to be 2.46 in 2050, with a 95% forecast interval ranging from 0.91 to 5.53. Life expectancy at birth for males was forecasted to be 75.5, with a 95% forecast interval ranging from 68.5 to 82.8. For fertility, these intervals are much larger than those found in the Census Bureau projections, which assumed that the total fertility rate would range only from 1.83 to 2.52 in 2050 (Cheeseman-Day 1992). For mortality, the interval widths are not much different than those reported by the Census Bureau, in which life expectancy at birth was projected to range between 75.3 and 87.6 in 2050.

Swanson and Beck (1994) developed a regression-based model for making short-term county population forecasts in the state of Washington. They compared the 2/3 forecast intervals associated with this model to census counts of Washington's 39 counties in 1970, 1980, and 1990. They found the forecast intervals to contain the 1970 census count in 30 counties (77%), the 1980 census count in 24 counties (62%), and the 1990 census count in 31 counties (79%). These results suggest that Swanson and Beck's 2/3 forecast intervals provided a reasonably accurate view of forecast uncertainty.

6.3.2 Empirically-Based Intervals

The second type of forecast interval is based on empirical analyses of errors from past forecasts rather than on an explicit stochastic model (Keyfitz 1981; Smith and Sincich 1988; Stoto 1983; Smith and Rayer 2012; Tayman et al. 1998). The empirical approach has some advantages over models that incorporate the stochastic nature of change and may generally be more useful for small areas. This approach is much less complex and within the capabilities of most agencies preparing forecasts. The problems implementing stochastic models for small areas are even more difficult because of the lack of time-series data and the lower reliability of rates and statistical parameters based on relatively small areal sizes. However, empirically-based probability intervals require past forecasts whose availability/usability may be an issue.

Keyfitz (1981) took approximately 1,100 national forecasts made between 1939 and 1968 and calculated the difference between the forecast annual growth rate and the rate actually occurring over time. He found this difference to be largely independent of the length of horizon over which the forecasts were made. He calculated the RMSE for the entire sample to be approximately 0.4% points and developed 2/3 forecast intervals by applying that error to the growth rates forecasted for each country. For example, if a country were forecast to grow by 2% per year for the next 20 years, the probability would be approximately 2/3 that the actual growth rate would be somewhere between 1.6 and 2.4%. He refined his analysis and found RMSE of 0.60 for rapidly growing countries, 0.48 for moderately growing countries, and 0.29 for slowly growing countries. He applied the 0.29 RMSE to U.S. growth rate of 0.79% per year projected by the Census Bureau, yielding a range of 245–275 million in 2000. He concluded that the odds were about 2 to 1 that this range would contain the U.S. population in that year.

Stoto (1983) followed a similar approach, but analyzed forecasts containing more temporal and geographic diversity. He differentiated between two components of error, one related to the launch year of the forecast and the other to seemingly random events (the residual). For more developed countries, he found the launch-year component to have a distribution that was stable over time and centered on zero, implying that the forecasts were unbiased. For less developed countries, he found the variance of the launch-year component to be stable, but that earlier sets of forecasts had a strong downward bias (although recent sets had little bias). The second component (the residual) was found to have a stable distribution but with occasional outliers. For both components, the variance was larger for less developed countries than more developed countries. He calculated the standard deviations for these two components of error and constructed forecast intervals of the U.S. population and estimated that there was about a 2/3 probability that an interval of 241–280 million would contain the actual population in 2000, and a 95% probability that an interval of 224–302 million would contain the population. He compared his results to projections produced by the Census Bureau, concluding that the Census Bureau's low and high series were very similar to a 2/3 forecast interval.

Smith and Sincich (1988) also used the distribution of past forecast errors to construct forecast intervals, but followed a different approach. They modified a technique developed by Williams and Goodman (1971), in which the predicted distribution of future forecast errors was based directly on the distribution of past forecast errors. An important characteristic of this technique is that it can accommodate any error distribution, including the asymmetric and truncated distributions typically found for absolute percent errors.

Using population data for states from 1900 to 1980, Smith and Sincich (1988) used four simple extrapolation methods to make a series of forecasts covering 10- and 20-year horizons. They calculated absolute percent errors for each target year by comparing forecasts with census counts, focusing on the 90% intervals for each set of forecasts (i.e., the absolute percent error larger than exactly 90% of all absolute percent errors). They investigated two approaches to constructing 90% forecast intervals, one using the 90% interval from the previous set of forecasts and the other using the 90% interval from all other sets of forecasts. They found both approaches to provide relatively accurate forecast intervals. For most individual target years, 88–94% of state forecast errors fell within the forecasted 90% interval. Summing over all target years, 91% of all forecast errors fell within the forecasted 90% interval. They concluded that stability in the distribution of absolute percent errors over time made it possible to construct useful forecast intervals for state forecasts.

Rayer, et al. (2009) constructed and tested forecast intervals for a large sample of counties in the U.S. using the Williams and Goodman (1971) approach. They constructed county forecasts covering 10-, 20-, and 30-year horizons and calculated forecast errors for target years covering decades from 1900 to 2000. Although the center of the error distributions shifted considerably from one decade to the next, their shape remained relatively constant over time. They evaluated the performance of 90% forecast intervals based on the distribution of absolute percent errors and found over all decades errors for 91% of the counties fell within the forecast intervals for all three horizons. Although there was some decade to decade variation, the proportion of errors falling within the intervals was usually between 88% and 93% and never varied by more than 10% points.

Smith and Rayer (2012) also constructed and tested forecast intervals for counties in Florida. Using forecast errors for target years 1985, 1990, and 1995, they constructed 2/3 forecast intervals for forecasts with launch years 1995, 2000, and 2005 and counted the number of counties in which the subsequent population counts or estimates fell within the forecast intervals. They found that 43 counties (64%) fell within the forecasted range for 5-year horizons and 49 counties (73%) for both 10- and 15-year horizons. These numbers were fairly close to the 45 counties implied by the forecast intervals. Given the year-to-year volatility of Florida's population growth, this reflects a reasonably good forecasting performance.

Tayman et al. (1998) developed statistically-based forecast intervals for subcounty population forecasts in San Diego County. They started by forecasting the population residing in 2000 ft. by 2000 ft. grid cells. These forecasts had 1980 as a launch year and 1990 as a target year. Using repeated sampling techniques and randomly selected grid cells, they developed forecasts for a large number of areas varying in size from 500 to 50,000. Forecast errors were calculated by comparing the 1990 forecasts with 1990 census counts. Rather than constructing forecast intervals for the population forecasts per se, they developed forecast intervals for the mean errors implied by those forecasts. Empirical forecast intervals for MAPEs and MALPEs were developed using an approach similar to that used by Williams and Goodman (1971) and Smith and Sincich (1988). For areas with 500 persons, they found a 95% forecast interval of 67.4%–80.3% for the MAPE. For areas with 50,000 or more, the interval was 9.7%–11.5%. For MALPE, the intervals were wider, but centered closer to zero.

6.4 Statistical Intervals for Cohort Change Ratios and Population Forecasts

6.4.1 *Statistical Inference and the Concept of a Super-Population*

The approach underlying the discussion of uncertainty found in this chapter is based on the concept of a super-population rather than a random sample. This concept can be traced at least back to 1941 in a paper entitled "On the interpretation of censuses as samples" (Deming and Stephan 1941). Although the concept of a super-population has been refined (Graubard and Korn 2002), the definition provided by Deming and Stephan (1941: 48) remains relevant:

> Even a complete census, for scientific generalizations, describes a population that is but one of the infinity of populations that will result by chance from the same underlying social and economic cause systems. The infinity of populations may itself be thought of as a population, and might possibly be called a super-population. A sample inquiry is then only a sample of a sample, a so-called 100 percent sample is simply a larger sample, but still only a sample. In order to study the underlying cause systems, it is necessary to study several members of this infinity of populations...

Not surprisingly, the idea of a super-population gained ground since 1941, largely due to the increased use of samples and other data to guide decision making. It has found a home in wildlife studies and other areas of research where sampling is widely used (because conducting a complete enumeration is either too costly or simply not feasible), but without the benefit of a sample frame. For example, a super-population represents the theoretical foundation underlying the statistical inference applied to capture/recapture studies used to estimate the size of a finite population. This type of study is also applied to evaluations of census accuracy and other forms of estimates involving human populations, where it is known, among other names, as "dual system estimation" (Andridge and Little 2010, Brown et al. 2011, and Wolter 1986).

6.4.2 *Hamilton-Perry Method*

As described in Chapter 4, the H-P method moves a population by age from time t to time $t + k$ using cohort change ratios (CCR) computed from data in the two most recent censuses:

$$ {}_{n}\mathrm{CCR}_{x,t} = {}_{n}\mathrm{P}_{x,t} / {}_{n}\mathrm{P}_{x-k,t-k.} \tag{6.1}$$

where,

${}_{n}P_{x,\,\mathrm{t}}$ is the population aged x to $x + n$ at the most recent census (t),

$_nP_{x\text{-}k,\ t-k}$ is the population aged $x - k$ to $x - k + n$ at the 2nd most recent census $(t - k)$, and

k is the number of years between the most recent census at time t and the one preceding it at time $t - k$.

The formula for moving the age cohorts of a population into the future is then:

$$_{\mathrm{n}}\mathrm{P}_{\mathrm{x+k,t+k}} = {}_{n}\mathrm{CCR}_{\mathrm{x,t}} \times {}_{\mathrm{n}}\mathrm{P}_{\mathrm{x,t}} \tag{6.2}$$

where,

$_nP_{x+k,\ t+\mathrm{k}}$ is the population aged $x + k$ to $x + k + n$ at time $t + k$, and
$_nCCR_{x,\ t}$ and $_nP_{x,\ \mathrm{t}}$ are as defined in Eq. 6.1.

Given the nature of the CCRs, 10–14 is the youngest 5-year age group for which forecasts can be made if there are 10 years between censuses. To forecast the populations aged 0–4 and 5–9 one can use the Child-Woman Ratio or more generally as a Child-Adult Ratio as previously discussed in Chapters 1 and 4. Another way to forecast the youngest age groups is to take their ratios (R) at two points in time and apply that ratio to the launch year age group. In the first step, the ratios are:

$$\text{Population } 0\text{–}4 : {}_5\mathrm{R}_{0,\mathrm{t}} = {}_5\mathrm{P}_{0,\mathrm{t}}/{}_5\mathrm{P}_{0,\mathrm{t-k}} \tag{6.3}$$

$$\text{Population } 5\text{–}9 : {}_5\mathrm{R}_{5,\mathrm{t}} = {}_5\mathrm{R}_{5,\mathrm{t}}/{}_5\mathrm{P}_{5,\mathrm{t-k}} \tag{6.4}$$

In the second step, the forecast population at $t + k$ is found by:

$$\text{Population } 0\text{–}4 : {}_5\mathrm{P}_{0,\mathrm{t+k}} = {}_5\mathrm{P}_{0,\mathrm{t}} \times {}_5\mathrm{R}_{0,\mathrm{t}} \tag{6.5}$$

$$\text{Population } 5\text{–}9 : {}_5\mathrm{P}_{5,\mathrm{t+k}} = {}_5\mathrm{P}_{5,\mathrm{t}} \times {}_5\mathrm{R}_{5,\mathrm{t}} \tag{6.6}$$

We prefer the ratio method since it is better suited for the regression-based method for creating intervals around forecasts for the two youngest age groups. It is better suited because the CAR values are substantially different than the CCRs, whereas the ratios are not. This means that the CAR values are potential outliers that could serve as influential observations that adversely affect model construction (Fox 1991).

Forecasts of the oldest open-ended age group also differ slightly from the forecasts for the age groups beyond age 10 up to the oldest open-ended age group. If for example the final closed age group is 70–74, with 75 years and older as the terminal open-ended age group, then calculations for the $CCR_{x+,t}$ require the summation of the three oldest age groups to get the population aged 65 years and older at time $t - k$:

$$_{\infty}\mathrm{CCR}_{75,\mathrm{t}} = {}_{\infty}\mathrm{P}_{75,\mathrm{t}}/{}_{\infty}\mathrm{P}_{65,\mathrm{t-k}}. \tag{6.7}$$

The formula for forecasting the population aged 75 years and older for the year $t + k$ is:

$$_{\infty}P_{75+,t+k} = {}_{\infty}CCR_{75,t} \times {}_{\infty}P_{65,t}. \tag{6.8}$$

6.4.3 *Incorporating Uncertainty into the Hamilton-Perry Method*

It is not surprising that the H-P Method is deterministic given its consistency with the fundamental demographic equation (See Appendix at the end of this chapter), which by its nature is an accounting method. However, we also know that population forecasting is subject to uncertainty since we do not precisely know the future components making up the fundamental equation. So, the question is how to introduce an element of statistical uncertainty into a forecasting method that is inherently deterministic. One answer to this question is found by employing a simple regression method to estimate CCRs and then applying the regression-estimated CCRs to the launch-year age groups to obtain forecasts by age group. Text was changed to explictily reference the Appendix at the end of this chapter

The CCRs for the most current census period ($_nCCR_{x,t}$) was given in Eq. 6.1 and we define the CCRs for the preceding census period as:

$$_{n}CCR_{x,t-k} = {}_{n}P_{x,t-k} / {}_{n}P_{x-k,t-2k} \tag{6.9}$$

We construct a regression model with $_nCCR_{x,t}$ as the dependent variable and $_nCCR_{x,t-k}$ as the independent variable. We note that for age groups 0–4, 5–9, and the terminal open-ended age group that the dependent and independent observations follow the equations provided earlier. The estimated CCRs at time t are as follows:

$$_{n}ECCR_{x,t} = a + b \times {}_{n}CCR_{x,t-k} \tag{6.10}$$

We then multiply $_nECCR_{x,t}$ and the corresponding population by age at time t to forecast the CCR at time $t + k$:

$$_{n}CCR_{x,t+k} = {}_{n}ECCR_{x,t} \times {}_{n}P_{x,t}. \tag{6.11}$$

Utilizing the regression measure of statistical uncertainty (the standard error of estimate) for the model along with the sample size and other characteristics of the data, we can generate forecast intervals around $_nCCR_{x,t+k}$. The approximate margin of error associated with a regression-based forecast is given by Hyndman and Athanasopoulos (2012):

$$\mathrm{moe} = t_{n-2} s_e \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{X})^2}{(n-1){s_x}^2}} \tag{6.12}$$

where

n is the total number of observations,
t_{n-2} is the t-distribution value corresponding to the probability level,
$\bar{X}$ is the mean of the observed x values,
${s_x}^2$ is the variance of the observed x values, and
s_e is the standard error of the regression.

These forecasts intervals around ${}_nCCR_{x,t+k}$ are translated directly to the actual population numbers forecasted for each age group (Espenshade and Tayman 1982; Swanson and Beck 1994).

We use data from 1980–2010 for Minnesota to illustrate the derivation of point and interval forecasts using regression combined with the H-P method. We begin by computing the CCRs and ratios for the two youngest age groups for 1980–1990 and 1990–2000 as shown in Table 6.1. We then estimate the these values for 1990–2000 by regressing the observed 1990–2000 values against the observed 1980–1990 values for each age group and solving the regression as follows:

$$\begin{aligned} &{}_n\mathrm{ECCR}_{x,1990-2000} = 0.1676667 + \left(0.8644256 \times {}_n\mathrm{CCR}_{x,1980-1990}\right) \\ &\mathrm{adj.}r^2 = 0.755 \text{ and } s_e = 0.07124. \end{aligned} \tag{6.13}$$

Under usual assumption in the H-P method that the launch year ratios are held constant, point forecasts in 2010 are computed by:

$${}_n\mathrm{Pop}_{x,2010} = {}_n\mathrm{ECCR}_{x,1990-2000} \times \mathrm{Pop}_{x,2000}, \text{ where x } (0-4 \text{ and } 5-9), \tag{6.14}$$

$${}_n\mathrm{Pop}_{x,2010} = {}_n\mathrm{ECCR}_{x,1990-2000} \times \mathrm{Pop}_{x-10,2000}, \text{ where x } (10-74), \text{ and} \tag{6.15}$$

$${}_n\mathrm{Pop}_{x,2010} = {}_n\mathrm{ECCR}_{x,1990-2000} \times \mathrm{Pop}_{65+,2000}, \text{ where x } (75+). \tag{6.16}$$

The 1990–2000 ${}_nECCR_x$ and point forecasts for population by age are shown in Table 6.1.

Table 6.2 shows the 66% forecast intervals for both the 1990–2000 ${}_nECCR_x$ and 2010 population. We first develop intervals around the 1990–2000 ${}_nECCR_x$ by:

$${}_n\mathrm{ECCR}_{x\ x,1990-2000} \pm \mathrm{moe} \tag{6.17}$$

where,

moe is the margin of error at a given probability level.

Equation 6.12 shows the forecast interval is wider when x is farther from $\bar{X}$ (or the average of the 1980–1990 CCRs). That is, we are more certain about our

Table 6.1 1980–1990 and 1990–2000 cohort change ratios and 2010 forecast population, Minnesota

	Population			Cohort change ratios[a]			
					1990–2000		
Age	1980	1990	2000	1980–1990	Observed	Estimated[b]	2010 Population[c]
0–4	307,249	336,800	329,594	1.09618	0.97860	1.11501	367,501
5–9	296,295	345,840	355,894	1.16722	1.02907	1.17641	418,677
10–14	333,378	313,297	374,995	1.01968	1.11341	1.04890	345,711
15–19	399,818	297,609	374,362	1.00443	1.08247	1.03572	368,607
20–24	393,566	316,046	322,483	0.94801	1.02932	0.98696	370,105
25–29	363,435	381,759	319,826	0.95483	1.07465	0.99286	371,689
30–34	313,104	397,984	353,312	1.01123	1.11791	1.04160	335,898
35–39	246,356	361,274	412,490	0.99405	1.08050	1.02675	328,381
40–44	202,860	304,810	411,692	0.97351	1.03444	1.00900	356,492
45–49	187,051	237,050	364,247	0.96223	1.00823	0.99925	412,181
50–54	193,199	191,410	301,449	0.94356	0.98897	0.98312	404,743
55–59	189,457	173,066	226,857	0.92523	0.95700	0.96727	352,325
60–64	170,638	171,220	178,012	0.88624	0.93000	0.93358	281,427
65–69	149,114	160,036	153,169	0.84471	0.88503	0.89769	203,647
70–74	121,034	134,486	142,656	0.78814	0.83317	0.84880	151,097
75+	209,416	252,412	298,441	0.52634	0.54566	0.62254	369,954
Total	4,075,970	4,375,099	4,919,479				5,438,435

(Sources: 2000, U.S. Census Bureau (http://factfinder2.census.gov). 1990, U.S. Census Bureau, Table 19, General Population Characteristics,1990 Decennial Census. 1980, U.S. Census Bureau, Table 19, General Population Characteristics,1980 Decennial Census)

[a] $P_{0–4,t}/P_{0–4,t-10}$ Ages 0–4
$P_{5–9,t}/P_{5–9,t-10}$ Ages 5–9
[a] $P_{x+10,t}/P_{x,t-10}$ Ages 10–74
$P_{75+,t}/P_{65+,t-10}$ Ages 75+

[b] Based on the regression equation, $0.1676667 + (0.8644256 \times {}_nECCR_{x\ 1980–1990})$

[c] ${}_nECCR_{0–4,1990–2000} \times P_{0–4,2000}$ Ages 0–4
${}_nECCR_{5–9,1990–2000} \times P_{5–9,2000}$ Ages 5–9
${}_nECCR_{x,1990–2000} \times P_{x-10,2000;}$ Ages 10–74
${}_nECCR_{75+,1990–2000\ +} \times P_{65+,2000}$ Ages 75+

Table 6.2 66% forecast intervals, Minnesota, 2010

	1990–2000 Cohort change ratios				2010 Population forecast	
Age	Point forecast[a]	Margin of error[b]	Lower limit[c]	Upper limit[d]	Lower limit[e]	Upper limit[e]
0–4	1.11501	0.07615	1.03886	1.19116	342,402	392,599
5–9	1.17641	0.07909	1.09732	1.25550	390,530	446,825
10–14	1.04890	0.07415	0.97475	1.12305	321,272	370,151
15–19	1.03572	0.07390	0.96182	1.10962	342,306	394,907
20–24	0.98696	0.07344	0.91352	1.06040	342,565	397,645
25–29	0.99286	0.07346	0.91940	1.06632	344,188	399,190
30–34	1.04160	0.07400	0.96760	1.11560	312,035	359,762
35–39	1.02675	0.07376	0.95299	1.10051	304,791	351,972
40–44	1.00900	0.07356	0.93544	1.08256	330,502	382,481
45–49	0.99925	0.07349	0.92576	1.07274	381,867	442,495
50–54	0.98312	0.07343	0.90969	1.05655	374,512	434,973
55–59	0.96727	0.07346	0.89381	1.04073	325,568	379,083
60–64	0.93358	0.07377	0.85981	1.00735	259,189	303,665
65–69	0.89769	0.07447	0.82322	0.97216	186,753	220,541
70–74	0.84880	0.07603	0.77277	0.92483	137,562	164,631
75+	0.62254	0.09091	0.53163	0.71345	315,930	423,979

[a]From Table 6.1
[b]Based on Eq. 6.12, using a t-value of 1.00 for a 66% forecast interval
[c]Point forecast – margin of error
[d]Point forecast + margin of error
[e]2000 population × upper and lower limits of the $_{n}ECCR_{x,1990–2000}$

forecasts when values of the predictor variable are close to its sample mean. For example, the largest margin of error is for ages 75 years and older (0.09091). The 1980–1990 CCR for that group (0.52634) is 44% below the average CCR. We then translate the intervals around the 1990–2000 $_{n}ECCR_{x}$ into population forecast intervals by applying Eqs. 6.14, 6.15 and 6.16 to the lower and upper limits determined by Eq. 6.17.

6.5 Evaluation

To test the regression-based method for developing intervals around population forecasts by age generated from the H-P Method, we selected a sample made up of one state from each of the four census regions in the United States. The states selected are Georgia (South Region), Minnesota (Midwest Region), New Jersey (Northeast Region) and Washington (West Region). We then assembled census data for these four states for each census year from 1900 to 2010 (U.S. Census Bureau 1973, 1982, 1992, 2000, 2010). The data provide nine points in time at which the forecast intervals can be evaluated 1930, 1940, 1950, 1960, 1970,

Table 6.3 Total population 1900 and 2010 and annual rate of change by decade, sample states

Census year	Georgia	Minnesota	New Jersey	Washington
1900[a]	2,209,974	1,747,292	1,879,890	511,844
1900–1910	1.64%	1.70%	2.99%	7.97%
1910–1920	1.05%	1.41%	2.19%	1.75%
1920–1930	0.05%	0.72%	2.47%	1.44%
1930–1940	0.72%	0.86%	0.30%	1.06%
1940–1950	0.98%	0.66%	1.50%	3.14%
1950–1960	1.35%	1.35%	2.27%	1.83%
1960–1970	1.52%	1.08%	1.67%	1.78%
1970–1980	1.74%	0.69%	0.27%	1.92%
1980–1990	1.70%	0.71%	0.48%	1.64%
1990–2000	2.34%	1.17%	0.85%	1.92%
2000–2010	1.68%	0.75%	0.44%	1.32%
2010	9,687,653	5,303,925	8,791,894	6,724,540

[a]The 1900 population totals exclude those for whom age was not reported

1980, 1990, 2000, and 2010. The terminal open-ended age group is reported differently over the period for which we assembled census data, so we used 75 years and older since it was the common denominator. This means there are 16 age groups used in the evaluation (0–4, 5–9,. . ., 70–74, and 75 years and older).

This sample provides a wide range of demographic characteristics in terms of variation in population size, age-composition, and rates of change. Table 6.3 provides an overview of this range by displaying the population of each state in 1900 and in 2010 and decennial rates of population change from 1900 to 2010. Although we do not show a summary of the changes in age composition by state and census year, they are extensive.

We proceed by constructing CCRs over two successive decennial periods (e.g., 1910–1920/1900–1910) over the entire period, using regression to estimate the CCR in the more current period (e.g., 1910–1920) from the CCR in the earlier period (e.g., 1900–1910). We then use the regression-based estimate of the CCR of the "current period" (e.g., 1910–1920) to forecast the CCRs to the next period, the "launch year" (e.g., 1920–1930) and develop forecast intervals around the forecasted CCRs, which are then translated into the forecasted age groups for the "target year" (e.g., 1930). The forecast intervals are then examined to see if they contain the census age groups for the target year.

6.5.1 Age Groups

How well does the regression approach based on the H-P method perform in its ability to predict the uncertainty of population forecasts? One way to address this

Table 6.4 Number of population counts falling within the 66% forecast intervals by state and target year, 2010

Target year	Georgia	Minnesota	New Jersey	Washington	Total	Percent (N/64)
1930	9	12	8	13	42	67%
1940	3	5	11	12	31	48%
1950	10	14	4	3	31	47%
1960	13	14	14	8	49	86%
1970	6	12	14	13	45	77%
1980	7	12	12	10	41	67%
1990	13	14	14	14	55	83%
2000	8	15	14	15	52	81%
2010	7	15	15	14	51	81%
Total	76	113	106	102	397	
Percent	53%	78%	74%	71%	69%	
	Percent (N/144)	Percent (N/144)	Percent (N/144)	Percent (N/144)	Percent (N/576)	

question is to determine the number of population counts that fall inside the forecast intervals (Tayman et al. 2007). In terms of the forecast interval probability, we selected 0.66 or 66% because of prior research indicating that "low" and "high" scenarios constructed for the cohort-component method corresponded empirically to 66% confidence intervals (Stoto 1983), as well as findings by Swanson and Beck (1994).

Table 6.4 provides a summary of the results for all four states at each of the nine census test points. The table shows the number of times (out of 16) that the 66% forecast interval contained the corresponding census number for a given age group. If the forecast intervals provide a valid measure of uncertainty, they will contain approximately 11 of the 16 observed population counts. The table also shows percent of the counts falling within the forecast intervals for all target years for each state (144 intervals), the percent falling within all states for each target year (64 intervals), and the single percent falling within all states for all target years (576 intervals).

In Georgia (South Census Region), we find that its population increased by almost fivefold between 1900 and 2010. In 1900 it had the largest population of any of the four sample states and it retains that position in 2010. Its annual average growth rates (by decade) ranged from 0.05% between 1920 and 1930 to 2.34% between 1990 and 2000. Changes in its age composition are substantial with large impacts associated with the great depression, World War II, the baby boom, and immigration to the Sunbelt states more recently. The 66% forecast intervals contain their corresponding age groups 76 times out of 144 observations, or 53%. Overall, Georgia has the lowest percent of census age groups within the 66% forecast intervals.

The population of Minnesota tripled from 1900 to 2010. Its average annual growth rates ranged from a low of 0.66% between 1940 and 1950 to a high of 1.70%

between 1900 and 1910, a period when the state was still receiving large numbers of immigrants from Europe. As is the case for Georgia, changes in its age composition are extensive, with big impacts associated with the restrictions placed on immigration during World War I and the great depression, World War II, the baby boom, and out-migration to Sunbelt states in more recent decades. The 66% forecast intervals contain their corresponding age groups 113 times out of 144 observations, or 78%. Overall, Minnesota has the highest percent of census age groups within the 66% forecast intervals.

For New Jersey, its population grew from 1,879,890 in 1900 to 8,791,894 in 2010. New Jersey had the second highest population in 1900 and again in 2010. Its average annual growth rates ranged from a low of 0.27% between 1970 and 1980 to a high of 2.99% between 1900 and 1910. As is the case for Georgia and Minnesota, changes in its age composition are extensive, with big impacts associated with the restrictions placed on immigration during World War I and the great depression, World War II, the baby boom, and out-migration to Sunbelt states in more recent decades. The 66% forecast intervals contain their corresponding census age groups 106 times out of 144 observations, or 74%. Overall, New Jersey has the second-highest percent of census age groups within the 66% forecast interval.

In 1900, Washington was largely a frontier state. It had the smallest population (511,844) of any of the four states in the sample. However, by 2010 it had grown to 6,724,540 which surpassed the population of Minnesota in 2010. Its annual rates of population change are somewhat more dramatic than the other states. Between 1900 and 1910 it posted an annual rate of 7.97%, the highest of any of the decennial growth rates in the sample. It also posted the second highest rate. Between 1940 and 1950 the state grew at an annual rate of 3.14%. The lowest rate of annual population change (1.06%) is found between 1930 and 1940. The 66% forecast intervals contain their corresponding census age groups 102 times, which represents 71% of the 144 observations.

6.5.2 Total Population

It should be clear that we are primarily interested in measuring uncertainty in forecasts of age groups. This is an important topic due to the role that the absolute and relative sizes of age groups play in both commerce (Gauthier et al. 2006; Martins et al. 2012, Murdock et al. 1997) and public policy (Bongaarts and Bulatao 2000, Murdock et al. 1997, Smith et al. 2013: 23, Tuljapurkar et al. 2005). We are aware that levels of uncertainty related to forecasts of the total population are important as well. In this regard, we note that technically the forecast intervals we generated here apply only to the age groups.

There are two ways in which intervals around age group forecasts can be used to place intervals around the total population forecast; one is informal while the other is formal. In the informal approach, we obtain 66% forecast intervals for the total population by adding the lower and upper boundaries of the intervals for each age group. We found that in 28 of the 36 forecasts (four states at each of nine time points) the summed lower and upper boundaries contained the actual total population, or 78%. By state, we find: Georgia's total population is contained in 5 of the 9 time points (56%); Minnesota's in 9 of the 9 time points; New Jersey's in 6 of the 9 time points (67%); and Washington's total population is contained in 8 of the 9 time points (89%). By target year, we find: 4 of 4 were contained in the 1960, 1970, and 1990 years; 3 of 4 were contained in the 1930, 1980, 2000, and 2010 years; and 2 of 4 in the 1940 and 1950 years.

The formal approach is called the "error propagation method" by Deming (1950: 127–134). In different forms it has been used by Alho and Spencer (2005), Espenshade and Tayman (1982), and Hansen et al. (1953), among others. In this application, the error propagation method involves summing the squared values of the forecast intervals by age, finding the square root of the summed forecast interval values and dividing this square root of the sample size (n = 16) to obtain an estimate of the standard error for the total population forecast. This standard error is then multiplied by the total population forecast (found by summing the point forecast for each age group) to obtain the margin of error. The margin of error is added to and subtracted from the total population forecast to obtain its 66% forecast interval. This approach assumes a simplifying assumption that the 16 age groups are independent (Espenshade and Tayman 1982). Using this approach, we found that in 29 of the 36 forecasts (four states at each of nine time points) the error propagation intervals contained the actual total population, or 81%. By state, we find: Georgia's total population is contained in 6 of the 9 time points (67%); Minnesota's is in 9 of the 9 time points; New Jersey's is in 6 of the 9 (67%), and Washington's is in 8 of the 9 time points (89%). By time point, we find: 4 of 4 were contained in the 1960, 1970, 1990, and 2010 target years; 3 of 4 were contained in the 1930, 1980, and 2000 target years; and 2 of 4 in the 1940 and 1950 target years.

6.6 Conclusions

Overall, the 66% intervals contain their corresponding census age groups in 397 cases, which represent 69% of the 576 total observations. In terms of the nine census target years, the overall results show that in five of them (1960, 1970, 1990, 2000, and 2010) the forecast intervals contain the census age groups substantially more than 66% of the time. In two target years (1930 and 1980), the intervals contain the census age groups 67% of the time. In the remaining two target years, 1940 and 1950, the intervals contain the census age groups 48% and 47% of the time, respectively. We note that the 1940 test point encompasses the economic boom experienced in the 1920s, the economic depression during the 1930s, and the large scale "baby bust" associated with it. The 1950 point encompasses the depression and baby bust period of the 1930s, the economic recovery stimulated by World War II, and the initial part of the large scale "baby boom" from 1946 to 1950.

Table 6.5 Number of population counts falling within the 66% forecast interval by age group, all states and target years, 2010

Age	Number	Percent (N/36)
0–4	9	25%
5–9	9	25%
10–14	26	72%
15–19	27	75%
20–24	24	67%
25–29	21	58%
30–34	19	53%
35–39	22	61%
40–44	26	72%
45–49	28	78%
50–54	30	83%
55–59	31	86%
60–64	30	83%
65–69	31	86%
70–74	33	92%
75+	31	86%
Total	397	69%

Table 6.5 contains a summary of the results by age group across all of the nine census target years and the four states. The table shows the number of times (out of 36) that the 66% forecast interval contained the corresponding census number for a given age group. If the forecast intervals provide a valid measure of uncertainty, they will contain approximately 24 of the 36 observed population counts. In general, the forecast intervals capture the population count at least 66% of the time for age groups 10–14, 15–19, 20–24, and 40–44 through 75 years and older. For age groups 0–4 and 5–9, the forecast intervals only encompass the population counts 25% of time. For age group 30–34 the count is encompassed 53% of the time, while for age group 25–29 the count is encompassed 58% of the time. The population counts are captured by the forecast intervals 61% of the time for age group 35–39.

Perhaps it should not be surprising that the cohort change method is better able to capture older age groups than the very youngest since births are not part of a cohort change ratio. In addition, migration likely comes into play because the population in the two youngest age groups (0–4 and 5–9) would be moving with their parents, who are likely to be in age range 25–39, the other age groups for which the forecast intervals encompassed the population counts less than 66% of the time. Overall, we find that these effects are consistent with theory regarding migration; that is, those who tend to move are less socially integrated into communities than those who tend not to move, and the aging of those who tend not to move increases community social integration (Goldscheider 1978).

Although not shown here, the average width of the forecast intervals appears to us to be reasonable at the 66% level in that they are neither so wide as to be meaningless nor too narrow to be overly-restrictive. This is largely consistent with prior work by Swanson and Beck (1994) on forecast intervals derived from regression-based forecasts, which found that the forecast intervals contained the actual numbers by age in 69% of the 576 observations; providing further support that 66% forecast intervals based on an regression-estimated CCR approach are both useful and feasible. We find these results encouraging.

At this point, we suggest caution using this method beyond a 10-year forecast horizon. This is consistent with observations about the use of the H-P method in general (Swanson et al. 2010) and as such is not a major limitation.[1] We also suggest that this approach to developing uncertainty measures be used with care when applied to small populations, such as those found at the county and subcounty levels. While our sample provides a wide range of demographic behavior in terms of size, age composition, and population changes, it is a sample of states, which means that greater variability in demographic characteristics found at substate levels is not present (Swanson et al. 2010). We suggest that further research using this approach would be useful by examining both longer forecast horizons and smaller populations and different probability intervals. Another area for further research would be to utilize root mean square errors in conjunction with the H-P Method (Keyfitz 1981).

The fact that the forecast intervals do not contain the population counts at least 66% of the time for neither the two youngest age groups (0–4 and 5–9) nor the age groups associated with those most likely to be the parents of these children (ages 25–39) should not be surprising. The dynamics of birth and migration are difficult to capture in the cohort-component method forecast and the H-P Method is a variant of this method (Smith et al. 2013: 177; Smith and Tayman 2003). Thus, work on these issues in regard to one of these two methods should be of use to the other.

Appendix

Cohort Change Ratios and the Fundamental Demographic Equation

It is important that a demographic technique satisfy various mathematical identities and, in particular, the demographic accounting identity known as the fundamental demographic equation:

$$P_{t+k} = \mathrm{P_t} + \text{Births}-\text{Deaths} + \text{In}-\text{migrants} - \text{Out}-\text{migrants} \tag{A.1}$$

[1]The ten-year horizon is also consistent with accuracy evaluations of the H-P method, which show that the method performs well for ten year forecasts (Smith and Tayman 2003; Swanson and Tayman 2013) and even 20 year forecasts (Smith and Tayman 2003).

This equation states that the population at a given point in time, P_{t+k}, must be equal to the population at an earlier time, P_t, plus the births and in-migrants and minus the deaths and out-migrants that occur between time t and time $t + k$.

The Cohort Change Ratio method moves a population by age from time t to time $t + k$ using cohort-change ratios (CCRs) computed from data in the two most recent censuses. It consists of two steps. The first step uses existing data to develop CCRs and the second step applies the CCRs to the cohorts of the launch year population to move them into the future. The formula for the first step, the development of a CCR is:

$$_{\mathrm{n}}\mathrm{CCR}_{\mathrm{x,t}} = {}_{\mathrm{n}}\mathrm{P}_{\mathrm{x,t}} / {}_{\mathrm{n}}\mathrm{P}_{\mathrm{x-k,t-k}} \tag{A.2}$$

where,

$_nP_{x,t}$ is the population aged x to $x + n$ at the most recent census (t),
$_nP_{x-k,t-k}$ is the population aged $x - k$ to $x - k + n$ at the 2nd most recent census ($t - k$), and
k is the number of years between the most recent census at time t and the one preceding it at time t-k.

The basic formula for the second step, moving the cohorts of a population into the future is:

$$_{\mathrm{n}}\mathrm{P}_{\mathrm{x+k,t+k}} = {}_{\mathrm{n}}\mathrm{CCR}_{\mathrm{x,t}} \times {}_{\mathrm{n}}\mathrm{P}_{\mathrm{x,t}} \tag{A.3}$$

where,

$_nP_{x+k,t+k}$ is the population aged $x + k$ to $x + k + n$ at time $t + k$, and
$_nCCR_{x,t}$ and $_nP_{x,t}$ are as defined in Eq. (6.2).

In terms of the CCR Method satisfying the fundamental demographic equation, let

$$_{\mathrm{n}}\mathrm{CCR}_{\mathrm{x,t}} = ({}_{\mathrm{n}}\mathrm{P}_{\mathrm{x-k,t-k}} + \mathrm{B} - \mathrm{D} + \mathrm{I} - \mathrm{O}) / ({}_{\mathrm{n}}\mathrm{P}_{\mathrm{x-k,t-k}}) \tag{A.4}$$

where,

$_nP_{x-k,t-k}$ is the population aged $x - k$ to $x - k + n$ at the 2nd most recent census ($t - k$),
B = Births between time $t - k$ and t,
D = Deaths between time $t - k$ and t,
I = In-migrants between time $t - k$ and t, and
Out-migrants between time $t - k$ and t.

Since,

$$ {}_{n}P_{x+k,t+k} = (({}_{n}P_{x-k,t-k} + B - D + I - O)/({}_{n}P_{x-k,t-k})) \times ({}_{n}P_{x,t}). \tag{A.5} $$

then,

$$ {}_{n}CCR_{x,t} = ({}_{n}P_{x-k,t-k} - D + I - O)/({}_{n}P_{x+k,t+k}), \tag{A.6} $$

where, $x + k >= 10$.

Thus, the CCR method expresses the individual components of change—births, deaths, and migration—in terms of cohort change ratios and satisfies the fundamental demographic equation.

References

Alders, M., Keilman, N., & Cruijsen, H. (2007). Assumptions for long-term stochastic population forecasts in 18 European countries. *European Journal of Population, 23*, 33–69.

Alho, J. (1984). Probabilistic forecasts, The case of population projections. *Scandinavian Housing and Planning Research, 1*, 99–105.

Alho, J., & Spencer, B. (1985). Uncertain population forecasting. *Journal of the American Statistical Association, 80*(390), 306–314.

Alho, J., & Spencer, B. (1990). Error models for official mortality forecasts. *Journal of the American Statistical Association, 85*(411), 609–616.

Alho, J., & Spencer, B. (1997). The practical specification of the expected error of population forecasts. *Journal of Official Statistics, 13*, 203–225.

Alho, J., & Spencer, B. (2005). *Statistical demography and forecasting*. Dordrecht: Springer.

Alkema, L., Raftery, A., Gerland, P., Clark, S., Pelletier, F., Buettner, T., & Heilig, G. (2011). Probabilistic projections of the total fertility rate for all countries. *Demography, 48*, 815–839.

Andridge, R., & Little, R. J. A. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review, 78*(1), 49–64.

Bongaarts, J., & Bulatao, R. (Eds.). (2000). *Beyond six billion: Forecasting the world's population*. Washington, DC: National Academy Press.

Brown, J., Abbott, O., & Smith, P. (2011). Design of the 2001 and 2011 census coverage surveys for England and Wales. *Journal of the Royal Statistical Society, Series A, 174*(4), 881–906.

Cameron, M., & Poot, J. (2011). Lessons from stochastic small-area population projections: The case of Waikato subregions in New Zealand. *Journal of Population Research, 28*, 245–265.

Campbell, P. (1996). *Population projections for states by age, sex, race, and Hispanic Origin: 1995 to 2050. PPL-47*. Washington, DC: U.S. Bureau of the Census.

Cheeseman-Day, J. (1992). *Population projections of the United States, by age, sex, race, and Hispanic Origin: 1992–2050*, *Current Population Reports P-25, No. 1092*. Washington, DC: U.S. Bureau of the Census.

Cohen, J. (1986). Population forecasts and confidence intervals for Sweden: A comparison of model-based and empirical approaches. *Demography, 23*(1), 105–123.

De Beer, J. (1993). Forecast intervals of net migration: The case of the Netherlands. *Journal of Forecasting, 12*, 585–599.

Deming, W. (1950). *Some theory of sampling*. New York: Dover Publications.

Deming, W. E., & Stephan, F. F. (1941). On the interpretation of censuses as samples. *Journal of the American Statistical Association, 36*(213), 45–49.

Espenshade, T., & Tayman, J. (1982). Confidence intervals for postcensal population estimates. *Demography, 19*(2), 191–210.

Fox, J. (1991). *Regression diagnostics, Sage Quantitative Applications in the Social Sciences, no. 79*. London: Sage Publications.

Gauthier, A., Chu, C., & Tuljapurkar, S. (Eds.). (2006). *Allocating public and private resources across generations: Riding the age waves* (Vol. 2). Dordrecht: Springer.

Goldscheider, C. (1978). *Modernization, migration, and urbanization*. Paris: International Union for the Scientific Study of Population.

Graubard, B., & Korn, E. (2002). Inference for superpopulation parameters using sample surveys. *Statistical Science, 17*(1), 73–96.

Hansen, N., Hurwitz, W., & Madow, W. (1953). *Sample survey methods and theory, volume I, methods and applications*. New York: John Wiley and Sons. (re-published in 1993).

Hyndman, R., & Athanasopoulos, G. (2012) *Forecasting: Principles and practice*. Retrieved from http://otexts.com/fpp/

Hyndman, R., & Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting, 24*, 323–342.

Keyfitz, N. (1972). On future population. *Journal of the American Statistical Association, 67*, 347–363.

Keyfitz, N. (1981). The limits of population forecasting. *Population and Development Review, 7*, 579–593.

Keyfitz, N. (1987). The social and political context of population forecasting. In W. Alonso & P. Starr (Eds.), *The politics of numbers* (pp. 235–258). New York: Russell Sage Foundation.

Lee, R. (1974). Forecasting births in post-transition populations: Stochastic renewal with serially correlated fertility. *Journal of the American Statistical Association, 69*, 607–617.

Lee, R. (1992). Stochastic demographic forecasting. *International Journal of Forecasting, 8*, 315–327.

Lee, R. (1993). Modeling and forecasting the time series of U.S. fertility: Age distribution, range, and ultimate level. *International Journal of Forecasting, 9*, 187–212.

Lee, R., & Carter, L. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association, 87*, 659–675.

Lee, R., & Tuljapurkar, S. (1994). Stochastic population projections for the United States: Beyond high, medium, and low. *Journal of the American Statistical Association, 89*, 1175–1189.

Linstone, H., & Turoff, M. (1975). *The Delphi method: Techniques and applications*. Reading: Addison-Wesley.

Lutz, W., Sanderson, W., & Scherbov, S. (1999). Expert-based probabilistic population projections. In W. Lutz, J. Vaupel, & D. Ahlburg (Eds.), *Frontiers of population forecasting* (pp. 139–155). New York: The Population Council. (A supplement to *Population and Development Review*, 24).

Martins, J., Yusuf, F., & Swanson, D. (2012). *Consumer demographics and behaviour: markets are people*. Dordrecht: Springer.

McNees, S. (1992). The use and abuses on 'consensus' forecasts. *Journal of Forecasting, 11*, 703–710.

McNown, R., Rogers, A., & Little, J. (1995). Simplicity and complexity in extrapolative population forecasting models. *Mathematical Population Studies, 5*, 235–257.

Moen, E. (1984). Voodoo Forecasting: Technical, political, and ethical issues regarding the projection of local population growth. *Population Research and Policy Review, 3*, 1–25.

Murdock, S., Hoque, M. N., Michael, M., White, S., & Pecotte, B. (1997). *The Texas challenge: Population change and the future of Texas*. College Station: Texas A&M University Press.

Pflaumer, P. (1988). Confidence intervals for population projections based on Monte Carlo methods. *International Journal of Forecasting, 4*, 135–142.

Pflaumer, P. (1992). Forecasting U.S. population totals with the Box-Jenkins approach. *International Journal of Forecasting, 8*, 329–338.

Pittenger, D. (1978). The role of judgment, assumptions, techniques, and confidence limits in forecasting population. *Socio-Economic Planning Sciences, 12*, 271–276.

Raftery, A., Alkema, L., & Gerland, P. (2013). *Bayesian population projections for the United Nations*. Unpublished report. Retrieved from https://www.stat.washington.edu/raftery/Research/PDF/AlkemaGerland2013Statsci/pdf

Rayer, S., Smith, S., & Tayman, J. (2009). Empirical prediction intervals for county population forecasts. *Population Research and Policy Review, 28*, 773–793.

San Diego County Water Authority. (2002). *Regional water facilities master plan. Appendix C: Development of probabilistic water demand forecasts*. San Diego: San Diego County Water Authority.

Ševčíková, H., Raftery, A., & Gerland, P. (2013). *Bayesian probabilistic population projections: Do it yourself*. Working Paper 13.2. Joint Eurostat/UNECE Work Session on Demographic Projections. Rome, Italy, 29–31 October. Retrieved from http://www.unece.org/stats/documents/2013.10.projections.html

Smith, S. K., & Bayya, R. (1992). An evaluation of population forecast errors for Florida and its counties. *Applied Demography, 7*, 1–5.

Smith, S. K., & Rayer, S. (2012). *An evaluation of population forecast errors for Florida and its counties, 1980–2010*. Paper presented at the Conference on Applied Demography, San Antonio, TX, 8–10, January.

Smith, S. K., & Sincich, T. (1988). Stability over time in the distribution of population forecast errors. *Demography, 25*, 461–474.

Smith, S. K., & Tayman, J. (2003). An evaluation of population projections by age. *Demography, 40*(4), 741–757.

Smith, S. K., Tayman, J., & Swanson, D. (2013). *A practitioner's guide to state and local population projections*. Dordrecht: Springer.

Spencer, G. (1989). *Projection of the population of the United States by age, sex, and race: 1988 to 2008*, *Current Population Reports, P-25, No. 1018*. Washington, DC: U.S. Bureau of the Census.

Stoto, M. (1983). The accuracy of population projections. *Journal of the American Statistical Association, 78*, 13–20.

Swanson, D., & Beck, D. (1994). A new short-term county population projection method. *Journal of Economic and Social Measurement, 20*, 1–26.

Swanson, D., & Tayman, J. (1995). Between a rock and a hard place: The evaluation of demographic forecasts. *Population Research and Policy Review, 14*, 233–249.

Swanson, D., & Tayman, J. (2013). *The accuracy of the Hamilton-Perry method for forecasting state populations by age*. Working paper 13–01, Center for Sustainable Suburban Development, University of California Riverside. Retrieved from http://cssd.ucr.edu/papers/pdfs/the%20accuracy%20of%20the%20hamilton perry%20method%20for%20forecasting%20state%20populations%20by%20age%20v6.pdf

Swanson, D., Schlottmann, A., & Schmidt, R. (2010). Forecasting the population of census tracts by age and sex: An example of the Hamilton–Perry method in action. *Population Research and Policy Review, 29*, 47–63.

Tayman, J. (2011). Assessing uncertainty in small area forecasts: State of the practice and implementation strategy. *Population Research and Policy Review, 30*, 781–800.

Tayman, J., Schaefer, E., & Carter, L. (1998). The role of population size in the determination and prediction of population forecast errors: An evaluation using confidence intervals for subcounty Areas. *Population Research and Policy Review, 17*(1), 1–20.

Tayman, J., Smith, S., & Lin, J. (2007). Precision, bias, and uncertainty for state population forecasts: An exploratory analysis of time series models. *Population Research and Policy Review, 26*(3), 347–369.

Tayman, J., Smith, S., & Rayer, S. (2011). Evaluating population forecast accuracy: A regression approach using county data. *Population Research and Policy Review, 30*, 235–262.

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

Torri, T., & Vaupel, J. (2012). Forecasting life expectancy in an international context. *International Journal of Forecasting, 28*, 519–531.

Tuljapurkar, S., Pool, I., & Prachuabmoh, V. (Eds.). (2005). *Population, resources and development: Riding the age waves* (Vol. 1). Dordrecht: Springer.

U.S. Census Bureau. (1973). *1970 census of population, general population characteristics (by State)*. Washington, DC: Government Printing Office.

U.S. Census Bureau. (1982). *1980 census of population, general population characteristics, (by State)*. Washington, DC: Government Printing Office.

U.S. Census Bureau. (1992). *1990 Census of population, general population characteristics (by State)*. Washington, DC: Government Printing Office.

U.S. Census Bureau. (2000). Table QT-P1 (By State). Retrieved from http://factfinder2.census.gov/faces/nav/jsf/pages/guided_search.xhtml

U.S. Census Bureau. (2010). Table QT-P1 (By State). Retrieved from http://factfinder2.census.gov/faces/nav/jsf/pages/guided_search.xhtml

Williams, W., & Goodman, M. (1971). A simple method for the construction of empirical confidence limits for economic forecasts. *Journal of the American Statistical Association, 66*, 752–754.

Wilson, T. (2012). Forecast accuracy and uncertainty of Australian Bureau of Statistics state and territory population projections. *International Journal of Population Research*. Retrieved http://www.hindawi.com/journals/ijpr/2012/419824/

Wilson, T., & Bell, M. (2004). Probabilistic regional population forecasts: The example of Queensland, Australia. *Geographical Analysis, 39*, 1–25.

Wilson, T., & Terblanche, W. (2017). The growth of Australia's very elderly population: Past estimates and probabilistic forecasts. In D. Swanson (Ed.), *The frontiers of applied demography* (pp. 115–136). Cham: Springer.

Wolter, K. (1986). Some coverage error models for census data. *Journal of the American Statistical Association, 81*(394), 338–346.

Chapter 7
Forecasting School Enrollment Size and Composition

7.1 Introduction

Demographics drive all aspects of school district management, and student population and enrollment forecasts represent integral components of the management process. In this chapter, we briefly review the history of the development of methods used for student population and enrollment forecasting before describing how cohort change ratios can be used in this process. We cover both short-term and long-term forecasting needs and describe methods that can be used in conjunction with the cohort change ratio approach, giving examples. We conclude the chapter with a discussion on the accuracy and utility of the CCR method regarding its use for forecasting student populations and enrollment.

"For reasons understood by every school administrator, a primary demographic concern centers on the size of a school district's student population. Size drives planning from every perspective relating to budget, program, staffing, space, and so forth." (Wood et al. 1995: 5–17). This observation says it all in regard to the importance of demographic information to school districts planning, and because planning is about the future one can see the importance of student population and enrollment forecasting to school districts. In this chapter, we show how the cohort change ratio (CCR) method can be used to forecast student populations and enrollments for school districts. In general, the chapter is aimed at public schools, grades kindergarten to twelve (K-12). We note, however, these methods can be used for private schools as well as post-secondary institutions such as community colleges and universities, both public and private. However, unlike the latter, which are not mandated to accept students, the K-12 system is mandated to accept students because of compulsory attendance laws and regulations. That is, in the K-12 system enrollments are virtually determined by student demographics, whereas enrollment in private schools and post-secondary institutions stop short of being virtually determined by student demographics because of admission policies.

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_7

These policies add an additional factor to the enrollment forecasting task (Swanson 2016: 25–34).

Student population and enrollment forecasting using demographic methods in use today can be traced to the early 1950s in the states of Washington (Swanson 2016: 23) and California (California Department of Finance 1954). Today, these methods and variants of them are used in school districts throughout the U.S. (Demographics Research Group 2014; Hussar and Bailey 2011; Lapkoff 2008; Rynerson and Chun 2015).

7.2 Short-Term Enrollment Forecasting by Grade

One useful method for short-term (year to year) enrollment forecasting is known as the Grade Progression Ratio (GPR). This method is basically the CCR method applied to grade levels in the K-12 system—those in 2nd grade this year were part of the cohort of first graders last year and so on. This method has a long history of use. Hauser and Kitagawa (1961) describe it as being simple to use and capable of yielding good short-range forecasts.

This method consists of two steps. In the first step, GPRs from the prior year to the current year are calculated; and in the second step, the GPRs are applied to enrollments by grade for the most current year to get a forecast by grade for next year. Typically, fall enrollments are used.

$$\mathrm{GPR}_{x,t} = \mathrm{G}_{x,t}/\mathrm{G}_{x-1,t-1} \quad (\text{Step 1}) \tag{7.1}$$

where,

$G_{x,t}$ is the (fall) enrollment in grade x for the current year, and
$G_{x-1,\ t-1}$ is the (fall) enrollment in grade $x - 1$ for the prior year.

Assuming that the school or school district in question offers Kindergarten, First Grade is typically the lowest grade for which a GPR can be calculated as shown in Eq. 7.1. We will address this issue shortly, but for now, we move on to the second step, forecasting

$$\mathrm{G}_{x+1,t+1} = \mathrm{GPR}_{x,t} \times \mathrm{G}_{x,t} \quad (\text{Step 2}) \tag{7.2}$$

where,

$G_{x,t}$ is the (fall) enrollment in grade x for the current year, and
$G_{x+1,\ t+1}$ is the forecasted (fall) enrollment in grade $x + 1$ for the next year.

Several methods are typically employed to forecast Kindergarten enrollment. One approach is to find the ratio of current Kindergarten enrollment to the number of births reported in the same area (e.g., the school district) 6 years ago. Then apply this ratio to the births reported 5 years ago to obtain a forecast of Kindergarten enrollment for next year. A variation on this approach is to compute an average of these ratios over the past 5 years (in terms of Kindergarten enrollment) and apply the average to births 5 years ago to obtain a forecast of Kindergarten enrollment for next year.

Another approach is to “look backwards” and compute a ratio composed of the Kindergarten enrollment in the current year (the numerator) to the first grade enrollment in the current year (the denominator). This ratio can then be applied to the number of first graders forecast for next year to obtain a forecast of Kindergarten enrollment for next year. Conceptually, this approach is similar to using a “Child-Adult Ratio” to forecast those in the age groups born since the last census. We use this approach to illustrate the GPR method discussed later.

Yet another approach is to take the ratio of current year Kindergarten enrollment (the numerator) to prior year Kindergarten enrollment (the denominator) and then apply this ratio to the current year’s Kindergarten enrollment to obtain a forecast of Kindergarten enrollment for the next year. Without exhausting all of the possibilities for obtaining a forecast of Kindergarten enrollment, it also is possible to use more than one of the approaches and average the results.

In many schools and school districts, there are special categories of students who do not fall in one of the grades from Kindergarten through twelfth grade. They are often classified as ungraded. In order to deal with these students, ratios such as described for forecasting Kindergarten enrollment may be used.

An empirical example of the GPR method is found in Table 7.1, where the K-12 enrollment by Grade of the Riverside (California) Unified School District is forecasted for fall 2015 (using fall 2013 and fall 2014 enrollment data). The GPRs are very close to one in grades 2 through 4, 8, and 9 through 12. The largest GPR shows there are 5.4% more Kindergarteners than first graders in 2014, the current year of this illustration. Most of the other GPRs show changes of between one and two percent. The enrollment forecast for 2015 for all grades increases by only 2 students from 2014 to 2015. Six grades show enrollment declines and six show enrollment increases, excluding the ungraded category. The largest decline (−84 students) occurs in grade 8 due in large part to the smaller number of students in grade 7 compared to grade 8 in 2014. The largest change (199 students) occurs in grade 4 as there are 178 more students in grade 3 than in grade 4 in 2014.

While the GPR method works well in the short-term, it addresses neither the long-term nor the student population from which enrollment is drawn. To deal with these issues, we turn to the CCR method, which can handle long-term student population forecasts and, with some augmentation, long term enrollment forecasts.

Table 7.1 Public school enrollment forecast, Riverside, California Unified School District, Fall 2015

Grade	Fall 2013	Fall 2014	GPR 2013–2014[a]	Fall 2015[b]	Change 2014–2015	
					Numeric	Percent
Pre K & K	3,251	3,278	1.05436	3,305	27	0.8%
First	3,162	3,109	0.95632	3,135	26	0.8%
Second	3,251	3,163	1.00032	3,110	−53	−1.7%
Third	3,082	3,280	1.00892	3,191	−89	−2.7%
Fourth	3,023	3,102	1.00649	3,301	199	6.4%
Fifth	3,147	3,054	1.01025	3,134	80	2.6%
Sixth	3,168	3,195	1.01525	3,101	−94	−2.9%
Seventh	3,326	3,242	1.02336	3,270	28	0.9%
Eighth	3,327	3,322	0.99880	3,238	−84	−2.5%
Ninth	3,370	3,389	1.01864	3,384	−5	−0.1%
Tenth	3,406	3,366	0.99881	3,385	19	0.6%
Eleventh	3,389	3,358	0.98591	3,319	−39	−1.2%
Twelfth	3,638	3,420	1.00915	3,389	−31	−0.9%
Ungraded (9–12)	47	61	1.29787	79	18	29.5%
Total	42,587	42,339		42,341	2	0.0%

Source: California Department of Education DataQuest (http://dq.cde.ca.gov/dataquest)
[a] $G_{x,t}/G_{x-1,t-1}$ Grades 1–12
$G_{k,t}/G_{1,t}$ Grade K
$G_{u,t}/G_{u,t-1}$ Ungraded
[b] $G_{x+1,t+1} = G_{x,t} \times GPR_{x,t}$ Grades 1–12
$G_{k,t+1} = G_{1,t+1} \times GPR_{k,t}$ Grade K
$G_{u,t+1} = G_{u,t} \times GPR_{u,t}$ Ungraded

7.3 Long-Term Student Population and Enrollment Forecasting by Grade

By using the CCR method to forecast a population by age, the student population of a given area (e.g., a school district) can be generated. Because the age groups generated by the CCR method are associated with K–12 enrollments, the latter can then be generated using an "enrollment rate" approach (Swanson and Tayman 2012: 127–128). In turn, future enrollment rates can be generated using several methods, including the shift method, which forecasts changes in the rates into the future (Swanson and Tayman 2012: 128–130). Because we describe the CCR method in detail elsewhere (Chapters 1, 4 and 8 for example) and give examples of population forecasts generated by it, we will move directly to a discussion of the participation rate method and then to the shift method. Next, we provide an example that illustrates all of the combined elements, the CCR method, the enrollment rate method, and the shift method.

By itself, the CCR method does not directly generate enrollment forecasts. However, by embedding "enrollment rates" in a CCR method forecast, enrollment forecasting is easily done (George et al. 2004). In the "Enrollment Rate" approach, current and historical data are used to construct proportions (the rate) of the population that have the characteristic of interest (e.g., enrollment in a given grade level). These ratios are forecast into the future. The forecast rates are then applied to population forecasts by age (and other characteristics) to obtain enrollment forecasts. Chapter 2 provides more detail on the participation rate forecasting method. The first step is to obtain an enrollment "rate." Because we are discussing enrollment by grade, we will simply use it as the characteristic in question:

$$R_{i,t} = E_{i,t}/P_{i,t} \quad (7.3)$$

where,

R_i is the Enrollment Rate for grade i,
P_i is the population age i (the age group most closely associated with the grade in question),
E_i is enrollment in grade i (the grade most closely associated with the corresponding population age group), and
t is time.

It is important to note that when 5 year age groups are used, the enrollment rate generated is for the set of grades with which given age groups correspond. We illustrate this point in the example given later for the Memphis, Tennessee School District.

With a set of participation rates, we are ready to generate a forecast. One assumption is to hold the rates constant, but this often yields less satisfactory results than using the shift method. The approach we show here is perhaps the simplest way to implement the shift method (Smith et al. 2013: 2006–211; Swanson and Tayman 2012: 128–131) as shown below:

$$S_{i,t} = R_{i,t}/R_{i,t-k} \quad (7.4)$$

where,

$S_{i,t}$ is the shift in the Enrollment Rate for grade i between time $t-k$ and time t,
$R_{i,t}$ is the Enrollment Rate for grade i at time t (the most recent census),
$R_{i,t-k}$ is the Enrollment Rate for grade i at time $t-k$ (the 2nd most recent census), and
k is the number of years between the years of the most recent census and 2nd most recent census.

Using the shift in the Enrollment Rate, we forecast the enrollment rate by:

$$R_{i,t+k} = S_{i,t} \times R_{i,t} \quad (7.5)$$

where,

$R_{i,t+k}$ is the forecasted enrollment rate for grade i at time $t+k$, and
$S_{i,t}$ and $R_{i,t}$ are defined as before.

Finally, enrollment is forecast by applying the forecast enrollment rate for a given grade to its corresponding age group generated by the CCR forecast:

$$E_{i,t+k} = R_{i,t+k} \times P_{i,t+k} \tag{7.6}$$

where,

$E_{i,t+k}$ is the forecast enrollment in grade i at time $t + k$,
$R_{i,t+k}$ is the forecast enrollment rate for grade i at time $t + k$,
$P_{i,t+k}$ is the forecast population age i (the age group most closely associated with the grade in question) at time $t + k$, and
t and k are as defined before.

As an example of the CCR participation rate and shift methods, we use public school enrollment in Memphis, Tennessee School District. These data consist of three grade groups for school years 1989–1990 and 1999–2000: (1) pre-kindergarten and kindergarten (PreK&K); (2) grades 1–8; and (3) grades 9–12. Fall public school enrollments represent the school years. That is, for 1989–1990 we have fall 1989 enrollment and for 1999–2000 we have fall 1999 enrollment. The fall enrollments work well with the census data since the latter are as of April 1st the following year. The CCR method uses population in 5 year age groups for the Memphis School District in 1990 and 2000. Age groups 0–4, 5–14, and 15–19 are associated with grades PreK&K, 1–8, and 9–12, respectively. Using these data and the CCR and enrollment rate methods, we forecast the population by age and enrollment by grade group to 2010. Selecting this "historical" forecast for 2010 allows us to conduct an ex post facto evaluation of the accuracy of the population and enrollment forecasts by comparing them to the corresponding 2010 data, respectively. This is done in the following Sect. 7.4.

First, we forecast the population of the Memphis, Tennessee School District by age. The 1990 and 2000 input data as well as the CCRs and forecast population for 2010 are shown in Table 7.2. The total population increased by 13.8% between 2000 and 2010. Growth in the school-age groups exceeds the percentage increase in the total population, except for ages 0–4. The fastest growing age group is 45–59, reflecting the aging of the baby boom cohorts.

Second, we calculate enrollment participation rates by grade group for fall 1989 and fall 1999 using the corresponding age groups from the 1990 and 2000 censuses. We use the shift method, described earlier, to forecast these rates to 2010 based on trends in the previous decade. Finally, we convert enrollments into expected 2010 enrollment levels by grade using the 2010 forecast for the school-age population. The results are shown in Table 7.3

We use grades 1–8 to illustrate the enrollment forecasting method. The enrollment rate for the population aged 5–14 found in grades 1 through 8 in 1990 is 0.67155 and by 2000 this rate increases to 0.70221. The ratio of these two rates (1.04566) is multiplied by the 2000 rate to obtain the forecast 2010 enrollment rate

Table 7.2 Population forecast, Memphis, Tennessee School District, 2010

Age	1990 Population	2000 Population	CCR[a]	2010 Population[b]	Change 2000–2010 Number	Change 2000–2010 Percent
0–4	49,394	50,175	0.32991	54,548	4,373	10.4%
5–9	45,864	52,315	0.34670	53,425	1,110	16.5%
10–14	42,989	50,620	1.02482	51,420	800	19.6%
15–19	44,945	47,180	1.02869	53,816	6,636	19.7%
20–24	49,093	50,720	1.17984	59,724	9,004	21.7%
25–29	55,132	53,960	1.20058	56,643	2,683	2.7%
30–34	55,054	47,405	0.96562	48,976	1,571	−11.0%
35–39	47,060	49,530	0.89839	48,477	−1,053	3.0%
40–44	39,358	49,520	0.89948	42,640	−6,880	8.3%
45–49	29,722	44,745	0.95081	47,094	2,349	58.4%
50–54	26,256	35,690	0.90680	44,905	9,215	71.0%
55–59	24,275	26,240	0.88285	39,503	13,263	62.7%
60–64	26,747	20,780	0.79144	28,246	7,466	5.6%
65–69	24,813	19,185	0.79032	20,738	1,553	−16.4%
70–74	18,803	17,215	0.64362	13,374	−3,841	−28.9%
75–79	14,349	16,310	0.65732	12,611	−3,699	−12.1%
80–84	9,126	9,710	0.51641	8,890	−820	−2.6%
85+	7,357	8,530	0.27666	9,559	1,029	29.9%
Total	610,337	649,830		694,589	44,759	13.8%
Annual growth rate		0.63%		0.67%		

Source: 1990 and 2000, National Center for Education Statistics, Special Census Tabulation for School Districts (https://nces.ed.gov/datatools/index.asp?DataToolSectionID=4)

[a] ${}_4P_{0,t}/{}_{15}P_{20,t}$ Ages 0–4
${}_9P_{5,t}/{}_{15}P_{25,t}$ Ages 5–9
$P_{x,t}/P_{x-10,t-10}$ Ages 10–84
$P_{85+,t}/P_{75+,t-10}$ Ages 85+

[b] ${}_4CCR_{0,t} \times {}_{15}P_{20,t+10}$ Ages 0–4
${}_9CCR_{5,t} \times {}_{15}P_{25,t+10}$ Ages 5–9
$CCR_{x,t} \times P_{x,t}$ Ages 10–84
$CCR_{85+,t} \times P_{75+,t}$ Ages 85+

(0.73427). This forecast rate is then multiplied by the forecast population aged 5–14 in 2010 (104,845) to obtain the forecasted enrollment in grades 1 through 8 (76,985).

Public school enrollment increases by 9,740 students between 1990 and 2000, an increase of 8.9%. The school-age population (student population) increases at a slower pace (6.5%). This difference is due to grades under 9, which show increasing participation rates between 1990 and 2000 that are assumed to continue for the next 10 years. Conversely, the participation rates for grades 9 through 12 declines by 2.4%, and its enrollment growth is slower than the population growth in the corresponding age group (15–19).

This example shows how both the student population by age and the enrollment by grade can be forecast using the CCR method in conjunction with participation rates by grade/age group and the shift method. To get individual age groups (e.g., age 6, age 7, age 8, and so forth) the same “age-splitting” methods cited in Chapter 9 and elsewhere (Smith et al. 2013:279–284; Judson and Popoff 2004) can be applied. These same methods can also create enrollment by individual

Table 7.3 Public school enrollment forecast, Memphis, Tennessee School District, 2009–2010

	Population[a]			Change 2000–2010	
Age	1990	2000	2010	Number	Percent
0–4	49,394	50,175	54,548	4,373	8.7%
5–14	88,853	102,935	104,845	1,910	1.9%
15–19	44,945	47,180	53,816	6,636	14.1%
Total	1,83,192	2,00,290	2,13,209	12,919	6.5%

Grade	Enrollment[b]		Enrollment Rate[c]				Enrollment[f]	Change 2000–2010	
	1989–1990	1999–2000	1989–1990	1999–2000	Shift[d]	2009-2010[e]	2009–2010	Number	Percent
Pre K & K	8,415	9,432	0.17036	0.18798	1.10343	0.20742	11,314	1,882	20.0%
1–8	59,669	72,282	0.67155	0.70221	1.04566	0.73427	76,985	4,703	6.5%
9–12	27,098	27,771	0.60291	0.58862	0.97630	0.57467	30,926	3,155	11.4%
Total	95,182	1,09,485					119,225	9,740	8.9%

[a]From Table 7.2
[b]National Center for Education Statistics, ELS*i table*Generator (http://nces.ed.gov/ccd/elsi/tableGenerator.aspx)
[c]$E_{i,t}/P_{i,t}$ and $E_{i,t-10} / P_{i,t-10}$
[d]$R_{i,t}/R_{i,t-10}$
[e]$R_{i,t} \times S_{i,t}$
[f]$R_{i,t+k} \times P_{i,t+k}$

grade level (e.g., grade 1, grade 2, grade 3, and so forth). Similarly, interpolation methods can be applied to obtain age groups and grade levels for the years between the launch year and the horizon year, which in this example would be 2001, 2002, and so on to 2010 (Smith et al. 2013: 273–278; Judson and Popoff 2004).

7.4 Evaluation

In this section we provide two evaluation examples. The first is for the 2015 forecast of enrollment by individual grade we prepared for the Riverside, California Unified School District. To evaluate this short-term enrollment forecast, we compared 2015 actual enrollments by grade to the forecast enrollments by grade for 2015. Table 7.4 provides this summary.

As can be seen in Table 7.4, the GPR method provides highly accurate results. For all grades, the forecast is low by only 121 students or −0.3%. In terms of the K-12 enrollments (excluding the ungraded and adult learners), the MAPE is 1. 9, indicating that on average the GPR method errs by only 1.9%. In terms of basis (tendency to over-forecast or under-forecast), the MALPE is −0.2%, which indicates a very slight overall under-forecast of the actual fall 2015 enrollment.

Table 7.4 Public school enrollment forecast error, Riverside, California Unified School District, Fall 2015

	Fall 2015		Error	
Grade	Actual[a]	Forecast[b]	Number	Percent
Pre K & K	3,393	3,305	−88	−2.6%
First	2,891	3,135	244	8.4%
Second	3,105	3,110	5	0.2%
Third	3,193	3,191	−2	−0.1%
Fourth	3,323	3,301	−22	−0.7%
Fifth	3,191	3,134	−57	−1.8%
Sixth	3,129	3,101	−28	−0.9%
Seventh	3,351	3,270	−81	−2.4%
Eighth	3,342	3,238	−104	−3.1%
Ninth	3,445	3,384	−61	−1.8%
Tenth	3,413	3,385	−28	−0.8%
Eleventh	3,292	3,319	27	0.8%
Twelfth	3,332	3,389	57	1.7%
Ungraded (9–12)	62	79	17	27.4%
Total	42,462	42,341	−121	−0.3%
MALPE (K-12 only)	−0.2%			
MAPE (K-12 only)	1.9%			

[a]California Department of Education DataQuest (http://dq.cde.ca.gov/dataquest/)
[b]From Table 7.1

Table 7.5 School-age population forecast error, Memphis, Tennessee School District, 2010

	2010		Error	
Age	Actual[a]	Forecast[b]	Number	Percent
5–9	44,404	53,425	9,021	20.3%
10–14	45,249	51,420	6,171	13.6%
15–19	49,833	53,816	3,983	8.0%
Total	139,486	158,661	19,175	13.7%
MALPE	14.0%			
MAPE	14.0%			

[a]Special tabulation of the 2010 census by Cropper GIS (croppergis.com)
[b]From Table 7.2

Table 7.6 Public school enrollment forecast error, Memphis, Tennessee School District, 2009–2010

	2009–2010		Error	
Grade	Actual[a]	Forecast[b]	Number	Percent
Pre K & K	10,521	11,314	793	7.5%
1–8	65,664	76,985	11,321	17.2%
9–12	33,115	30,926	−2,189	−6.6%
Total	109,300	119,225	9,925	9.1%
MALPE	6.0%			
MAPE	10.4%			

[a]National Center for Education Statistics, ELS*i* *table*Generator (http://nces.ed.gov/ccd/elsi/tableGenerator.aspx)
[b]From Table 7.3

For the long-term forecast, we first look at the forecast of the school-age population of the Memphis School District for the fall of 2010. This result is found in Table 7.5

Overall, the forecast of the student population (people aged 5–19) is reasonably accurate, being high by 13.7%. One item of interest is that all three age groups (5–9, 10–14, and 15–19) are over-forecast. Thus the MALPE is the same as the MAPE, indicating that the average error over the three age groups is 13.7% too high.

Table 7.6 shows the public school enrollment forecast for the three grade groups, PreK&K, grades 1–8, and grades 9–12. The MAPE is 10.4%, which indicates that on average, the forecast by grade level errs by about 10%. Unlike the forecast of the population age groups, not all of the grade groups are over-forecast. The forecast of grades 9–12 is −6.6% less than the actual enrollment. The other two grade groups are, however, over-forecast. Grades 1–8 are over-forecast by 17.2% and PreK&K by 7.5%. Taken as a whole, total public school enrollment is over-forecasted by 9.1%.

7.5 Conclusions

In this chapter, we provide examples of how the Cohort Change Ratio method can be used for both short- and long-term student population and enrollment forecasts. There is a great deal of work that goes into a good enrollment forecast and we have only touched the surface. A good source for details on methods is Hussar and Bailey (2011). Other sources include materials at the websites of demographic organizations that specialize in school enrollment and related forecasts. Four examples, two in the private sector and two in the public sector, are given below:

1. Lapkoff and Gobalet Demographic Research Inc. (http://www.demographers.com/about_us.htm);
2. McKibben Demographic Research, LLC (http://www.mckibbendemographics.com);
3. Population Research Center, Portland State University (https://www.pdx.edu/prc/about-prc); and
4. Demographic Research Unit, California Department of Finance (http://www.dof.ca.gov/Forecasting/Demographics/).

References

California Department of Finance. (1954). *Projections of public school enrollment in California to 1960 and 1965*. Sacramento: California Department of Finance.

Demographics Research Group. (2014). *Public school enrollment trends in Virginia*. Charlottesville: Demographics Research Group, Weldon Cooper Center for Public Service, University of Virginia.

George, M., Smith, S. K., Tayman, J., & Swanson, D. (2004). Population projections. In J. Siegel & D. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 561–602). San Diego: Elsevier Academic Press.

Hauser, P., & Kitagawa, E. (1961). On estimation of school population. *Proceedings of the American Statistical Association, 1961*, 64–70.

Hussar, W., & Bailey, T. (2011). *Projections of education statistics to 2020*. Washington, DC: Institute of Education Science, National Center for Education Statistics, U.S. Department of Education.

Judson, D., & Popoff, C. (2004). Selected general methods. In J. Siegel & D. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 677–732). San Diego: Elsevier Academic Press.

Lapkoff, S. (2008). *K-12 school enrollment projections study*. Report prepared for the Washington State Office of Superintendent of Public Schools. Retrieved from http://www.k12.wa.us/SchFacilities/Publications/pubdocs/EnrollProjectionMethodologiesFinalReport2008.pdf

Rynerson, C., & Chun, N. (2015). *North Clackamas school district and enrollment forecasts 2015–16 to 2024–25*. Portland, OR: Population Research Center, Portland State University. Retrieved from http://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=1092&context=enrollmentforecasts .

Smith, S. K., Tayman, J., & Swanson, D. (2013). *A practitioner's guide to state and local population projections*. Dordrecht: Springer.

Swanson, D. (2016). *The Washington state census board and its demographic legacy*. Dordrecht: Springer.
Swanson, D., & Tayman, J. (2012). *Subnational population estimates*. Dordrecht: Springer.
Wood, R., Thompson, D., Picus, L., & Tharpe, D. (1995). *Principles of school business management* (2nd ed.). The Association of School Business Officials International: Reston.

Chapter 8
Forecasting Other Characteristics

8.1 Introduction

Chapter 7 focused on forecasts of school enrollment size and composition using a grade progression rate variant of the cohort change rate and participation rates. Our focus now turns to forecasts of other population-related characteristics like households, family structure, labor force, poverty, obesity, and disability needed for planning, budgeting, policy analysis, and program administration. Socioeconomic and health characteristics possess a feature that distinguishes them from strictly demographic characteristics; namely, they are "achieved" rather than "ascribed." Ascribed characteristics such as age, sex, and race/ethnicity are largely set at birth, while achieved characteristics such as marital status, labor force status, and health status change over time (Stark 2007). This distinction is not totally clear-cut, however, because a person's sex or gender classification can be altered and his/her racial and ethnic identity may vary according to the prevailing social context (Alba and Islam 2009; Kaneshiro et al. 2011).

Because achieved characteristics can change substantially over time they are more difficult to forecast accurately than strictly demographic characteristics. Many achieved characteristics are strongly affected by population size and demographic composition, but they are influenced by other factors as well. Forecasts of a population's age structure (and, to a lesser extent, its sex and race/ethnicity structure) provide a basis for forecasting achieved characteristics. This chapter uses the participation-rate (or prevalence-rate method) in which forecasts of achieved characteristics are derived from forecasts of demographic characteristics through the use of rates. The participation rate method is discussed in Chapter 2. We present several studies that have used the participation-rate method to develop forecasts of the U.S. population with disabilities, obesity, and cardiovascular disease. We then develop, using the participation-rate method, population-related forecasts of alcohol consumption, diabetes, cigarette use and consumption, labor force, and households and related variables.

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_8

8.2 Studies Using the Participation-Rate Method

8.2.1 Disability in the United States

The older population in the U.S. is large and growing rapidly. There were 48 million persons aged 65 and older in 2015, representing 15% of the total population. This population is forecast to reach almost 88 million by 2050, or 22% of the total population (U.S. Census Bureau 2014). Since many types of disability rates rise with age, the aging of the population is likely to bring substantial increases in the number of disabled persons and associated costs (Bhattacharya et al. 2004).

Smith, Rayer, and Smith (2008) used the participation-rate method to forecast the number of persons with a particular type of disability; namely, mobility limitations. First, they constructed mobility limitation rates by age and sex using population data from the 2000 census and data on mobility limitations from the 2000 census microdata files. Two definitions of disability were analyzed. One definition related to persons with long-lasting conditions that substantially limit one or more physical activities (DIS-1); and the other related to persons with self-care limitations lasting 6 months or more (DIS-2). They developed three scenarios regarding changes in those rates between 2000 and 2050. Under the medium scenario, rates were forecast to remain constant through 2050. Under the low and high scenarios, they were forecast to fall or rise by 5% per decade, respectively. They applied the forecasted rates to forecasts of the U.S. population and households by age and sex developed from the ProFamy model (Zeng et al. 2006).

Under the medium scenario, the number of persons with mobility limitations was forecast to grow by 106% (DIS-1) and 127% (DIS-2) between 2000 and 2050. Even under the low scenario, the number of disabled persons grew more rapidly than the total population (46%); 59% for DIS-1 and 76% for DIS-2. Under the high scenario, the number of disabled persons grew by 163% (DIS-1) and 190% (DIS-2). They also forecasted an increase in length of time at least one disabled person will reside in a single family residence; rising from an average of 17.6 years in 2000 to an average of 21.2 years in 2050.

8.2.2 Obesity in the United States

Obesity is a common health issue associated with poorer mental health outcomes, diabetes, heart disease, strokes, and some types of cancer.[1] Obesity rates for adults aged 20 rose from 22.9% between 1988 and 1994 to 34.9% between 2011 and 2012 and the extremely obese rate more than doubled from 2.8% to 6.4% over the same period (Fryar et al. 2014). Roughly 17% of children aged 2–19 were obese in

[1]Obese is defined as a body mass index greater than or equal to 30.0 kg/m^2. Extremely obese is a body mass index greater than or equal to 40 kg/m^2.

2011–2012 (Ogden et al. 2014). Childhood obesity rates have stabilized in the last decade, but no stabilization has occurred in adult rates. There is a link between rising obesity rates and rising medical expenses. In 1998 medical costs of obesity were estimated to be $78.5 billion and may have reached $147 billion a decade later (Finkelstein et al. 2009); although, these estimates may be far too low (Cawley and Meyerhoefer 2012). Some believe that current trends in obesity could lessen future gains in life expectancy or cause life expectancy to decline in the United States (Olshansky et al. 2005; Walls et al. 2012).

Arterburn et al. (2004) used the participation-rate method to develop short-term forecasts of obesity in elderly (ages 60 years and older) Americans for 2010. Age- and sex- specific estimates of obesity prevalence were obtained from national surveys conducted from 1960 to 2000, which enabled them to track changes in the prevalence rates by four birth cohorts over time.[2] Three scenarios were used to provide a range of forecast obesity prevalence: 1) the Best-Case scenario assumed the obesity prevalence would change at the lowest absolute rate observed over the four birth cohorts; 2) the Middle-Case scenario assumed the obesity prevalence would change at the mean rate over the four birth cohorts; and 3) the Worst-Case scenario assumed that the dramatic changes occurring in the 1990s would continue until 2010. To conduct sensitivity analyses, the above scenarios were applied to low, middle, and high 2010 national population forecasts developed by the U.S. Census Bureau.

Thirty-two percent of the population aged 60 years and older was obese in 2000. By 2010 the obesity prevalence increased to between 33.6% (Best-Case) and 39.6% (Worst-Case). Using middle series population projections, the number of obese Americans increased from 14.6 million in 2000 to 18.8 million (Best-Case) and 22.2 million (Worst-Case) in 2010. Even under the best case scenario the number of obese elderly increased by more than four million. The Worst-Case scenario for obesity prevalence (39.6%) turned out to be a very accurate forecast for 2010 when compared to the estimated prevalence for 2009–2010 of 39.7% (Ogden et al. 2012).

8.2.3 Cardiovascular Disease in the United States

Death rates from cardiovascular disease (CVD) have declined precipitously; the rate (per 100,000) decreased from 520.4 in 1969 to 169.1 in 2013, a drop of 67.5% (Ma et al. 2015). Despite this progress, cardiovascular disease remains the number one cause of death for both men and women. Around 610,000 people die of CVD each year in the United States (about 1 in every 4 deaths), and each year 735,000

[2]The birth cohorts were: 1911–1920, 1921–1930, 1931–1940, and 1941–1950. The surveys analyzed were the National Health Examination Survey 1959–1962, National Health and Nutrition Examination Survey (NHANES) I 1971–1973, NHANES II 1976–1980, NHANES III 1988–1994, and NHANES 1999–2000.

Americans have a heart attack (Centers for Disease Control and Prevention 2015a; Mozaffarian et al. 2015). Since the prevalence of CVD is heavily concentrated in those aged 60 and over (especially those aged 80 years and older) (Mozaffarian et al. 2015), the aging of the population is likely to bring substantial increases in the number of persons affected by CVD and the associated costs.

Heidenreich et al. (2011) used the participation-rate method to forecast the future of CVD in the United States for both prevalence and cost to the year 2030. They generated historical estimates of CVD prevalence by age (18–44 years, 45–64 years, 65–79 years, and 80 years and older), sex, and race/ethnicity (white non-Hispanic, white Hispanic, black, and other) using data from the 1996 to 2006 National Health and Nutrition surveys. CVD prevalence rates were held constant over the forecast horizon. These rates were applied against the 2008 U.S. Census Bureau forecast for the years 2010–2030 to forecast the number of people with CVD by age, sex, and race/ethnicity.

Forecasts were prepared for both direct and indirect expenses of CVD. The main data sources for direct medical expenses were the 2001–2005 Medical Panel Surveys (MEPS). Logistic regression was used to estimate per capita medical expenses by age, sex, and race/ethnicity (from MEPS) stratified by CVD condition (e.g., hypertension, congestive heart failure). Adjustments were made to remove double counting of expenses (e.g., when more than one condition is treated during a visit) and to account for nursing home expenses (from the 2004 National Nursing Home Survey and National Health Accounts). Per capita direct expenses were adjusted by assuming the same future growth rate (annual rate of 3.6%) to 2030 in health care expenses above and beyond those due to aging and population growth forecasted by the Congressional Budget Office. Future direct expenses for CVD were obtained by applying forecasted per capita expenses against the forecasted number of people with CVD.

Indirect costs represented the value of foregone earnings due to lost productivity from morbidity and premature mortality by age, sex, and race/ethnicity. Morbidity expenses included three components: 1) work loss among employed persons; 2) home productivity loss; and 3) work loss among people too sick to work. Work loss expenses were derived from per capita work loss days due to CVD (from MEPS), adjusted for double counting. Per capita work loss days multiplied by the probability of employment given CVD (from MEPS) and the mean per capita earning per day (from the Current Population Survey) provided estimates of the work loss expenses. Home loss expenses were derived from per days in bed due to CVD (from MEPS) multiplied by the dollar value of a day of housework. Work loss expenses for sickness due to CVD were derived from estimates of the number of people too sick to work who would have been employed except for their CVD multiplied by their mean annual earnings.

Mortality costs were based on 2006 death rates due to CVD by age, sex, and race/ethnicity. These rates multiplied by population forecasts determined the number of deaths due to CVD to 2030. Total mortality expenses were based on the forecast deaths multiplied by remaining life time earnings by age, sex, and race/

ethnicity. Indirect expenses of CVD were assumed to grow at the assumed annual rate of real earnings growth to 2030 (1.4%).

By 2030, 116 million people are forecasted to have some form of CVD, an increase of 42 million from 2010 (36%). The crude prevalence rate increased from 36.9% of U.S. adults having CVD in 2010 to 40.5% in 2030. This forecast assumed no change in CVD prevalence rates but reflected the demographics of an aging population. Direct expenses are forecasted to triple from $272.5 billion in 2010 to $818.1 billion in 2030. The aging of the population had less impact on indirect expenses than on direct expenses because of lower employment rates among persons aged 60 years and older. Indirect expenses are forecast to increase by 61% from 171.7 billion in 2010 to $275.8 billion in 2030. By 2030, the forecast total expenses for CVD exceeded $1 trillion.

8.3 Developing Population-Related Forecasts

8.3.1 Alcohol Consumption in the United States

Alcohol use is very common and can increase the risk of many harmful health conditions. Excessive alcohol consumption either in the form of heavy drinking or binge drinking leads to an average of 80,000 deaths in the United States each year and cost $223.5 billion in 2006 (Sacks et al. 2013). The prevalence of binge drinking in the United States has been relatively stable, remaining in a range of 14–17% of adults from 1993 to 2013, while the prevalence of heavy drinking rose from 3% to 6% of adults from 1993 to 2002 and has been stable since that time (Centers for Disease Control and Prevention (2015b).

We prepare a forecast of alcohol use in the United States for the year 2025 based on 2013 prevalence use rates by age for current users, binge users, and heavy users (U.S. Department of Health and Human Services (2014).[3] The prevalence use rates represent the percent of the population in each age group that falls into each alcohol user category. The binge and current use categories are not mutually exclusive; heavy use is included in the estimates of binge use, and heavy use and binge use are included in the estimate of current use. To avoid double and triple counting of alcohol users, we create a mutually exclusive set of prevalence use rate. Binge users are separated by subtracting the heavy use rate from the published binge use rate, and current users are separated by subtracting the adjusted binge use and heavy use rates from the published current use rate.

[3]Current use is at least one drink in the past 30 days. Binge use is five or more drinks on the same occasion (i.e., at the same time or within a couple of hours of each other) on at least one day in the past 30 days. Heavy use is five or more drinks on the same occasion each of five or more days in the past 30 days.

Table 8.1 Alcohol consumption, United States, 2025

	2013 Use Rate (per 100 pop)				2025 Alcohol Consumption			
Ages	Current[a]	Binge[b]	Heavy[c]	2025 Pop	Current[d]	Binge[e]	Heavy[f]	Total Use
12–13	1.3	0.7	0.1	8188,490	106,450	57,319	8,188	171,957
14–15	5.0	3.8	0.7	8,259,785	412,989	313,872	57,818	784,679
16–17	9.6	10.4	2.7	8,429,685	809,250	876,687	227,601	1,913,538
18–20	14.7	20.6	8.5	13,005,283	1,911,777	2,679,088	1,105,449	5,696,314
21–25	26.0	30.2	13.1	22,359,370	5,813,436	6,752,530	2,929,077	15,495,043
26–29	29.0	28.8	11.2	18,473,712	5,357,376	5,320,429	2,069,056	12,746,861
30–34	28.4	24.7	10.5	24,449,875	6,943,765	6,039,119	2,567,237	15,550,121
35–39	30.6	22.1	7.5	23,585,874	7,217,277	5,212,478	1,768,941	14,198,696
40–44	35.0	18.4	7.3	22,290,783	7,801,774	4,101,504	1,627,227	13,530,505
45–49	33.1	18.9	6.9	20,613,122	6,822,943	3,895,880	1,422,305	12,141,128
50–54	36.9	17.4	5.6	20,062,792	7,403,170	3,490,926	1,123,516	12,017,612
55–59	36.6	12.0	3.9	20,293,826	7,427,540	2,435,259	791,459	10,654,258
60–64	39.5	9.4	4.7	21,265,225	8,399,764	1,998,931	999,466	11,398,161
65+	32.6	7.0	2.1	65,919,552	21,489,774	4,614,369	1,384,311	27,488,454
12+				297,197,374	87,917,285	47,788,391	18,081,651	153,787,327
				2013	78,318,166	44,491,006	16,878,719	139,687,891
				2013–2025	12.3%	7.4%	7.1%	10.1%

Sources: U.S. Census Bureau (2014); U.S. Department of Health and Human Services (2014).
[a]No more than 4 drinks on one occasion in the past 30 days
[b]Five or more drinks on one to four occasions in the past 30 days
[c]Five or more drinks on five or more occasions in the past 30 days
[d]2013 current use rate × 2025 population
[e]2013 binge use rate × 2025 population
[f]2013 heavy use rate × 2025 population

The 2013 current use rates are shown in Table 8.1. The binge and heavy use rates have similar age-specific patterns. They both peak at ages 21–25 and decline consistently for ages above 25 years. The current use rate is similar in magnitude to the binge use rate to ages 21–25, but the current use rate shows a continuing increase until ages 60 to 64 and then declines for the oldest age group (65 years and older). Table 8.1 shows the alcohol use forecast for 2025, which is derived by applying the 2013 mutually exclusive prevalence rates to the most recent U.S. population forecast (U.S. Census Bureau 2014). Since the prevalence rates are held constant, this forecast shows the impact of the aging population on alcohol use.

There are 139.7 million alcohol users in 2013 (52.2% of the population aged 12 years or older).[4] In 2025, there are 153.8 million users of alcohol, an increase of 10.1% from 2013. By comparison, the total population increases by 11.1%. Current alcohol users increase the fastest by 12.3% between 2013 and 2025, while binge and heavy users show similar and lower increases of 7.4% and 7.1%, respectively. Prevalence rates for current users are relatively high in ages 60 years and older which show the fastest population growth, while the population declines or grows much slower in the age groups (21–29) with the highest binge and heavy use prevalence rates. Consequently, 51.7% of the population aged 12 years and older uses alcohol in 2025, down slightly from 52.2% in 2013; the percentage of current users rises from 29.3% in 2013 to 29.6% in 2025; and the percentage of binge and heavy use decreases from 22.9% to 22.1%.

8.3.2 Diabetes in the United States

There are significant personal impacts in terms of reduced quality of life and suffering of people with diabetes. Additionally diabetes places a substantial economic burden in the form of direct medical costs and indirect costs from work-related absenteeism, reduced productivity and labor force participation, and premature mortality. In 2012, the total economic cost in the U.S. from diabetes was estimated at $245 billion, including $176 billion in direct expenses and $69 billion in indirect costs (American Diabetes Association 2013). As Fig. 8.1 shows, the prevalence of the population with diagnosed diabetes continues to rise. The prevalence rate was relatively stable in each age group from 1980 to 1990, but it has shown a substantial increase over the last 20 years. On average, the prevalence rate grew by around 4% annually from 1990 to 2011 in every age group.

[4]The number of alcohol users from the 2013 Drug Survey (136.9 million) is about two percent lower than our estimate. The Drug Survey estimate is based on the non-institutional population, whereas our estimate is based on the total population. The total population in the United States is roughly two percent higher than the non-institutional population.

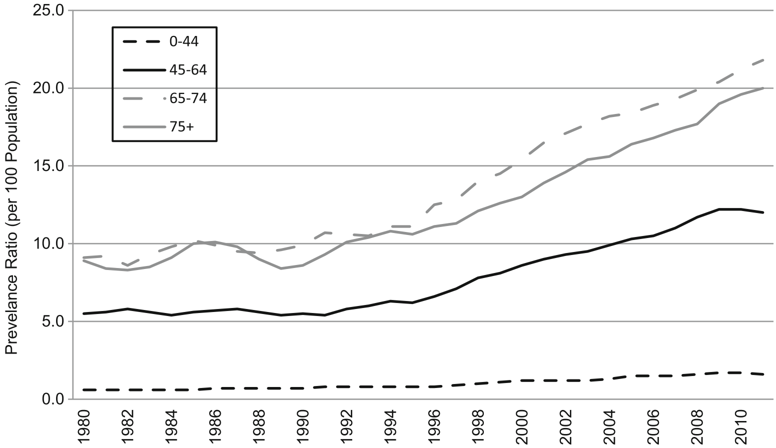

Fig. 8.1 Diagnosed diabetes prevalence rate by age, United States, 1980–2011 (Source: National Center for Health Statistics 2015)

We prepare a forecast of persons with diagnosed diabetes in the United States for the year 2025 based on prevalence rates by age (National Center for Health Statistics 2015). The prevalence rate represents the percent of the population diagnosed with diabetes in each age group. We create two scenarios based on different assumptions about the prevalence rate. The first scenario (Constant) holds the prevalence rate constant for each age group at their 2013 level.[5] The second scenario (Trend) assumes the prevalence rate changes linearly from 2013 to 2025 based on the 1980–2013 base period:

$$2025PR_a = 2013PR_a \times (1 + ((PR2013_a - PR1980_a) \times 12/33) \tag{8.1}$$

where,

PR is the prevalence rate,
a is the age group,
12 is the length of the forecast horizon (2013–2025), and
33 is the length of the base period (1980–2013).

As Table 8.2 shows, persons diagnosed with diabetes continues to increase in the United States to the year 2025, increasing by almost 4.4 million (20.3%) in the Constant scenario and nearly doubling in the Trend scenario (93.1%). By far the

[5]The latest prevalence rates provided by the National Center for Health Statistics were for 2011. To estimate rates for 2013, we adjusted the 2011 prevalence rates so when applied to the 2013 population by age the result would match the latest estimate of the number of persons diagnosed with diabetes in the United States from the Centers for Disease Control and Prevention (CDCP), adjusted upward for the difference in population definition (21.6 million). CDCP numbers are based on non-institutional population and we are using total population.

Table 8.2 Persons with diagnosed diabetes, United States, 2025

				Persons with diagnosed diabetes				
	2025 Preval. Rate				2025		Pct. Chg. 2013–2025	
Ages	Constant[a]	Trend[b]	2025 Pop	2013[c]	Constant[d]	Trend[e]	Constant	Trend
0-44	1.5	2.4	199,180,395	2,825,112	2,987,706	4,780,329	5.8%	69.2%
45–64	11.6	18.6	82,234,965	9,637,740	9,539,256	15,295,703	−1.0%	58.7%
65–74	21.1	33.9	37,093,437	5,320,738	7,826,715	12,574,675	47.1%	136.3%
75+	19.5	31.3	28,826,115	3,800,025	5,621,092	9,022,574	47.9%	137.4%
Total			347,334,912	21,583,615	25,974,769	41,673,281	20.3%	93.1%

Sources: National Center for Health Statistics (2015); U.S. Census Bureau (2014)
[a]2013 prevalence rate
[b]Linear extrapolation of the 1980–2013 trend applied to the 2013 prevalence rate.
[c]2013 prevalence rate × 2013 population
[d]Constant rate × 2025 population
[e]Trend rate × 2025 population

largest increase in persons diagnosed with diabetes occurs in the elderly population (ages 65 years and older). In the Constant scenario, 98.5% of the increase occurs in the elderly population, while in the Trend scenario the corresponding figure is 62.1%. In 2013 6.2% of the total population is diagnosed with diabetes, rising to 7.5% and 12.0% by 2025 in the Constant and Trend scenarios, respectively. By 2025, roughly 20% of the elderly population will be diagnosed with diabetes in the Constant scenario. In the Trend scenario that percentage increases to almost one-third.

8.3.3 Cigarette Use and Consumption in the United States

Cigarette consumption in the U.S. grew from 2.5 billion in 1900 to a peak of 640 billion in 1981, and then declined to a level of under 300 billion in 2013 (ISH Global Inc. 2014). The decline over the last four decades is due to a number of factors including the increased recognition of the adverse health impacts of smoking (including second-hand smoke), proliferation of programs and treatments to help people quit smoking, indoor smoking bans that spread across the United States, and increases in federal and state taxes on cigarettes. Both the percent of the population using cigarettes and daily cigarette volume varies by age (see Fig. 8.2). For cigarette use the pattern is N-shaped; the percentage reaches a peak at ages 20 and 21, stays relatively stable until ages 26–29, and then declines steadily to ages 65 years and older. For daily consumption, there is a direct relationship with age, with the highest consumption at ages 50–64. From 2003 to 2013 both usage and volume declined in every age group, with the greatest declines occurring in ages 18–25.

Our 2025 forecast of persons using cigarettes and cigarette consumption in the U.S. is based on prevalence use rates by age and daily cigarette volume by age (Substance Abuse and Mental Health Services Administration 2014). The prevalence use rate represents the percent of the population using cigarettes in the last month. The consumption rate represents the average daily number of cigarettes (or sticks) consumed by cigarette users. We create two scenarios based on different assumptions about the prevalence and consumption rates. The first scenario (Constant) holds the prevalence and consumption rates constant for each age group at their 2013 level. The second scenario (Trend) assumes the prevalence rate and consumption rate changes linearly from 2013 to 2025 based on the 2003–2013 trend, using the same logic shown in Eq. 8.1, except the base period is now 10 years while the forecast horizon is still 12 years. We prepare the forecast in two stages. First, we forecast the number of cigarette users from the assumed prevalence rate and the 2025 forecast of the U.S. population. Second, we forecast 2025 cigarette consumption using the cigarette use forecast from the first stage and the assumed consumption rate.

Assuming no change in the prevalence rate, cigarette users increase from 55.6 million in 2013 to 60.0 million in 2025 (see Table 8.3). This increase offsets the

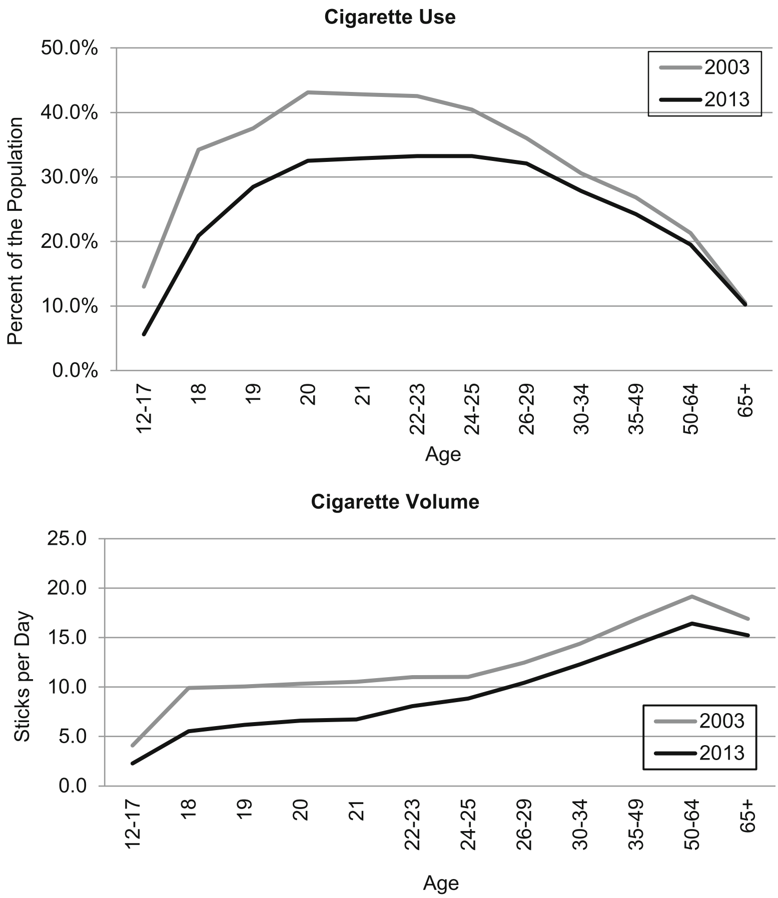

Fig. 8.2 Cigarette use and consumption by age, United States, 2003 and 2013 (Source: Substance Abuse and Mental Health Services Administration 2014)

decline in cigarette users from 2003 to 2013. The Trend scenario shows a continued decline in cigarette users, reaching 52.6 million in 2025. In both scenarios, the change in cigarette users varies by age group, with declines generally seen for ages under 30 years and with the largest increases in cigarette users 65 years and older. One-fifth (20.8%) of the population aged 12 and older uses cigarettes in 2013. By 2025, the percentage declines slightly to 20.2% in the Constant scenario and more steeply to17.7% in the Trend scenario.

The Constant scenario shows a reversal to declining trend of cigarette consumption, with consumption rising by 23 billion cigarettes (9%) between 2013 and 2025 (see Table 8.4). Over one-half of the increase is due to smokers aged 65 and older. The Trend scenario shows continued decline in cigarette consumption from 254.9 billion in 2013 to 221.5 billion in 2025 (17%). In this scenario, cigarette consumption declines in every age group, except for those aged 65 years and older.

Table 8.3 Cigarette Use, United States, 2025

	2025 Use Rate			Persons Using Cigarettes				
					2025		Pct. Chg. 2013–2025	
Ages	Constant[a]	Trend[b]	2025 Pop	2013[c]	Constant[d]	Trend[e]	Constant	Trend
12–17	5.6%	2.4%	24,877,960	1,401,010	1,393,166	597,071	−0.6%	−57.4%
18	20.9%	12.8%	4,331,448	895,429	905,006	554,425	1.1%	−38.1%
19	28.5%	21.6%	4,326,904	1,245,816	1,231,668	934,611	−1.1%	−25.0%
20	32.5%	24.5%	4,346,931	1,445,431	1,413,311	1,064,998	−2.2%	−26.3%
21	32.9%	25.3%	4,389,574	1,502,863	1,443,951	1,110,562	−3.9%	−26.1%
22–23	33.2%	26.0%	8,801,554	3,092,390	2,923,965	2,288,404	−5.4%	−26.0%
24–25	33.2%	27.3%	9,168,242	2,939,315	3,046,830	2,502,930	3.7%	−14.8%
26–29	32.1%	28.6%	18,473,712	5,519,980	5,925,996	5,283,482	7.4%	−4.3%
30–34	27.8%	25.3%	24,449,875	5,912,886	6,798,659	6,185,818	15.0%	4.6%
35–49	24.3%	21.9%	66,489,779	14,957,024	16,128,335	14,561,262	7.8%	−2.6%
50–64	19.5%	17.9%	61,621,843	12,081,035	12,031,478	11,030,310	−0.4%	−8.7%
65+	10.2%	9.9%	65,919,552	4,558,492	6,721,843	6,526,036	47.5%	43.2%
Ages 12+			297,197,374	55,551,671	59,964,208	52,639,909	7.9%	−5.2%

Sources: Substance Abuse and Mental Health Services Administration (2014); U.S. Census Bureau (2014)

[a]2013 use rate (per 100 persons)

[b]Linear extrapolation of the 2003–2013 trend applied to the 2013 use rate.

[c]2013 use rate × 2013 population

[d]Constant use rate × 2025 population

[e]Trend use rate × 2025 population

Table 8.4 Cigarette Consumption, United States, 2025

	Constant Scenario				
	2025		Consumption (in 000"s)		
Ages	Rate[a]	Users[b]	2013[c]	2025[d]	Pct. Chg. 2013–2025
12–17	2.28	1,393,166	1,163,832	1,159,393	−0.4%
18	5.53	905,006	1,808,150	1,826,709	1.0%
19	6.18	1,231,668	2,809,172	2,778,274	−1.1%
20	6.60	1,413,311	3,481,828	3,404,666	−2.2%
21	6.73	1,443,951	3,693,707	3,546,993	−4.0%
22–23	8.06	2,923,965	9,102,920	8,602,013	−5.5%
24–25	8.84	3,046,830	9,488,391	9,830,902	3.6%
26–29	10.42	5,925,996	20,997,200	22,538,341	7.3%
30–34	12.28	6,798,659	26,504,785	30,472,949	15.0%
35–49	14.32	16,128,335	78,187,100	84,299,581	7.8%
50–64	16.41	12,031,478	72,344,800	72,064,342	−0.4%
65+	15.23	6,721,843	25,342,600	37,366,389	47.4%
Ages 12+		59,964,208	254,924,485	277,890,552	9.0%
	Trend Scenario				
	2025		Consumption (in 000"s)		
Ages	Rate[a]	Users[b]	2013[c]	2025[d]	Pct. Chg. 2013–2025
0	1.27	597,071	1,163,832	276,772	−76.2%
12–17	3.09	554,425	1,808,150	625,308	−65.4%
18	3.80	934,611	2,809,172	1,296,305	−53.9%
19	4.22	1,064,998	3,481,828	1,640,416	−52.9%
20	4.30	1,110,562	3,693,707	1,743,027	−52.8%
21	5.90	2,288,404	9,102,920	4,928,078	−45.9%
22–23	7.09	2,502,930	9,488,391	6,477,207	−31.7%
24–25	8.71	5,283,482	20,997,200	16,796,982	−20.0%
26–29	10.49	6,185,818	26,504,785	23,684,569	−10.6%
30–34	12.18	14,561,262	78,187,100	64,735,002	−17.2%
35–49	14.05	11,030,310	72,344,800	56,566,187	−21.8%
50–64	13.74	6,526,036	25,342,600	32,728,723	29.1%
Ages 12+		52,639,909	254,924,485	211,498,576	−17.0%

Source: Substance Abuse and Mental Health Services Administration (2014)
Trend is the linear extrapolation of the 2003–2013 trend applied to the 2013 rate.
[a]Constant is the 2013 daily consumption rate of cigarettes.
[b]From Table 8.3
[c]2013 daily consumption rate × 2013 users × 365
[d]2025 daily consumption rate × 2025 users × 365

8.3.4 Civilian Labor Force Forecast for San Diego County, California

The U.S. civilian labor force has gone through substantial changes in its size and demographic composition. The labor force grew rapidly from 1970 to 1990 due to rising female labor force participation and the baby boom generation entering the labor market. Since then, demographic changes and social forces have dampened labor force growth. The female labor force participation rate (LFPR) peaked in 1999 and baby boomers are starting to exit the workforce due to retirement. Moreover, the severe economic impacts of the 2007–2009 recession caused disruptions in the labor market. Going forward, the aging of the population poses challenges for Social Security, Medicare, and pension programs as the working population is expected to grow much slower than the non-working elderly population. Forecasts of the size and composition of the labor force are useful for understanding the future impacts of these challenges and for other types of economic planning.

Most long-term labor force forecasts use the participation-rate method (Frees, 2006; Loichinger 2015). The Bureau of Labor Statistics (BLS) uses the participation-rate method to forecast the labor force in the United States (Toossi, 2013) from population forecasts developed by the U.S. Census Bureau in conjunction with labor force participation rates developed by BLS. The same method can be used for state and local labor force forecasts. We illustrate this method by preparing a 2025 labor force forecast by age for San Diego County, California. We derive LFPRs by age in 2014 from civilian population in the denominator (San Diego Association of Governments 2014) and persons in the labor force in the numerator (U.S. Census Bureau 2015). We create two scenarios based on different assumptions about the LFPR. The first scenario (Constant) holds the LFPR constant for each age group at its 2014 level. The second scenario (Trend) uses the synthetic method (Smith et al. 2013: 65), which assumes San Diego's LFPR changes at the same rate as the national LFPR (forecast by the BLS from 2012 to 2022).

Comparing the Constant and Trend LFPRs found in Table 8.5, between 2014 and 2025 the LFPR rate declines in all age groups under the age of 55 years, with the largest percentage declines in the ages 15–19 and 20–24 (−22.3% and −5.5%, respectively). LFPRs increase in all ages 55 years and older, with larger percentage increases as the labor force ages. The LFPR in ages 55–59 increases by 4.6% and it increases by 41.3% in ages 85 years and older.

The total labor force increases by 8.1% in the Constant scenario and 9.0% in the Trend scenario. Percentage changes in the labor force show a similar pattern in terms of direction across ages in both scenarios, but are generally of greater magnitude (ignoring the sign) in the Trend scenario. The impacts of changes on the labor force age composition from changes in the civilian population age composition are seen by looking at the Constant scenario. In that scenario, the labor force increases in all age groups, except for 15–24 and 50–59. As expected the largest increases in the labor force occur in the older ages. The median age of the labor force increases from 40.9 years in 2014 to 41.0 years in the Constant scenario and to 42.4 years in the Trend scenario.

Table 8.5 Civilian Labor Force, San Diego County, California, 2025

	2025 LFPR			Civilian Labor Force				
					2025		Pct. Chg. 2014–2025	
Ages	Constant[a]	Trend[b]	2025 Civ. Pop	2014[c]	Constant[d]	Trend[e]	Constant	Trend
15-19	28.2%	21.9%	207,507	65,315	58,517	45,444	−10.4%	−30.4%
20–24	70.5%	66.6%	220,936	160,045	155,760	147,143	−2.7%	−8.1%
25–29	80.0%	79.4%	238,161	177,073	190,529	189,100	7.6%	6.8%
30–34	81.2%	80.5%	244,586	181,211	198,604	196,892	9.6%	8.7%
35–39	82.0%	81.1%	247,234	162,662	202,732	200,507	24.6%	23.3%
40–44	82.3%	81.4%	222,927	167,235	183,469	181,463	9.7%	8.5%
45–49	81.1%	80.8%	211,133	164,716	171,229	170,595	4.0%	3.6%
50–54	77.3%	77.0%	199,335	166,817	154,086	153,488	−7.6%	−8.0%
55–59	69.8%	73.0%	195,752	140,060	136,635	142,899	−2.4%	2.0%
60–64	53.3%	58.2%	211,113	89,484	112,523	122,868	25.7%	37.3%
65–69	30.7%	37.2%	185,907	40,637	57,073	69,157	40.4%	70.2%
70–74	17.8%	22.3%	157,598	15,937	28,052	35,144	76.0%	120.5%
75–79	12.1%	16.2%	115,581	7,932	13,985	18,724	76.3%	136.1%
80–84	7.3%	10.4%	67,235	3,946	4,908	6,992	24.4%	77.2%
85+	4.6%	6.5%	71,837	2,930	3,305	4,669	12.8%	59.4%
Ages 15+			2,796,842	1,546,000	1,671,407	1,685,085	8.1%	9.0%

Sources: San Diego Association of Governments (2014); U.S. Census Bureau (2015)
[a]2014 labor force participation rate (per 100 persons)
[b]Synthetic forecast based on trend in U.S. LFPR from 2012 to 2022
[c]2014 LFPR × 2014 civilian population
[d]Constant LFPR × 2025 civilian population
[e]Trend LFPR × 2025 civilian population

8.3.5 Other Population and Housing Variables for San Diego County, California

Households (occupied housing units) and their populations are important consumers in the goods and services markets and are an important determinant of housing market trends. Forecasts can show the extent to which households and their occupants will change in the future in both number and composition. As such, they provide basis for other forecasts such consumer durables, the need for in-home nursing care services and assisted living, and the need for municipal services. Persons per household (PPH) can be derived from forecasts of households and household population; and housing units (or supply) forecasts can be derived from household forecasts. Uses for these other population and housing variables might include developing economic strategies, designing service delivery programs, evaluating future housing and transportation needs, and preparing marketing and business plans.

Our approach begins with a participation-rate method for developing group quarters and household population forecasts. The household population forecast is then combined with householder rates to forecast households. From households, household population, and assumptions about the housing vacancy rate, we derive forecasts of PPH and housing units. The general formulae for this approach are:

$$\mathrm{GQPop_c} = \mathrm{TPop_c} \times \mathrm{GQR_c}, \tag{8.2}$$

$$\mathrm{HHPop_c} = \mathrm{Tpop_c} - \mathrm{GQPop_c}, \tag{8.3}$$

$$\mathrm{HH_c} = \mathrm{HHPop_c} \times \mathrm{HHR_c}, \tag{8.4}$$

$$\mathrm{PPH} = \sum \mathrm{HHPop_c} / \sum \mathrm{HH_c}, \tag{8.5}$$

$$\mathrm{HU} = \mathrm{HH}/(1 - \mathrm{VR}), \tag{8.6}$$

where,

c is the demographic characteristic
GQpop is population living in group quarters,
Tpop is the total population,
GQR is the group quarters population participation rate,
HHPOP is the population living in households,
HH is the number of households,
HHR is the householder rate,
PPH is the average number of persons per household,
HU is the number of housing units (occupied and vacant), and
VR is the housing vacancy rate.

Housing units can also be forecast directly from data on housing trends, zoning requirements, the amount of buildable land, and other relevant factors (Smith et al. 2013: 296–297). A population forecast can be derived from the housing unit forecast by applying the widely-used housing unit method (Swanson and Tayman 2012: 137–164):

$$\text{Tpop} = \text{HU} \times (1 - \text{VR}) \times \text{PPH} + \text{GQpop}. \tag{8.7}$$

8.3.5.1 Group Quarters and Household Population

We prepare 2025 group quarters and household population forecasts by age for San Diego County, California. We split the group quarters population into civilian and military. In places with a large military presence, like San Diego County, it is useful to make this division. Not only do the demographic characteristics differ between the military and civilian group quarters populations, but the military group quarters population is not likely to be affected by changes in demographic composition of the local population like, say, the nursing home population.[6] Civilian group quarters participation rates (CivGQR) by age in 2014 are developed from the civilian population in the denominator (San Diego Association of Governments 2014) and persons living in civilian group quarters in the numerator (U.S. Census Bureau 2015). The 2014 military group quarters population and CivGQR are held constant in this forecast.

Table 8.6 shows the CIVGQR is relatively similar in size for most ages, generally between 1.5% and 2.5% of the civilian population. The highest rates occur in ages 75 years and older, reflecting the population in elder care facilities. The civilian group quarters population increases by 9,014 (14.2%) from 2014 to 2025, slightly faster than the 12.5% increase for the civilian population. Most of this increase (7,499) occurs in ages 65 years and older, reflecting the aging of the civilian population. The household population reaches 3.49 million in 2025, up from 3.11 million in 2014.

8.3.5.2 Households, Household Size, and Housing Units

We conclude our examples by preparing 2025 forecasts of households by age, PPH, and housing units for San Diego County, California. When a participation-rate is used to forecast households, it is often called the headship or householder rate (HHR) since the principal person (or householder) is usually reported by age (as well as by sex and by ethnicity and race, and so on). The HHR approach is widely used to develop forecasts of households (Goodman et al. 2015; Holmans 2012; Kono 1987; McCue 2014; Reardon and Hari 2014: Appendix B). We derive

[6]If an area has large college group quarters or prison/jail populations, they should be forecast separately from other civilian group quarters for the same reasons the military and civilian group quarters populations handled distinctly.

Table 8.6 Group quarters and household population, San Diego County, California, 2025

Ages	TotPop	UmilPop[a]	CivPop[b]	CivGQ Rate[c]	CivGQPop[d]	MilGQPop[e]	HHPop[f]
0–4	257,384	0	257,384	0.00%	0	0	257,384
5–9	233,226	0	233,226	0.00%	0	0	233,226
10–14	209,896	0	209,896	0.00%	0	0	209,896
15–19	218,267	10,760	207,507	2.01%	4,171	9,828	204,268
20–24	263,870	42,934	220,936	3.22%	7,114	26,237	230,519
25–29	261,757	23,596	238,161	2.25%	5,359	4,675	251,723
30–34	257,042	12,456	244,586	2.48%	6,066	879	250,097
35–39	255,690	8,456	247,234	2.33%	5,761	233	249,696
40–44	226,706	3,779	222,927	2.04%	4,548	40	222,118
45–49	212,458	1,325	211,133	1.88%	3,969	0	208,489
50–54	199,696	361	199,335	1.84%	3,668	0	196,028
55–59	195,895	143	195,752	1.49%	2,917	0	192,978
60–64	211,113	0	211,113	1.84%	3,884	0	207,229
65–69	185,907	0	185,907	2.35%	4,369	0	181,538
70–74	157,598	0	157,598	2.62%	4,129	0	153,469
75–79	115,581	0	115,581	3.70%	4,276	0	111,305
80–84	67,235	0	67,235	8.15%	5,480	0	61,755
85+	71,837	0	71,837	9.24%	6,638	0	65,199
Total	3,601,158	103,810	3,497,348		72,349	41,892	3,486,917

Sources: San Diego Association of Governments (2014); U.S. Census Bureau (2015)
[a]2014 uniformed military population
[b]Total population – uniformed military population
[c]2014 civilian group quarters population /civilian population × 100
[d]Civilian population × CivGQrate
[e]2014 military group quarters population
[f]Tpop – CivGQPop – MilGQpop

HHR by age in 2014 from household population in the denominator (San Diego Association of Governments 2014) and householders in the numerator (U.S. Census Bureau 2015). We create two scenarios based on different assumptions about the HHR. The first scenario (Constant) holds the HHR constant for each age group at its 2014 level. The second scenario (Trend) uses the synthetic method (Smith et al. 2013: 65), which assumes San Diego's HHR changes at the same rate as the national HHR (forecast by Goodman et al. (2015) from 2010 to 2025). To derive housing units in 2025, we assume that the 2014 vacancy rate (San Diego Association of Governments 2014) does not change.

HHR are lowest in ages under 30 years, then rise until ages 45–49, stay relatively constant through ages 75–79, and then decline for ages 80 years and older (see Table 8.7). Comparing the Constant and Trend HHR, between 2014 and 2025, the HHR rate declines in all age groups, with the largest percentage declines in the ages 15–19 and 20–24 (−13.0%), 25–29 (−7.3%), and 30–34 (6.3%). The slowest declines (around 2%) occur in ages 35–54 and the decline increases to around 3% for those 55 years and older. More households are formed under the Constant scenario between 2014 and 2024. Households increase by 165,198 (14.8%) in the Constant scenario compared to 115,578 (10.3%) in the Trend scenario. As a result, under the Constant scenario the PPH in 2025 is −2.3% lower than it is in 2014, but the PPH increases by 1.7% in the Trend scenario. Roughly 52,000 more housing units would have to be built in San Diego County to accommodate the demand under the Constant scenario (173,747) compared to the Trend scenario (121,559).

8.4 Conclusions

Population-related forecasts can be used to address a broad array of socioeconomic and health-related issues. They play an important role in many types of real-world decision making. The participation-rate method described in this chapter is conceptually simple and relatively easy to apply. More complex methods for forecasting households, health status, employment, and other population-related variables have also been developed (Barnichon and Nekarda 2012; Christiansen and Keilman 2013; IHS Global Inc. 2014; Finkelstein et al. 2012; Huang et al. 2009; Lindh and Malmberg 2007; Rowley and Bezold 2012; Zeng et al. 2006). Complex methods draw on a greater variety of inter-relationships among variables and provide a richer array of detailed characteristics than simpler methods; for some purposes, they will be more useful than the methods described here. However, the methods presented in this chapter require considerably less data and can be applied more easily than complex methods. These are important advantages when resources are scarce and time is short. Their relatively small data requirements are particularly important for small-area forecasts because many types of data are not available for them.

Table 8.7 Households, household size, and housing units, San Diego County, California, 2025

	2025 HHRate			Households				
					2025		Pct. Chg. 2014–2025	
Ages	Constant[a]	Trend[b]	2025 HHPop[c]	2014	Constant[d]	Trend[e]	Constant	Trend
0-4	0.0%	0.0%	257,384	0	0	0	0.0%	0.0%
5–9	0.0%	0.0%	233,226	0	0	0	0.0%	0.0%
10–14	0.0%	0.0%	209,896	0	0	0	0.0%	0.0%
15–19	8.6%	7.5%	204,268	19,642	17,606	15,310	−10.4%	−22.1%
20–24	23.5%	20.4%	230,519	55,517	54,124	47,065	−2.5%	−15.2%
25–29	37.3%	34.6%	251,723	87,820	93,988	87,111	7.0%	−0.8%
30–34	45.5%	42.7%	250,097	104,355	113,837	106,721	9.1%	2.3%
35–39	49.9%	48.9%	249,696	100,713	124,553	122,159	23.7%	21.3%
40–44	54.5%	53.5%	222,118	110,533	121,121	118,791	9.6%	7.5%
45–49	55.0%	54.0%	208,489	110,453	114,769	112,603	3.9%	1.9%
50–54	54.7%	53.7%	196,028	116,052	107,198	105,175	−7.6%	−9.4%
55–59	55.5%	53.9%	192,978	109,864	107,126	103,928	−2.5%	−5.4%
60–64	55.0%	53.1%	207,229	90,658	114,013	110,014	0.0%	0.0%
65–69	56.3%	54.5%	181,538	72,898	102,293	98,993	0.0%	0.0%
70–74	54.2%	52.4%	153,469	47,184	83,116	80,435	0.0%	0.0%
75–79	56.6%	54.8%	111,305	35,870	62,998	60,965	75.6%	70.0%
80–84	53.8%	52.1%	61,755	26,694	33,220	32,149	24.4%	20.4%
85+	51.3%	49.6%	65,199	29,929	33,418	32,341	11.7%	8.1%
Total			3,486,917	1,118,182	1,283,380	1,233,760	14.8%	10.3%
			Persons per HH[f]	2.78	2.72	2.83	−2.3%	1.7%
			Vacancy rate	4.9%	4.9%	4.9%	0.0%	0.0%
			Housing units[g]	1,176,046	1,349,793	1,297,605	14.8%	10.3%

Sources: Goodman et al. (2015); San Diego Association of Governments (2014); U.S. Census Bureau (2015)
[a]2014 householder rate (per 100 persons)
[b]Synthetic forecast based on trend in U.S. Householder rate from 2010 to 2025.
[c]From Table 8.6.
[d]Constant rate × 2025 household population
[e]Trend rate × 2025 household population
[f]Household population/households
[g]Households/(1 - vacancy rate)

We believe there are many circumstances in which the methods described here will provide useful forecasts of population-related variables.

Although the participation-rate method is widely used, the usefulness of the forecast it produces will depend on the validity of its underlying assumptions. The illustrations presented in this chapter depict several different approaches to forecasting future participation rates. In the alcohol consumption and group quarters population forecasts, we assumed constant launch year rates. Launch year rates were also held constant in the other forecasts (cigarette use and consumption, diabetes, labor force and other population and housing), but an alternate assumption regarding future rates was also applied for these variables. In the labor force and household forecasts, county rates were assumed to change at the same rate as national rates; and in the other projections, historical changes in the rates were extrapolated into the future. Regardless of the approach, developing reasonable assumptions regarding future participation rates is an important part of the development of any set of population-related forecasts. Thorough knowledge of historical trends and the factors affecting the variables of interest is essential. Although the participation-rate method is capable of producing reasonably accurate forecasts, there is no guarantee that it will actually do so.

References

Alba, R., & Islam, T. (2009). The case of disappearing Mexican Americans: An ethnic-identity mystery. *Population Research and Policy Review, 28*, 109–121.

American Diabetes Association. (2013). Economic costs of diabetes in the U.S. in 2012. *Diabetes Care, 36*(4), 1033–1046.

Arterburn, D. E., Crane, P. K., & Sullivan, S. D. (2004). The coming epidemic of obesity in elderly Americans. *Journal of the American Geriatrics Society, 52*, 1907–1912.

Barnichon, R., & Nekarda, C. (2012). *The ins and outs of forecasting unemployment: Using labor flows to forecast the labor market. Brookings papers on economic activity*. Washington, DC: The Brookings Institute.

Bhattacharya, J., Cutler, D., Goldman, D., Hurd, M., Joyce, G., Lakdawalla, D., et al. (2004). Disability forecasts and future medicare costs. *NBER Frontiers in Health Policy Research, 7*, 75–94.

Cawley, J., & Meyerhoefer, C. (2012). The medical costs of obesity: An instrumental variables approach. *Journal of Health Economics, 31*(1), 219–230.

Centers for Disease Control and Prevention. (2015a). *Underlying cause of death 1999–2013*. Retrieved from file:///C:/Users/owner/Documents/Cohort%20Change%20Book/Chapter%207_characteristics/Heart%20Disease%20Facts%20&%20Statistics%20_%20cdc.gov.html

Centers for Disease Control and Prevention. (2015b). *Behavioral Risk Factor Surveillance System*. Retrieved from http://www.cdc.gov/brfss/brfssprevalence/index.html

Christiansen, S., & Keilman, N. (2013). Probabilistic household forecasts based on register data-the case of Denmark and Finland. *Demographic Research, 28*, 1263–1302.

Finkelstein, E., Trogdon, J., Cohen, J., & Dietz, W. (2009). Annual medical spending attributable to obesity: Payer- and service-specific estimates. *Health Affairs, 28*(5), 822–831.

Finkelstein, E., Khavjou, O., Thompson, H., Trogdon, J., Pan, L., Sherry, B., & Dietz, W. (2012). Obesity and severe obesity forecast through 2030. *American Journal of Preventive Medicine, 42*(6), 563–570.

Frees, E. (2006). Forecasting labor force participation rates. *Journal of Official Statistics, 22*(3), 453–485.

Fryar, C., Carroll, M., & Ogden, C. (2014). *Prevalence of overweight, obesity, and extreme obesity among adults: United States trends 1960–1962 through 2011–2012*, *NCHS Health E-Stat*. Hyattsville: National Center for Health Statistics.

Goodman, L., Pendall, R., & Zhu, J. (2015). *Headship and homeownership: What does the future hold*. Washington, DC: Urban Institute.

Huang, E., Basu, A., O'Grady, M., & Capretta, J. (2009). Projecting the future diabetes population size and related costs for the U.S. *Diabetes Care, 32*(13), 2225–2229.

Heidenreich, P., Trogdon, J., Khavjou, O., Butler, J., Dracup, K., Ezekowitz, M., et al. (2011). Forecasting the future of cardiovascular disease in the United States a policy statement from the American Heart Association. *Circulation, 123*, 933–944.

Holmans, A. (2012). *Household projections in England: Their history and uses*. Cambridge, UK: Cambridge Center for Housing and Planning Research, University of Cambridge.

ISH Global Inc. (2014). *A forecast of U.S. cigarette consumption (2014–2040) for the Niagara Tobacco Asset Securitization Corporation*. Philadelphia, PA.

Kaneshiro, M., Martinez, A., & Swanson, D. (2011). Disappearing hispanics? The case of Los Angeles County, California: 1990–2000. In R. Verdugo (Ed.), *The demography of the hispanic population: selected essays* (pp. 95–122). Charlotte: Information Age Publishing.

Kono, S. (1987). The headship rate for projecting households. In J. Bongaarts, T. Burch, & K. Wachter (Eds.), *Family demography: Methods and their applications* (pp. 287–308). New Oxford: Oxford University Press.

Lindh, T., & Malmberg, B. (2007). Demographically based global income forecasts up to the year 2050. *International Journal of Forecasting, 23*, 553–567.

Loichinger, E. (2015). Labor force projections up to 2055 for 26 EU countries, by, age, sex, and highest level of educational attainment. *Demographic Research, 32*, 443–486.

Ma, J., Ward, E., Siegel, R., & Jemal, A. (2015). Temporal trends in mortality in the United States, 1969-2013. *Journal of the American Medical Association, 314*(16), 1731–1739.

McCue, D. (2014). *Baseline household projections for the next decade and beyond*, *Report W14–1*. Cambridge, MA: Joint Center for Housing Studies, Harvard University.

Mozaffarian, D., Benjamin, E., Go, A., Arnett, D., Blaha, M., Cushman, M., et al. (2015). Heart disease and stroke statistics–2015 update. *Circulation, 131*, 29–322.

National Center for Health Statistics. (2015). Number (in millions) of civilian, non-institutionalized persons with diagnosed diabetes, United States, 1980-2014. Retrieved from https://www.cdc.gov/diabetes/statistics/prev/national/figpersons.htm.

Ogden, C., Carroll, M., Kit, B., & Flegal, K. (2014). Prevalence of childhood and adult obesity in the United States, 2011-2012. *Journal of the American Medical Association, 311*(8), 806–814.

Ogden, C., Carroll, M., Kit, B., & Flegal, K. (2012). *Prevalence of obesity in the United States, 2009–2010*, *NCHS Data Brief No. 82*. Hyattsville: National Center for Health Statistics.

Olshansky, S., Passaro, D., Hershow, R., Layden, J., Carnes, B., Brody, J., et al. (2005). A potential decline in life expectancy in the United States in the 21st century. *New England Journal of Medicine, 352*(11), 1138–1145.

Reardon, T., & Hari, M. (2014). *Population and housing demand projections for Metro Boston: Regional projections and provisional municipal forecasts*. Boston: Metropolitan Area Planning Council.

Rowley, W., & Bezold, W. (2012). Creating public awareness: State 2025 diabetes forecasts. *Population Health Management, 15*(4), 194–200.

Sacks, J., Roeber, J., Bouchery, E., Gonzales, K., Chaloupka, F., & Brewer, R. (2013). State costs of excessive alcohol consumption, 2006. *American Journal of Preventive Medicine, 45*(4), 474–485.

San Diego Association of Governments. (2014). *Current population and housing unit estimates*. Retrieved from http:// http://datasurfer.sandag.org

Smith, S. K., Rayer, S., & Smith, E. (2008). Aging and disability: Implications for the housing industry and housing policy in the United States. *Journal of the American Planning Association, 74*, 289–306.

Smith, S. K., Tayman, J., & Swanson, D. (2013). *A practioner's guide to state and local population projections*. Dordrecht: Springer.

Stark, R. (2007). *Sociology* (10th ed.). Independence: Cengage Learning.

Substance Abuse and Mental Health Services Administration. (2014). *Population Data/NSDUH*. Retrieved from http://www.samhsa.gov/data/population-data-nsduh/reports.

Swanson, D., & Tayman, J. (2012). *Subnational population estimates*. Dordrecht: Springer.

Toossi, M. (2013). Labor force projections to 2022: The labor force participation rate continues to fall. *Monthly Labor Review*, December.

U.S. Census Bureau. (2015). *2014 1-year ACS PUMS*. Retrieved from http://www2.census.gov/acs2014_1yr/pums.

U.S. Census Bureau. (2014). *2014 National Population Projections*. Retrieved from Table 1. http://www.census.gov/population/projections/data/national/2014/summarytables.html.

U.S. Department of Health and Human Services. (2014). *Results from the 2013 national survey on drug use and health*. Substance Abuse and Mental Health Services Administration: Rockville.

Walls, H., Backholer, K., Proietto, J., & McNeil, J. (2012). Obesity and trends in life expectancy. *Journal of Obesity*. Published online doi: 10.1155/2012/107989.

Zeng, Y., Land, K., Wang, Z., & Gu, D. (2006). U.S. family household momentum and dynamics: An extension and application of the ProFamy method. *Population Research and Policy Review, 25*, 1–41.

Chapter 9
Estimating Population Size and Composition

9.1 Introduction

In a sense, population estimates are like population forecasts in that they are done in lieu of a census. For both future and past populations it is not possible to conduct a census; and while it may be possible in principle to conduct a census for a current point in time, it may not be feasible. In this chapter, we explore the use of the CCR method for generating estimates for a current point in time (In the following chapter, we will extend this discussion to past points in time). Before proceeding, recall the definition from Chapter 2 of a population *estimate*: the determination of the size or the characteristics of a population at a current or past date in the absence of census data for the same date.

When used to make population estimates, the CCR method falls into the second of three categories of estimation identified by Swanson and Tayman (2012: 3–4), namely that of "mathematical models that use census data." Methods falling into this category have generally been developed by and for academic demographers, most of whom work at universities and research institutes. To a large extent, this is not the case for the CCR method, in that it has become widely known and used among applied demographers as a forecasting method. This chapter shows that the CCR method can be used to generate current estimates (and the following chapter shows that the CCR method can be used to generate historical estimates). In order to accomplish this, however, the CCR method needs to be used in conjunction with a general class of demographic methods known as interpolation, the subject to which we now turn.

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_9

9.2 Interpolation Methods

Interpolation methods are a well-established technique in the field of demography and have a wide range of uses (Judson and Popoff 2004). In a sense, they are methods of estimation, in and of themselves. Here, we focus on a general class of interpolation methods that have two uses: (1) splitting age groups in to single years of age; and (2) assembling an annual series of numbers between the last census and a current point in time. We note at the outset that there is a wide range of interpolation methods in this class that could be used, but we focus on only three of them: (1) Waring's formula (Judson and Popoff 2004: 685–686, 2) The Karup-King method (Judson and Popoff 2004: 688, 726); and (3) Interpolation by Prorating (Judson and Popoff 2004 696–697). In the examples provided later, we show how to use each of these methods in conjunction with the CCR method and at the same time illustrate how two of them may be combined for use in conjunction with the CCR method.

Waring's formula (Judson and Popoff 2004: 685–686) is a form of polynomial interpolation based on four known points. Simpler forms of it can be used, however, with three and two known points, respectively. We use the version that is based on two known points. This allows us to generate a current estimate using only data from the last census and a forecast made using the CCR method. Interpolation methods can also be used to develop a forecast for point in time that is between a current estimate and the target year for a given CCR method forecast (e.g., if we have a 2020 forecast and the current point in time is 2017, we can use interpolation to obtain 2018 and 2019 data).

In general terms, this version of Waring's formula is defined as:

$$P_{t+x} = P_t \times (k - x/k) + P_{t+k} \times (x/k) \tag{9.1}$$

where,

x is the point in time for which an estimate is desired and *t < x < t* + *k*,
t is the year of the most recent census (from which the CCR method forecast was launched),
k is the horizon length of the CCR method forecast (typically 10 years), and
P is the population.

Equation 9.1 can be applied to age groups, age groups by sex, age groups by age, race and sex, and whatever other characteristics that the CCR method forecast generated.

The Karup-King method (Judson and Popoff 2004: 688, 726) is also a form of polynomial interpolation, but it is an "osculatory" approach. This means that it can create a smooth junction at the point where two ranges of interpolated numbers meet (e.g., if we interpolate age group 5–9 and age group 10–14 into single years of age, respectively, then the number for age 9 will intersect smoothly with the number for age 10). This method is implemented using coefficients (Judson and Popoff 2004:

726). It can be used with four known "points" or three.[1] We provide an example using three. The method disaggregates grouped data into fifths, which makes it well-suited for obtaining single years of age from five-year age groups.

The third method, "Interpolation by Prorating" (Judson and Popoff 2004: 696–697) uses applicable known data to subdivide grouped data. For example, we could take a distribution of enrollment for grades 9, 10, 11, and 12, and apply it to an estimate of enrollment in grades 9–12 to obtain enrollment in grades 9, 10, 11, and 12 consistent with the estimate. We show a variation of this approach in regard to disaggregating enrollment by groups based on single years of age obtained using the Karup-King method.

9.3 Examples

Our first example uses a 2020 forecast of the population by age of Riverside County, California by age using the CCR method (see Table 9.1). We then use Waring's 2 point formula and produce "current" estimates of this population by age for each year from 2011 to 2019 (see Table 9.2).

As specific examples, here is how the 2011, 2015, and 2019 estimates for age group 10–14 were generated, respectively. For 2011, the estimated number of 183, 604 = ((0.9 × 177,644) +0.1 × 237,247)), where 177,644 is the population aged 10–14 in 2010 and 237, 247 is the forecasted population aged 10–14 in 2020. For 2015, the estimated number of 207,446 = ((0.5 × 177,644) + (0.5 × 237,247)), where 177,644 and 237,247 are previously defined. For 2019 the estimated number is 231,287 = ((0.1 × 177,644) + (0.9 × 237,247)), where 177,644 and 237,247 are previously defined.

Waring's 2-point formula weights each endpoint, with the sum of the weights equal to 1.0. Estimates closer to the census point receive larger weights on the census number, which diminish as the estimates get closer to the forecast number. As can be seen in Table 9.2, the estimated numbers across the years from 2010 to 2020 are consistent with one another, as are the numbers across the age groups within a given year and across all of the years. This is a highly desirable feature that is an outcome of using the CCR method in conjunction with Waring's 2-point formula.

Figure 9.1 shows the population in selected 5 year age groups (0–4, 5–9, 10–14, 15–19, and 20–24) forecasted by the CCR method developed in Chapter 7 for the Memphis School District in 2010. Table 9.3 shows the population by individual year of age from 5 to 19 using the Karup-King method for interpolating grouped

[1] First and last interval coefficients are available for interpolating the youngest (0–4) and terminal age groups (e.g., 85 years and older). Interpolation for these age groups is not as reliable as interpolation for other age groups because they use information from only one side of the relevant age group.

Table 9.1 Population forecast by age, Riverside County, California, 2020

Age	2000	2010	CCR[a]	2020[b]	Change 2010–20	
					Number	Percent
0–4	121,629	162,438	0.37171	242,421	79,983	49.2%
5–9	139,468	167,065	0.39184	260,112	93,047	55.7%
10–14	133,886	177,644	1.46054	237,247	59,603	33.6%
15–19	119,725	187,125	1.34171	224,153	37,028	19.8%
20–24	96,374	154,572	1.15450	205,090	50,518	32.7%
25–29	95,621	143,992	1.20269	225,053	81,061	56.3%
30–34	108,602	138,437	1.43646	222,036	83,599	60.4%
35–39	124,260	143,926	1.50517	216,732	72,806	50.6%
40–44	117,910	149,379	1.37547	190,416	41,037	27.5%
45–49	96,484	152,722	1.22905	176,892	24,170	15.8%
50–54	79,538	140,016	1.18748	177,385	37,369	26.7%
55–59	61,880	114,765	1.18947	181,658	66,893	58.3%
60–64	54,046	98,974	1.24436	174,230	75,256	76.0%
65–69	52,309	78,495	1.26850	145,579	67,084	85.5%
70–74	50,845	62,103	1.14908	113,729	51,626	83.1%
75–79	44,184	49,003	0.93680	73,534	24,531	50.1%
80–84	27,542	36,793	0.72363	44,940	8,147	22.1%
85–89	14,399	22,399	0.50695	24,842	2,443	10.9%
90+	6,685	9,793	0.20139	13,893	4,100	41.9%
Total	1,545,387	2,189,641		3,149,942	960,301	43.9%

Source: 2000 and 2010, U.S. Census Bureau (http://factfinder2.census.gov)

[a] ${}_4P_{0,t}/{}_{15}P_{20,t}$ Ages 0–4
${}_9P_{5,t}/{}_{15}P_{25,t}$ Ages 5–9
$P_{x,t}/P_{x-10,t-10}$ Ages 10–89
$P_{90+,t}/P_{80+,t-10}$ Ages 90+

[b] ${}_4CCR_{0,t} \times {}_{15}P_{20,t+10}$ Ages 0–4
${}_9CCR_{5,t} \times {}_{15}P_{25,t+10}$ Ages 5–9
$CCR_{x,t} \times P_{x,t}$ Ages 10–89
$CCR_{80+,t} \times P_{80+,t}$ Ages 90+

data, which requires the population in the age groups shown in Fig. 9.1. In implementing the Karup-King method for this purpose, we took each age group 5–9, 10–14, and 15–19, as the "middle panel," respectively (Judson and Popoff 2004: 726, Table C.13.b).

As examples of how we use the Karup-King method: (1) the number in Table 9.3 for those aged 5 is 10,789 = (0.064 × 54,548) + (0.152 × 53,425) + (−0.016 × 51,420), where 0.064, 0.152, and −0.016 are middle panel Karup-King coefficients from Table C13.b, respectively applied to the total number aged 0–4 (54,548), 5–9 (53,425), and 10–14 (51,420); (2) the number shown for age 12 is 10,178 = (−0.024 × 53,425) + (0.248 × 51,420) + (−0.024 × 53,816), where −0.024, 0.248, and −0.024 are the Karup-King coefficients respectively applied to the total number aged 5–9 (53,425), 10–14 (51,420), and 15–19 (53,816); and (3) the number shown for age 16 is 10,555 = (0.008 × 51,420)

Table 9.2 Annual population by age, Riverside County, California, 2010–2020[a]

Age	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
0–4	162,438	170,436	178,435	186,433	194,431	202,430	210,428	218,426	226,424	234,423	242,421
5–9	167,065	176,370	185,674	194,979	204,284	213,589	222,893	232,198	241,503	250,807	260,112
10–14	177,644	183,604	189,565	195,525	201,485	207,446	213,406	219,366	225,326	231,287	237,247
15–19	187,125	190,828	194,531	198,233	201,936	205,639	209,342	213,045	216,747	220,450	224,153
20–24	154,572	159,624	164,676	169,727	174,779	179,831	184,883	189,935	194,986	200,038	205,090
25–29	143,992	152,098	160,204	168,310	176,416	184,523	192,629	200,735	208,841	216,947	225,053
30–34	138,437	146,797	155,157	163,517	171,877	180,237	188,596	196,956	205,316	213,676	222,036
35–39	143,926	151,207	158,487	165,768	173,048	180,329	187,610	194,890	202,171	209,451	216,732
40–44	149,379	153,483	157,586	161,690	165,794	169,898	174,001	178,105	182,209	186,312	190,416
45–49	152,722	155,139	157,556	159,973	162,390	164,807	167,224	169,641	172,058	174,475	176,892
50–54	140,016	143,753	147,490	151,227	154,964	158,701	162,437	166,174	169,911	173,648	177,385
55–59	114,765	121,454	128,144	134,833	141,522	148,212	154,901	161,590	168,279	174,969	181,658
60–64	98,974	106,500	114,025	121,551	129,076	136,602	144,128	151,653	159,179	166,704	174,230
65–69	78,495	85,203	91,912	98,620	105,329	112,037	118,745	125,454	132,162	138,871	145,579
70–74	62,103	67,266	72,428	77,591	82,753	87,916	93,079	98,241	103,404	108,566	113,729
75–79	49,003	51,456	53,909	56,362	58,815	61,269	63,722	66,175	68,628	71,081	73,534
80–84	36,793	37,608	38,422	39,237	40,052	40,867	41,681	42,496	43,311	44,125	44,940
85–89	22,399	22,643	22,888	23,132	23,376	23,621	23,865	24,109	24,353	24,598	24,842
90+	9,793	10,203	10,613	11,023	11,433	11,843	12,253	12,663	13,073	13,483	13,893
Total	2,189,641	2,285,672	2,381,702	2,477,731	2,573,760	2,669,797	2,765,823	2,861,852	2,957,881	3,053,911	3,149,942

[a]Interpolations for 2011–2019 based on Waring's two point formula

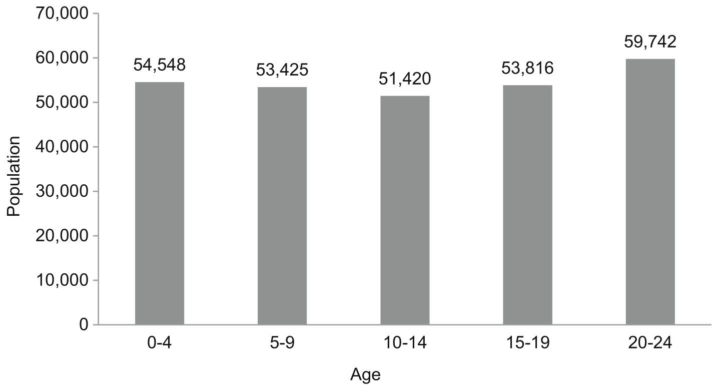

Fig. 9.1 Population forecast by selected age groups, Memphis, Tennessee School District, 2010 (Source: Chapter 7, Table 7.2)

Table 9.3 School age population forecast by single years of age, Memphis, Tennessee School District, 2010[a]

Associated Grade	Age	Population		
Pre K	5	10,789		
K	6	10,758		
1	7	10,706		
2	8	10,633		
3	9	10,539	53,425	Ages 5–9
4	10	10,374		
5	11	10,223		
6	12	10,178		
7	13	10,239		
8	14	10,405	51,419	Ages10–14
9	15	10,515		
10	16	10,555		
11	17	10,679		
12	18	10,887		
	19	11,180	53,816	Ages 15–19

[a]Age splitting based on the Karup-King middle panel coefficients

+ (0.224 × 53,816) + (−0.032 × 59,724), where 0.008, 0.224, and −0.032 are the Karup-King coefficients respectively applied to the total number aged 10–14 (51,420), 15–19 (53,816), and 20–24 (59,724).

The Karup-King method is self-normalizing in that the sum of the populations for the single years of age within an age group will sum to the population of that age group. This occurs because the weights for the age group being interpolated (e.g. 5–9) sum to 1.0 and the weights for the younger age group (e.g., 0–4) and older age group (e.g. 10–14) sum to zero.

Table 9.4 Public school enrollment by individual grade, Memphis, Tennessee School District, 2010

Grade	Population Share[a]	Enrollment[b]		
Pre K & K	n/a	11,314		
1	0.12853	9,895		
2	0.12765	9,827		
3	0.12652	9,740		
4	0.12454	9,588		
5	0.12273	9,448		
6	0.12219	9,407		
7	0.12292	9,463		
8	0.12492	9,617	76,985	Grades 1–8
9	0.24662	7,627		
10	0.24756	7,656		
11	0.25047	7,746		
12	0.25535	7,897	30,926	Grades 9–12
Total		119,225		

[a]Derived from Table 9.3
Grades 1–8 represent shares of the population aged 7–14
Grades 9–12 represent shares of the population aged 15–18
[b]Grades 1–8 are the population share times the total enrollment in grades 1–8
Grades 9–12 are the population shares times the total enrollment in grades 9–12

Our third example uses "Interpolation by Prorating" (Judson and Popoff 2004: 696–697) to estimate public school enrollment by individual grades (1–12) in the Memphis, Tennessee School District based on enrollment forecasts by grade groups (1–8 and 9–12) developed in Chapter 7. The Pre-Kindergarten and Kindergarten group does not require any further disaggregation. In prorating, a distribution is taken from a similar group (i.e., population) that has satisfactory detail to split a known total for a given group (i.e., enrollment by grade group).

The first step is to find the school age population by single years of age, which we did in Table 9.3. In the second step, we develop population shares that based on two aggregated age groups (7–14) and (15–18), which are associated with grades 1–8 and 9–12, respectively. These shares are shown in Table 9.4 in which the sum of the population shares associated with grades 1–8 and grades 9–12 each sum to 1.0. In the third step, we split the aggregated enrollment into individual grades by multiplying the proportion at given age by the aggregated enrollment in the group with which it is associated. For example, the 10,706 children aged 7 shown in Table 9.3 are associated with grade 1 and represent 0.12853 of the total children aged 7–14 (10,706/83,297). Multiplying this proportion by the total number enrolled in grades 1–8 provides the estimated number of first graders, $9895 = 0.12853 \times 76{,}985$. This same logic was used to obtain the estimated enrollment for grade levels 1 thru 12. These estimates along with the "forecasted" number for Pre-Kindergarten/Kindergarten are displayed in Table 9.4.

9.4 Conclusions

In this chapter we have shown several examples to illustrate a few of the many applications there are for estimating "current " populations and their characteristics by combining the CCR method with interpolation methods. Of course, these same interpolation methods can be applied to populations forecasted using the CCR method in order to obtain detailed information such as the population by individual age group and enrollments by individual grade level (Smith et al. 2013: 272–284). By extension, the same approach can be used to estimate (or forecast) a very wide range of demographic, social, and economic characteristics. In the next chapter we provide several examples, one of which uses interpolation and the CCR method run in reverse, to create historical estimates.

References

Judson, D., & Popoff, C. (2004). Selected general methods. In J. Siegel & D. Swanson (Eds.), *The Methods and Materials of Demography* (2nd ed., pp. 677–732). San Diego: Elsevier Academic Press.

Smith, S. K., Tayman, J., & Swanson, D. (2013). *A practitioner's guide to state and local population projections*. Dordrecht: Springer.

Swanson, D., & Tayman, J. (2012). *Subnational population estimates*. Dordrecht: Springer.

Chapter 10
Estimating Historical Populations

10.1 Introduction

Using the discussion of developing current estimates via the CCR method found in Chapter 9 as a point of departure, in this chapter we show how the CCR method can be run in reverse to generate historical population estimates, a procedure known as backcasting. This is followed by a section that provides three examples and also shows how life table survival rates are related to CCRs. The chapter concludes with a brief discussion and summary.

10.2 Reverse Cohort Change Ratios

Running the Cohort Change Ratio method in reverse provides a way to generate historical estimates using data from the two relevant censuses so that we can move a population by age (and sex) backwards, from time t to time $t-k$ using reverse cohort-change ratios (RCCR). This takes place in two steps. First, we calculate an RCCR:

$$ {}_{n}RCCR_{x,t} = {}_{n}P_{x,t} / {}_{n}P_{x+k,t+k} \quad (Step1) \tag{10.1}$$

where,

${}_{n}P_{x,t}$ is the population aged x to $x + n$ at the time of the census (t) just following the period for which an historical estimate is desired, and
${}_{n}P_{x+k,t+k}$ is the population aged $x + k$ to $x + k + n$ at the time of the census ($t + k$), which follows the census at time (t), and
k is the number of years between the two censuses.

In the second step, we move the population into the past:

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_10

$$_{n}P_{x-k,t-k} = {}_{n}RCCR_{x,t} \times {}_{n}P_{x,t} \quad \text{(Step 2)} \tag{10.2}$$

where,

$_{n}RCCR_{x,t}$ and $_{n}P_{x,t}$ are defined above.

One advantage of RCCRs is that we can backcast age groups 0–4 and 5–9 (from those aged 10–14 to 15–19, 10 years later, respectively). This is not possible for the forward-looking CCR method discussed in Chapters 1 and 4. A backcast of the oldest age group (the terminal, open-ended age group, e.g., 75 years and older), however, requires some adjustments. The initial two steps are straightforward, but as one goes back in time an important adjustment is required. As an example, suppose the final closed age group is 80–84, with 85 years and older as the terminal open-ended age group in the census following the earliest census, then calculating the terminal age group RCCR is:

$$RCCR_{85+,t} = P_{75+,t}/P_{85+,t+k}. \tag{10.3}$$

The formula for estimating the population 75+ for the year t−k is:

$$P_{75+,t-k} = RCCR_{85+} \times P_{85+,t}. \tag{10.4}$$

Notice that the population aged 85 years and older in the census used to launch the backcast becomes 75 years and older in the target year given that k = 10. If we apply the RCCRs to our initial backcast in order to backcast another k years, the terminal, open-ended age group would be 65+. Every 10 years we went back in time, the terminal, open-ended age group would be 10 years younger until, finally, we would be left with 5 years and older as the only population age group. For example, if one takes the ratio of the population aged 80 years and older in the 2000 census to the population aged 90 years and older in the 2010 census and applies this to the population aged 90 years and older in the 2000, the population aged 80 years and older is backcasted for the year 1990. This is now the "new" terminal open-ended age group, so an RCCR for 80+/70+ must be applied to this age group, which in turn, generates the population 70 years and older for the year 1980. By the time the backcasting process reaches 1910, the only age information would be for the total population aged 0 and above and 1910 would be the terminal point of the backcast.

As the preceding discussion indicates, an adjustment is needed for the terminal open-ended age group because every 10 years in the past (in the U.S. Census context) this group would be 10 years younger and, as such, successively providing less information about the age structure of the population in question. To avoid this, proportions of the closed age groups that make up a given open-ended age group are calculated and applied to the backcasted number in the terminal open-ended age group. For example, in the 2010 census one can redefine the terminal open-ended age group not only as 90 years and older but also as 80 years and older, and the latter would have three associated age groups, 80–84, 85–89 and 90 years and older.

These proportions are used to maintain a constant definition of the terminal open-ended age group as the backcast proceeds. That is, as soon as one backcasts the population 80 years and older for the year 2000 from the population aged 90 years and older in 2010, the proportions can be applied to the backcasted 80 years and older population so that the 2000 population aged 80–84, 85–89 and 90 years and older can be estimated.

10.3 Examples

10.3.1 1910 Native Hawaiian Population Estimates in Hawai'i

In our first example, we show an estimate taken from Swanson and Tayman (2012: 348–351). It uses 1930 and 1920 age-sex census data on Native Hawaiians in Hawai'i to develop RCCRs and then backcasts the 1920 Native Hawaiian population to generate population estimates by age and sex for Native Hawaiians in Hawai'i for 1910. The input data, calculations, and results are shown in Table 10.1.

Because Native Hawaiians were counted in the 1910 census, we can compare our estimates of them to the enumerated numbers to get an idea of the method's accuracy. These comparisons are found in Table 10.2. The RCCR method underestimates the total population of Native Hawaiians in 1910 by 930 people (−3.6%). The total numbers of males and females are underestimated by the same percentage. The MAPE is 7.1 for the estimates by age group for both sexes combined, with the estimates for males being less accurate than the estimates for females. The absolute percent error distributions are right-skewed as the MEDAPE (median APE) is substantially less than the MAPE for both sexes combined and separately. The average level of downward bias is similar, with MALPEs ranging for −4.1 to −3.5.

10.3.2 1770 to 1900 Native Hawaiian Population Estimates in Hawai'i

Our second example extends the historical estimates of the Native Hawaiian population in Hawai'i back to 1778, the year of first European contact. Here, we provide estimates by age for both sexes combined. In this example, the 1920 and 1910 U.S. Census data are used to define the RCCRs using 5 year age groups, 0–4, 5–9, 10–14,. . ., 70–74, with a terminal open-ended age group of 75 years and older. This means that the ratio of the population aged 65 years and older in 1910 to the population aged 75 years and older in 1920 is used to generate the terminal open-ended age group of 65 years and older, with the latter having 65–69, 70–74 and 75 years and older as its three associated age groups. The proportions for these three

Table 10.1 Native Hawaiian population estimates by age and sex, Hawai'i, 1910[a]

Age in 1930	1930 Male	Age in 1920	1920 Male	RCCR[b]	Age in 1910	1910[a] Male	
10–14	1,161	0–4	1,266	1.09044	0–4	1,223	
15–19	1,127	5–9	1,219	1.08163	5–9	1,180	
20–24	952	10–14	1,122	1.17857	10–14	1,223	
25–29	760	15–19	1,091	1.43553	15–19	1,380	
30–34	728	20–24	1,038	1.42582	20–24	1,098	
35–39	748	25–29	961	1.28562	25–29	1,196	
40–44	710	30–34	770	1.08527	30–34	665	
45–49	631	35–39	930	1.47444	35–39	1,255	
50–54	553	40–44	613	1.10800	40–44	573	
55–59	466	45–49	851	1.82496	45–49	878	
60–64	371	50–54	517	1.39471	50–54	633	
65–69	266	55–59	481	1.80700	55–59	669	
70–74	153	60–64	454	2.97096	60–64	377	
75+	197	65–69	370	3.41624	65+	601	
		70–74	127				
		75+	176				
Total	11,299[d]		11,986			12,951	
Age in 1930	**1930 Female**	**Age in 1920**	**1920 Female**	**RCCR[b]**	**Age in 1910**	**1910[c] Female**	**1910 Total Population**
10–14	1,222	0–4	1,298	1.06219	0–4	1,282	2,505
15–19	1,071	5–9	1,209	1.12885	5–9	1,242	2,422
20–24	1,031	10–14	1,207	1.17071	10–14	1,290	2,513
25–29	915	15–19	1,100	1.20219	15–19	1,273	2,653
30–34	794	20–24	1,102	1.38791	20–24	1,137	2,235
35–39	876	25–29	1,059	1.20890	25–29	1,059	2,255
40–44	642	30–34	819	1.27570	30–34	856	1,521
45–49	625	35–39	876	1.40174	35–39	1,036	2,291
50–54	522	40–44	671	1.28529	40–44	598	1,171
55–59	399	45–49	739	1.85126	45–49	718	1,596
60–64	296	50–54	465	1.57194	50–54	486	1,119
65–69	202	55–59	388	1.92020	55–59	413	1,082
70–74	118	60–64	309	2.62003	60–64	296	673
75+	166	65–69	215	2.92169	65+	459	1,060
		70–74	113				
		75+	157				
Total	11,308[b]		11,727			12,145	25,096

Sources: The Bureau of the Census (1922, 1932)

[a]Based on the reverse cohort change ratio method and 1930 to 1920 CCRs.

[b]$P_{x,t}/P_{x+k,t+k}$ Ages 0–64

$P_{65+,t}/P_{75+,t+k}$ Ages 65+

[c]$RCCR_{x,t} \times P_{x,t}$ Ages 0–64

$RCCR_{75+,t} \times P_{75+,t}$ Ages 65+

[d]Includes 2477 males aged 0–9 not shown in table

[e]Includes 2429 females aged 0–9 not shown in table

Table 10.2 Native Hawaiian population estimation error by sex, Hawai'i, 1910

	Total population		Error	
Age	Actual	Estimate[a]	Number	Percent
0–4	2,713	2,505	−208	−7.7%
5–9	2,509	2,422	−87	−3.5%
10–14	2,528	2,513	−15	−0.6%
15–19	2,657	2,653	−4	−0.2%
20–24	2,267	2,235	−32	−1.4%
25–29	2,213	2,255	42	1.9%
30–34	1,784	1,521	−263	−14.7%
35–39	2,049	2,291	242	11.8%
40–44	1,468	1,171	−297	−20.2%
45–49	1,575	1,596	21	1.3%
50–54	1,242	1,119	−123	−9.9%
55–59	1,049	1,081	32	3.1%
60–64	651	673	22	3.4%
65+	1,320	1,060	−260	−19.7%
Total	26,025	25,096	−929	−3.6%
MAPE	7.1%			
MEDAPE	3.5%			
MALPE	−4.0%			

(continued)

Table 10.2 (continued)

	Males				Females			
			Error				Error	
Age	Actual	Estimate[a]	Number	Percent	Actual	Estimate[a]	Number	Percent
0–4	1,368	1,223	−145	−10.6%	1,345	1,282	−63	−4.7%
5–9	1,253	1,180	−73	−5.8%	1,256	1,242	−14	−1.1%
10–14	1,307	1,223	−84	−6.4%	1,221	1,290	69	5.7%
15–19	1,343	1,380	37	2.7%	1,314	1,273	−41	−3.1%
20–24	1,129	1,098	−31	−2.8%	1,138	1,137	−1	−0.1%
25–29	1,123	1,196	73	6.5%	1,090	1,059	−31	−2.8%
30–34	837	665	−172	−20.5%	947	856	−91	−9.6%
35–39	1,043	1,255	212	20.3%	1,006	1,036	30	3.0%
40–44	734	573	−161	−22.0%	734	598	−136	−18.6%
45–49	841	878	37	4.4%	734	718	−16	−2.1%
50–54	638	633	−5	−0.8%	604	486	−118	−19.6%
55–59	611	669	58	9.4%	438	413	−25	−5.7%
60–64	407	377	−30	−7.3%	244	296	52	21.3%
65+	800	601	−199	−24.8%	520	459	−61	−11.8%
Total	13,434	12,951	−483	−3.6%	12,591	12,145	−446	−3.5%
MAPE	10.3%				7.8%			
MEDAPE	7.0%				5.2%			
MALPE	−4.1%				−3.5%			

Source: The Bureau of the Census (1913)
[a]From Table 10.1

Table 10.3 RCCRs and allocation proportions for generating decennial estimates of the Native Hawaiian population, Hawai'i, 1920–1910

1920		1910		
Age in	Population	Age in	Population	RCCR[a]
10–14	2,329	0–4	2,713	1.16488
15–19	2,191	5–9	2,509	1.14514
20–24	2,140	10–14	2,528	1.18131
25–29	2,020	15–19	2,657	1.31535
30–34	1,589	20–24	2,267	1.42668
35–39	1,806	25–29	2,213	1.22536
40–44	1,284	30–34	1,784	1.38941
45–49	1,590	35–39	2,049	1.28868
50–54	982	40–44	1,468	1.49491
55–59	869	45–49	1,575	1.81243
60–64	763	50–54	1,242	1.62779
65–69	585	55–59	1,049	1.79316
70–74	240	60–64	651	2.71250
75+	333	65+	1,320	3.96396
		Proportions for Allocating Pop 65+[b]		
		65–69	0.42518	
		70–74	0.23684	
		75+	0.33798	
			1.00000	

Sources: The Bureau of the Census (1913, 1922)
[a] $P_{x,t}/P_{x+k,t+}$ Ages 0–64
[b] $P_{65+,t}/P_{75+,t+k}$ Ages 65+

age groups were found by averaging the proportions for them found in the 1930, 1920, and 1910 census counts for Native Hawaiians in Hawai'i.

The RCCRs and the adjustments were initially applied to the 1910 census by age to generate a set of backcasted 1900 estimates by age for the Native Hawaiian population in Hawai'i. The same RCCRs were then applied to the 1900 estimates by age to generate a set of backcasted 1890 estimates by age. This process was repeated until the 1770 population of Native Hawaiians by age was generated for Hawai'i. As should be clear, the backcasting proceeded in decennial cycles from 1900 to 1770. The 1920 and 1910 input data and the 1920–1910 RCCRs used to generate the estimates are shown in Table 10.3.

As shown in Table 10.4, the total population estimates of Native Hawaiians track well with the 1900 U.S. census count, two census counts done by the Kingdom of Hawai'i for 1890 and 1860, and a carefully prepared estimate done by Adams et al. (1925) for1850. The estimate of 683,200 for 1778 is found by calculating the rate of change between 1770 and 1780 and then applying that rate of change to the 1770 estimate. As mentioned at the end of Chapter 9, the estimate for 1778 is found using an interpolation method.

In addition to tracking well with the census counts, it also is important to note that the RCCRs are all in excess of one. This means that their corresponding reciprocals, the CCRs, are all less than one. This makes sense for Native Hawaiians

Table 10.4 Total population of Native Hawaiians, Hawai'i, 1900–1770

Year	Estimate	Census[a]
1900	29,336	29,799
1890	33,457	34,436
1880	39,711	n/a
1870	48,579	n/a
1860	61,931	67,084[b]
1850	80,574	82,035[c]
1840	110,948	n/a
1830	149,297	n/a
1820	200,018	n/a
1810	267,971	n/a
1800	359,010	n/a
1790	480,978	n/a
1780	644,383	n/a
1778[d]	683,200	n/a
1770	863,302	n/a

[a]Schmitt (1968)

[b]Included Chinese living in Honolulu and part Hawaiians (Schmitt 1968:74)

[c]Estimate by Adams (Schmitt 1968:43)

[d]$683{,}200 = 863{,}302 \times e^{r \times 8,}$ where $r = -0.02925 = [\ln(644{,}383/863{,}302)]/10$

since there is virtually no migration into Hawai'i of Native Hawaiians, which means the CCRs are generated only by out migration and mortality. Evidence suggests that while out-migration did occur, it was not extensive among Native Hawaiians. To the extent any appreciable out-migration—and return in-migration— occurred, it was largely confined to young adult males (Adams et al. 1925: 10–12; Kana'iupuni and Malone 2006; Schmitt 1968: 38–40; Schmitt 1977: 90–91; Schmitt and Nordyke 2001: 5).

These RCCRs also indicate high levels of mortality in the Native Hawaiian population in the early part of the twentieth century. In this regard, the RCCRs are consistent with survival rates generated from the life tables constructed for Native Hawaiians in the early part of the twentieth century by Park et al. (1979: 14), who estimate Native Hawaiian male and female life expectancy at birth in 1920 as 34.21 and 32.90 years, respectively.

Although we do not go into details here, it is worth noting that there is a link between the application of RCCRs and the subject of Chapter 12 that uses CCRs to analyze stable population theory. The major point in Chapter 12 is that when a constant set of CCRs is applied to a given population, a stable population will eventually result (whereby once stability is achieved, the relative age distribution of the population remains constant over time). Given this, the application of a constant set of RCCRs to a given population should also yield a stable population at some point in time. In fact, this was found for the Native Hawaiian population in Hawai'i, which reached stability by 1820. Also worth noting is that the relative 1820 age structure not only remains constant back to 1770 (and will do so beyond), but that it is different than the 1910 relative age structure. This finding also is consistent with stable population theory in that the initial 1910 age structure is "forgotten" by the time stability is reached in 1820.

10.3.3 CCRs and Life Table Survival Rates

In this section, we build on the idea that the RCCRs used in the preceding example approximate the inverse of survival rates. In the context of a life table (Kintner 2004: 322–324), a CCR is known as a "survivorship ratio." We discuss the application of the CCR method to life tables and survivorship in Chapter 11, but the context in that chapter is on developing life expectancy and mortality estimates rather than population estimates. The latter topic is our focus here. In preparation for this example, we note that the survivorship rates computed from the "${}_nL_x$" column (years lived in a given age interval) of a life table are equivalent to the CCRs calculated for age groups of a specific width, while the survivorship rates computed from the "T_x" column (years lived at this and all subsequent ages) are equivalent to the CCRs calculated for open-ended terminal age groups (for a discussion of ${}_nL_x$ and T_x, see Kintner 2004: 322–323). The relationship between survivorship rates calculated from T_x and CCRs calculated from open-ended, terminal age intervals brings up a way in which the RCCR method can be used to estimate an historical population.

We first consider the relationship between T_0 and T_x in a life table as follows:

$$ {}_xS_0 = T_x/T_0 \qquad (10.5) $$

where,

${}_xS_0$ is the survivorship rate from birth to the open-ended terminal age group, *x*
T_x is the years lived in the open-ended, terminal age group, and
T_0 is the years lived at birth and all subsequent age groups.

Re-arranging the terms in Eq. 9.5, we see that:

$$ T_x = {}_xS_0 \times T_0, \qquad (10.6) $$

and further that:

$$ T_0 = (T_x/{}_xS_0), \qquad (10.7) $$

The preceding equations suggest that a RCCR can be constructed such that a total population can be estimated. First, note that:

$$RCCR_{k+,t} = P_{0+,t}/P_{k+,t+k}. \tag{10.8}$$

Second, that the formula for estimating the total population 0+ of area *i* for the year $t-k$ is:

$$P_{0+,t\text{-}k} = RCCR_{k+} \times P_{k+,t}. \tag{10.9}$$

To illustrate how Eq. 10.8 and Eq. 10.9 can be used to estimate a population, we again turn to the historical data on the Native Hawaiian population in Hawai'i using an example from Swanson and Tayman (2012: 350–351). Here, 1930 and 1910 data for Native Hawaiians is used to estimate a RCCR for age group 20 years and older. We then apply this RCCR to the Native Hawaiian population aged 20 years and older in 1910 in order to estimate the total number of Native Hawaiians in 1890. There are 13,120 Native Hawaiians aged 20 years and over in 1930, while in 1910 there are 25,095 Native Hawaiians in total, of whom 15,001 are aged 20 and over. From these data, we find:

$$RCCR_{20+,1910} = 1.9127 \ (25,095/13,120), \text{and}$$
$$P_{0+,1890} = 28,693 \ (1.9127 \times 15,001).$$

So, our 1890 estimate of the Native Hawaiian population of Hawai'i is 28,693. This estimate is 16.7% less than the number reported by Schmitt (1977: 25) from the Hawaiian Kingdom's 1890 census (34,436). Given that migration of Native Hawaiians was not a major factor of its population change (Schmitt, 1968: 183), it appears that mortality rates were dramatically higher for this population between 1890 and 1910 than they were between 1910 and 1930, which the available evidence suggests was the case (Nordyke 1989; Schmitt 1968; and Schmitt 1977). The correct $RCCR_{20+}$ for estimating the total number of Native Hawaiians in 1890 from those aged 20 and over in 1910 would be 2.2956.

10.3.4 Multi-racial Population Estimates for San Bernardino and Riverside Counties

The third example uses the RCCR method to estimate the 1990 multi-racial population of Riverside and San Bernardino Counties, California. This estimate is of interest because the U.S. Census Bureau started the practice of providing multi-racial respondents with the opportunity to identify themselves as multi-racial in the 2000 census. In 1990, respondents had to choose which single race category best fit them. The RCCR method provides an opportunity to construct RCCRs for a multi-racial category of interest (e.g., Asian and one or more other races) found in the 2000 and 2010 censuses in order to estimate the number of people in the category of interest for 1990.

Table 10.5 Estimation of the multi-racial population by age, Riverside and San Bernardino counties, California, 1990

2010		2000			1990	
Age in	Population	Age in	Population	RCCR[a]	Age in	Population[b]
10–14	26,522	0–4	21,266	0.80183	0–4	13,165
15–19	24,268	5–9	20,271	0.83528	5–9	11,288
20–24	17,040	10–14	16,418	0.96352	10–14	9,788
25–29	13,461	15–19	13,514	1.00395	15–19	8,988
30–34	12,073	20–24	10,159	0.84144	20–24	6,623
35–39	10,880	25–29	8,952	0.82281	25–29	6,557
40–44	10,319	30–34	7,871	0.76276	30–34	5,380
45–49	9,525	35–39	7,969	0.83668	35–39	4,499
50–54	7,914	40–44	7,053	0.89122	40–44	3,668
55–59	5,955	45–49	5,377	0.90293	45–49	2,534
60–64	7,914	50–54	4,116	0.52008	50–54	1,107
65–69	4,115	55–59	2,806	0.68196	55–59	1,183
70–74	1,982	60–64	2,128	1.07356	60–64	1,397
75+	3,105	65–69	1,735	1.58758	65+	3,006
		70–74	1,301		65–69	1,201
		75+	1,894		70–74	720
					75+	1,085
Total	212,132[c]		132,830			79,183
		Proportions for allocating population 65+[d]				
		65–69	0.39956			
		70–74	0.23966			
		75+	0.36078			
			1.00000			

Sources: 2000 and 2010, U.S. Census Bureau (http://factfinder2.census.gov)

[a] $P_{x,t}/P_{x+k,t+k}$ Ages 0–64

$P_{65+,t}/P_{75+,t+k}$ Ages 65+

[b] $RCCR_{x,t} \times P_{x,t}$ Ages 0–64

$RCCR_{75+,t} \times P_{75+,t}$ Ages 65+

1990 P65+ × proportions Ages 65–69, 70–74, and 75+

[c] Includes 57,059 population aged 0–9 not shown in table.

[d] Average of the share of the population 65+ from 2000 and 2010

The 2010 and 2000 input, the RCCRs calculated from them, and the 1990 estimates of the multi-racial population of these two counties are shown in Table 10.5. The multi-racial population has increased substantially in San Bernardino and Riverside counties since 1990 (132,949 persons and 68%). Between 2000 and 2010 the multi-racial population increased by 79,302 or 59.7%. From 1990 to 2010, the largest percentage changes are seen in the population aged 45 years and older. However, almost 70% of the numeric change in the multi-racial population occurs in persons younger than 45 years of age (data not shown).

The change in the multi-racial population of these two counties between 2000 and 2010 is not only due to demographic factors (births, deaths, and migration), but also social factors. As a great deal of research shows, ethnicity and race are social constructs and fluid (Cornell and Hartmann 2007; Goldstein and Morning 2000; Nagel 1994, 1995; Omi and Winant 2015; Perez and Hirschman 2009; Yamashiro 2011). As such, a population defined on the basis of race or ethnicity is subject to change from factors that are not demographic. One desirable feature of the CCR method is that both demographic and non-demographic changes affecting race and other social constructs are captured across censuses.

10.4 Conclusions

In addition to the RCCR method, the CCR method itself can be used to construct certain types of historical population data. For example, by using 1850 and 1860 census counts to construct CCRs, the 1860 white male population aged, say, 15–44 (by 5 year age groups) of a given state such as Virginia could be projected to 1870 and compared with the 1870 census count of white males aged 25–54 (by 5 year age groups) to estimate the demographic impact of the Civil War on this population. This population would, of course, been most likely to have served in the Confederate army during the Civil War so the comparison would provide an estimate not only of casualties, but of migration as well. Casualty counts are available for the Civil War, but for the Confederate army they are widely believed to be understated (Hacker 2011).

Swanson and Verdugo (2016) conducted this type of analysis and estimated that there were nearly 25% fewer white males aged 20–54 than expected in the 1870 census results for all of the 11 Confederate states due to the combined effects of mortality and net out-migration between 1860 and 1870. They obtained this estimate by subtracting the 1870 expected number (1,393,125) for age group 20–54 generated by the CCR method (using 1950–1860 cohort change ratios) from the 1870 actual (census) number (1,047,323). Swanson et al. (2009) used this type of approach to estimate the demographic impact of Hurricane Katrina by zip code and found that by 2007 there were 311,250 people fewer than expected in the absence of Katrina, which struck in August of 2005. Swanson (2009) used the same approach to estimate the effect of Hurricane Katrina on the potential client populations in the service area associated with two medical facilities on the Mississippi Gulf Coast. The results showed that Katrina had an adverse impact on the client base of the two facilities.

Both the CCR method and the RCCR method could be used and averages could be taken across the CCR and RCCR results in a manner similar to what is done when one uses the "forward-reverse survival rate procedure" to estimate an intercensal population (Bryan 2004: 537–538). Thus, either the CCR and RCCR approaches, or a combination of the two, can be used to provide new perspectives on the demographic impact of wars and other forms of human conflict, as well as natural and man-made disasters. Whether one is using RCCRs, CCRs or both to generate historical estimates, keep in mind what was covered in Chapter 9, namely, that when CCRs (or RCCRs) are used in conjunction with interpolation methods, a wide range of possibilities opens up.

In this chapter we have provided several examples to illustrate the many applications for estimating "historical" populations and their characteristics using the RCCR method. In addition, the interpolation methods described and discussed in Chapter 9 also can applied to population's backcasted using the RCCR method in order to obtain a very wide range of historical demographic, social, and economic characteristics.

References

Adams, R., Livesay, T., & Van Winkle, E. (1925). *The peoples of Hawaii*. Honolulu: Institute of Pacific Relations.

Bryan, T. (2004). Population estimates. In J. Siegel & D. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 523–560). San Diego: Elsevier Academic Press.

Cornell, S., & Hartmann, D. (2007). *Ethnicity and race: Making identities in a changing world*. Thousand Oaks: Pine Forge Press.

Goldstein, J., & Morning, A. (2000). The multiple-race population of the United States: Issues and estimates. *Proceedings of the National Academy of Sciences of the United States of America, 97*(11), 6230–6235.

Hacker, J. A. (2011). Census-based count of the civil war dead. *Civil War History, 57*, 307–348.

Kana'iupuni, S., & Malone, N. (2006). This land is my land: The role of place in Native Hawaiian identity. In J. Frazier & E. Tetty-Fio (Eds.), *Race, ethnicity and place in a changing America* (pp. 291–305). New York: Global Press.

Kintner, H. J. (2004). The life table. In J. Siegel & D. Swanson (Eds.), *The methods and materials of demography* (pp. 301–340). Amsterdam: Elsevier Academic Press.

Nagel, J. (1994). Constructing ethnicity: Creating and recreating ethnic identity and culture. *Social Problems, 41*(1), 152–176.

Nagel, J. (1995). American Indian ethnic renewal: Politics and resurgence of identity. *American Sociological Review, 60*(6), 947–965.

Nordyke, E. (1989). *The peopling of Hawai'i* (2nd ed.). Honolulu: University of Hawai'i Press.

Omi, M., & Winant, H. (2015). *Racial formation in the United States* (3rd ed.). New York: Routledge.

Park, C., Gardner, R., & Nordyke, E. (1979). *Life tables by ethnic group for Hawai'i, 1920 to 1970. R and S Report*. Honolulu: Hawai'i State Department of Health, Research, and Statistics Office.

Perez, A., & Hirschman, C. (2009). The changing racial and ethnic composition of the U.S. population: Emerging American identities. *Population and Development Review, 35*(1), 1–51.

Schmitt, R. C. (1968). *Demographic statistics of Hawai'i, 1778–1965*. Honolulu: University of Hawai'i Press.

Schmitt, R. C. (1977). *Historical statistics of Hawai'i*. Honolulu: University of Hawai'i Press.

Schmitt, R., & Nordyke, E. (2001). Death in Hawai'i: The epidemics of 1848–1849. *Hawaiian Journal of History, 35*, 1–13.

Swanson, D., & Tayman, J. (2012). *Subnational population estimates*. Dordrecht: Springer.

Swanson, D. (2009). Hurricane Katrina: A case study of its impacts on medical service providers and their client populations. *The Open Demography Journal, 2*, 8–17.

Swanson, D., & Verdugo, R. (2016). *The Civil War's demographic impact on non-Hispanic white males in the 11 Confederate states: An analysis by state and selected age groups*. Paper presented at the annual conference of the Southern Demographic Association, Athens, GA, 12–14, October.

Swanson, D., Forgette, R., McKibben, J., Van Boening, M., & Wombold, L. (2009). The socio-demographic and environmental effects of Katrina: An impact analysis perspective. *The Open Demography Journal, 2*, 36–46.

The Bureau of the Census. (1913). *Thirteenth census of the United States, taken in the year 1910: Vol. 3 Population, 1910, Nebraska–Wyoming, Alaska, Hawaii, and Puerto Rico*. Washington, DC: Government Printing Office.

The Bureau of the Census. (1922). *Fourteenth census of the United States, taken in the year 1920: Vol. 3 Population, 1920, Composition and characteristics of the populations by state*. Washington, DC: Government Printing Office.

The Bureau of the Census. (1932). *Fifteenth census of the United States, 1930, outlying territories and possessions*. Washington, DC: Government Printing Office.

Yamashiro, J. (2011). Racialized national identity construction in the ancestral homeland: Japanese American migrants in Japan. *Ethnic and Racial Studies, 34*(9), 1502–1521.

Chapter 11
Estimating Life Expectancy

11.1 Introduction

Census survival methods are the oldest and most widely applicable methods of estimating mortality, and for populations with negligible migration they can provide accurate estimates. As opposed to other data and analytically intensive methods, CCR methods have minimal data requirements; use available census data; and do not require a great deal of judgment or "data-fitting" techniques to implement. In this chapter we demonstrate that life expectancy at birth can be computed by using CCRs in combination with a protocol in which the life table radix set to one. We compare our estimates of life expectancy at birth using CCRs against U.S. Census Bureau estimates and find the CCR method works reasonably well. We discuss the benefits of the CCR method for estimating life expectancy and believe that it is a viable alternative in populations that experience negligible migration.

11.2 Estimating Life Expectancy

As noted in *Methods for Estimating Adult Mortality from Census Data* (United Nations 2002: 5), "Census survival methods are the oldest and most widely applicable methods of estimating adult mortality… and can provide excellent results for populations that experience negligible migration…" The reason for the ubiquity of this approach is threefold: (1) data requirements are minimal in that only two successive age distributions are needed; (2) the two successive age distributions are usually easily obtained from census counts; and (3) the method is straightforward in that it requires neither a great deal of judgment nor "data-fitting" techniques to implement. This ubiquity is in contrast to other methods, such as "Model Life Tables," which require more data as well as judgment and, often, data

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_11

fitting (United Nations 1982: 16–27). Our purpose in this chapter, however, is not to debate the relative merits of these and other approaches (e.g., Swanson 1989; Swanson and Palmore 1976; Swanson and Stanford 2012; Swanson et al. 1977, 2009), but to demonstrate another way of calculating life expectancy from census survival rates that is less involved than existing methods.

11.3 Life Expectancy: The United Nations Census Survival Method

Census survival methods require two population age distributions at two points in time (generally, two successive census enumerations). Ideally, the interval between the census enumerations (e.g., 10 years) is either equal to the width of the age groups (e.g., the age groups are given in 10 year increments, 0–9, 10–19,. . .,75–84, 85 years and older) or a whole number multiple thereof (e.g., age groups given in 5 year increments, 0–4, 5–9,. . .,80–84, through the final open-ended age group (e.g., 85 years and older).

The United Nations (2002: 6) shows that using the census survival method, expectation of life at age x can be computed as:

$$\mathrm{e_x} = (\mathrm{T_x}/\mathrm{l_{(n/2)}})/(\mathrm{l_x}/\mathrm{l_{(n/2)}}) = \mathrm{T_x}/\mathrm{l_x} \tag{11.1}$$

where,

e_x = life expectancy (average years remaining) at age x,
x is age,
n is the width of the age groups (up to, but not including the terminal, open-ended age group)
T_x is the total person years remaining to persons age x,
l_x is the number reaching age x, and
$l_{(n/2)}$ are persons aged x to $x + n$ assumed to be concentrated at the mid-point of the age group, and

$$\mathrm{l_{(x+n/2)}}/\mathrm{l_{(x-n/2)}} = \mathrm{P2_{(x,n)}}/\mathrm{P1_{(x-n,n)}} \tag{11.2}$$

where,

$P2_{(x,n)}$ are the number of persons counted in the second census in age group x to $x + n$, and
$P1_{(x-n,n)}$ are the number of persons counted in the first census in age group $x - n\ to\ n$.

In general, then, the life-table probability of surviving from the mid-point of one age group to the next ($\mathrm{l_{(x+n/2)}}/\mathrm{l_{(x-n/2)}}$) is approximated by the census survival ratio ($\mathrm{P2_{(x,n)}}/\mathrm{P1_{(x-n,n)}}$). The cumulative multiplication of the probabilities shown in

Eq. 11.2 gives the conditional survival schedule ($l_x/l_{(n/2)}$) (United Nations 2002: 5–6). From the conditional l_x values given by Eq. 11.2, the conditional estimates of the number of person years lived in each age group ($_nL_x$) can be calculated as:

$$_nL_x/l_{(n/2)} = (n/2) \times \left[\left(l_x/l_{(n/2)}\right) + \left(l_{(x+n)}/l_{(n/2)}\right)\right] \tag{11.3}$$

where,

$_nL_x$ is the number of person years lived in each age group.

Given a value of $T_x/l_{(n/2)}$ for some initial age x, the United Nations (2002) shows that total remaining years expected at age x (Tx) can be calculated as:

$$T_{(x-n)}/l_{(n/2)} = T_x/l_{(n/2)} + {}_nL_{(x-n)}/l_{(n/2)}, \tag{11.4}$$

or Eq. 11.1 for the expectation of life at age x.

11.4 Estimating Life Expectancy from Cohort Change Ratios

In the proposed CCR method, we start with the radix of a life table (l_x) equal to one and life expectancy at birth can be computed directly from the expression:

$$e_0 = S_0 + (S_0 \times S_1) + (S_0 \times S_1 \times S_2) + \ldots + (S_0 \times S_1 \times S_2 \ldots \times S_x) \tag{11.5}$$

where,

e_0 is life expectancy at birth,
S_0 is the survivorship from $t = 0$ (e.g., birth) to t = 1 (e.g., age 1), S_1 = survivorship from $t = 1$ (e.g., age 1) to $t = 2$ (e.g., age 2), and so on through S_x, and
$S_x = {}_nL_x/{}_nL_{(x-n)}$.

Equation 11.5 represents single years of age. However, we can generalize that equation to other age groups ($_nS_x = {}_nL_x/{}_nL_{(x-n)}$), so that:

$$e_0 = {}_nS_0 + ({}_nS_0 \times {}_nS_1) + ({}_nS_0 \times {}_nS_1 \times {}_nS_2) + \ldots + ({}_nS_0 \times {}_nS_1 \times {}_nS_2 \ldots \times {}_nS_x). \tag{11.6}$$

As Eqs. 11.5 and 11.6 imply, the fundamental life table function is inherent in our method; that is, via the $_nS_x$ values, we have $_nq_x$ values. The Appendix at the end of this chapter shows a derivation of the relationship between survivorship rates and life expectancy shown in Eq. 11.5 and generalized in Eq. 11.6.

We also use census survival rates, although we prefer to use the more general term "cohort change ratios" (CCRs). Following Smith et al. (2013: 177) and using notation from Eq. 11.2, a CCR can be generally defined as:

$$_{n}CCR_{x} = P2_{(x,n)}/P1_{(x-n,n)}. \tag{11.7}$$

Survivorship rates can approximated by CCRs as follows:

$$_{n}S_{x} = {}_{n}L_{x}/{}_{n}L_{(x-n)} \approx P2_{(x,n)}/P1_{(x-n,n)}. \tag{11.8}$$

We can determine life expectancy at birth by substituting CCR values for $_{n}S_{x}$ values in either Eq. 11.5 (for single years of age) or Eq. 11.6 (for age groups).

As with the more involved United Nations (2002) approach, or methods will only work for populations for which migration is negligible, but there are many areas around the world where this is the case, or approximately so (United Nations 2002). The world as a whole meets this requirement. Countries with negligible migration include North Korea and Burma, among others. Other such populations are found in the historical record—the former Soviet Union, Albania from 1950 to 1980, and the Peoples Republic of China from 1950 through 1970, for example. Still others may be defined by race and ethnicity or other 'rules' of membership (e.g., Indigenous populations in Australia and Canada, Native Hawaiians; native-born populations).

Broadly speaking, the method can be applied to any population subject to renewal through a single increment (birth) and extinction through a single decrement (death), where there are at least two successive census counts that provide the population by age and other characteristics if desired. We also note that unlike the United Nations' method, the CCR method can be used to yield estimates of life expectancy at birth. Our method is also not subject to the limitations imposed by stationary or even stable population requirements.

11.5 Empirical Examples and Evaluation

We developed life expectancy estimates directly from cohort change ratios constructed for the world as a whole and Burma, using 5 year age groups. We compare the CCR-based life expectancy at birth estimates to U.S. Census Bureau (CB) estimates for the period 1950–1955 to 2045–2050 for the world and 1975–1980 to 2010–2015 for Burma. The data for implementing the CCR method and the CB e_0 estimates were obtained from the international database (U.S. Census Bureau 2010). We computed e_0 estimates by male, female, and both sexes, but we only present and evaluate the estimates for both sexes.[1]

[1]The absolute and relative differences by sex are similar to the results for both sexes combined.

Table 11.1 World life expectancy at birth estimates for both sexes, 1950–1955 to 2045–2050

Years	CCR e_0 Estimate	CB e_0[a] Estimate	Difference Number[b]	Percent[c]
1950–1955	51.8	46.6	5.2	11.2%
1955–1960	54.0	49.5	4.5	9.1%
1960–1965	56.0	52.4	3.6	6.9%
1965–1970	58.4	56.1	2.3	4.1%
1970–1975	60.4	58.2	2.2	3.8%
1975–1980	61.8	60.2	1.6	2.7%
1980–1985	62.6	61.7	0.9	1.5%
1985–1990	63.5	63.2	0.3	0.5%
1990–1995	64.1	64.0	0.1	0.2%
1995–2000	64.9	65.2	−0.3	−0.5%
2000–2005	65.8	66.4	−0.6	−0.9%
2005–2010	66.7	67.6	−0.9	−1.3%
2010–2015	67.7	68.9	−1.2	−1.7%
2015–2020	68.6	70.1	−1.5	−2.1%
2020–2025	69.4	71.1	−1.7	−2.4%
2025–2030	70.1	72.1	−2.0	−2.8%
2030–2035	70.8	73.1	−2.3	−3.1%
2035–2040	71.4	73.9	−2.5	−3.4%
2040–2045	72.0	74.8	−2.8	−3.7%
2045–2050	72.6	75.5	−2.9	−3.8%
			MALPD	0.7%
			MAPD	3.3%

[a]U.S. Census Bureau (2010)
[b]CCR est. – CB est.
[c](CCR est.– CB est.)/CB est. × 100

Table 11.1 contains the CCR and CB estimates of world e_0 for both sexes for the period 1950–1955 to 2045–2050. The Mean Absolute Percent Difference (MAPD) between our estimates and those made by the CB over the entire period is 3.3%, while the Mean Algebraic Percent Difference (MALPD) is 0.7% indicating only a slight upward bias compared to the CB estimates. During the first 50 years (1950–2000), the CCR estimates have a distinct upward bias relative to the CB estimates (MALPD of 4.0%). From the year 2000 forward, however, all CCR estimates are lower than the CB estimates (MALPD of −2.5%). These summary measures of difference indicate rather close agreement between the two sets of estimates.

Table 11.2 shows e_0 estimates for Burma for the period 1975–1980 to 2005–2010. As the Table shows, all but one of the CCR e_0 estimates are less than those produced by the CB, which means the MALPD of −4.1 percent is the same as the MAPD ignoring the sign. The one exception is the 1985–1990 period when the CCR and CB estimates are the same. The CCR estimates of e_0 remain almost constant at age 60 from 1995–2000 to 2005–2010, while the CB estimates increase

Table 11.2 Life expectancy at birth estimates for both sexes, Burma, 1950–1955 to 2045–2050

Years	CCR e_0 Estimate	CB e_0[a] Estimate	Difference Number[b]	Percent[c]
1975–1980	50.0	54.0	−4.0	−7.4%
1980–1985	52.0	56.0	−4.0	−7.1%
1985–1990	56.0	56.0	0.0	0.0%
1990–1995	57.0	59.0	−2.0	−3.4%
1995–2000	60.0	61.0	−1.0	−1.6%
2000–2005	61.0	63.0	−2.0	−3.2%
2005–2010	61.0	65.0	−4.0	−6.2%
			MALPD	−4.1%
			MAPD	4.1%

[a]U.S. Census Bureau (2010)
[b]CCR est. – CB est.
[c](CCR est.– CB est.)/CB est. × 100

from 61 to 65 years. However, we again find that the summary measures suggest reasonably close agreement.

11.6 Conclusions

Despite some nuances (e.g., converting CCRs into survival ratios may require additional refinements) and cautions (e.g., the population data by age may be faulty), we find benefits in using this approach to estimate life expectancy, including the ability to develop estimates of average remaining life at any age (not shown here). We suggest that the technique is worthy of consideration for use in estimating life expectancy in populations experiencing negligible migration, given the cautions we discuss. As such, we believe that this approach adds another dimension to census survival methods—which, as we noted at the outset, are "... the oldest and most widely applicable methods of estimating adult mortality... (and can) provide excellent results (for) populations that experience negligible migration..." (United Nations 2002: 5).

Appendix

Relation Between Survival Rates and Life Expectancy

Any particular set of age-specific survival rates implies a specific life expectancy. As an example using a complete life table where x is a single age, the relationship between a set of survival rates (S_x) and, the corresponding entries in the "years lived" column of the life table (L_x) is $S_x = L_x/L_{x-1}$ for ages 1 and over, while $S_0 = L_0$ for survivors from birth to the age zero. Thus, in a life table with a radix = 1.0, life expectancy at birth can be expressed as $e_0 = L_1 + L_2 + L_3 +, \ldots,$ $+ L_x$. That is, life expectancy can be expressed as the sum of the L_x values. It is readily seen that:

$$L_0 = S_0.$$
$$L_1 = S_0 \times S_1, \text{and}.$$
$$L_2 = S_0 \times S_1 \times S_2, \ldots L_x = S_0 \times S_1 \times S_2, \ldots, S_{x-1} \times S_x.$$

Substituting S_x for L_x in the preceding yields Eq. 11.5:

$$e_0 = S_0 + (S_0 \times S_1) + (S_0 \times S_1 \times S_2) + , \ldots, + (S_0 \times S_1 \times S_2, \ldots, \times S_x).$$

Equation 11.5 can be generalized to apply to an abridged life table and expressed as Eq. 11.5a:

$$e_0 = {}_nS_0 + ({}_nS_0 \times {}_nS_1) + ({}_nS_0 \times {}_nS_1 \times {}_nS_2) + , \ldots, + ({}_nS_0 \times {}_nS_1 \times {}_nS_2, \ldots, \times {}_nS_x).$$

References

Smith, S. K., Tayman, J., & Swanson, D. (2013). *A practitioner's guide to state and local population projections*. Dordrecht: Springer.

Swanson, D. (1989). A state-based regression model for estimating substate life expectancy. *Demography, 26*(1), 161–170.

Swanson, D., & Palmore, J. (1976). Two parameter regression estimates of current life expectancy at birth: Part I. *Asian & Pacific Census Forum, 3*(November), 5–10.

Swanson, D., & Sanford, A. (2012). Socio-economic status and life expectancy in the United States, 1990–2010: Are we reaching the limits of human longevity? *Population Review, 51*(2), 16–41.

Swanson, D., Palmore, J., & Sundaram, C. (1977). Two parameter regression estimates of current life expectancy at birth: Part II. *Asian & Pacific Census Forum, 5*(May), 5–10.

Swanson, D., McGehee, M., & Hoque, N. (2009). Socio-economic status and life expectancy in the United States, 1970–1990. *Population Review, 48*(1), 39–63.

United Nations. (1982). *Model life tables for developing countries*. New York: Population Division, United Nations.

United Nations. (2002). *Methods for estimating adult mortality*. New York: Population Division, United Nations.

U.S. Census Bureau. (2010). *World population statistics*. Retrieved from http://www.census.gov/ipc/www/idb/informationGateway.php

Chapter 12
Stable Population Theory

12.1 Introduction

Stable population theory underpins much of our intuition about population dynamics and it continues to have a fundamental influence on research in demography. Classical presentations of the theory focused upon analyzing the long-term implications of a stable set of vital rates within a closed population. Under these conditions, it has been shown that any population—regardless of its initial population structure and growth rate—will converge upon a stable equilibrium over time characterized by a constant rate of growth and a stable proportional age-structure (Fisher 1930; Keyfitz and Caswell 2005; Lotka 1956; Schoen 2010). Any population characterized by stable vital rates may be considered to be in the process of converging upon a stable equilibrium (Kim and Schoen 1993; Schoen 2010). Once this convergence is reached, the population will change smoothly in step with an exponential birth series dictated by a constant fertility regime and a consistent proportion of the population comprised of women of reproductive age (Coale 1972; Coale and Demeny 1966). More recently, it has been demonstrated that these findings also hold under conditions of migration (Espenshade 1986; Espenshade et al. 1982; Mitra and Cerone 1986; Schoen 2010), thus suggesting the use of cohort change ratios to consider questions in stable population theory (Swanson et al. 2016).

In this chapter, we present the classical stable population model in terms of cohort change ratios (CCRs), illustrate the classical findings of the theory using CCR-based demographic forecasts, demonstrate how these findings continue to hold when net-migration is considered within cohort change ratios, and evaluate the effect of the components of population change on convergence to a stable population. We conclude by showing how CCRs ratios can lead to novel analyses aimed at traditional questions in stable population theory.

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_12

12.2 Cohort Change Ratios and the Stable Population Model

The classical stable population model applies a constant set of age-specific fertility and survival rates to any age-specific population. If enough time passes that initial population reaches a stable state with a constant rate of change and stable proportionate age structure. This classical model has been extended to show that even with the inclusion of a constant set of age-specific migration rates a population will eventually reach a stable state. Cohort change ratios (CCRs) combine the effects of mortality and migration. As such they provide a tool for examining the transient dynamics of a population as it moves toward the stable equivalent captured in most formal demographic models based on asymptotic population dynamics (Swanson et al. 2016). This application uses a Leslie matrix containing an initial population, an invariant set of CCRs and an invariant set of age-specific fertility rates.

CCRs are calculated by dividing the population aged x in year t by the population aged $x–k$ in year $t–k$. For the analysis conducted in this chapter, $k = 5$, indicating a 5-year time period between censuses. Given the nature of the CCRs in this instance, 0 to 4 is the youngest five-year age group for which a forecast can be made. Children younger than age 5 are forecast using age-specific fertility rates (ASFR). Equations 12.1 and 12.2[1] show the forecast calculation sequence:

$$ {}_{\mathrm{n}}\mathrm{P}_{\mathrm{x+5,t+5}} = {}_{\mathrm{n}}\mathrm{CCR}_{\mathrm{x,t}} \times {}_{\mathrm{n}}\mathrm{P}_{\mathrm{x,t}} \quad \text{and} \quad (\text{Ages } 5+) \tag{12.1} $$

$$ {}_{4}\mathrm{P}_{0,\mathrm{t+5}} = \sum {}_{30}\mathrm{P}_{15,\mathrm{t}} \times {}_{30}\mathrm{ASFR}_{15} \quad (\text{Ages } 0-4). \tag{12.2} $$

A convenient way to express these equations to employ a matrix population model (Caswell 2001; Cushing 1998; Lefkovitch 1971; Leslie 1945, 1948; Schoen 2010; Sykes 1969) to further develop these relationships and to relate cohort change ratios to stable population theory (Swanson et al. 2016). To accomplish this, let us consider ${}_{n}P_{x,t}$ as the initial population count vector, the *ASFRs* are contained in a single top row of the matrix and the CCRs are in the sub-diagonal to create a forecasting matrix (*A)* that is conformable for multiplication with the population count vector. Thus, Eqs. 12.1 and 12.2 and be recast into forecasting equation based on a matrix population model:

$$ {}_{\mathrm{n}}\mathrm{P}_{\mathrm{x+5,t+5}} = \mathrm{A}_{\mathrm{n}}\mathrm{P}_{\mathrm{x,t}}. \tag{12.3} $$

Using Eq. 12.3, a population can be sequentially forecast in 5-year time intervals. If the initial population ${}_{n}P_{x,t}$ is multiplied against the matrix A for a significant number of iterations, ${}_{n}P_{x,t}$ will converge upon a stable population with a constant rate of growth and unchanging proportional age structure (Caswell 2001; Swanson et al. 2016). This has been shown to be true in cases including assumptions of

[1]The births are adjusted for infant and child survivorship probabilities.

no-migration (in which the cohort change ratio is equivalent to a survival rate) as well as under conditions where this assumption is relaxed (Cerone 1987; Espenshade 1986; Espenshade et al. 1982; Mitra 1983, 1990; Swanson et al. 2016). The Perron-Frobenius theorem (Gantmacher 1959; Shores 2007) implies that any forecasting matrix for which all entries are non-zero and positive will converge into a stable population equilibrium. CCRs, which are always positive, meet the requirements of the Perron-Frobenius theorem.

12.3 Illustration of Stable Populations with and without Migration

Demographic forecasts in which rates are held constant over a sufficiently long time provide a method for gaining insights about stable population models (Cushing 1998; Swanson et al. 2016). To arrive at a stable population, we successively apply Eq. 12.3 to the base population producing a forecast every 5-years until stability is reached. We define stability as the point in time where the change from one forecast interval (five years) to the next produces no measurable impact in either the proportion of persons in any specific age category or the observed five-year exponential growth rate. In this illustration a population is deemed stable when the Index of Stability (***S***) (Index of Dissimilarity), introduced by Swanson et al. (2016), between two successive age structures in time (e.g., 2100 and 2105) equals 0.000, or a less than a 0.04% difference.

We show two alternative paths to stability. One uses fertility and survival rates and assumes zero migration, while the other uses fertility rates and CCRs that combine the effect of mortality and net effect of migration. The first alternative (NoMig) holds the ASFRs and 5-year life-table survival rates constant until stability is reached, assuming zero population change due to migration. The 5-year survival rates are in the sub-diagonal of the Leslie matrix. The second alternative (MIG) allows for migration and holds the ASFRs and CCRs constant until stability is reached. The CCRs replace the 5-year survival rates in the sub-diagonal of the Leslie matrix used in the NoMig alternative.

We illustrate these alternatives using data from Greece starting in 2005 (see Table 12.1). The female population of Greece in 2005 is 5.4 million and has a slow growth rate. Between 2000 and 2005 the population increased by 71,300, or an average of only 0.3% per year. Females in Greece have a relatively old age structure. Their median age is 41.6 years; 20.6% of the population is 65 years and older; and 13.7% of the population is under the age of 20. These characteristics are indicative of a very low total fertility rate (1.33). Female life expectancy at birth is 81.8 years. For the forecasts, the two adjustments are made to the ASFR rates. First, male births are removed using the 2005 proportion of births that are female (0.484) (Hellenic Statistical Authority 2016); and second, these adjusted rates are

Table 12.1 Female demographics, Greece, 2005

Age	Population[a]	Population share	2005/2000 CCR	5-year[b] survival rate	Fertility rate Intital[a]	Fertility rate Adjusted[c]
Child				0.995180		
0–4	248,465	4.6%	1.0024	0.999311		
5–9	244,554	4.5%	1.0039	0.999489		
10–14	251,159	4.6%	1.0086	0.999179		
15–19	275,802	5.1%	1.0313	0.998675	0.0116	0.0281
20–24	344,984	6.4%	1.0453	0.998571	0.0554	0.1341
25–29	396,902	7.3%	1.0324	0.998298	0.0957	0.2316
30–34	410,594	7.6%	1.0167	0.997541	0.0791	0.1914
35–39	416,009	7.6%	1.0090	0.996370	0.0301	0.0728
40–44	385,419	7.1%	1.0035	0.993940	0.0058	0.0014
45–49	373,201	6.9%	0.9986	0.990749		
50–54	343,818	6.3%	0.9917	0.986085	TFR 1.39	
55–59	331,449	6.1%	0.9817	0.979553		
60–64	286,497	5.3%	0.9624	0.965937		
65–69	312,159	5.7%	0.9262	0.936690		
70–74	293,305	5.4%	0.8644	0.875901		
75–79	246,909	4.5%	0.7683	0.768619		
80–84	149,470	2.8%	0.5332	0.486671		
85+	120,245	2.2%				
Total	5,430,941	100.0%				
	Median Age	41.6		e_0 81.8		

[a]U.S. Census Bureau International Data Base (http://www.census.gov/population/international/data/idb/informationGateway.php)
[b]2000–2004 Female Life Table for Greece. Human Mortality Database (http://www.mortality.org)
[c]Adjusted to remove male births and to represent a 5-year period (2005ASFR × 0.484 × 5).

multiplied by 5 to reflect the number of births that occur over the 5-year forecast horizon. Forecasted female births are reduced by the child survival rate shown in Table 12.1.

Table 12.2 compares the number and share of the population by age in 2005 and 2290, the year in which stability is reached. The absence of migration does not impact the year stability is reached, which for both alternatives is the year 2290. Between 2000 and 2005, net migration of females was only 68,314 or a total population net migration rate of (0.013 per person).[2] This low level of migration is embedded in the CCRs used in the migration alternative. As such, it is not surprising that

[2]Net migration from 2000 to 2005 was computed using the residual method by subtracting female natural increase from the female population change. Births and deaths were obtained from the Hellenic Statistical Authority (2016).

Table 12.2 Stable population structure without and with migration, Greece, 2005 and 2290

	Population			Share of population				
		2290			2290		2005–2290	
	2005	No Mig.	Migration	2005	No Mig.	Migration	No Mig.	Migration
0–4	248,465	6,134	20,039	4.6%	3.5%	3.6%	−1.1%	−1.0%
5–9	244,554	6,539	20,999	4.5%	3.7%	3.8%	−0.8%	−0.7%
10–14	251,159	6,977	22,049	4.6%	3.9%	4.0%	−0.7%	−0.6%
15–19	275,802	7,443	23,265	5.1%	4.2%	4.2%	−0.9%	−0.9%
20–24	344,984	7,935	25,091	6.4%	4.5%	4.5%	−1.9%	−1.9%
25–29	396,902	8,453	27,410	7.3%	4.8%	5.0%	−2.5%	−2.3%
30–34	410,594	8,996	29,559	7.6%	5.1%	5.3%	−2.5%	−2.3%
35–39	416,009	9,569	31,402	7.6%	5.4%	5.7%	−2.2%	−1.9%
40–44	385,419	10,176	33,136	7.1%	5.7%	6.0%	−1.4%	−1.1%
45–49	373,201	10,804	34,803	6.9%	6.1%	6.3%	−0.8%	−0.6%
50–54	343,818	11,434	36,369	6.3%	6.4%	6.6%	0.1%	0.3%
55–59	331,449	12,032	37,702	6.1%	6.8%	6.8%	0.7%	0.7%
60–64	286,497	12,561	38,645	5.3%	6.9%	7.0%	1.6%	1.7%
65–69	312,159	12,926	38,827	5.7%	7.3%	7.0%	1.6%	1.3%
70–74	293,305	12,913	37,589	5.4%	7.3%	6.8%	1.9%	1.4%
75–79	246,909	12,084	34,020	4.5%	6.8%	6.2%	2.3%	1.7%
80–84	149,470	9,933	27,383	2.8%	5.6%	5.0%	2.8%	2.2%
85+	120,245	10,727	34,462	2.2%	6.0%	6.2%	3.8%	4.0%
Total	5,430,941	177,636	552,750	100.0%	100.0%	100.0%	0.0%	0.0%
			Median age	41.6	52.5	51.2		
			% Under 20	13.7%	11.1%	11.4%		
			% 65+	20.6%	33.0%	31.2%		

migration does not affect the time to stability in Greek females. This finding is not universal and is a function of the particular characteristics of this population. However, given the same fertility and mortality conditions, high levels (positive/ negative) of migration will lengthen / shorten the time to stability (Swanson et al. 2016).

In the NoMig alternative, the population of Greek women shrinks from 5,430,941 in 2005 to 177,636 in 2290, a decline of −96.7%. The population growth rate has stabilized at −1.3% per year, down from 0.2% in 2005.There is also a remarkable shift is the female age structure. Median age rises from 41.6 in 2005 to 52.5 in 2290, caused by drop in the share of the population under 20 years of age (13.7% vs 11.1%) and an even more dramatic increase in the population 65 years an older (20.6% vs 33.0%). The overall absolute average difference in the population shares in 2290 and 2005 across the age groups is 1.7 percentage points, but the differences in the populations aged 80 to 84 and 85 years and older are 2.8% and 3.8%, respectively.

Allowing for migration, the population of Greek women shrinks from 5,430,941 persons in 2005 to 552,750 in 2290, a decline of −89.8%. The population growth rate has stabilized at −0.9% per year, 0.4% higher (lower rate of decline) compared to the NoMig alternative. The effect of migration results in 375,114 more Greek females in 2290. As shown in Table 12.1, net-migration is positive in the 20 to 34 year age groups because the CCRs are above 1.00 and mortality rates are very low in these ages. These ages also have the highest fertility rates, indicating that migration is reinforcing the impact of fertility on convergence. Although migration has a sizable impact on the total population size, the stable age composition is similar in both alternatives. In 2290, the median age is 1.3 years higher, the percent of the population 65 years and older is 1.8 percentage points higher, and the percent of the population under age 20 is 0.3 percentage point lower than the corresponding figures in the NoMig alternative. We compared the two stable age distribution using the Index of Dissimilarity and find a value of 2.0%, indicating a very close match. In 15 age groups the difference is 0.3 percentage points or less and in 3 age groups (70 to 84) the differences range between 0.5 and 0.6 percentage points.

The results for the NoMig alternative are consistent with stable population theory and would occur in any population subjected to constant birth and survival rates. However, the stable age structure and growth rate that results is uniquely determined by the specific observed rates involved. This finding reflects the principle of ergodicity in which the long-term, asymptotic population dynamics are guaranteed to occur indifferent of the initial population age structure or growth rate (Cushing 1998; Tuljapurkar 1982; Wachter 2014). We have demonstrated ergodicity is the context of a forecast, but ergodicity may also be demonstrated analytically using the eigenvalues of the forecast matrix (Cohen 1979; Caswell 2000, 2001; Schoen 2010).

The comparison between the NoMig and MIG alternatives also are in line with expectations. With positive migration the population would be larger due to the migration itself and its positive impact on births. The population would also age less due to migration's greater effect on younger adult age groups (ages 20 to 44). The total population in the MIG alternative is 211% larger than the total population in the NoMig scenario. However in ages 25 to 44, the population is between 224% and 228% larger. The impact of migration of Greek females is not sufficient to offset the effect of the continuation of very low fertility rates: as the MIG alternative shows, the population also diminishes considerably.

12.4 Impact of Demographic Components of Change on Convergence

The focus of classical population stable population models on a closed population is a simplifying assumption that allows ignores the complexities of migration, which complicates the analysis of stable populations (Caswell 2001; Fisher 1930; Lotka 1907; Wilson and Bossert 1971). While classical stable population theory has focused on the convergence toward a stable proportional age-structure and population growth rate, a corollary prediction is the general fertility rate should stabilize and the time-series of births should be exponentially increasing or decreasing in a smooth manner. At stability, variation in fertility should be the primary determinant of population growth (Cushing 1998; Preston et al. 2001; Wachter 2014).

In our example of Greek females, the observed convergence of the general fertility rate, the time series of births, and exponential trajectory of population growth rates are in accordance with the predictions of stable population. In Fig. 12.1, we observe the stabilization of the general fertility rate (GFR) at each forecast year to the year 2290.[3] While the amplitude of GFR fluctuations is strong in the earliest portion of the forecast, over time the pattern of oscillation dampens progressively, settling to a low that at stability is produced by rounding error in calculations rather than real fluctuations. The MIG alternative has slightly higher GFRs and births (shown below); suggesting that over the forecast horizon migration would tend to add proportionately fewer females in the child-bearing ages than the additional births they produce.

Figure 12.2 shows how convergence plays out in the time-series of births (over a five- year period). While their numbers are falling throughout the period (indeed at this rate Greek women are in danger of eventually going extinct!), they smooth to an equilibrium decline that is similar whether migration is present or absent in the forecast. The trend in births is much smoother compared to the trend in GFRs. This equilibrium trend in births is at a higher level for the MIG alternative in accordance with the slightly higher population growth rate (lower rate of loss), which is also seen to stabilize similarly over time in both forecast alternatives (see Fig. 12.3). The positive impact of migration itself and its impact of migration on births explains why the MIG alternative has 375,114 more persons than the NoMig alternative in the year 2290.

[3]The general fertility rate is computed by dividing female births by females aged 15 to 44.

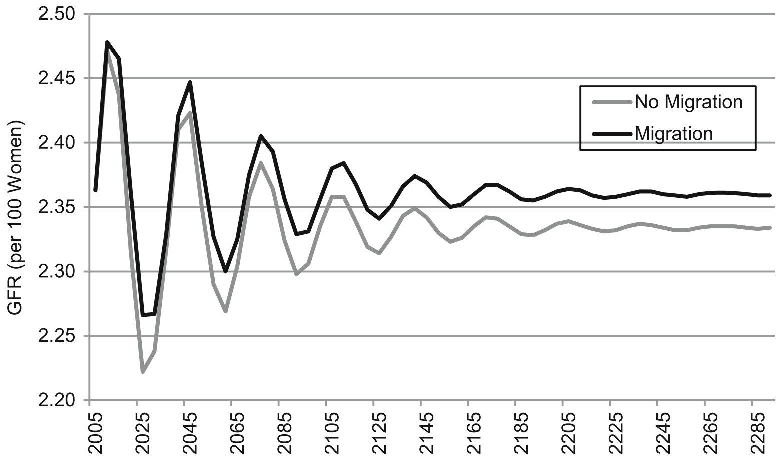

Fig. 12.1 Female general fertility rate, Greece, 2005–2290

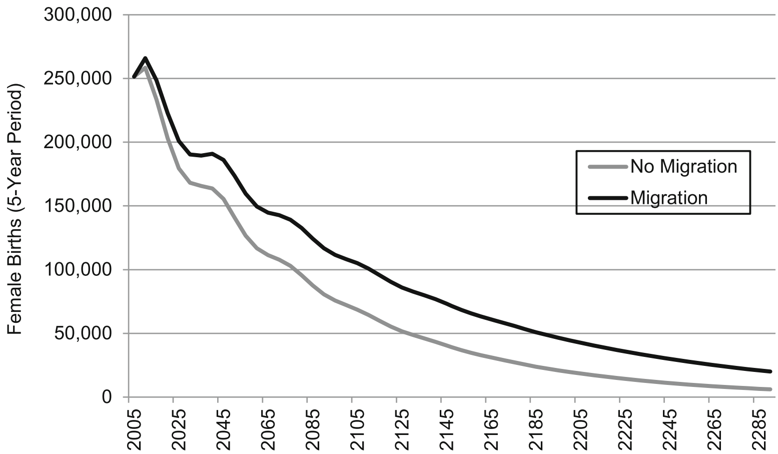

Fig. 12.2 Female births, Greece, 2005–2290

These results suggest that variation in demographic components impact the process of convergence, both in terms of transient and long-term or asymptotic dynamics (Caswell 2001; Schoen 2010; Swanson et al. 2016). Stable population researchers have analyzed the time required for a population to converge to stability in light of variation in fertility or survival rates (Coale 1972; Coale and Trussell 1974; Keyfitz 1977; Kim and Schoen 1993; Schoen 2010) and have considered how patterns of oscillations are determined by them (Caswell and Werner 1978; Lefkovitch 1971; Longstaff 1984; Rago and Goodyear 1987). By comparison relatively few studies have focused on the role and impact of migration on the path to population stability (Alho 2008; Bratadan 2016). To date, a discussion of the complex, multidimensional interaction of a nearly infinite space occupied by

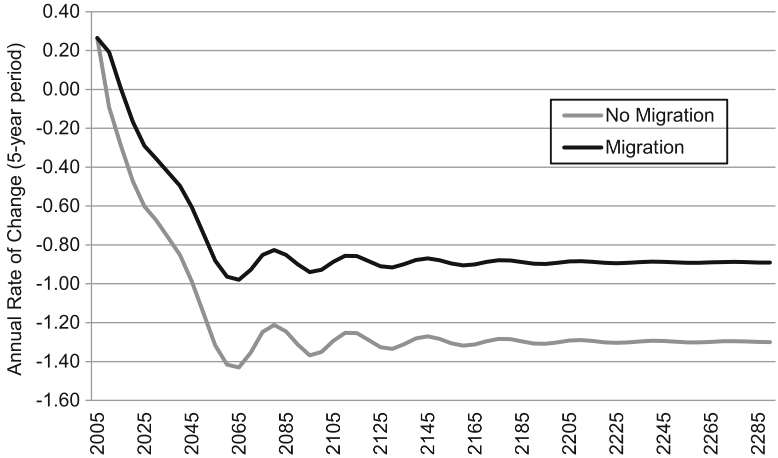

Fig. 12.3 Annual rate of female population change, Greece, 2005–2290

combinations of fertility, mortality, and migration rates has been all but absent from the literature (Swanson et al. 2016).

To address these issues, Swanson et al. (2016) examined the relationship between fertility (total fertility rate), mortality (life-expectancy at birth), and migration (the average CCR across the 20 to 34 year old age groups) and measures of population convergence derived from CCR-based demographic forecasts. Using CCRs from 2000 to 2005 or 2001 to 2006 and age-specific fertility rates at the middle of the base period for 62 countries, they forecasted the base populations forward to stability using the same approach previously described for the migration alternative for Greek females, except they defined stability using a smaller ***S*** of 0.000000. Swanson et al. (2016) regressed the years to stability on the components of change variables previously described.

These regression results showed that both life expectancy at birth and migration are positively related to the time to stability, while fertility had an inverse relationship. That is, increases in fertility hasten the time to convergence, while increases in life expectancy and net-migration slow it. Using standardized regression coefficients, they found that migration plays the largest role in determining the time to stability (beta = 0.428), fertility the second largest (beta = −0.333), and life expectancy the least (beta = 0.238). These values suggest that the time to stability is longer for a population with low mortality, low fertility, and high net in-migration than it is for a population with high mortality, high fertility, and low net in-migration.

12.5 Other Strategies to Analyze Convergence

While developments in stable population theory suggest that relaxing assumptions of population closed to migration do not invalidate the basic findings associated with this theory, the widespread availability of cohort change ratios can be used to further explore stable population theory. We believe that an important opportunity exists to build upon the results of Swanson et al. (2016) and to compare and contrast convergence paths associated with asymptotic and transient dynamics (Caswell 2001; Caswell and Werner 1978). This section outlines three strategies that may prove useful for further research and that explore the use of cohort change ratios in examining the impact of demographic components of change on population convergence.

12.5.1 Clarifying Measures of Convergence: Transient or Asymptotic Dynamics

As a general rule, measures of convergence may be either asymptotic (long-term) or transient (short-term). Examples of asymptotic measures of convergence include the damping ratio (Anderson and May 1979; Caswell 2001) and the force of convergence (Kim and Schoen 1993; Schoen 2010). Both measures reflect a time to stability that is determined solely by the rates contained in a projection matrix, which is important because these rates are independent of the observed state of the population (Caswell 2001; Schoen 2010). They also suggest that the process of convergence should follow a path of exponential decay from a maximum difference to a minimum. Using unpublished data from Swanson et al. (2016), Fig. 12.4 illustrates this exponential pattern of decay for Albania, Russia, and Tajikistan. This figure shows showing the path of the Index of stability ***(S)*** to convergence over a 290 year forecast horizon. The negative exponential decay to convergence shown for these countries is not unique and occurs in the other 59 counties their study.

In cases where transient dynamics are to be studied through forecasts, variation in the rate of convergence might be directly modeled as an exponential rate of decay with observed oscillations in that overall rate considered as the population approaches stability. For instance, ***S*** could be tracked over time and rates of convergence at each time step recorded to establish these oscillations with a baseline exponential decay model providing a basis for comparison. In this way, a logically consistent set of asymptotic and transient measures of population convergence could be considered in a single study from both perspectives (Caswell 2001; Caswell and Werner 1978). What asymptotic measures cannot tell us about the specific trajectory of a population as it moves toward stability might be observed using forecast-based (transient) measures of convergence (Caswell 200; Swanson et al. 2016). However, forecast-based observations begin at a specific point under a specific demographic structure. As such, we may expect differences

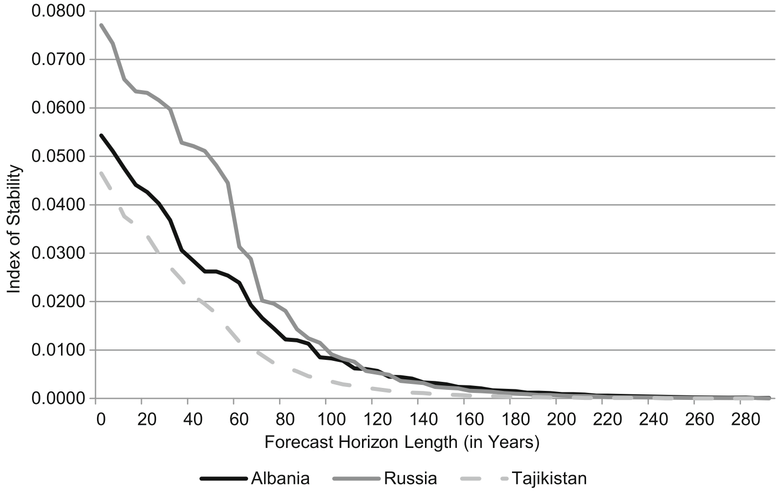

Fig. 12.4 Exponential decay in population convergence to stability, Albania, Russia, and Tajikistan

between transient measures because initial conditions may impact time to convergence, and asymptotic ones that depend solely on the rates contained in a projection matrix. What transient measures cannot tell us, due to limitations inherent to where we begin observation of differences between current and forecasted stable structures, we may derive insights based on asymptotic measures.

We believe progress in studying stable population theory using data containing CCRs would benefit from a clarification between the two types of convergence measures. While both transient and asymptotic approaches should produce convergence pathways that are distributed according to the negative exponential model, the estimated rates of convergence for each approach may differ in important and unknown ways. For example, Swanson et al. (2016) indicated a strong correlation between forecasted and asymptotic estimates of the time to stability. However non-trivial differences between these two solutions were observed, suggesting that "jump-off bias" could be present and may be a confounding influence. We believe that much can be learned from both asymptotic and forecasting approaches using CCRs in the study of stable populations.

12.5.2 Components of Change: Interactions and Convergence

To our knowledge, Swanson et al. (2016) was the first attempt to explore the multidimensional space in which combinations of fertility, mortality, and migration schedules interact to determine the timing of convergence as well as the amplitude and frequency of oscillations experienced along the path to stability. Creating usable referent categories of interactions for analyzing these dynamics occurring in such a complex space is a daunting challenge and subject to misclassification biases (Agresti 2013; Christensen 1997; Hastie et al. 2009). In this section, we suggest a scheme for organizing and analyzing these interactions using unpublished data from the Swanson et al. (2016) study.

Imagine a simplified space in which countries are stratified into low or high life expectancy. Within these groupings, countries are classified into combinations of high or low fertility and low and high migration. Within each category of high and low life expectancy, there are four separate categories of high-low fertility and high-low migration. These eight combinations represent one simplification over the alternative of using a continuous space. Table 12.3 defines this multidimensional space where. Taking the order of letter options H (high) and L (low) as life expectancy/fertility/net-migration, the 3-letter groups such as HLL (high life expectancy, low fertility, and low migration) depict the eight combinations covering the multidimensional space.

Using the unpublished data from the Swanson et al. (2016), we stratified the 62 countries based on their relationship to the averages for life expectancy at birth (e_0), fertility (TFR), and net-migration (meanCCR_20 to34).[4] Countries above the mean on any indicator were considered “high” and those equal to or below the mean were considered “low. Countries were assigned to the eight strata based on a binary categorization for each component of change. In this example, the dependent variable is the years required to reach stability (Years).

Table 12.4 shows the average of Years in relationship to the components of change categories. The top panel shows these averages for the survivorship, fertility, and migration dichotomies separately. These averages not only are consistent with the regression results previously discussed, they provide additional information as to their numeric impact on Years. High survivorship increases Years; countries with high life expectancy take on average 161 more Years than countries with low life expectancy. The same directional relationship is seen for migration; countries with high migration take on average of 158 more Years than countries with low migration. The effect of fertility is in the opposite direction than that found for migration; countries with high fertility take on average 83 fewer Years than countries with low fertility.

The bottom panel of Table 12.4 shows the average of years for the 3-way, cross-classification of the components of change to analyze their interaction on Years. With a sample of 62, the cells contain relatively few observations; four cells contain less than five countries and the other four cells contain between 11 and 14 countries. In countries with low survivorship and low fertility, migration has a modest impact; with high migration needing 33 more Years than low migration. This impact is significantly reduced in countries with low survivorship and high fertility, with high

[4]Three countries had outlying TFRs greater than 3 (Tajikistan (3.28), Saudi Arabia (3.43), and Guatemala (4.23)). We excluded these countries from the average used to classify the TFRs. This resulted in a more even distribution between high and low TFR countries (28 High and 34 Low). If the unadjusted average was used 6 countries would have been from High to Low TFR.

Table 12.3 Multidimensional space combinations of components of change

	Low fertility		High fertility	
	Low migration	High migration	Low migration	High migration
Low Life Expectancy	LLL	LLH	LHL	LHH
High Life Expectancy	HLL	HLH	HHL	HHH

Table 12.4 Average number of years to stability by components of change, selected countries

	Surviorship[a]	Fertility[b]	Migration[c]
Low[d]	408	526	417
	(31)	(34)	(34)
High[e]	569	443	575
	(31)	(28)	(28)

	Low fertility		High fertility	
	Low migration	High migration	Low migration	High migration
Low survivorship	432	465	378	385
	(14)	(2)	(14)	(1)
High survivorship	496	637	437	532
	(4)	(14)	(2)	(11)

Sample sizes in parentheses
[a]Life expectancy at birth
[b]Total fertility rate
[c]CCR 20–24 to 30–34
[d]Less than or equal to the average rate for all countries
[e]Greater than the average rate for all countries

migration needing only 12 more Years or 1.8%; the smallest difference of any comparison in the table.

Regardless of fertility level, migration has a large impact in high survivorship countries. However, it is more noticeable in high survivorship countries when fertility is low. In high survivorship countries with low fertility, those with high migration need an average of 141 more Years than those with low migration. However, in high survivorship countries with high fertility, those with high migration need an average of 95 more Years than those with low migration to reach stability. The shortest average Years (378) occur in countries with low survivorship, high fertility, and low migration, while the longest average Years (637) occurs in countries with the opposite characteristics (high survivorship, low fertility, high migration); a difference of 68.5% between these two extremes.

In classical stable population models, fertility will have a much more dramatic impact on population structure, growth, and convergence than mortality (Coale and Demeny 1966; Preston et al. 2001 Wachter 2014). CCRs play an important role in this regard because they permit an evaluation of migration along with fertility and mortality in the context of stable populations. In the preceding example, a much more complicated relationship is suggested in which all three components of change interact in shaping both transient and asymptotic population dynamics.

A larger sample and finer grade cross-classifications will allow more precise statements to be made about the levels of fertility, mortality, and migration and their impacts on stable populations. Also, more rigorous statistics methods such loglinear modeling (Agresti 2013) or regression models using continuous variables and interaction terms (Mitchell 2012) may be useful analytical strategies in this regard. In Chapter 13, we explore a decomposition of the components of change that may provide a more comprehensive analysis of the referent categories useful for additional investigation into population models based on cohort change ratios.

12.5.3 Perturbation Analysis and the Life Table Response Experiment Framework

In seeking to understand the impact of demographic components of change on stable population trajectories, one might consider an analytical approach using the stable population mathematical framework. Using this framework, responses of a population convergence measure, such as the damping ratio, could be measured in light of changes in the *(ij)* elements of forecasting matrix using perturbation analysis (Caswell 2000; de Kroon et al. 2000; McPeek and Kalisz 1993). One such strategy would be to compute damping ratios associated with a forecast matrix holding fertility and mortality levels constant and adjusting migration levels using bootstrap resampling of the *(ij)* elements of the matrix (Baker et al. 2015; Brault and Caswell 1993; Caswell 2000; Lewontin and Cohen 1969; Sykes 1969).

Perturbing the forecast in a pre-specified manner while considering the effects across strata reflects different sets of referent categories would allow replication of the Life Table Response Experiment (LTRE) Caswell (2000, 2010). The LTRE measures the sensitivity of overall population growth to changes in each *(ij)* element of a forecast matrix along meaningful "experimental" strata. In this context, one might decompose CCRs into survivorship and net-migration components (see Chapter 13), forecast these factors using stochastic simulation (Gardiner 1983; Graham and Talay 2013; Lemieux 2009; Taylor and Karlin 1998), reconstitute the matrix, and re-compute the damping ratio. Although computationally intensive, such an approach would measure the sensitivity of a measure of convergence to shifts in components of change. These shifts could be formulated to test specific hypotheses about the relationship between demographic components of change and population convergence across relevant demographic factors.

The LTRE approach is well-established in population ecology (Brault and Caswell 1993; Caswell 2000; Caswell and Kayne 2001) and similar simulation-based alternatives have been proposed as well (Wisdom and Mills 1997; Wisdom et al. 2000). At least one application of both LTRE and simulation-based studies has been done in the field of anthropology (Baker et al. 2015), where the similar concept of natural experiments is widely utilized in studies of growth and development (Bogin and Loucky 1997; Lasker 1969). Equivalent approaches have also been used in the field of toxicology (Gentile et al. 1982; Marshall 1962) and in operations research (Coleman and Montgomery 1993; Taguchi 1986).

12.6 Conclusions

In this chapter, we illustrate the principal findings of stable population theory using CCR-based demographic forecasts, examined the role of demographic components of change on population convergence, and sketched future possibilities for expanding such research. We believe that the CCR-based demographic approach provides new insights and capabilities for stable population analysis. The CCR-based demographic approach is also more flexible than classical approaches in dealing with migration. Because CCRs are always greater than zero and encompass both net in-migration and net out-migration, they can be used in a Leslie matrix with assurance that a given population will converge to stability. Other approaches that have examined migration as part of the process to convergence have only allowed for net in-migration in order to provide assurance that a given population would converge (Espenshade 1986, Sivamurthy 1982). The CCR-based demographic forecast approach does not require the mathematical sophistication of classical stable population theory making its implementation more widely accessible.

References

Agresti, A. (2013). *Categorical data analysis* (3rd ed.). New York: Wiley.

Alho, J. (2008). Migration, fertility, and aging in stable population. *Demography, 43*(3), 641–650.

Anderson, R., & May, R. (1979). Population biology and infectious diseases: Part I. *Nature, 280* (02), 361–367.

Baker, J., Hill, K., Hurtado, A. M., Alcantara, A., Hunsinger, E., & Sprague, W. (2015). Microdemographic determinants of population recovery among the Northern Ache. *Human Biology, 87*(1), 5–18.

Bratadan, C. (2016). Emigration and the stable population model: Migration effects on the demographic structure of the sending country. In R. Schoen (Ed.), *Dynamic demographic analysis* (pp. 217–225). Dordrecht: Springer.

Bogin, B., & Loucky, J. (1997). Plasticity, political economy, and the physical growth status of Guatemala Maya children living in the United States. *American Journal of Physical Anthropology, 102*(1), 17–32.

Brault, S., & Caswell, H. (1993). Pod-specific demography of killer whales (*Orcinus orca*). *Ecology, 74*, 1444–1454.

Caswell, H. (2000). Prospective and retrospective perturbation analyses: Their roles in conservation biology. *Ecology, 8*, 619–627.

Caswell, H. (2001). *Matrix population models: Construction, analysis, and interpretation* (2nd ed.). Sunderland: Sinauer Associates, Inc..

Caswell, H. (2010). Life table response experiment analysis of the stochastic growth rate. *Journal of Ecology, 98*, 324–333.
Caswell, H., & Kayne, T. (2001). Stochastic demography and conservation of *Lomatium bradshawii* in a dynamic fire region. *Advances in Ecological Research, 32*, 1–51.
Caswell, H., & Werner, P. (1978). Transient behavior and life history analysis. *Ecology, 59*, 53–66.
Cerone, P. (1987). On stable population theory with immigration. *Demography, 24*, 431–438.
Christensen, R. (1997). *Log-linear models and logistic regression* (2nd ed.). New York: Springer-Verlag.
Coale, A. J. (1972). *The growth and structure of human populations: A mathematical investigation*. Princeton: Princeton University Press.
Coale, A. J., & Demeny, P. (1966). *Regional model life tables and stable populations*. Princeton: Princeton University Press.
Coale, A. J., & Trussell, J. (1974). Model fertility schedules: Variations in the age structure of childbearing in human populations. *Population Index, 40*(2), 185–258.
Cohen, J. (1979). Ergodic theorems in demography. *Bulletin of the American Mathematical Society, 1*(2), 275–295.
Coleman, D., & Montgomery, D. (1993). A systematic approach to planning for a designed industrial experiment. *Technometrics, 35*(1), 1–12.
Cushing, J. (1998). *An introduction to structured population dynamics*. Philadelphia: Society for Industrial and Applied Mathematics.
de Kroon, H., Groenendael, J., & Ehrlen, J. (2000). Elasticities: A review of methods and model limitations. *Ecology, 81*, 607–618.
Espenshade, T. (1986). Population dynamics with immigration and low fertility. *Population and Development Review, 12*, 248–261.
Espenshade, T., Bouvier, L., & Arthur, W. (1982). Immigration and the stable population model. *Demography, 19*, 125–133.
Fisher, R. (1930). *On the genetical theory of natural selection*. New York: Plenum.
Gantmacher, F. (1959). *Matrix theory*. New York: Chelsea.
Gardiner, C. (1983). *A handbook of stochastic methods for physics, chemistry, and the natural sciences*. New York: Springer.
Gentile, J., Gentile, M., Hairston, N., & Sullivan, B. (1982). The use of life tables for evaluating chronic toxicity of pollutants to *Mysidopsis bahia*. *Hydrobiologia, 93*(1), 179–187.
Graham, C., & Talay, D. (2013). *Stochastic simulation and Monte Carlo methods*. Dordrecht: Springer.
Hastie, T., Tibshirani, R., & Friedman, J. (2009). *Elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York: Springer.
Hellenic Statistical Authority. (2016). *Statistics: Health and demography*. Retrieved from http://www.statistics.gr/en/statistics/pop
Keyfitz, N. (1977). *Introduction to the mathematics of population*. New York: Addison-Wesley.
Keyfitz, N., & Caswell, H. (2005). *Applied mathematical demography*. New York: Springer.
Kim, Y., & Schoen, R. (1993). On the intrinsic force of convergence to stability. *Mathematical Population Studies, 4*(2), 89–102.
Lasker, G. (1969). Human biological adaptability: The ecological approach in physical anthropology. *Science, 166*(3912), 1480–1486.
Lefkovitch, L. P. (1971). Some comments on the invariants of population growth. *International Symposium of Stat Ecology New Haven, V2*, 337–360.
Lemieux, C. (2009). *Monte Carlo and quasi-Monte Carlo sampling*. New York: Springer.
Leslie, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika, 33*, 183–212.
Leslie, P. H. (1948). Some further notes on the use of matrices in population mathematics. *Biometrika, 35*, 213–245.
Lewontin, R., & Cohen, D. (1969). On population growth in a randomly varying environment. *Proceedings of the National Academies of Science, 62*, 1056–1060.

Longstaff, B. (1984). An extension of the Leslie matrix model to include a variable immature period. *Austral Ecology, 9*(3), 289–293.

Lotka, A. J. (1907). Relation between birth rates and death rates. *Science (New Series), 26*(653), 21–22.

Lotka, A. J. (1956). *Elements of mathematical biology*. New York: Dover.

Marshall, J. (1962). The effects of continuous gamma radiation of the intrinsic rate of natural increase in *Daphia pulex*. *Ecology, 43*(4), 598–607.

McPeek, M., & Kalisz, S. (1993). Population sampling and bootstrapping in complex designs: Demographic analysis. In S. Scheiner & J. Gurevitch (Eds.), *Design and analysis of ecological experiments* (pp. 232–252). New York: Chapman Hall.

Mitchell, M. (2012). *Interpreting and visualizing regression models using Stata*. College Station: Stata Press.

Mitra, S. (1983). Generalization of immigration and the stable population model. *Demography, 20*, 111–115.

Mitra, S. (1990). Immigration, below-replacement fertility, and long-term national population trends. *Demography, 9*(4), 595–612.

Mitra, S., & Cerone, P. (1986). Migration and stability. *Genus, 42*(1-2), 1–12.

Preston, S., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes*. Malden: Blackwell Publishing.

Rago, P., & Goodyear, C. (1987). Recruitment mechanisms of striped bass and Atlantic salmon: Comparative liabilities of alternative life histories. *American Fish Society Symposium, 1*, 402–416.

Schoen, R. (2010). *Dynamic population models*. New York: Springer.

Shores, T. (2007). *Linear algebra and its applications*. New York: Prentice-Hall.

Sivamurthy, M. (1982). *Growth and structure of human population in the presence of migration*. London: Academic Press.

Swanson, D., Tedrow, L., & Baker, J. (2016). Exploring stable population concepts from the perspective of cohort change ratios: Estimating the time to stability and intrinsic r from initial information and components of change. In R. Schoen (Ed.), *Dynamic demographic analysis* (pp. 227–258). Dordrecht: Springer.

Sykes, Z. M. (1969). Stochastic versions of the matrix model of population dynamics. *Journal of the American Statistical Association, 64*(325), 111–130.

Taguchi, G. (1986). *An introduction to quality engineering: Designing quality into products and processes*. Tokyo: Asian Productivity Agency.

Taylor, H., & Karlin, S. (1998). *An introduction to stochastic modeling* (3rd ed.). New York: Academic Press.

Tuljapurkar, S. (1982). Why use population entropy? It determines the rate of convergence. *Journal of Mathematical Biology, 13*, 325–337.

Wachter, K. (2014). *Essential demographic methods*. Cambridge, MA: Harvard University Press.

Wilson, E., & Bossert, W. (1971). *A primer of population biology*. Sunderland: Sinauer Publishers.

Wisdom, M., & Mills, S. (1997). Sensitivity analysis to guide population recovery: Prairie chickens as an example. *Journal of Wildlife Management, 61*, 302–312.

Wisdom, M., Mills, S., & Doak, D. (2000). Life-stage simulation analysis: Estimating vital rate effects on population growth for conservation. *Ecology, 81*, 628–641.

Chapter 13
Decompositions

13.1 Introduction

Decomposition methods in demography attempt to partition rates of population change into constituent contributors such as population structure or other characteristics that might confound an overall comparison (Canudas-Romo 2003). For example, when comparing the change in the crude fertility rate over time, observed differences might be due to changes in the female age-structure, changes in age-specific rates of childbearing, or a combination of the two. A decomposition of the crude birth rate can ascertain whether its change is due to increasing or decreasing fertility rates or changes in the population age composition (Canudas-Romo 2003; Das Gupta 1978; Kitagawa 1955). This form of decomposition is similar to standardization methods in epidemiology and demography that attempt to adjust comparisons of population-level indicators differences in demographic composition (Aschengrau and Seage 2003; Palmore and Gardner 1994).

Similarly, one might consider components of change within the fundamental equation of demography (see Eq. 2.1 in Chapter 2 and the Appendix) to be a decomposition of an overall rate of population change, such as the cohort change ratio. This form of decomposition is directly related to the procedure of "controlling" for confounding variables in biostatistics (Agresti 2013; Aschengrau and Seage 2003; Canudas-Romo 2003; Hastie et al. 2009) and to the analysis of sensitivities in population ecology (Caswell 2000; de Kroon et al. 2000; Tuljapurkar 1982). Decomposition, then, ought to be possible for cohort change ratios (CCRs) in order to analyze subgroup contributions to growth or to decompose population change into its components related to survivorship and migration. In this chapter, we present algebraic derivations for these types of decompositions and illustrate this architecture by analyzing 2010–2020 population change in the state of California. We then discuss additional possibilities for applying decompositions to demographic analyses involving CCRs.

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_13

13.2 Decompositions

13.2.1 Subgroup Decomposition

Often in demographic analysis, one wants to assess the proportional contribution of population growth in a specific subgroup or set of subgroups to overall population change. A decomposition of the cohort change ratio into components contributed by subgroups is one way to approach this question. We begin with the overall CCR:

$$ {}_{n}CCR_{x,t} = {}_{n}P_{x,t} / {}_{n}P_{x-k,t-k} \tag{13.1}$$

where,

${}_{n}P_{x,\,t}$ is the population aged x to $x+n$ at the most recent census (t),
${}_{n}P_{x-k,\,t-k}$ is the population aged $x-k$ to $x-k+n$ at the 2nd most recent census $(t-k)$, and
k is the number of years between the most recent census at time t and the one preceding it at time $t-k$.

The overall CCR can be represented as the weighted average of subgroup CCRs:

$$ {}_{n}CCR_{x,t} = \frac{n_{1,t-k}\ {}_{n}CCR_{x,1,t} + n_{2,t-k}\ {}_{n}CCR_{x,2,t} + \cdots + n_{g,t-k}\ {}_{n}CCR_{x,g,t}}{N_{t-k}} \tag{13.2}$$

where,

g is a subgroup;
n is the population for each subgroup at time t−k; and
N is the total population of that cohort at time t−k.

The weighting for each subgroup, then, is proportional to its contribution to the entire cohort population (N), such that we may rewrite this decomposition as the convex combination of population weights (this means that the weights are all positive and sum to 1.00) and CCRs as:

$$ {}_{n}CCR_{x,t} = w_{1n}CCR_{x,1,t} + w_{2n}CCR_{x,2,t} + \ldots + w_{gn}CCR_{x,s,t} \tag{13.3}$$

where,

w is the population weight for a subgroup based on n / N.

The proportional contribution of each subgroup's CCR to total population CCR is captured in the weighting term. This relationship permits the decomposition of an overall CCR into the contribution of its subgroup's CCRs as:

$$ p_g = w_{g,n}CCR_{x,g,t} / {}_{n}CCR_{x,t} \times 100 \tag{13.4}$$

where,

p is the relative contribution of a subgroup CCR to the overall CCR.

13.2.2 Components of Change Decomposition

As discussed in Chapter 1, a CCR represents a geometric rate of change that captures information on both mortality and migration in a single factor and as we know from the Appendix, a CCR is algebraically equivalent to the fundamental population theorem. As such, it may be decomposed into its constituent elements using standard demographic approaches (Canudas-Romo 2003; Das Gupta 1978; Kitagawa 1955). The numerator of the CCR (${}_nP_{x,t}$) can be decomposed into survivorship and net migration components as follows:

$$ {}_{n}P_{x,t} = ({}_{n}P_{x-k,t-k} \times {}_{n}SR_{x,}) + ({}_{n}P_{x-k,t-k} \times {}_{n}NMR_{x}) \tag{13.5} $$

where,

SR is the survival rate,
NMR is the net migration rate.

The quantity inside the first set of parentheses is the survived population at time *t* and quantity inside the second set of parentheses is the net migration from time $t-k$ *to* time *t*. Therefore, the CCR can be re expressed as:

$$ {}_{n}CCR_{x,t} = (({}_{n}P_{x-k,t-k} \times {}_{n}SR_{x}) + ({}_{n}P_{x-k,t-k} \times {}_{n}NMR_{x}))/{}_{n}P_{x-k,t-k}. \tag{13.6} $$

By dividing the right side of de of Eq. 13.6 by ${}_nP_{x-k,t-k}$, we see that the CCR is the sum of the survivorship rate and net migration rate:

$$ {}_{n}CCR_{x,t} = {}_{n}SR_{x} + {}_{n}NMR_{x}. \tag{13.7} $$

The net migration rate can then derived from by subtracting ${}_nS_x$ from both sides of Eq. 13.7:

$$ {}_{n}NMR_{x} = {}_{n}CCR_{x,t} - {}_{n}SR_{x}. \tag{13.8} $$

Deaths and net migration for persons aged *x* at time $t-k$ who are aged $x+k$ at time *t* are found by:

$$ {}_{n}D_{x} = {}_{n}P_{x-k,t-k} \times (1 - {}_{n}SR_{x}) \text{ and} \tag{13.9} $$

$$ {}_{n}NM_{x} = {}_{n}P_{x-k,t-k} \times {}_{n}NMR_{x} \tag{13.10} $$

where,

D is the deaths over the period between time $t-k$ and time *t*, and
NM is net migrants over the period between time $t-k$ and time *t*.

Equations 13.9 and 13.10 together are equivalent to the "forward-survival rate" (FSR) method of indirect estimation of net migration (Siegel 2002: 22–23). These equations measure the deaths and net migration of persons alive at time $t-k$. It does not account for births, deaths, or net migration of persons born after time $t-k$. We will show how these components of change in aggregate can be measured using results from the CCR decomposition of survivorship and net migration. Like the FSR method, any error in the census counts or survival rates is transmitted to the net migration rates and numbers.

13.2.3 Subgroup and the Components of Change Decomposition

In the preceding section, we presented a method for deriving the components of total population change from the CCR and life table survival rates. Here, we elaborate on this decomposition by proposing a method that determines the contribution of the components of change within each subgroup to overall population change reflected in the CCR. Specifically, our aim in this section is to determine the relative contribution (RelCon) of subgroup specific deaths and migration to the total population change.

We begin by calculating the components of change for each subgroup by modifying Eqs. 13.8, 13.9, and 13.10 to make them specific for each subgroup (g):

$$ {}_{n}NMR_{x,g} = {}_{n}CCR_{x,t,g} - {}_{n}SR_{x,g}, \tag{13.11} $$

$$ {}_{n}D_{x,g} = {}_{n}P_{x-k,t-k,g} \times (1 - {}_{n}SR_{x,g}), \text{ and} \tag{13.12} $$

$$ {}_{n}NM_{x,g} = {}_{n}P_{x-k,t-k,g} \times {}_{n}NMR_{x,g}. \tag{13.13} $$

Therefore, the total population change is derived by:

$$ {}_{n}PChg_{x} = \sum {}_{n}NM_{x,g} - {}_{n}D_{x,g}, \text{ where, } \sum \text{ is the sum across the subgroup } (g). \tag{13.14} $$

The RelCon's for deaths and migration are derived by:

$$ {}_{n}RelConD_{x,g} = -{}_{n}D_{x,g} / \pm {}_{n}PChg_{x} \text{ and} \qquad \text{(Deaths)} \tag{13.15} $$

$$ {}_{n}RelConNM_{x,g} = \pm {}_{n}NM_{x,g} / \pm {}_{n}PChg_{x}. \qquad \text{(Migration)} \tag{13.16} $$

Normally a RelCon would represent a proportion that ranges from zero to one, but in this application some RelCons may be less than zero and others greater than one. When the total population decreases, the RelCon for deaths and negative migration will be positive and the RelCon for positive migration will be negative. The opposite occurs when the total population increases; the RelCon for deaths and negative migration will be negative and the RelCon for positive migration will be positive. In this application, the only condition where the RelCons will fall between zero and one is if the total population decreases and the migration for all subgroups is negative.

A RelCon smaller than −1.00 or larger than 1.00 means that changes in one or more categories will have to overcompensate for changes in the other categories. For example, Hispanic and non-Hispanic deaths of 1920 and 1760, respectively, and Hispanic and non-Hispanic migration of 131,590 and −12,708, respectively, leads to a total population change of 115,192 (131,590 – 12,708 – 1920 – 1760). In this instance the RelCon for Hispanic migration is 1.142 (131,590 / 115,192), while the RelCon's for the other categories are all less than zero.

13.3 Applications

13.3.1 Contribution of Subgroup CCRs to the Total CCR

One often is confronted with questions about the proportional contribution of different subgroups to overall population growth, such as the contribution of Hispanics to overall population growth in the United States (Krogstad et al. 2015; Passel and Cohn 2008). Between 2000 and 2010, Hispanics had the largest growth of any ethnic minority in the Unites States (Passel et al. 2011). New Mexico had the highest percentage of Hispanics in 2010 (46.3%) and California was the state with the largest overall Hispanic population (14,013,719). A majority of the Hispanic population growth between 2000 and 2010 has been attributed to increases in Hispanic responses to the census (Ennis et al. 2011), raising questions about the overstating the impact Hispanics on overall future U.S. population dynamics.

A decomposition of the cohort change ratio for the overall population into Hispanic and non-Hispanic components provides a means for analyzing the question in the context of CCRs. Table 13.1 provides this decomposition based on Eq. 13.4 for the Hispanic and non-Hispanic populations of the State of California from 2010 to 2020. The 2010 age-specific proportions (Hispanic proportion of the total population) show the youth of the Hispanic population. There is a clear inverse relationship between the Hispanic proportion and the age of the population; the proportion ranges from 0.532 for ages 0 to 4 to 0.159 for ages 75 years and older. For populations under the age of 35, the Hispanic proportion exceeds that for the total Hispanic population (0.377). For Hispanics over the age of 35 years, the proportion declines rapidly and is substantially below the proportion for the total Hispanic population. In the oldest age group, Hispanics comprise just 0.159 of the population aged 75 years and older compared to 0.377 for the total Hispanic population.

Table 13.1 Relative contribution of Hispanics and non-Hispanics to population change by age, California, 2010–2020

	2010 population			Age-specific proportions		Cohort change ratio (2010/2000)			Contribution to total CCR	
Age	Hispanic	Non-Hispanic	Total	Hispanic[a]	Non-Hispanic[b]	Hispanic	Non-Hispanic	Total	Hispanic[c]	Non-Hispanic[b]
0–4	1,340,017	1,180,995	2,521,012	0.532	0.468	1.09676	0.98775	1.03981	56.1%	43.9%
5–9	1,289,578	1,207,569	2,497,147	0.516	0.484	1.09831	0.98686	1.03743	54.7%	45.3%
10–14	1,302,813	1,283,187	2,586,000	0.504	0.496	1.17665	1.01356	1.07975	54.9%	45.1%
15–19	1,358,358	1,469,548	2,827,906	0.480	0.520	1.18744	1.07941	1.12303	50.8%	49.2%
20–24	1,227,751	1,548,104	2,775,855	0.442	0.558	1.08416	1.08838	1.08657	44.1%	55.9%
25–29	1,174,951	1,577,459	2,752,409	0.427	0.573	1.01283	0.98039	0.99357	43.5%	56.5%
30–34	1,107,661	1,479,772	2,587,433	0.428	0.572	0.99110	0.97830	0.98308	43.2%	56.8%
35–39	1,058,789	1,498,361	2,557,149	0.414	0.586	0.99249	0.94047	0.95646	43.0%	57.0%
40–44	982,838	1,627,584	2,610,422	0.377	0.623	0.97109	0.95116	0.95651	38.2%	61.8%
45–49	853,903	1,822,998	2,676,901	0.319	0.681	0.99046	0.93437	0.94707	33.4%	66.6%
50–54	700,256	1,869,481	2,569,737	0.273	0.727	0.96152	0.92690	0.93374	28.1%	71.9%
55–59	525,690	1,695,438	2,221,128	0.237	0.763	0.95538	0.88338	0.89656	25.2%	74.8%
60–64	377,110	1,477,631	1,854,741	0.203	0.797	0.91180	0.83985	0.85268	21.7%	78.3%
65–69	256,546	1,058,934	1,315,480	0.195	0.805	0.82097	0.73272	0.74724	21.4%	78.6%
70–74	186,510	791,381	977,891	0.191	0.809	0.77630	0.68929	0.70214	21.1%	78.9%
75+	314,826	1,663,345	1,978,171	0.159	0.841	0.43818	0.34544	0.35503	19.6%	80.4%
Total	14,057,596	23,251,786	37,309,382	0.377	0.623					

Source: 2000 and 2010, U.S. Census Bureau (http://factfinder2.census.gov)
[a]Hispanic / Total
[b]1 – Hispanic
[c]Hispanic proportion × Hispanic CCR / total CCR × 100

The age-specific proportion trends for the non-Hispanic population are the reverse of those for the Hispanic population, which is expected since the Hispanic and non-Hispanic proportions are complements. A clear direct relationship is seen between the non-Hispanic proportion and the age of the population; the proportion ranges from 0.468 for ages 0 to 4 to 0.841 for ages 75 years and older. For the population over the age of 44 years, the non-Hispanic proportion exceeds that for the total non-Hispanic population (0.623). For non-Hispanics under the age of 44 years, the proportion declines steadily and is below the proportion for the total non-Hispanic population. In the youngest age group, non-Hispanics make up 0.468 of ages 0 to 4 population compared 0.623 for the total non-Hispanic population.

Hispanic CCRs are highest in the age groups 0 to 4 through 25 to 29, falling below 1.0 (indicating a decline in the age cohort) at ages 30 to 44 years and for subsequent age groups. Non-Hispanic CCRs are generally lower than Hispanic CCRs, except for ages 20 to 24, which is not surprising because in California non-Hispanics have lower survivorship and considerably less net migration than Hispanics. Among non-Hispanics, we see the same general pattern across age groups as for Hispanics. However, the non-Hispanic CCR is less than 1.0 in the two youngest age groups and they fall below 1.0 at ages 25 to 29, 5 years earlier than the Hispanic CCR.

These last two columns of Table 13.1, based on Eq. 13.4, show the relative contributions (in percentage terms) of Hispanics and non-Hispanics to age-specific total population change from 2010 to 2020 embedded in the total CCRs. The contribution to the total CCR is greater for Hispanics than non-Hispanics in the three youngest age group (ages 0 to14), and the contributions of the two subgroups are roughly equal for ages 15 to19. For the other age groups, the contribution of Hispanics declines from 44.1% for ages 20 to 24 to 19.6% for ages 75 years and older. Given that the Hispanic and non-Hispanic contribution are complements, declines in the Hispanic contribution are reflected by increases in the non-Hispanic contribution.

Because there are fewer Hispanics than non-Hispanics in each age-group, their relative contribution declines with age despite the fact that Hispanic CCRs remain higher than non-Hispanic CCRs for all age groups. However for each age group, the Hispanic contribution to the total CCR between 2010 and 2020 is greater than their share of the population in the year 2010. Conversely, the non-Hispanic contribution is less than their share of the 2010 population.

13.3.2 Indirect Forecasts of the Components of Change

While the relative impact of Hispanic CCRs and non-Hispanic CCRs on total CCRs is evident, it says nothing about the reasons for the age-specific population changes; that is, changes due to survivorship and net migration. Of course, the survivorship component will decrease the population and the net migration will either increase or decrease the population. Some very general statements about net migration can be ascertained from the CCRs. A CCR of 1.0 suggests a positive net migration sufficient enough to exactly offset any declines due to survivorship. A CCR exceeding 1.0 suggests a positive net migration that offsets any declines due to survivorship. The interpretation of a CCR less than 1.0 is not definitive about the direction of the migration. A CCR less than one could indicate positive net migration that is not sufficient to offset declines due to survivorship or a negative net migration that accentuates the declines due to survivorship.

What can we say about net migration from the CCRs for Hispanics and non-Hispanics in California? Net-migration is positive and large for Hispanics during childhood and young adulthood (ages 0 to 24), which suggests family migration in which young adults bring dependents (See Fig. 13.1). For non-Hispanics, the CCRs suggest positive net-migration among young adults aged 10 to 24, but without a corresponding positive migration of youngest dependents (ages 0 to 9). The high survivorship probabilities for ages 30 to 59 suggest negative net migration in this age range for both Hispanics and non-Hispanics, with non-Hispanics showing a greater loss due to migration. Because of declining survivorship probabilities in the older age groups (e.g., ages 60 years and older), CCRs will generally be less than 1.0; this is the case for both Hispanic and non-Hispanics.[1] All that can be inferred about net migration in

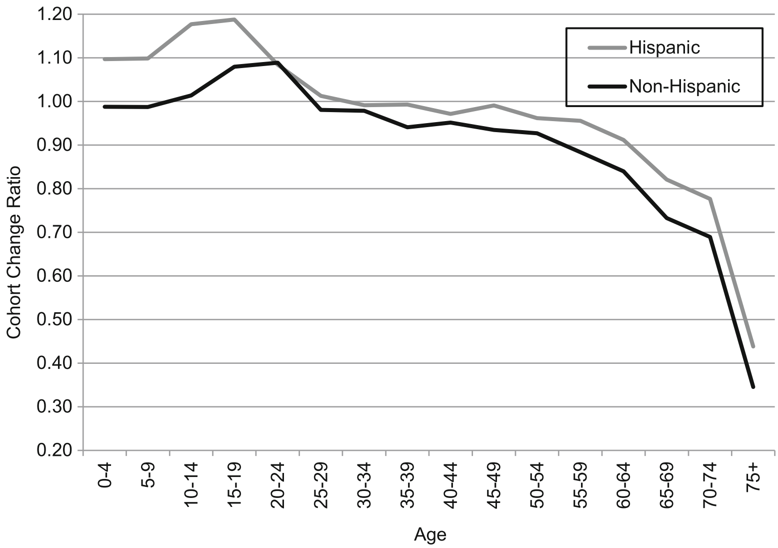

Fig. 13.1 Cohort change ratios by age and Hispanic origin, California, 2000–2010 (Source: U.S. Census Bureau, 2000 and 2010 censuses (http://factfinder2.census.gov))

[1]The notable exception to this generalization will be in places with large retirement related in-migration. For example in Maricopa County, Arizona, the CCRs are 1.10, 1.09, and 1.02 for the five-year age groups from 60 to 74. Even the CCR of 0.87 for ages 75 to 79 suggests net in-migration given the relatively low survivorship probabilities in this age group.

these age groups is that any net migration will be greater (more positive or less negative) for Hispanics because they have larger CCRs than the non-Hispanics.

More precise statements about impact of components of change on the dynamics of Hispanic versus non-Hispanic population growth could be made from a direct estimate of the survivorship and net-migration components of growth using Eqs. 13.8, 13.9, and 13.10. These computations require age-specific survival rates, ideally computed from a life table. Our example is uses a 10-year forecast horizon, so 10-year survival rates are required. That is, the probability of those aged 0 to 4 surviving to ages 10 to 14, the probability of those aged 5 to 9 surviving to ages 15 to19 and so forth until the last survival rate for the opened-end age group (e.g., the probability of those aged 75 years and older surviving to ages 85 years and older. Smith et al. (2013: 58–60) provide details for constructing life table survival rates.

Survival rates for Hispanics and non-Hispanics were developed from 2010 abridged life tables for California. Life tables were created using age-specific death rates (under 1 year of age, 1 to 4, 5 to 9,. . ..., 80 to 84, and 85 years and older) for Hispanics and non-Hispanics. The 2010 census furnished denominators for the rates and the deaths were obtained from the California Department of Public Health (2016: Tables 5-4 and 5-5). Non-Hispanic deaths were obtained by subtracting Hispanic deaths from total deaths by age. The Hispanic death data had broad ranges for some age groups (1 to 14, 15 to 24, 25 to 34, 35 to 44, and 45 to 54). Deaths for these age groups were partitioned into the finer classification used shares from total deaths. We adjusted the 2010 age-specific death rates so the life tables would conform to 2020 life expectancy assumptions from the latest forecasts developed by the California Department of Finance (2014).

Based on Eqs. 13.8, 13.9, and 13.10, Table 13.2 provides information about the components of change for the Hispanic population in the State of California from 2010 to 2020. Specific age ranges mentioned pertain to the 2020 age groups. Hispanics alive in 2010 experienced 465,677 deaths and a net in-migration of 932,710, for a population gain of 457,033. With exception of age groups 50 to 54, 75 to79, and 85 years and older net-migration is positive in the other age groups. The highest net migration rates (0.180 and 0.193) are found in ages 20 to 29. These age groups have a net migration of 497,757 or 53.4% of the total Hispanic net migration. Net migration rates of around 0.10 seen in ages 10 to19 and 30 to 34 indicate significant net in-migration of Hispanic children and teenagers as well as other young adults. Positive net migration rates and numbers in other age groups are small by comparison.

The positive net migration and relatively high survival rates results in positive population change for ages 10 to 39. The population declines in the other age groups as net migration is not enough to offset the impact of decreasing survival

Table 13.2 Decomposition of Hispanic population growth into components of change, California, 2010–2020

2010 age	2020 age	2010[a] population	Cohort[a] change ratio	10-year survival rate	10-year net[b] migration rate	2010–2020 Deaths[c]	2010–2020 Net[d] migration	2010–2020 Population[e] change
0–4	10–14	1,340,017	1.09676	0.99856	0.09820	1,930	131,590	129,660
5–9	15–19	1,289,578	1.09831	0.99843	0.09988	2,025	128,803	126,778
10–14	20–24	1,302,813	1.17665	0.99618	0.18047	4,977	235,119	230,142
15–19	25–29	1,358,358	1.18744	0.99409	0.19335	8,028	262,638	254,610
20–24	30–34	1,227,751	1.08416	0.99365	0.09051	7,796	111,124	103,328
25–29	35–39	1,174,951	1.01283	0.99307	0.01976	8,142	23,217	15,075
30–34	40–44	1,107,661	0.99110	0.99106	0.00004	9,902	44	−9,858
35–39	45–49	1,058,789	0.99249	0.98678	0.00571	13,997	6,046	−7,951
40–44	50–54	982,838	0.97109	0.97892	−0.00783	20,718	−7,696	−28,414
45–49	55–59	853,903	0.99046	0.96659	0.02387	28,529	20,383	−8,146
50–54	60–64	700,256	0.96152	0.94995	0.01157	35,048	8,102	−26,946
55–59	65–69	525,690	0.95538	0.92824	0.02714	37,724	14,267	−23,457
60–64	70–74	377,110	0.91180	0.89529	0.01651	39,487	6,226	−33,261
65–69	75–79	256,546	0.82097	0.84141	−0.02044	40,686	−5,244	−45,930
70–74	80–84	186,510	0.77630	0.75825	0.01805	45,089	3,367	−41,722
75+	85+	314,826	0.43818	0.45494	−0.01676	171,599	−5,276	−176,875
Total		14,057,596				475,677	932,710	457,033

[a] From Table 13.1
[b] ${}_nCCR_x - {}_nSR_x$
[c] ${}_nP_{x,t-k} \times (1 - {}_nSR_x)$
[d] ${}_nP_{x,t-k} \times {}_nNMR_x$
[e] ${}_nNM_{x,t-k \text{ to } t} - {}_nD_{x,t-k \text{ to } t}$

probabilities. Between 2010 and 2020, the Hispanic population grew by 3,761,528.[2] Of that about 25% is due to net migration of persons alive in 2010 (932,710). The gain due to migration is offset by 475,677 deaths. So, the net impact on total Hispanic population growth from 2010 to 2020 of persons alive in 2010 is 457,033 (or 12.2%). Therefore, the bulk of the total population change in the Hispanic population in California is due to persons born after the year 2010 (3,304,495 = 3,761,528 – 457,033). This figure includes births that occurred in California, net migration of children ages 0 to 9, and population loss due to mortality, so it is not a pure measure of the fertility component of change. However, it seems clear that Hispanic fertility is the primary reason for growth in the Hispanic population in California between 2010 and 2020.

Table 13.3 provides information about the components of change for the non-Hispanic population in the State of California from 2010 to 2020. Non-Hispanics alive in 2010 experienced 2,116,806 deaths and a net out-migration of −182,692, for a population loss of −2,299,498. With exception of those aged 10 to 24, net migration is negative in the other age groups. The highest net migration rates (0.086 and 0.096) are found in ages 25 to 34. These age groups have a net migration of 274,248, which is not enough to offset the net migration loss in the other age groups. Non-Hispanic net in-migration rates in these age groups are roughly 50% lower than the largest net in-migration rates seen in the Hispanic population. Also, the peak positive net migration occurs 5 years later (ages 30 to 34) in the non-Hispanic population. The largest non-Hispanic net-out migration occurs in ages 75 to 79 (−85,339). The lowest amount of net out-migration ($< -20{,}000$ net migrants) occurs in ages 10 to 19, 34 to 44, and 80 to 84.

The positive net migration and relatively high survival rates results in positive population change for a narrow range of ages, 10 to 34. The population decreases in the other age groups because the losses due to net migration are further accentuated by deaths. This is especially evident in the population 50 years and older, where deaths exceed the net out-migration in every age group. The population loss is −79,491 for ages 50 to 54 and the population loss dramatically increases, reaching −1,088,759 in persons 85 years and older. Between 2010 and 2020, the Hispanic population grew by only 149,941. The net impact of migration (−182,692) and deaths (−2,116,806) on persons alive in 2010 is −2,299,498. The only reason there is a positive change in the non-Hispanic total population in California (149,941) between 2010 and 2020 is due to the births after the year 2010, which offset population loses due to net out-migration and deaths (2,449,439 = 149,941 – (−2,299,498)).

[2]The 2020 Hispanic and non-Hispanic populations aged 0 to 9 were based on 2010 Child-Adult Ratios (CADs). Chapter 5 illustrates the use of CADs in forecasting. The 2020 total population forecasts were computed by adding the population aged 0 to 9 to the population aged 10 to 85 years and older obtained using CCRs.

Table 13.3 Decomposition of non-Hispanic population growth into components of change, California, 2010–2020

						2010–2020		
2010 age	2020 age	2010[a] population	Cohort[a] change ratio	10-year survival rate	10-year net[b] migration rate	Deaths[c]	Net[d] migration	Population[e] change
0–4	10–14	1,180,995	0.98775	0.99851	−0.01076	1,760	−12,708	−14,468
5–9	15–19	1,207,569	0.98686	0.99830	−0.01144	2,053	−13,815	−15,868
10–14	20–24	1,283,187	1.01356	0.99579	0.01777	5,402	22,802	17,400
15–19	25–29	1,469,548	1.07941	0.99351	0.08590	9,537	126,234	116,697
20–24	30–34	1,548,104	1.08838	0.99277	0.09561	11,193	148,014	136,821
25–29	35–39	1,577,459	0.98039	0.99128	−0.01089	13,755	−17,179	−30,934
30–34	40–44	1,479,772	0.97830	0.98804	−0.00974	17,698	−14,413	−32,111
35–39	45–49	1,498,361	0.94047	0.98177	−0.04130	27,315	−61,882	−89,197
40–44	50–54	1,627,584	0.95116	0.97150	−0.02034	46,386	−33,105	−79,491
45–49	55–59	1,822,998	0.93437	0.95752	−0.02315	77,441	−42,202	−119,643
50–54	60–64	1,869,481	0.92690	0.93816	−0.01126	115,609	−21,050	−136,659
55–59	65–69	1,695,438	0.88338	0.91267	−0.02929	148,063	−49,659	−197,722
60–64	70–74	1,477,631	0.83985	0.87531	−0.03546	184,246	−52,397	−236,643
65–69	75–79	1,058,934	0.73272	0.81331	−0.08059	197,692	−85,339	−283,031
70–74	80–84	791,381	0.68929	0.71398	−0.02469	226,351	−19,539	−245,890
75+	85+	1,663,345	0.34544	0.37938	−0.03394	1,032,305	−56,454	−1,088,759
Total		23,251,786				2,116,806	−182,692	−2,299,498

[a] From Table 13.1
[b] ${}_{n}CCR_{x} - {}_{n}SR_{x}$
[c] ${}_{n}P_{x,t-k} \times (1 - {}_{n}SR_{x})$
[d] ${}_{n}P_{x,t-k} \times {}_{n}NMR_{x}$
[e] ${}_{n}NM_{x,t-k \text{ to } t} - {}_{n}D_{x,t-k \text{ to } t}$

13.3.3 Contribution of Subgroup Components of Change to Total Population Change

In the prior section, we illustrated the effects of mortality and migration on the population change by age group for Hispanic and non-Hispanic separately. In this section, we illustrate the impact of the subgroup population dynamics on the total population growth across the age intervals. Does the strong and positive migration among young Hispanics drive strong total population growth among these age groups? How does the relatively large numbers of non-Hispanic deaths in persons aged 65 years and older impact the total population change in these age groups? The greatest positive change due to migration for both Hispanics and non-Hispanics occurs in ages 20 to 34, so which subgroup contributes the most to the total population change in these age groups? To answer these and other similar types of questions, we use the age-specific components of change for each subgroup (Hispanic deaths, Hispanic migration, non-Hispanic deaths and non-Hispanic migration) and relate them to the age-specific total population change rather than the age- and subgroup-specific population change shown in Tables 13.2 and 13.3. These relative contributions (RelCon) are expressed in proportionate terms that sum to 1.0 over these four comparison categories or any mutually set of comparison categories. In this illustration, we focus only on changes to the population alive in the launch year (i.e., 2010) and not persons born over the forecast horizon (i.e., 2010–2020).

Interpretation of the RelCon is somewhat involved because it may involve both negative and positive population changes. Deaths always remove people, and migration can either add or remove people depending on its direction, and total population change can be either positive or negative. Since the RelCon sums to 1.00 across categories, some RelCon's may be less than zero and others greater than 1.00, depending on the direction of the change in total population. As noted earlier, the RelCon for deaths and losses due to migration will be positive if the total population change is negative and negative if the total population change is positive. Conversely, for gains due to migration the RelCon will be positive if the total population change is positive and negative if the total population change is negative.

We begin by examining the impact of the components of change and Hispanic origin separately on the total population change by age group (see Table 13.4). For reference, the table includes the total population change. We first examine the relative impact of deaths and migration. The population for all ages decreases by 1.84 million persons due to 2.59 million deaths (1.407), which are offset by a positive migration of 0.75 million (−0.407). Population growth is positive for persons under 34 years of age (age refers to age in 2020) and migration has a much greater influence than deaths. Like for all ages, population change in ages 35 to 39 is more heavily influenced by deaths than migration. Persons age 40 and above show a decline in total population and a population loss due to migration. With the exception of ages 45 to 49, deaths have a larger impact than migration,

Table 13.4 Proportionate contribution of deaths, net-migration, and Hispanic origin to population change, California, 2010–2020[a]

2010 age	2020 age	Population change	Deaths	Net migration	Hispanic	Non-Hispanic
0–4	10–14	115,192	−0.032	1.032	1.125	−0.125
5–9	15–19	110,910	−0.037	1.036	1.143	−0.144
10–14	20–24	247,542	−0.042	1.042	0.930	0.070
15–19	25–29	371,307	−0.048	1.047	0.685	0.314
20–24	30–34	240,149	−0.079	1.079	0.431	0.569
25–29	35–39	−15,859	1.380	−0.381	−0.951	1.950
30–34	40–44	−41,969	0.658	0.342	0.235	0.765
35–39	45–49	−97,148	0.425	0.575	0.082	0.918
40–44	50–54	−107,905	0.622	0.378	0.263	0.737
45–49	55–59	−127,789	0.829	0.170	0.063	0.936
50–54	60–64	−163,605	0.921	0.079	0.164	0.836
55–59	65–69	−221,179	0.840	0.160	0.106	0.894
60–64	70–74	−269,904	0.829	0.171	0.123	0.877
65–69	75–79	−328,961	0.725	0.275	0.140	0.860
70–74	80–84	−287,612	0.944	0.056	0.145	0.855
75+	85+	−1,265,634	0.952	0.049	0.140	0.861
	Total	−1,842,465	1.407	−0.407	−0.248	1.248

[a]Derived from Table 13.5

with proportions ranging from 0.622 for ages 50 to 54 to 0.952 for ages 85 years and older. For ages 45 to 49, migration accounts for 0.575 of the total population loss.

We now look at the impact of Hispanics and non-Hispanics on total population change. The population loss for all ages is due completely to the non-Hispanic population, which declines by 2.30 million (1.248). This decline is offset by growth in the Hispanic population of 0.46 million (−0.248). For persons under the age of 30 years, Hispanics contribute more to the total population growth than non-Hispanics. However, for ages 30 to 34, the last age group with a population increase, non-Hispanics contribute more (0.569) than Hispanics (0.431). For the remaining age groups non-Hispanics are the major factor in their population losses. For ages 35 to 39, non-Hispanics decline by 30,934, which is double the gain seen in the Hispanic population (15,075). For persons 40 years and older, the outsized role of the non-Hispanic population is evidence, with RelCon's ranging from 0.737 for ages 50 to 54 to 0.936 for ages 55 to 59.

Table 13.5 presents the combined effects of the components of change specific to each Hispanic origin subgroup on the total population change. Non-Hispanic deaths are the primary reason for the population loss for all ages (1.149), followed by Hispanic deaths (0.258). Non-Hispanic negative migration accounts for just under 1/10 of the total population loss. These three factors more than compensate for the positive migration of Hispanics (932,710). Hispanic migration is the only reason that the population under 20 years of age increases between 2010 and 2020, offsetting the small loses due to deaths and a loss of 12,708 non-Hispanics due to

Table 13.5 Proportionate contribution of deaths and net-migration by Hispanic origin to population change, California, 2010–2020[a]

2010 age	2020 age	Population change	Deaths Hispanic	Deaths Non-Hispanic	Net migration Hispanic	Net migration Non-Hispanic
0–4	10–14	115,192	−0.017	−0.015	1.142	−0.110
5–9	15–19	110,910	−0.018	−0.019	1.161	−0.125
10–14	20–24	247,542	−0.020	−0.022	0.950	0.092
15–19	25–29	371,307	−0.022	−0.026	0.707	0.340
20–24	30–34	240,149	−0.032	−0.047	0.463	0.616
25–29	35–39	−15,859	0.513	0.867	−1.464	1.083
30–34	40–44	−41,969	0.236	0.422	−0.001	0.343
35–39	45–49	−97,148	0.144	0.281	−0.062	0.637
40–44	50–54	−107,905	0.192	0.430	0.071	0.307
45–49	55–59	−127,789	0.223	0.606	−0.160	0.330
50–54	60–64	−163,605	0.214	0.707	−0.050	0.129
55–59	65–69	−221,179	0.171	0.669	−0.065	0.225
60–64	70–74	−269,904	0.146	0.683	−0.023	0.194
65–69	75–79	−328,961	0.124	0.601	0.016	0.259
70–74	80–84	−287,612	0.157	0.787	−0.012	0.068
75+	85+	−1,265,634	0.136	0.816	0.004	0.045
	Total	−1,842,465	0.258	1.149	−0.506	0.099

[a]Derived from Tables 13.2 and 13.3

migration. Hispanic positive migration (0.950 and 0.707) is more important that non-Hispanic positive migration (0.092 and 0.340) in explaining the population growth in ages 20 to 29. However, non-Hispanic positive migration (0.616) contributes more than Hispanic positive migration (.463) to the population growth in ages 30 to 34.

In general, for the population 35 years and older, non-Hispanic deaths and migration are the largest contributors to the population losses in these age groups; although there are some variations. Non-Hispanic negative migration is the most important factor in ages 35 to 39 (1.083) and ages 45 to 49 (0.637), but non-Hispanic deaths are most important in the other age groups. Hispanic migration even when negative (positive RelCon) contributes little to the total population loss in the population 35 years and older, with impacts ranging from 0.004 for the population ages 85 years and older to 0.071 for ages 50 to 54. Hispanic deaths play a modest role in the total population losses, with RelCon ranging from 0.136 for the population ages 85 years and older to 0.223 for the population aged 55 to 59. For the population 55 years and older, non-Hispanic deaths account for most of the population loss in these ages, with RelCon ranging from 0.606 for ages 55 to 59 to 0.816 for the population 85 years and older. In the two oldest age groups, Hispanic deaths have a greater impact on the total population loss (0.157 and 0.136) than does the loss of population due to non-Hispanic migration (0.068 and 0.045).

13.4 Conclusions

The decompositions presented here provide important insights into population dynamics using CCRs. Certainly, these types of decompositions do not apply where current or historical period CCRs are involved or in forecasts that use a cohort component model where all components of change are explicitly represented. So why use CCRs for this type of analysis? Perhaps the most compelling reason for doing so revolves around the use of CCRs in demographic forecasting where a complete cohort component model is not feasible or cannot be implemented due to time and resource constraints (Smith et al. 2013; Smith and Shahidullah 1995; Swanson et al. 2010). In these instances, CCR-based forecasts have been used successfully with the caveat that no information was available on the components of change.

The decomposition methods shown here may help address this shortcoming in CCR-based forecasting models because they require only survival rates to accompany the CCRs. From a practical standpoint, it is much easier and feasible to create survival rates as opposed to fertility rates and, especially, migration rates for many different forecasting applications. This is particularly relevant for areas with relatively small populations where data availability is problematic. In this context, one could produce scenario-based forecasts by adjusting the CCRs and survival rates and analyze not only analyze the impact of population changes, but be able to make statements about the causes of such changes. Information on the impact of shifts in any subgroup's cohort change ratio or an ability to decompose the impact of survivorship and migration on forecasted population dynamics would be valuable information to this end.

If one makes the additional leap to incorporate cohort change ratios into matrix-based population models as shown in Chapter 12, the links between these decompositions and the analysis of demographic sensitivity becomes immediate as well (Baker et al. 2015; Tuljapurkar 1982). A major goal behind decomposition appears to be to gain an understanding into the sensitivity of overall population dynamics to sub-dynamics (Canudas-Romo 2003; Das Gupta 1978; Keyfitz 1971); a question that is directly related to examining the sensitivity of overall population change to variation in the *(ij)* elements of a forecast matrix (Caswell 2000). In a forecasting matrix encompassing multiple subgroups (Schoen 1986; Rogers 1995), the sensitivity of an *(ij)* element in such a matrix provides direct information on the proportional contribution of overall population growth to survivorship or migration within any subgroup desired. By applying the decompositions presented here within a forecast matrix model framework, one could efficiently describe and analyze both the impact of population structure using decompositions and shifts in using elasticity measures within a single analysis.

The difficulties of operationalizing combinations of sensitivity analysis and population structure impacts (as in decompositions) is not trivial and has confounded demographic analyses in both human populations (Baker et al. 2015; Preston and Coale 1982) and animal populations (Wisdom and Mills 1997; Wisdom et al. 2000). However, it has been shown that extensions of this analysis could link decomposition and sensitivity analysis within a matrix framework without necessarily resorting to complex computational and simulation-based frameworks (Baker et al. 2015; Wisdom and Mills 1997; Wisdom et al. 2000).

The decompositions presented are easily generalizable in terms of numbers and categorization schemes for subgroups. Provided that the subgroups are appropriately defined, the methods presented in this chapter are scalable and can accommodate more than the two subgroups analyzed. The framework of contained in Eqs. 13.2, 13.3, and 13.4 is simply a weighted average. As such, there is no limitation to the groups that might be compared as long as they are mutually exclusive and exhaustive categories of the overall population (Agresti 2013; Christensen 1997; Witmer and Samuels 1998). For example, one might imagine expanding the decomposition to include gender or to further decompose Hispanic ethnicity into the complex racial designations. This capability is, of course, tempered by the quality and reliability of the available data.

Decomposition is a challenging, but important area of analysis within demography (Canudas-Romo 2003). In the context of population analysis and forecasting using CCRs, decomposition provides a means for analyzing important numerical determinants of population dynamics including shifts in composition and components of change (Das Gupta 1978; Keyfitz and Caswell 2005) as well as rates (Caswell 2000; Caswell and Werner 1978; de Kroon et al. 2000; Tuljapurkar 1982). The decompositions presented here explore both types of decompositions and provide an example of how to combine them in a way that provides insights into the basic components of population dynamics. Extensions outlined here, including melding decomposition methods with sensitivity analysis, will likely provide fruitful avenues for further research on the applications of CCRs in demographic analysis.

References

Agresti, A. (2013). *Categorical data analysis* (3rd ed.). New York: Wiley.

Aschengrau, A., & Seage, G. (2003). *Essentials of epidemiology in public health*. Sudbury: Jones and Bartlett.

Baker, J., Hill, K., Hurtado, A., Alcantara, A., Hunsinger, E., & Sprague, W. (2015). Microdemographic determinants of population recovery among the Northern Ache. *Human Biology* 87(1), Paper 70.

California Department of Finance. (2014). *State and county population projections by race/ethnicity, sex, and age 2010-2060*. Retrieved from http://www.dof.ca.gov/Forecasting/Demographics/Projections/

California Department of Public Health. (2016). *Death statistical data tables*. Retrieved from https://www.cdph.ca.gov/data/statistics/Pages/DeathStatisticalDataTables.aspx

Canudas-Romo, V. (2003). *Decomposition methods in demography* (Doctoral dissertation). Retrieved from http://www.rug.nl/research/portal/files/10068144/thesis.pdf

Caswell, H., & Werner, P. (1978). Transient behavior and life history analysis of teasel. *Ecology, 59*, 53–66.

Caswell, H. (2000). Prospective and retrospective perturbation analysis: Their roles in conservation biology. *Ecology, 81*, 619–627.

Christensen, R. (1997). *Log-linear models and logistic regression* (2nd ed.). New York: Springer-Verlag.

Das Gupta, P. (1978). A general method for decomposing a difference between two rates into several components. *Demography, 15*(1), 99–112.

de Kroon, H., Groenendael, J., & Ehrlen, J. (2000). Elasticities: A review of methods and model limitations. *Ecology, 81*, 607–618.

Ennis, S., Rios-Vargas, M., & Albert, N. (2011). *The Hispanic population: 2010*, *2010 census briefs*. Washington, DC: US Census Bureau.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York: Springer.

Keyfitz, N. (1971). On the momentum of population growth. *Demography, 8*, 71–80.

Keyfitz, N., & Caswell, H. (2005). *Applied mathematical demography*. New York: Springer.

Kitagawa, E. (1955). Components of a difference between two rates. *Journal of the American Statistical Association, 50*(272), 1168–1194.

Krogstad, J., Lopez, M., & Rohal, M. (2015). *English proficiency on the rise among Latinos: U.S. born driving language changes*. Washington, DC: Pew Research Center.

Palmore, J., & Gardner, R. (1994). *Measuring mortality, fertility, and natural increase: A self-teaching guide to elementary measures* (5th ed.). Honolulu: East-West Center.

Passel, J., & Cohn, D. (2008). *U.S. population projections: 2005-2050*. Washington, DC: Pew Research Center.

Passel, J., Cohn, D., & Lopez, M. (2011). *Hispanics account for more than half of the nation's growth in past decade*. Washington, DC: Pew Research Center.

Preston, S., & Coale, A. (1982). Age structure, growth, attrition, and accession: A new synthesis. *Population Index, 48*, 217–259.

Rogers, A. (1995). *Multiregional demography: Principles, methods, and extensions*. New York: Wiley.

Schoen, R. (1986). The two sex multiethnic stable population model. *Theoretical Population Biology, 29*(3), 343–364.

Siegel, J. (2002). *Applied Demography: Applications to business, government, law, and public policy*. San Diego: Academic Press.

Smith, S., & Shahidullah, M. (1995). An evaluation of population projection errors for census tracts. *Journal of the American Statistical Association, 90*(429), 64–71.

Smith, S., Tayman, J., & Swanson, D. (2013). *A practitioner's guide to state and local population projections*. Dordrecht: Springer.

Swanson, D., Schlottmann, A., & Schmidt, B. (2010). Forecasting the population of census tracts by age and sex: An example of the Hamilton-Perry method in action. *Population Research and Policy Review, 29*(6), 797–818.

Tuljapurkar, S. (1982). Why use population entropy? It determines the rate of convergence. *Journal of Mathematical Biology, 13*, 325–337.

Wisdom, M., & Mills, S. (1997). Sensitivity analysis to guide population recovery: Prairie chickens as an example. *Journal of Wildlife Management, 61*(2), 302–312.

Wisdom, M., Mills, S., & Doak, D. (2000). *Ecology, 81*, 628–641.

Witmer, J., & Samuels, M. (1998). *Statistics for the life sciences* (2nd ed.). Upper Saddle River: Prentice Hall.

Chapter 14
Forecasting with Spatial Dependencies

14.1 Introduction

In terms of estimates and forecasts, applied demography can be viewed as essentially a geographic field. Data utilized in producing estimates and forecasts are nearly always reported in geographically-bounded summary units (Swanson and Tayman 2012: 22–28 and 74–79; Voss 2007) and it is now well-known that population data are spatially-dependent (Baker et al. 2014; Hogan and Tchernis 2004; Pace and Gilly 1997; Pattachini and Zenou 2007). As such, it has been suggested that demographic estimation and forecasting methods move in the direction of incorporating population dynamics across dimensions of both space and time (Baker et al. 2008, 2012; Chi and Voss 2011; Chi and Zhu 2008; Tayman 1996). The explosion of computerized mapping technology over the last 30 years has put this possibility within the grasp of many practitioners and has ushered in a new era of applied spatial demography.

In this chapter, we examine both the promise and pitfalls associated with this possibility and then present a coherent method for making spatially-explicit demographic forecasts using an extension of the Hamilton-Perry (H-P) method (Baker et al. 2014). This method provides a simple way to leverage the power of spatial dependencies in the forecasting process while avoiding the pitfalls associated with using geocoded data to make small area forecasts (Baker et al. 2012, 2013, 2014). Our approach in involves three steps: (1) constructing a spatial weighting matrix of geographic neighbors that directly captures spatial dependencies to the geographic area of interest; (2) creating weighted averages of cohort change ratios (CCRs); and (3) using the weighted CCRs to produce a forecast for the geographic area of interest using the H-P method (as described in Chapter 4 and elsewhere in this book).

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_14

14.2 Issues with Georeferenced Data

Using georeferenced data (such as address points) in a computerized mapping interface should allow demographic estimates and forecasts to be made for any subcounty geographic level from municipalities (Cai 2007), to census tracts (Swanson et al. 2010), to census block groups (Baker et al. 2015; Zandbergen and Ignizio 2010), and even down to assessor parcels (Jarosz 2008; Waddell 2012). This has, understandably, generated a considerable amount of excitement around the possibilities of extending standard demographic methods to small geographic units. However, known defects in the geocoding create missing data bias or incorrect geographic assignments when using geocoded data (Flotow and Burson 1996; Karimi et al. 2004; Zandbergen 2009).

The general direction of this bias should be negative and the amount of bias associated with incomplete geocoding may be considerable. Geocoding error might be as high as 9% over a ten year period for total population estimates for block groups in Albuquerque, New Mexico (Baker et al. 2012). This study indicated that geocoding errors were associated with demographic and socioeconomic factors related to ethnicity, income, education, and spatial residency dynamics, and that a statistically-significant clustering of "missing data bias" was directly attributable to defects in the geocoding process itself. Baker et al. (2013) suggested that even larger errors are associated with age/sex-specific estimates made using geocoded data. Other studies have found that geocoding-based errors are spatially clustered along lines of race/ethnicity and rural/urban residence (Gilboa 2006; Goldberg et al. 2007; Oliver et al. 2005; Zandbergen 2009).

While the challenges associated with using geocoded data are not to be minimized, it also is important to recognize that demographic data show spatially-explicit clustering patterns (Vasan et al. 2015). As such, a perspective suggested by Chi and Zhu (2008), Chi and Voss (2011) and Baker et al. (2014) is that patterns of spatial dependencies and relationships might be leveraged to both improve forecast accuracy and minimize bias (Fotheringham et al. 2002; Getis 2009; Hogan and Tchernis 2004; Pace and Gilly 1997; Pattachini and Zenou 2007). Chi and Zhu (2008), Chi and Voss (2011), and Chi and Wang (2017) have focused on regression-based methods that incorporate spatial relationships, while Baker et al. (2014) introduced spatial dependencies into standard demographic models; specifically, the H-P method based on CCRs.

14.3 Modeling Spatial Dependencies: Spatial Weights Matrices

Spatial dependency—the tendency of things close in space to have similar characteristics—may be modeled in a variety of ways depending on the type of data (Fotheringham et al. 2002; Tobler 1979, Turnbull 1976). Geographically-referenced data generally fall into three types: point, polygon, and raster (Getis 2009; Getis and Aldstadt 2004). Point data is straightforward—a good example would be address points that capture latitude/longitude (xy) coordinates of a housing unit; as they say, X marks the spot (along with Y in the case of a coordinate system). Geographic units such as census tracts, block groups, and blocks are typological examples of polygon data. Raster data refers to information captured in pixel densities, such as aerial imagery, terrain maps, digital elevation models, and the like. Our focus here is on modeling spatial dependency when individual data are aggregated into geographic units represented by polygons, which is by far a more common representation than either point or raster representations in terms of small area demographic forecasts.

In the case of polygon-based data, spatial dependency is usually captured by one of three processes: (1) defining "neighborhoods" based on contiguity (sharing of boundaries), which is found by measuring distances between centroids (centers) of the polygons (Fotheringham et al. 2000: 20–21); (2) by strategies of "overlaying" viewing windows to define an alternative number of neighboring geographic units (Kuldorff 1997, 1999; Turnbull 1976); and (3) by selecting a variable number of nearest neighbors visually or algorithmically (Steinberg and Steinberg 2015: 278–279). In this example, we provide a relatively simple approach based on contiguity. While we focus on visual selection for the small example presented here, computerized mapping software such as ESRI's Arc-GIS provides algorithms for ready-automation of these approaches.

14.3.1 Defining a Geographic Neighborhood

Perhaps two of the simplest forms of spatial weighting revolve around the use of rook and queen contiguity. In rook contiguity, polygons within a geographic "neighborhood" share only sides, while in queen contiguity, the polygons may share sides and/or corners (Getis 2009; Getis and Aldstadt 2004). A visual example provides an aid for understanding the two relationships. We use adjacent sets of census tracts within the city of Albuquerque to illustrate rook and queen contiguity. Figure 14.1 shows rook contiguity using census tract 1.14 as the area for defining adjacency. Rook contiguity defines the geographic neighborhood of census tract 1.14 to include census tracts 1.10 to the north, 1.21 to the west, 1.15 to the south, and 1.13 to the east. In constructing a spatial weights matrix to reflect this relationship, values from each of the four census tracts along with census tract 1.14 would be utilized in the calculation of a smoothed value for census tract 1.14.

Figure 14.2 shows a more liberal definition is applied when using queen contiguity. The queen criterion for neighborhood membership adds three additional census tracts to the neighborhood, 1.16 to the southeast, 1.11 to the northeast, and 1.20 to the northwest. While under rook contiguity only four neighbors were included, relaxing the membership criteria to include corners as well resulted in a total of eight census tracts within the neighborhood. In queen contiguity, the

Fig. 14.1 Rook contiguity, census tract 1.14, Albuquerque, New Mexico (Sources: Google Earth and Geospatial and Population Studies at the University of New Mexico (http://gps.unm.edu))

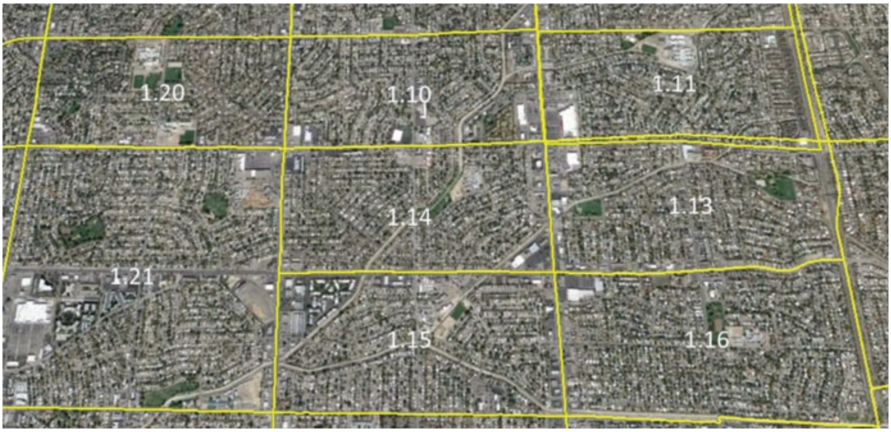

Fig. 14.2 Queen contiguity, census tract 1.14, Albuquerque, New Mexico (Sources: Google Earth and Geospatial and Population Studies at the University of New Mexico (http://gps.unm.edu))

contribution of the focal census tract to its own value would be reduced in comparison to the more restrictive rook contiguity. In constructing a "spatial weights matrix" to reflect this relationship, values from each of the seven census tracts neighbors along with census tract 1.14 would be utilized in the calculation of a smoothed value for census tract 1.14. Baker et al. (2014) have noted that this difference in neighborhood definition could impact forecast accuracy; a point discussed in the Conclusion to this chapter.

14.3.2 Constructing and Using a Spatial Weights Matrix

The strategy for capturing spatial dependency through weighting is contained within a spatial weights matrix that determines how the value of each geographic area (e.g., census tracts) will contribute to the forecast of the focal census tract (i.e. census tract 1.14). A spatial weights matrix is a rule for defining the contribution of each census tract in the designated neighborhood to the variable or variables of interest in the focal census tract. Such a matrix allows each census tract to "share in one another's fortune" in a specific way (Baker et al. 2014; Fotheringham et al. 2002; Le Sage and Pace 2004; Pace and Gilly 1997). The standard H-P forecast might be seen as inherently "ultra-local" because it tracks trends specific to the geographic unit being examined. However, the geographic patterning of population growth is regulated in some sense by housing unit density-dependent effects (Baker et al. 2008; Herold et al. 2003; Ward et al. 2000). A spatially-weighted forecast allows the preservation of spatial dependencies as well as incorporating hierarchical effects in which groups of census tracts interact in terms of their growth patterns.

Table 14.1 presents how: (1) a spatial weights matrix is conceptualized; (2) weights are determined and a corresponding matrix is constructed; and (3) the weighted average CCR is determined based on rook contiguity. A total of five census tracts (including the focal census tract of 1.14) contribute to a spatially weighted forecast. The geographic configuration is presented in the upper left hand portion of the table, while the corresponding weighting is captured in the upper right hand portion. Typically, the weighting is equal across all of the census tracts, dictating that each census tract will contribute 1/5th or 0.20 to the CCR of the focal tract. Finally, the bottom section shows how the weighted average CCR for census tract 1.14 is determined.

Table 14.2 presents the same set of representations for queen contiguity. The main difference, of course, is that with queen contiguity a greater number of census tracts is permitted to join the neighborhood, thus decreasing the "self-contribution" of census tract 1.14 to its own forecast. Under queen contiguity, each census tract would contribute 1/8th or 0.125 to the CCR of the focal tract.

In these examples (and in practice), non-weighted values are used in all calculations. That is, when a specific census tract contributes to its neighbor it does so using its non-weighted value rather than its weighted value based on calculations from nearby neighborhoods it also belongs to. This computational strategy ensures independence of calculations and minimizes potential distortions propagating from one neighborhood to another. The weighting also reflects an equal contribution of each census tract in the defined neighborhood; however, Baker et al. (2014) suggest that the weights could be defined through an optimization method where the weights are chosen to minimize some loss function similar to the estimation of portfolio weightings in quantitative finance (DeMiguel et al. 2009; Markowitz 1952).

Table 14.1 Spatial weights matrix and weighted CCR under rook contiguity

Census tract geographic configuration			Census tract weights		
	1.10			0.20	
1.21	**1.14**	1.13	0.20	**0.20**	0.20
	1.15			0.20	
Census tract	Count	Weight	Weighted CCR		
1.10	$CCR_{1.10}$	0.20	$CCR_{1.10} \times w_{1.10}$		
1.13	$CCR_{1.13}$	0.20	$CCR_{1.13} \times w_{1.13}$		
1.14	$\mathbf{CCR_{1.14}}$	**0.20**	$CCR_{1.14} \times w_{1.14}$		
1.15	$CCR_{1.15}$	0.20	$CCR_{1.15} \times w_{1.15}$		
1.21	$CCR_{1.21}$	0.20	$CCR_{1.21} \times w_{1.21}$		
			$\sum (CCR_i \times w_i)$ Weighted CCR		

Table 14.2 Spatial weights matrix and weighted CCR under queen contiguity

Census tract geographic configuration			Census tract weights		
1.20	1.10	1.11	0.125	0.125	0.125
1.21	**1.14**	1.13	0.125	**0.125**	0.125
	1.15	1.16		0.125	0.125
Census tract	Count	Weight	Weighted CCR		
1.10	$CCR_{1.10}$	0.125	$CCR_{1.10} \times w_{1.10}$		
1.11	$CCR_{1.11}$	0.125	$CCR_{1.11} \times w_{1.11}$		
1.13	$CCR_{1.13}$	0.125	$CCR_{1.13} \times w_{1.13}$		
1.14	$\mathbf{CCR_{1.14}}$	0.125	$CCR_{1.14} \times w_{1.14}$		
1.15	$CCR_{1.15}$	0.125	$CCR_{1.15} \times w_{1.15}$		
1.16	$CCR_{1.16}$	0.125	$CCR_{1.16} \times w_{1.16}$		
1.20	$CCR_{1.20}$	0.125	$CCR_{1.20} \times w_{1.20}$		
1.21	$CCR_{1.21}$	0.125	$CCR_{1.21} \times w_{1.21}$		
			$\sum (CCR_i \times w_i)$ Weighted CCR		

In this process, keep in mind we weight the CCRs themselves rather than the forecasted population values. The rationale for weighting the CCRs is provided by Baker et al. (2014), who observed significant biases when the weights were applied to the forecasts themselves where two “neighbor” census tracts had radically different population sizes; larger census tracts were under-forecasted and small census tracts were over-forecasted.

14.4 Spatially-Weighted Hamilton-Perry Forecast

Using census tract 1.14, we illustrate the construction of a rook contiguity spatially weighted forecast using the H-P method for the year 2010. This forecast uses CCRs for the 1990–2000 decade shown in Table 14.3. Along with census tract 1.14, the table contains CCRs for the four other census tracts defined by rook contiguity. The weighted CCR represents the arithmetic average of the five census tracts and will be used in the spatially weighted H-P forecast that borrows additional information from adjacent census tracts. The last column of the table show there are substantial differences in many age groups between the CCRs for census tract 1.14 and its weighted average. The average absolute percent difference across age groups is around 11.0% and they range from –22.1% for ages 30 to 44 to 0.4% for ages 25 to 29. Relative to the other census tracts, census tract 1.14 has decidedly lower CCRs

Table 14.3 Weighted total population cohort change ratios under rook contiguity, census tract 1.14, Albuquerque, New Mexico[a]

	Census tract						
Age	1.10	1.13	1.14	1.15	1.21	Weighted[b] CCR	Percent[c] difference
0–4	0.21	0.34	**0.28**	0.29	0.33	**0.29**	2.1%
5–9	0.25	0.32	**0.29**	0.26	0.27	**0.28**	−4.3%
10–14	1.24	0.69	**1.02**	0.84	1.03	**0.96**	−6.1%
15–19	1.30	0.81	**0.81**	0.76	1.12	**0.96**	15.8%
20–24	1.35	0.88	**0.67**	1.15	1.19	**1.05**	35.7%
25–29	1.29	1.10	**1.19**	1.26	1.14	**1.20**	0.4%
30–34	0.92	0.94	**1.26**	0.95	1.08	**1.03**	−22.1%
35–39	0.81	0.88	**0.96**	0.77	0.95	**0.87**	−10.1%
40–44	0.85	0.74	**0.68**	0.71	0.86	**0.77**	11.2%
45–49	0.90	0.80	**0.79**	0.74	1.04	**0.85**	7.6%
50–54	0.87	0.82	**0.75**	0.75	0.98	**0.83**	9.7%
55–59	0.77	0.81	**0.85**	0.83	0.97	**0.85**	−0.5%
60–64	0.88	0.86	**0.76**	0.83	0.88	**0.84**	10.1%
65–69	0.89	0.73	**0.76**	0.73	0.81	**0.79**	3.0%
70–74	0.68	0.86	**0.79**	0.68	0.90	**0.78**	−1.6%
75–79	0.69	0.71	**0.56**	0.72	0.82	**0.70**	19.5%
80–84	0.62	0.63	**0.48**	0.63	0.65	**0.60**	19.6%
85+	0.31	0.29	**0.24**	0.29	0.29	**0.28**	14.2%

Source: Geospatial and Population Studies, University of New Mexico. 1990 and 2000 data are normalized to 2010 geographic boundaries.

[a]CCRs based on the 1990–2000 decade.

${}_4P_{0,t}/{}_{15}P_{20,t}$ Ages 0–4 (Child-Adult Ratio)

${}_9P_{5,t}/{}_{15}P_{25,t}$ Ages 5–9 (Child-Adult Ratio)

$P_{x,t}/P_{x-10,t-10}$ Ages 10–84

$P_{85+,t}/P_{75+,t-10}$ Ages 85+

[b]Average of the CCRs for all census tracts, assuming equal weighting

[c](Weighted – standard)/weighted * 100

in three age groups 15 to 24, ages40 to 44, and ages 75 years and older and decidedly higher CCRs in age groups 30 to 39.

In preparing a 2010 forecast using the H-P method for census tract 1.14, we used both the standard CCRs and spatially weighted CCRs, and evaluated them against the 2010 census (see Table 14.4). Both forecasts show a decline in the total population, but it is much steeper using standard CCRs. The total population forecast error is more than double using the standard CCRs compared to spatially weighted CCRs, and the spatially weighted CCRs also perform better in forecasting the age composition. The spatially weighted forecast has less bias and greater accuracy compared to forecast using standard CCRs with MALPEs of -7.8% vs -12.5% and MAPEs of 17.1% and 23.2%, respectively. Ignoring signs, the spatially weighted forecast has lower errors in 13 of the 18 age groups.

It may be difficult to see from this simple example how one might make forecasts for a large set of census tracts (or other geographic) areas using spatial weighting. However, ESRI's Arc-GIS contains modules for constructing spatial weights matrices under specified neighborhood inclusion rules such as the queen and rook contiguity models presented here. These tools can create a spatial weights matrix that can be overlaid over a larger section of census tracts simultaneously. A key point is that the weighting scheme would yield weighted CCRs for census tracts prior to any spatial weight-based adjustment being made on any census tract. As noted earlier, calculations of neighborhoods are kept independent of each another, which is thought to minimize distortions introduced by the order of calculations.

14.5 Alternative Spatial Approaches

The method presented in this chapter provides a relatively simple technique for directly incorporating spatial effects into demographic forecasts using the H-P method. There are, however, other techniques that include spatial relationships into small area forecasts (Harper et al. 2003; Jarosz 2008; Smith et al. 2013: 203–207 and 228–237; Tayman 1996). Although these methods do not explicitly incorporate spatial dependence into their algorithms, they are based on factors known to be spatially-dependent such as housing, historical shares, employment, or other socioeconomic factors (Hammer et al. 2004; Hauer et al. 2015; Fotheringham et al. 2002; Hogan and Tchernis 2004; Pattachini and Zenou 2007). Since these methods are based on a spatially-dependent pattern of symptomatic indicators or historical ratios their forecasts are influenced, at least to some degree, by spatial dependencies and relationships (Pagliara et al. 2010; Voss 2007; Waddell 2012; White et al. 2015). One of the most well-studied of methods with an embedded spatial component is the ratio-correlation regression model (Swanson and Tayman 2015), which links symptomatic indicators at a smaller geographic scale to symptomatic indicators at a larger geographic scale within a regression framework.

Table 14.4 Standard and spatially weighted H-P model forecast error, census tract 1.14, Albuquerque, New Mexico, 2010

Age	2000 Population	Standard CCR[a]	Weighted CCR[a,b]	2010 Forecast		2010 Census	Percent error	
				Standard[c]	Weighted[c]		Standard	Weighted
0–4	209	0.28358	0.28980	209	235	171	22.2%	37.4%
5–9	247	0.28957	0.27751	225	197	195	15.4%	1.0%
10–14	313	1.02288	0.96425	214	202	184	16.3%	9.8%
15–19	230	0.80986	0.96126	200	237	213	−6.1%	11.3%
20–24	201	0.67224	1.04622	210	327	324	−35.2%	0.9%
25–29	259	1.19355	1.19805	275	276	288	−4.5%	−4.2%
30–34	277	1.25909	1.03142	253	207	220	15.0%	−5.9%
35–39	317	0.96061	0.87277	249	226	202	23.3%	11.9%
40–44	277	0.68227	0.76793	189	213	231	−18.2%	−7.8%
45–49	267	0.78761	0.85252	250	270	241	3.7%	12.0%
50–54	194	0.75194	0.83314	208	231	248	−16.1%	−6.9%
55–59	136	0.85000	0.84582	227	226	239	−5.0%	−5.4%
60–64	106	0.75714	0.84182	147	163	211	−30.3%	−22.7%
65–69	125	0.76220	0.78557	104	107	151	−31.1%	−29.1%
70–74	122	0.79221	0.78011	84	83	120	−30.0%	−30.8%
75–79	84	0.56376	0.69994	70	87	129	−45.7%	−32.6%
80–84	45	0.48387	0.60214	59	73	100	−41.0%	−27.0%
85+	26	0.24299	0.28317	38	44	91	−58.2%	−51.6%
Total	3,435			3,211	3,404	3,558	−9.8%	−4.3%
						MALPE	−12.5%	−7.8%
						MAPE	23.2%	17.1%

Source: Geospatial and Population Studies, University of New Mexico. 1990 and 2000 data are normalized to 2010 geographic boundaries.

[a]From Table 14.3

[b]Under rook contiguity

[c]$_{4}CCR_{0,t} \times {}_{15}P_{20,t+10}$ Ages 0–4

$_{9}CCR_{5,t} \times {}_{15}P_{25,t+10}$ Ages 5–9

$CCR_{x,t} \times P_{x,t}$ Ages 10–84

$CCR_{75+,t} \times P_{75+,t}$ Ages 85+

Additionally, there are several new and very promising approaches that could be applied in the context of a CCR-based model. Ordorica-Mellado and Garcia-Guerrero (2016) have proposed Kalman filters as a small-area demographic modeling tool. A Kalman filter is an allocation model based on a stochastic process estimated using a least-squares algorithm. As such, it introduces some degree of spatial dependency through its use of the relationship between smaller and larger geographic areas based on historical trends. This method could be applied to an H-P forecast by relating the CCRs of smaller areas to those of larger ones and it has the added advantage of simultaneously forecasting a small area's population while controlling it to a population for a larger geographic area. Kalman filters have been utilized in a variety of settings ranging from aeronautics to economics (Grewal and Andrews 1993), but to our knowledge the work of Ordorica-Mellado and Garcia-Guerrero (2016) is the first attempt to utilize the method in demographic analysis.

In a similar vein to the method presented here, Inoue (2017) has proposed a smoothing method for CCRs and child-woman ratios (CWRs) for use in the n method. In particular, mean and median smoothing routines that use CCRs and CWRs from neighborhoods or adjacent small areas. Unlike our method, which is based solely on adjacency, his approach is based on two principles: (1) the demographic of a target area are similar to areas closest to it; and (2) if different areas are the same distance from the target area the more populated area has greater influence. So the smoothing weight applied to the CCRs and CWWs in the target area gets larger when neighboring areas are closer and more populated. Implementation of these principals is implemented using a measure known as population potential (PP) (Stewart 1947). Calculating PP requires detailed data on the distance between the small areas and complex calculation. Inoue (2017) developed an approach to simplify the calculations and make the implementation more practical. Simulations were performed using a 2000 to 2005 base period and the H-P method to produce a 2010 forecast that was subsequently compared to the 2010 census for small areas in the Shibuya Ward, Tokyo, Japan. These simulation shows that his method based on PP outperformed a method where no smoothing was applied.

There is also the alternative of using spatial regression models in demographic forecasting (Chi and Voss 2011; Chi and Wang 2017; Chi and Zhu 2008). Spatial regression methods directly utilize spatial weighting algorithms (hence the umbrella term of "geographically weighted least squares") and provide another way to incorporate spatial dependencies and relationships into demographic forecasts. Such models incorporate both space and time effects in a "spatial lag" framework. Using this framework, CCRs could be forecasted using a spatial regression model as another way to incorporate spatial dependencies into the H-P model. Swanson and Tayman (2014) provide an example of using CCRs in a non-spatial regression framework that could form a basis for investigating this possibility.

14.6 Boundary Changes

Like all demographic forecasts, those that utilize spatial weighting will be subject to unavoidable sources of errors such as those in the decennial census (Fellegi 1968; Hogan 1993, 2003; Hogan and Mulry 2015) or reversals in historical growth trends that cannot easily be captured in methods using CCRs (Baker et al. 2013; Hoque 2010; Smith et al. 2013: 1–2). Spatially detailed demographic forecasts are also prone to specific types of errors not always relevant when making small-area demographic forecasts. The most important of these involves geographic normalization, which is the process of harmonizing population counts when geographic boundaries change over time (Lloyd 2017; Tobler 1979; Voss et al. 1999; Zandbergen and Ignizio 2010). Generally, when the change involves adding up smaller areas that have been subsumed into larger ones—such as a set of census tracts that have been annexed by a municipality—the error is generally small (Fisher and Langford 1995; Sadahiro 2000). When, however, a large geographic area is split into smaller areas significant distortions in historical data can occur (Baker et al. 2015; Simpson 2002; Zandbergen and Ignizio 2010). Unfortunately, this latter form of geographic splitting is very common and geographic normalization will frequently be required prior to making spatially detailed demographic forecasts.

When such normalization is required, adjustments have relied upon various methods of areal interpolation that assume the population is proportional to the area split into a new geography (Flowerdrew and Green 1992; Tobler 1979; Voss et al. 1999). This assumption has been shown to introduce large amounts of error, perhaps as much as 47%, for example, when geocoding to streets (Zandbergen and Ignizio 2010). Significant improvements to areal interpolation may be made by using ancillary data on housing, such as may be garnered from aerial photography, E911 structure points, or assessors' parcel data (Baker et al. 2015; Jarosz 2008; Sylvester 2013).

Where census data require adjustment for boundary changes, we can use the Census Bureau's "tract relationship files," which are described in Chapter 3. These relationship files provide areal weighting factors based on the Bureau's Master Address File (MAF). The MAF forms the basis for census data collection (NRC 2011) and while its quality is largely unknown (Dobson et al. 2011), it is likely at its highest quality at the time of the decennial census when a block canvassing effort has double-checked its completeness (GAO 2015; NRC 2011; Swanson and Walashek 2011). Even at the time of a decennial census, the MAF may have misallocation errors of between five and ten percent when aggregated to census tracts (Ratcliffe 2001). However, that study is 15 years old and it is possible that the MAF has improved since. Those seeking to create spatially detailed population forecasts should be aware of errors due to spatial normalization and build review tools and allow sufficient time to create geographically normalized datasets (Swanson et al. 2010).

14.7 Conclusions

The possibilities for incorporating spatial effects into demographic forecasting models are promising, but there is little research on its effectiveness. We know of only one study that has specifically compared the accuracy of forecasts with and without spatial dependencies (Baker et al. 2014). That study, which used the H-P model, found that introducing either form of contiguity (rook or queen) improved forecast accuracy dramatically in total population forecasts as well as in forecasts of age and sex composition in urban census tracts in Albuquerque, New Mexico. This same study also found that queen contiguity reduced error even further than rook contiguity-based weighting schemes, suggesting that incorporating larger neighborhoods may be preferable to smaller ones. Importantly, it also found that either form of spatial weighting cut errors in half in census tracts whose change was the greatest either in terms of gains or loss; growth categories for which H-P and other forecast methods tend to underperform (Smith and Shahidullah 1995; Swanson et al. 2010; Baker et al. 2013).

These improvements strongly suggest the efficacy of introducing spatial dynamics into small area forecasting models and those producing such forecasts should consider these methods. The relatively simple contiguity models based on CCRs, shown in this chapter provide a flexible, relatively inexpensive, and accessible accurate method for modeling spatial population dynamics. There is, however, a need for more research on forecast accuracy of models with and without spatial dependencies. How would such models perform under different demographic conditions in different parts of the county? At what geographic level, if any, does the incorporation of spatial dependencies no longer impact forecast error? Are spatial models better suited for subcounty forecasts or do they help lower forecast errors in larger geographies such as counties, metropolitan areas or states? These and other questions will help us learn more about the strengths and weakness of spatial forecasting models.

References

Baker, J., Ruan, X. M., Alcantara, A., Jones, T., McDaniel, M., Frey, M., & Watkins, K. (2008). Density-dependence in urban housing unit growth: An evaluation of the Pearl-Reed model for predicting housing unit stock at the census tract level. *Journal of Economic eand Social Measurement, 33*(2–3), 155–163.

Baker, J., Alcantara, A., Ruan, X. M., & Watkins, K. (2012). The impact of incomplete geocoding on small area population estimates. *Journal of Population Research, 29*, 91–112.

Baker, J., Alcantara, A., Ruan, X. M., Watkins, K., & Vasan, S. (2013). A comparative evaluation of error and bias in census tract-level age/sex-specific population estimates: Component I (Net-migration) vs Component III (Hamilton-Perry). *Population Research and Policy Review, 32*, 919–942.

Baker, J., Alcantara, A., Ruan, X. M., Watkins, K., & Vasan, S. (2014). Spatial weighting improves accuracy in small-area demographic forecasts of urban census tract populations. *Journal of Population Research, 31*(4), 345–359.

Baker, J., Alcantara, A., Ruan, X. M., Ruiz, D., & Crouse, N. (2015). Sub-county population estimates using administrative records: A municipal-level case study in New Mexico. In N. Hoque & L. Potter (Eds.), *Emerging techniques in applied demography* (pp. 63–80). New York: Springer.

Cai, Q. (2007). New techniques in small area population estimates by demographic characteristics. *Population Research and Policy Review, 26*(2), 203–218.

Chi, G., & Voss, P. (2011). Small-area population forecasts: Borrowing strength across space and time. *Population, Space, and Place, 17*, 505–520.

Chi, G., & Wang, D. (2017). Small area population forecasting: A geographically weighted regression approach. In D. Swanson (Ed.), *The frontiers of applied demography* (pp. 449–472). Switzerland: Springer.

Chi, G., & Zhu, J. (2008). Spatial regression models for demographic analysis. *Population Research and Policy Review, 27*, 17–42.

DeMiguel, V., Garlappi, L., Nogales, F., & Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science, 55*(5), 798–812.

Dobson, M., Cowen, D., & Guptill, S. (2011). *Reporting the state and anticipated future directions of addresses and addressing: A report to the geography division*. U.S. Census Bureau, Geographic Support System Initiative-GSS. Retrieved from http://www2.census.gov/geo/pdfs/gssi/research/GSS%20Initiative%20Addresses%20ACCEPTED.pdf

Fotheringham, A., Brundson, C., & Charlton, M. (2000). *Quantitative geography: Perspectives on spatial data analysis*. London: Sage.

Fotheringham, A., Brundson, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially-varying relationships*. West Sussex: Wiley.

Fellegi, I. P. (1968). *Coverage check of the 1961 Census of Population*, *Technical memorandum census evaluation series no. 2*. Ottawa: Dominion Bureau of Statistics.

Fisher, P. F., & Langford, M. (1995). Modeling the errors in areal interpolation between zonal systems using Monte Carlo simulation. *Environment and Planning A, 27*, 212–214.

Flowerdrew, R., & Green, M. (1992). Developments in areal interpolation methods and GIS. *The Annals of Regional Science, 26*, 67–78.

Flotow, M., & Burson, R. (1996). *Allocation errors of birth and death records to subcounty geography*. Paper presented at the annual meeting of the Population Association of America, New Orleans, LA, 11–15, May.

GAO. (2015). *Geospatial data: Progress needed on identifying expenditures, building and utilizing a data infrastructure, and reducing duplicative efforts*, *Report # GAO-15-193*. Washington, DC: Government Accountability Office.

Getis, A. (2009). Spatial weight matrices. *Geographical Analysis, 41*(4), 404–410.

Getis, A., & Aldstadt, J. (2004). Constructing spatial weights matrix using a local statistic. *Geographical Analysis, 36*(2), 90–104.

Gilboa, S. M. (2006). Comparison of residential geocoding methods in population-based study of air quality and birth defects. *Environmental Research, 101*, 256–262.

Goldberg, D., Wilson, J., & Knoblock, C. (2007). From text to geographic coordinates: The current state of geocoding. *URISA Journal, 19*(1), 33–46.

Grewal, M., & Andrews, A. (1993). *Kalman filtering: Theory and practice*. Englewood Cliffs: Prentice-Hall.

Hammer, R., Stewart, S., Winkler, R., Radeloff, V., & Voss, P. (2004). Characterizing dynamic spatial and temporal residential density patterns from 1940-1990 across the North Central United States. *Landscape and Urban Planning, 69*, 183–199.

Harper, G., Coleman, C., & Devine, J. (2003). *Evaluation of the 2000 subcounty population estimates*. Washington, DC: Population Division, U.S. Census Bureau.

Hauer, M., Evans, J., & Alexander, C. (2015). Sea level rise and sub-county population projections in coastal Georgia. *Population and the Environment, 37*, 44–62.

Herold, M., Goldstein, N., & Clark, K. (2003). The spatiotemporal form of urban growth: Measurement, analysis, and modeling. *Remote Sensing of Environment, 86*, 286–302.

Hogan, H. (1993). The 1990 post-enumeration survey: Operations and Results. *Journal of the American Statistical Association, 88*, 1047–1060.

Hogan, H. (2003). The accuracy and coverage evaluation: Theory and design. *Survey Methodology, 29*(2), 129–138.

Hogan, H., & Mulry, M. (2015). Assessing accuracy in postcensal estimates: Statistical properties of different measures. In N. Hoque & L. Potter (Eds.), *Emerging techniques in applied demography* (pp. 119–136). New York: Springer.

Hogan, H., & Tchernis, J. (2004). Bayesian factor analysis for spatially-correlated data, with application to summarizing area-level material deprivation from census data. *Journal of the American Statistical Association, 99*, 314–324.

Hoque, N. (2010). An evaluation of small-area population estimates produced by component method II, ratio correlation, and housing unit methods for 1990. *The Open Demography Journal, 3*, 18–30.

Inoue, T. (2017). A new method for estimating small area demographics and its application to long-term population projection. In D. Swanson (Ed.), *The frontiers of applied demography* (pp. 431–448). Switzerland: Springer.

Jarosz, B. (2008). Using assessor parcel data to maintain housing unit counts for small area population estimates. In S. Murdock & D. Swanson (Eds.), *Applied demography in the 21st century* (pp. 89–101). New York: Springer.

Karimi, H., Durcik, M., & Rasdorf, W. (2004). Evaluation of uncertainties associated with geocoding techniques. *Computer-Aided Civil and Infrastructure Engineering, 19*(3), 170–185.

Kuldorff, M. (1997). A spatial scan statistic. *Communication in Statistics: Theory and Methods, 26*, 1481–1496.

Kuldorff, M. (1999). An isotonic spatial scan statistic for geographical disease surveillance. *Journal of the National Institute of Public Health, 48*, 94–101.

Le Sage, J., & Pace, K. R. (2004). Models for spatially-dependent missing data. *Journal of Real Estate Finance and Economics, 29*(2), 233–254.

Lloyd, C. (2017). Creating population surfaces for the analysis of small area. In D. Swanson (Ed.), *The frontiers of applied demography* (pp. 431–448). Switzerland: Springer.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance, 7*(1), 77–91.

NRC. (2011). Change and the 2020 census: Not whether but how. In T. Cook, J. Norwood, & D. Cork (Eds.), *Panel to review the 2020 census*. Washington, DC: National Academies of Science Press.

Oliver, M., Matthews, K., Siadaty, K., Hauck, F., & Pickle, L. (2005). Geographic bias related to geocoding in epidemiologic studies. *International Journal of Health Geographics, 4*(29). doi:10.1186/1476-072X-4-29.

Ordorica-Mellado, M., & Garcia-Guerrero, V. (2016). Estimating the demographic dynamic of small areas with the Kalman filter. In R. Schoen (Ed.), *Dynamic demographic analysis* (pp. 261–271). New York: Springer.

Pace, K., & Gilly, O. R. (1997). Using the spatial configuration of data to improve estimation. *The Journal of Real Estate Finance and Economics, 14*(3), 330–340.

Pagliara, F., Preston, J., & Simmonds, D. (2010). *Residential location choice: Models and applications*. Berlin: Springer.

Pattachini, E., & Zenou, Y. (2007). Spatial dependence in local unemployment rates. *Journal of Economic Geography, 7*(2), 169–191.

Ratcliffe, J. H. (2001). On the accuracy of Tiger-type geocoded address data in relation to cadastral and census area units. *International Journal of Geographic Information Science, 15*, 473–485.

Sadahiro, Y. (2000). Accuracy of count data transferred through the areal weighting interpolation method. *International Journal of Geographical Information Science, 14*, 25–50.

Simpson, L. (2002). Geography conversion tables: A framework for conversion of data between geographical units. *International Journal of Population Geography, 8*, 69–82.
Smith, S., & Shahidullah, M. (1995). An evaluation of population projections errors for census tracts. *Journal of the American Statistical Association, 429*(90), 64–71.
Smith, S., Tayman, J., & Swanson, D. (2013). *A practitioner's guide to state and local population projections*. Dordrecht: Springer.
Steinberg, S., & Steinberg, S. (2015). *GIS research methods: Incorporating spatial perspectives*. Redlands: ESRI Press.
Stewart, J. (1947). Empirical and mathematical rules concerning the distribution and equilibrium of population. *Geographical Review, 37*(3), 461–485.
Swanson, D., & Tayman, J. (2012). *Subnational population estimates*. New York: Springer.
Swanson, D., & Tayman, J. (2014). *Measuring uncertainty in population forecasts: A new approach employing the Hamilton-Perry method*. Paper presented at the annual meeting of British Society for Population Studies, Winchester, UK, 8-10, September.
Swanson, D., & Tayman, J. (2015). On the ratio-correlation regression method of population estimation and its variants. In N. Hoque & L. Potter (Eds.), *Emerging techniques in applied demography* (pp. 93–118). Dordrecht: Springer.
Swanson, D., & Walashek, P. (2011). *CEMAF as a census method: A proposal for a redesigned census and independent Census Bureau*. Dordrecht: Springer.
Swanson, D., Schlottmann, A., & Schmidt, B. (2010). Forecasting the population of census tracts by age and sex: An example of the Hamilton-Perry method in action. *Population Research and Policy Review, 29*(1), 47–63.
Sylvester, J. (2013). *The use of cadastral and CAMA data to estimate unincorporated subcounty population*. Paper presented at the annual meeting of the Population Association of America, New Orleans, LA, 11–13, April.
Tayman, J. (1996). The accuracy of small-area population forecasts based on a spatial interaction land use modeling system. *Journal of the American Planning Association, 62*(1), 85–98.
Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association, 74*, 519–530.
Turnbull, B. (1976). The empirical distribution function with arbitrarily grouped, censored, and truncated data. *Journal of the Royal Statistical Society Series B (Methodological), 38*(3), 290–295.
Vasan, S., Alcantara, A., Nefertari, N., Ruan, X. M., & Baker, J. (2015). Geography is destiny: Spatial correlations in poverty and educational attainment in a New Mexico school district. In N. Hoque & L. Potter (Eds.), *Emerging techniques in applied demography* (pp. 225–246). New York: Springer.
Voss, P. (2007). Demography as spatial social science. *Population Research and Policy Review, 26*, 457–476.
Voss, P., Long, D., & Hammer, R. (1999). *When census geography doesn't work: Using ancillary information to improve the spatial interpolation of demographic data*, *Working paper no. 99-26*. Madison: Center for Demography and Ecology, University of Wisconsin.
Waddell, P. (2012). Parcel-level microsimulation of land use and transportation: The walking scale of urban sustainability. In Chandra Bhat and Ran Pendyala (Eds.) *Travel behavior research in and evolving world* (pp. 77–102). International Association for Travel Behavior Research.
Ward, D., Murray, A., & Phinn, S. (2000). A stochastically constrained cellular model of urban growth. *Computers, Environment, and Urban Systems, 24*(6), 539–558.
White, R., Engelen, G., & Uljee, I. (2015). *Modeling cities and regions as complex systems: From theory to planning applications*. Cambridge, MA: MIT Press.
Zandbergen, P. (2009). Geocoding quality and implications for spatial analysis. *The Geography Compass, 3*(2), 647–680.
Zandbergen, P., & Ignizio, D. (2010). Comparison of dasymetric mapping techniques for small area population estimates. *Cartography and Geographic Information Science, 37*(3), 199–214.

Chapter 15
The Utility of Cohort Change Ratios

15.1 Introduction

In this chapter we first describe the utility of cohort change ratios (CCRs) and relate it to the field of demography, particularly to applied demography. Next ,we provide a measure of utility relative to the cohort change ratio (CCR) method (also referred to in this book as the Hamilton-Perry method) and provide examples of its use. We conclude with a brief discussion that suggests how the concept of utility relative to the CCR method can be extended.

15.2 The Concept of Utility

In preparation for our discussion of utility, we start with the argument by Swanson et al. (1996) that the guiding principle in applied demography is to do only as much as necessary for the immediate problem at hand. We note that properly applied, this can lead to efficiency, but poorly applied, this principle and lead to mediocrity. Underlying this principle is the "Triple Constraint" (TC) perspective (Rosenau 1981; Swanson et al. 1996), which consists of three dimensions:

1. a performance specification—the explanatory/predictive precision sufficient to support a given decision-making situation (PS);
2. time—the scheduling requirements under which the performance specification must be accomplished (TS); and
3. cost—the budget requirements under which the performance specification must be accomplished (CS).

Using this perspective, for example, we can see that a high performance specification, such as a very high degree of accuracy for a total population number, generally requires a great deal of time and a high cost, such as those required for a

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_15

complete census. A lower performance specification requires much less time and lower cost, such as those required to generate a population estimate.

The triple constraint (TC) perspective is embedded within a distinctly different context when applied to basic (academic) demography. For basic demography, the context involves the goal of maximizing the performance dimension. That is, the goal is to maximize explanatory power and precision. Thus, on the one hand, basic demography tends to view time and resources as barriers to surmount in order to maximize explanatory power and precision; for applied demography, on the other hand, the context is to set the performance dimension at a level that is just sufficient to support a given decision-making process in order to minimize the use of time and resources. Given this, we argue that utility is an important concept in applied demography, but that it has less importance in basic demography.

One way in which utility can be examined is presented by Tayman and Swanson (1996) who measured the utility of population forecasts as a complement to measuring forecast accuracy. As a first step in developing this measure, they asked how much "value added" knowledge is gained by a forecast over and above the knowledge gained by either not doing a forecast or using a no-cost alternative. Next, they asked if useful and valid generalizations could be made about forecast utility in specific forecasting situations. In answering these two questions, they argued that a proportionate reduction in error measure (PRE), introduced by Costner (1965) and employed by Swanson and Tayman (1995), could be used to quantify the "value added" component of a given forecast.

Using the basic idea of PRE, we propose to measure the performance dimension, the time dimension, and the cost dimension is such a way that the three measures can be simply (algebraically) combined to obtain an overall score or a composite measure of utility. We start with the formula for PRE:

$$\mathrm{PRE} = ((\mathrm{A} - \mathrm{B})/\mathrm{A}) \qquad (15.1)$$

where,

A is an Alternative method; and
B is the CCR method.

PRE will be positive if the TC dimension of the Alternative method is greater than the TC dimension of the CCR method, and a larger value would indicate greater value added by the CCR method. Conversely, PRE will be negative if the TC dimension of the Alternative method is smaller than the TC dimension of the CCR method, and a smaller value (larger negative) would indicate more value added by the Alternative method. A PRE of zero would indicate that value of the TC dimension is the same in the Alternative and CCR methods

In the case of the performance dimension (error), "*A*" would represent the error level of an estimate or forecast generated by an Alternative method, while "*B*" would represent the error of an estimate or forecast generated by the CCR method. In terms of time, "*A*" would represent the time needed to generate the forecast or estimate using the Alternative method, while "*B*" would represent the time needed using the CCR method. Finally, in terms of the cost dimension, "*A*" would represent the cost needed to generate the forecast or estimate using the Alternative method, while "*B*" would represent the cost needed to use the CCR method. These dimensions are averaged obtain a composite measure of utility over all TC dimensions:

$$\mathrm{U} = \left(\mathrm{PRE_{ps}} + \mathrm{PRE_{ts}} + \mathrm{PRE_{cs}}\right)/3 \tag{15.2}$$

where,

U is the TC or composite utility of the CCR method;
PRE_{ps} is the performance (error) of the CCR method relative to an Alternative method;
PRE_{ts} is the time required to implement the CCR method relative to an Alternative method; and
PRE_{cs} is the cost of the CCR method relative to an Alternative method.

As a starting point each TC score is implicitly given a weight of 0.333. If we want to place more importance on, say, cost compared to error, we could re-arrange the weights. For example, we could give a weight of 0.363 to cost and remove 0.030 from the error weight.

As a hypothetical example (we will shortly provide real-life examples), suppose that we have $PRE_{ps} = 0.10$, $PRE_{ts} = 0.20$ and $PRE_{cs} = 0.15$. In this case, the U is 0.15 (0.45 / 3), indicating that the CCR method has more composite utility compared to the Alternative method. A negative U would indicate that the Alternative method has more composite utility than the CCR method. A zero U would indicate that the CCR and Alternative methods have the same composite utility. Finally, a positive U would indicate the CCR method provides more composite utility than the Alternative method.

How should one evaluate the U itself relative to the performance of the CCR method? For example, is a U of 0.60% "excellent", or is it "good"? Is a U of 0.20 "poor"? Swanson and Tayman (1995) suggest the following guidelines: less than zero, bad; 0–0.25, poor; 0.26–0.50, average; 0.51–0.75, good; 0.76–1.00, excellent. These guidelines are not cast in stone and are likely to be specific to the historical context, size, and growth rate of the geographic areas under consideration.

15.3 Utility and the Cohort Change Ratio Method

We present two separate sets of examples. In the first set, we examine four ex post tests of the TC dimensions of population forecasts using 2010 census counts as the accuracy benchmark. In these examples, the accuracy of the total population is the "performance specification." Following this set, we look at the TC dimensions of age group forecasts as the performance specification and present two assessments of the accuracy of age group forecasts using 2010 census counts.

In the first set, we compare a 2010 forecast of the total population for the state of Nevada using 1990–2000 CCRs to a forecast from a cohort-component model. In the second example of the first set, we compare the 2010 CCR forecast of the total population to a simple exponential extrapolation forecast using 1990 and 2000 census data. In the third example, we repeat the first comparison, using Inyo County, California. In the fourth comparison, we repeat the second comparison using Inyo County. By using the state of Nevada and Inyo County, we can get a feel for the utility assessments involving a relatively large population (Nevada) and a relatively small population (Inyo County, California).

In regard to time and cost, the cohort-component forecasts took 20 person-hours to assemble; the CCR method forecast took approximately 2 person-hours; and the exponential extrapolation took 0.5 hour. The costs of these forecasting methods were completely determined on the basis of personnel time, which was set at $100 per hour. Therefore, the costs were $2000, $200, and $50 for the cohort-component, CCR method, and exponential extrapolation respectively.

Table 15.1 provides the 1990 and 2000 input data, the 1990–2000 CCRs, the 2010 forecast, the 2010 census counts, and the error statistics for the CCR forecast of Nevada. Table 15.2 provides the 2010 forecast using the cohort-component method along with the 2010 census data, and the error statistics for this forecast. Details underlying the cohort-component method forecast for Nevada are provided in Table A.1 at the end of the chapter.

As can be seen in Table 15.1, overall, the 2010 CCR forecast of the state of Nevada is too high by 20.4%. Table 15.2 shows that the 2010 cohort-component forecast is also too high by 23.5%. The U for the CCR method relative to the Cohort-Component method for Nevada is:

$$0.644 = (0.132 + 0.900 + 0.900)/3$$

where,

PRE_{ps} is 0.132 (0.235–0.204) / 0.235;
PRE_{ts} is 0.900 (20–2) / 20; and
PRE_{cs} is 0.900 ($2000 – $200) / $2000.

This U reveals that relative to the cohort-component method, the CCR method has far more utility in this total population forecast for the State of Nevada. A U of 0.553 suggests a good performance of the CCR method in terms of its composite utility relative to the cohort-component model.

The 2010 exponential extrapolation forecast of the total population for Nevada yields a forecast error of 21.8% as follows:

P2010	$3{,}289{,}281 = 1{,}988{,}257 \times e^{((\ln(1{,}988{,}257\ /\ 1{,}201{,}833))\ \times\ 10)}$,
Census 2010	2,700,551, and
Pct. Error	$21.8\% = ((3{,}289{,}281–2{,}700{,}551)/2{,}700{,}551) \times 100$.

Table 15.1 Population forecast error CCR method, Nevada, 2010

Age	1990 Population	2000 Population	CCR[a]	2010 Population		Forecast error		
				Forecast[b]	Actual	Number	Percent	Allocation[c]
0–4	92,217	135,817	0.31107	225,995	187,478	38,517	20.5%	0.010%
5–9	85,562	149,322	0.31601	224,861	183,077	41,784	22.8%	0.138%
10–14	76,008	139,193	1.50941	205,004	183,173	21,831	11.9%	0.477%
15–19	74,160	127,169	1.48628	221,934	182,600	39,334	21.5%	0.065%
20–24	87,946	130,006	1.71043	238,080	177,509	60,571	34.1%	0.751%
25–29	110,583	148,726	2.00547	255,034	196,644	58,390	29.7%	0.563%
30–34	111,444	157,885	1.79525	233,393	190,642	42,751	22.4%	0.120%
35–39	101,268	165,910	1.50032	223,137	191,652	31,485	16.4%	0.233%
40–44	90,997	156,051	1.40026	221,080	191,391	29,689	15.5%	0.287%
45–49	74,228	140,214	1.38458	229,716	193,790	35,926	18.5%	0.110%
50–54	61,772	128,836	1.41583	220,942	182,737	38,205	20.9%	0.030%
55–59	54,681	105,057	1.41533	198,449	164,575	33,874	20.6%	0.010%
60–64	53,336	85,142	1.37833	177,579	150,924	26,655	17.7%	0.126%
65–69	50,112	71,387	1.30552	137,154	115,501	21,653	18.7%	0.058%
70–74	35,673	60,388	1.13222	96,399	82,280	14,119	17.2%	0.082%
75–79	22,564	44,851	0.89502	63,893	57,503	6,390	11.1%	0.164%
80–84	11,819	25,314	0.70961	42,852	38,888	3,964	10.2%	0.122%
85+	7,463	16,989	0.40599	35,384	30,187	5,197	17.2%	0.029%

(continued)

Table 15.1 (continued)

Age	1990 Population	2000 Population	CCR[a]	2010 Population		Forecast error		
				Forecast[b]	Actual	Number	Percent	Allocation[c]
Total	1,201,833	1,988,257		3,250,886	2,700,551	550,335	20.4%	
						MAPE	19.3%	
						MALPE	19.3%	
						IOD[d]		1.7%

Sources: 2000 and 2010, U.S. Census Bureau (http://factfinder2.census.gov)
1990, U.S. Census Bureau, Table 19, General Population Characteristics,1990 Decennial Census

[a] ${}_4P_{0,t}/{}_{15}P_{20,t}$ Ages 0–4
${}_9P_{5,t}/{}_{15}P_{25,t}$ Ages 5–9
$P_{x,t}/P_{x-10,t-10}$ Ages 10–84
$P_{85+,t}/P_{75+,t-10}$ Ages 85+

[b] ${}_4CCR_{0,t} \times {}_{15}P_{20,t+10}$ Ages 0–4
${}_9CCR_{5,t} \times {}_{15}P_{25,t+10}$ Ages 5–9
$CCR_{x,t} \times P_{x-10,t}$ Ages 10–84
$CCR_{85+,t} \times P_{75+,t}$ Ages 85+

[c] $abs((P_{x,t+10} / \sum P_{x,t+1-}) - (A_{x,t+10} / \sum A_{x,t+10}))$

[d] Index of Dissimilarity

Table 15.2 Population forecast error cohort-component method, Nevada, 2010

	2010 Population		Forecast error		
Age	Forecast	Actual	Number	Percent	Allocation[a]
0–4	202,591	187,478	15,113	8.1%	0.625%
5–9	166,171	183,077	−16,906	−9.2%	1.598%
10–14	181,216	183,173	−1,957	−1.1%	1.132%
15–19	179,509	182,600	−3,091	−1.7%	1.164%
20–24	186,596	177,509	9,087	5.1%	0.755%
25–29	257,809	196,644	61,165	31.1%	0.757%
30–34	333,477	190,642	142,835	74.9%	3.339%
35–39	346,438	191,652	154,786	80.8%	3.706%
40–44	291,062	191,391	99,671	52.1%	1.989%
45–49	247,137	193,790	53,347	27.5%	0.530%
50–54	212,086	182,737	29,349	16.1%	0.154%
55–59	172,045	164,575	7,470	4.5%	0.730%
60–64	143,801	150,924	−7,123	−4.7%	1.105%
65+	287,072	324,359	−37,287	−11.5%	3.060%
Total	3,207,010	2,700,551	506,459	18.8%	
			MAPE	23.5%	
			MALPE	19.4%	
				IOD[b]	10.3%

Source: 2010, U.S. Census Bureau (http://factfinder2.census.gov)
[a]abs($(P_{x,t+10}/\sum P_{x,t+10}) - (A_{x,t+10}/\sum A_{x,t+10})$)
[b]Index of Dissimilarity

Relative to the exponential extrapolation method, the U for the CCR method is:

$$-1.98 = (0.064 + -3.00 + -3.00)/3$$

where,

PRE_{ps} is 0.064 (0.218–0.204)/0.218;
PRE_{ts} is −3.00 (0.5–2.0)/0.5; and
PRE_{cs} is −3.00 ($50 – $200)/$50.

Clearly, if we were only interested in a total population forecast of Nevada, the exponential extrapolation method has far more utility than does the CCR method, which performed poorly.

In regard to our example of a small population, Tables 15.3 and 15.4 provide the same information for Inyo County, California that Tables 15.1 and 15.2 did for Nevada. Details underlying the cohort-component method forecast for Inyo County are provided in Appendix Table A.2 at the end of the chapter.

Table 15.3 shows that the 2010 CCR forecast of Inyo County is too low by 7.4%. Table 15.4 shows us that the 2010 cohort-component forecast is also too low by 6.8%. This yields a PRE_{ps} of −0.088 ((−6.8 – (−7.4)) / −6.8). Using this PRE_{ps} in conjunction with the time and cost PREs shown earlier, we find the U for the CCR method relative to the cohort-component method for Inyo County is:

Table 15.3 Population forecast error CCR method, Inyo County, California, 2010

Age	1990 Population	2000 Population	CCR[a]	2010 Population		Forecast Error		
				Forecast[b]	Actual	Number	Percent	Allocation[c] (%)
0–4	1,196	961	0.44367	1,053	1,070	−17	−1.6	0.362
5–9	1,350	1,184	0.43450	1,038	985	53	5.4	0.733
10–14	1,261	1,360	1.13712	1,093	1,134	−41	−3.6	0.249
15–19	977	1,236	0.91556	1,084	1,087	−3	−0.3	0.450
20–24	687	673	0.53370	726	865	−139	−16.1	0.437
25–29	1,071	644	0.65916	815	1,041	−226	−21.7	0.868
30–34	1,367	849	1.23581	832	979	−147	−15.0	0.435
35–39	1,531	1,232	1.15033	741	977	−236	−24.2	0.954
40–44	1,378	1,482	1.08413	920	992	−72	−7.3	0.008
45–49	1,109	1,597	1.04311	1,285	1,367	−82	−6.0	0.111
50–54	954	1,314	0.95356	1,413	1,594	−181	−11.4	0.368
55–59	929	1,101	0.99279	1,585	1,581	4	0.3	0.704
60–64	1,080	883	0.92558	1,216	1,339	−123	−9.2	0.140
65–69	1,071	892	0.96017	1,057	1,035	22	2.1	0.574
70–74	888	898	0.83148	734	783	−49	−6.3	0.052
75–79	665	731	0.68254	609	676	−67	−9.9	0.099
80–84	418	493	0.55518	499	520	−21	−4.0	0.102

85+	349	415	0.28980	475	521	−46	−8.8		0.044
Total	18,281	17,945		17,175	18,546	−1,371	−7.4		
						MAPE	8.5		
						MALPE	−7.6%		
						IOD[d]			3.3%

Sources: 2000 and 2010, U.S. Census Bureau (http://factfinder2.census.gov)
1990, U.S. Census Bureau, Table 19, General Population Characteristics,1990 Decennial Census

[a] ${}_4P_{0,t}/{}_{15}P_{20,t}$ Ages 0–4
${}_9P_{5,t}/{}_{15}P_{25}$ Ages 5–9
$P_{x,t}/P_{x-10,t-10}$ Ages 10–84
$P_{85+,t}/P_{75+,t-10}$ Ages 85+

[b] ${}_4CCR_{0,t} \times {}_{15}P_{20,t+10}$ Ages 0–4
${}_9CCR_{5,t} \times {}_{15}P_{25,t+10}$ Ages 5–9
$CCR_{x,t} \times P_{x-10,t}$ Ages 10–84
$CCR_{85+,t} \times P_{75+,t}$ Ages 85+

[c] $abs((P_{x,t+10}/\sum P_{x,t+10}) - (A_{x,t+10}/\sum A_{x,t+10}))$

[d] Index of Dissimilarity

Table 15.4 Population forecast error cohort-component method, Inyo County, California, 2010

	2010 Population		Forecast error		
Age	Forecast	Actual	Number	Percent	Allocation[a](%)
0–4	943	1,070	−127	−11.9	0.314
5–9	880	985	−105	−10.7	0.220
10–14	809	1,134	−325	−28.7	1.434
15–19	1,018	1,087	−69	−6.3	0.028
20–24	1,134	865	269	31.1	1.897
25–29	647	1,041	−394	−37.8	1.870
30–34	263	979	−716	−73.1	3.757
35–39	536	977	−441	−45.1	2.167
40–44	1,282	992	290	29.2	2.068
45–49	1,589	1,367	222	16.2	1.822
50–54	1,696	1,594	102	6.4	1.217
55–59	1,706	1,581	125	7.9	1.345
60–64	1,273	1,339	−66	−4.9	0.145
65+	3,509	3,535	−26	−0.7	1.240
Total	17,285	18,546	−1,261	−6.8	
			MAPE	22.1%	
			MALPE	−9.2%	
			IOD[b]		9.8%

Source: 2010, U.S. Census Bureau (http://factfinder2.census.gov)
[a]abs(($P_{x,t+10}/\sum P_{x,t+10}$) – ($A_{x,t+1-}/\sum A_{x,t+10}$))
[b]Index of Dissimilarity

$$0.571 = (-0.088 + 0.900 + 0.900)/3.$$

This U reveals that relative to the cohort-component method, the CCR method has far more utility than the cohort-component methods in this forecast of Inyo County, California. Like the Nevada example, the CCR has a good performance relative to the cohort-component model in Inyo County.

The 2010 exponential extrapolation forecast of the total population for Nevada yields a forecast error of −5.0% as follows:

P2010	$17{,}615 = 17{,}945 \times e^{((\ln(17{,}945/18{,}281)) \times 10)}$,
Census 2010	18,546, and
Pct. Error	$-5.0\% = ((17{,}615–18{,}546)/18{,}546) \times 100.$

With the CCR method having a percent error of −7.4%, the PRE_{ps} of the CCR method relative to the exponential extrapolation method is −0.480 ((−5.0 – (−7.4))/ −5.0). Using this PRE_{ps} in conjunction with the time and cost PREs shown earlier, we find the U for the CCR method relative to the exponential extrapolation method for Inyo County is:

$$-2.16 = ((-0.480) + (-3.00) + (-3.00))/3$$

If we were only interested in a total population forecast for Inyo County, the exponential extrapolation method has far more utility than the CCR method. Like for the Nevada example, however, the CCR method has a poor performance relative to the extrapolation method.

In the second set, we set the performance specification as the accuracy of age group forecasts, operationalized as the Mean Absolute Percent Error (MAPE). The first example compares the age group accuracy of the CCR and cohort-component methods for Nevada. In the second example, we do the same comparison for Inyo County, California.

In regard to the Nevada 2010 forecast, the MAPE for the CCR method is 19.3% while the MAPE for the same forecast resulting from the cohort-component method is 23.5%. This yields a PRE_{ps} of 0.178 ((23.5–19.3) / 23.5). Using this PRE_{ps} in conjunction with the time and cost PRE's shown earlier, we find that the U for the CCR method relative to the cohort-component method is:

$$0.659 = ((0.178) + (.900) + (.900))/3.$$

This U reveals that relative to the cohort-component method the CCR method has more utility in forecasting the 2010 population by age for the Nevada. Given the U for total population was 0.591, the CCR method has greater utility than the cohort-component method in forecasting the population by age than in forecasting total population.

Looking at the comparison for Inyo County, the MAPE for the CCR method is 8.5% while the MAPE for the cohort-component method is 22. 1%. This yields a PRE_{ps} of 0.615 ((22.1–8.5) / 22.1). Using this PRE_{ps} in conjunction with the time and cost PRE's shown earlier, we find that the U for the CCR method relative to the cohort-component method is:

$$0.805 = ((.615) + (.900) + (.900))/3.$$

This U reveals that relative to the cohort-component method the CCR method has far more utility in forecasting the 2010 forecast of the population by age for Inyo County, California. This represents the largest U of any example and indicates an excellent performance of the CCR method compared to the cohort-component method in term of composite utility.

15.4 Conclusions

In this chapter we have discussed the concept of utility and applied an operationalized definition of it to the CCR method. Using two sets of examples, we illustrated how the utility can be measured and interpreted. There are, of course, a number of variations that can be applied to the concept as identified, as well as multiple ways to operationalize it.

The main objective of this chapter was to show that the CCR method can have a great deal of utility relative to the information it provides. For example, if one is interested in population forecasts by age (and sex and race, etc.), the CCR method can have a much higher composite utility than the cohort-component method for short-term forecasts, such as ten years. Smith and Tayman (2003) reported a similar finding in their analysis of the 50 states and the 67 counties in Florida. In their analysis, the CCR and cohort-component methods produced similar forecasts by age and sex; neither approach consistently produced more accurate forecasts. Although they did not measure composite utility, the much lower time and costs of the CCR method would have increased its utility relative to the cohort-component method along these two TC dimensions.

With modifications (discussed in the previous chapter), the CCR method has the potential for higher utility than the cohort-component method for long-term forecasts as well. As discussed in Chapter 6, the CCR method can be used to generate formal measures of uncertainty around the forecasts it produces. This feature is not found in forecasts made using the cohort-component method, which requires much more work and subtlety in order to generate measures of uncertainty around the forecasts it produces (Alho and Spencer 2005; Hyndman and Booth 2008; Lutz et al. 1999; Ševčíková et al. 2013). As illustrated in this analysis, if one is only interested in the total number of people at some future date, the CCR method does not have as much utility as a simple extrapolation method, all things equal.

We have used "accuracy" as a performance specification in this chapter. Other performance specifications are possible, along with other methods of scoring this dimension of utility. For example, it would be complicated and impractical to use a cohort-component method to forecast the population of a set of census tracts by age, something that is very feasible using the CCR method. In this case, the scoring for the performance specification could be a numeric coding for "yes" and "no" in regard to the ease of which each method could be implemented for a set of census tract forecasts by age.

Table A.1 Population forecast by age and sex, Nevada, 2005 and 2010

	Female forecast, 2005								
	2000	2005		Net migration 2000–2005		Births 2000–2005			2005
Age	Females	Survival[a] rate	Survived population	Rate[b]	Number	5-Year[c] ASFR	Births	Survived[d] births	Females[e]
0–4	71,017	0.99278	70,504	0.11501	8,168				73,549
5–9	72,775	0.99842	72,660	0.12347	8,986				78,672
10–14	67,570	0.99919	67,515	0.07251	4,900				81,646
15–19	61,301	0.99839	61,202	0.25340	15,534	0.33157	20,326		72,415
20–24	61,993	0.99718	61,818	0.62500	38,746	0.65901	40,854		76,736
25–29	71,492	0.99682	71,265	0.58820	42,052	0.69498	49,686		100,564
30–34	74,832	0.99594	74,528	0.47723	35,712	0.39937	29,886		113,317
35–39	78,876	0.99384	78,390	0.26049	20,546	0.11163	8,805		110,240
40–44	75,379	0.99051	74,664	0.20049	15,113	0.01860	1,402		98,936
45–49	69,074	0.98541	68,066	0.16049	11,086		150,959	150,010	89,777
50–54	64,320	0.97763	62,881	0.10000	6,432				79,152
55–59	52,557	0.96577	50,758	0.08000	4,205	Male[f]	76,989	76,461	69,313
60–64	42,059	0.94788	39,867	0.05342	2,247	Female	73,970	73,549	54,963
65+	116,961	0.70953	82,987	0.02430	2,842				127,943
Total	980,206		937,105		216,569				1,227,223

(continued)

Table A.1 (continued)

	Male forecast, 2005							
	2000	2005		Net migration 2000–2005		2005		
Age	Males	Survival[a] rate	Survived population	Rate[b]	Number	Males	Both sexes	
0–4	74,800	0.99266	74,251	0.11501	8,603	76,461	150,010	
5–9	76,547	0.99843	76,427	0.12347	9,451	82,854	161,526	
10–14	71,623	0.99888	71,543	0.07251	5,193	85,878	167,524	
15–19	65,868	0.99696	65,668	0.25340	16,691	76,736	149,151	
20–24	68,013	0.99388	67,597	0.62500	42,508	82,359	159,095	
25–29	77,234	0.99284	76,681	0.58820	45,429	110,105	210,669	
30–34	83,053	0.99278	82,453	0.47723	39,635	122,110	235,427	
35–39	87,034	0.99094	86,245	0.26049	22,671	122,088	232,328	
40–44	80,672	0.98670	79,599	0.20049	16,174	108,916	207,852	
45–49	71,140	0.97982	69,704	0.16049	11,417	95,773	185,550	
50–54	64,516	0.96938	62,541	0.10000	6,452	81,121	160,273	
55–59	52,500	0.95368	50,068	0.08000	4,200	68,993	138,306	
60–64	43,083	0.93040	40,084	0.05342	2,301	54,268	109,231	
65+	101,968	0.70298	71,681	0.02430	2,478	116,544	244,487	
Total	1018,051		974,542		233,203	1,284,206		2,511,429

Female forecast, 2010

Age	2005 Females	2010 Survival[a] rate	Survived population	Net migration 2005–2010 Rate[b]	Net migration 2005–2010 Number	Births 2005–2010 5-Year[c] ASFR	Births	Survived[d] births	2010 Females[e]
0–4	73,549	0.99278	73,018	0.11501	8,459				99,329
5–9	78,672	0.99842	78,548	0.12347	9,714				81,477
10–14	81,646	0.99919	81,580	0.07251	5,920				88,262
15–19	72,415	0.99839	72,298	0.25340	18,350	0.33157	24,011		87,500
20–24	76,736	0.99718	76,520	0.62500	47,960	0.65901	50,570		90,648
25–29	100,564	0.99682	100,244	0.58820	59,152	0.69498	69,890		124,480
30–34	113,317	0.99594	112,857	0.47723	54,078	0.39937	45,255		159,396
35–39	110,240	0.99384	109,561	0.26049	28,716	0.11163	12,306		166,935
40–44	98,936	0.99051	97,997	0.20049	19,836	0.01860	1,840		138,277
45–49	89,777	0.98541	88,467	0.16049	14,408		203,872	202,591	117,833
50–54	79,152	0.97763	77,381	0.10000	7,915				102,875
55–59	69,313	0.96577	66,940	0.08000	5,545	Male[f]	103,975	103,262	85,296
60–64	54,963	0.94788	52,098	0.05342	2,936	Female	99,897	99,329	72,485
65+	127,943	0.70953	90,779	0.02430	3,109				148,922
Total	1,227,223		1,178,288		286,098				1,563,715

(continued)

Table A.1 (continued)

	Male forecast, 2010						
	2005	2010		Net migration 2005–2010		2010	
Age	Males	Survival rate[a]	Survived population	Rate[b]	Number	Males	Both sexes
0–4	76,461	0.99266	75,900	0.11501	8,794	103,262	202,591
5–9	82,854	0.99843	82,724	0.12347	10,230	84,694	166,171
10–14	85,878	0.99888	85,782	0.07251	6,227	92,954	181,216
15–19	76,736	0.99696	76,503	0.25340	19,445	92,009	179,509
20–24	82,359	0.99388	81,855	0.62500	51,474	95,948	186,596
25–29	110,105	0.99284	109,317	0.58820	64,764	133,329	257,809
30–34	122,110	0.99278	121,228	0.47723	58,275	174,081	333,477
35–39	122,088	0.99094	120,982	0.26049	31,803	179,503	346,438
40–44	108,916	0.98670	107,467	0.20049	21,837	152,785	291,062
45–49	95,773	0.97982	93,840	0.16049	15,371	129,304	247,137
50–54	81,121	0.96938	78,637	0.10000	8,112	109,211	212,086
55–59	68,993	0.95368	65,797	0.08000	5,519	86,749	172,045
60–64	54,268	0.93040	50,491	0.05342	2,899	71,316	143,801
65+	116,544	0.70298	81,928	0.02430	2,832	138,150	287,072
Total	1,284,206		1232,451		307,582	1,643,295	3,207,010

[a]Rong et al. (2012)

[b]1995–2000 net migration rates using the forward survival rate method

[c]2000 Nevada fertility rates multiplied by 5

[d]Infant and child survival rates of 0.99314 and 0.99431 for males and females, respectively

[e]Survived births (Ages 0–4)

2005 survived population + net migration (Ages 5–9 to 60–64)

(e.g., 2005 population aged 5–9 uses the survived population and net migration of the population aged 0–4 in 2000)

2005 survived population + net migration population aged 60–64 and 65+ (Ages 65+)

[f]Assumes that 0.51 of total births are male. Female births equal total births minus male births

Table A.2 Population forecast by age and sex, Inyo County, California, 2005 and 2010

	Female forecast, 2005								
	2000	2005		Net migration 2000–2005		Births 2000–2005			2005
Age	Females	Survival[a] rate	Survived population	Rate[b]	Number	5-Year[c] Rate	Births	Survived[d] births	Females[e]
0–4	468	0.99491	466	−0.07235	−34				468
5–9	589	0.99903	588	−0.08590	−51				432
10–14	658	0.99944	658	−0.05711	−38				537
15–19	606	0.99894	605	−0.11331	−69	0.26271	159		620
20–24	345	0.99809	344	−0.40525	−140	0.90343	312		536
25–29	319	0.99793	318	−0.33558	−107	0.59380	189		204
30–34	432	0.99756	431	0.26774	116	0.48664	210		211
35–39	660	0.99653	658	0.19955	132	0.11722	77		547
40–44	722	0.99450	718	0.08704	63	0.01658	12		790
45–49	802	0.99048	794	0.07323	59		959	954	781
50–54	674	0.98543	664	0.02666	18				853
55–59	541	0.97921	530	−0.01081	−6	Malef	489	486	682
60–64	424	0.96863	411	−0.01517	−6	Female	470	468	524
65+	1,944	0.76441	1,486	−0.01000	−19				1,872
Total	9,184		8,671		−82				9,057

(continued)

Table A.2 (continued)

	Male forecast, 2005						
	2000	2005		Net migration 2000–2005		2005	
Age	Males	Survival[a] rate	Survived population	Rate[b]	Number	Males	Both sexes
0–4	493	0.99374	490	−0.07235	−36	486	954
5–9	595	0.99887	594	−0.08590	−51	454	886
10–14	702	0.99924	701	−0.05711	−40	543	1,080
15–19	630	0.99778	629	−0.11331	−71	661	1,281
20–24	328	0.99409	326	−0.40525	−133	558	1,094
25–29	325	0.99386	323	−0.33558	−109	193	397
30–34	417	0.99423	415	0.26774	112	214	425
35–39	572	0.99287	568	0.19955	114	527	1,074
40–44	760	0.98986	752	0.08704	66	682	1,472
45–49	795	0.98415	782	0.07323	58	818	1,599
50–54	640	0.97566	624	0.02666	17	840	1,693
55–59	560	0.96543	541	−0.01081	−6	641	1,323
60–64	459	0.95071	436	−0.01517	−7	535	1,059
65+	1,485	0.73649	1,094	−0.01000	−15	1,508	3,380
Total	8,761		8,275		−101	8,660	17,717

Female forecast, 2010

	2005	2010		Net migration 2005–2010		Births 2005–2010			2010
Age	Females	Survival[a] rate	Survived population	Rate[b]	Number	5-Year[c] ASFR	Births	Survived[d] births	Females[e]
0–4	468	0.99491	466	−0.07235	−34				463
5–9	432	0.99903	432	−0.08590	−37				432
10–14	537	0.99944	537	−0.05711	−31				395
15–19	620	0.99894	619	−0.11331	−70	0.26271	163		506
20–24	536	0.99809	535	−0.40525	−217	0.90343	484		549
25–29	204	0.99793	204	−0.33558	−68	0.59380	121		318
30–34	211	0.99756	210	0.26774	56	0.48664	103		136
35–39	547	0.99653	545	0.19955	109	0.11722	64		266
40–44	790	0.99450	786	0.08704	69	0.01658	13		654
45–49	781	0.99048	774	0.07323	57		948	943	855
50–54	853	0.98543	841	0.02666	23				831
55–59	682	0.97921	668	−0.01081	−7	Male[b]	483	480	864
60–64	524	0.96863	508	−0.01517	−8	Female	465	463	661
65+	1,872	0.76441	1,431	−0.01000	−19				1,912
Total	9,057		8,556		−177				8,842

(continued)

Table A.2 (continued)

	Male forecast, 2010						
	2005	2010		Net migration 2005–2010		2010	
Age	Males	Survival[a] rate	Survived population	Rate[b]	Number	Males	Both sexes
0–4	486	0.99374	483	−0.07235	−35	480	943
5–9	454	0.99887	453	−0.08590	−39	448	880
10–14	543	0.99924	543	−0.05711	−31	414	809
15–19	661	0.99778	660	−0.11331	−75	512	1,018
20–24	558	0.99409	555	−0.40525	−226	585	1,134
25–29	193	0.99386	192	−0.33558	−65	329	647
30–34	214	0.99423	213	0.26774	57	127	263
35–39	527	0.99287	523	0.19955	105	270	536
40–44	682	0.98986	675	0.08704	59	628	1,282
45–49	818	0.98415	805	0.07323	60	734	1,589
50–54	840	0.97566	820	0.02666	22	865	1,696
55–59	641	0.96543	619	−0.01081	−7	842	1,706
60–64	535	0.95071	509	−0.01517	−8	612	1,273
65+	1,508	0.73649	1,111	−0.01000	−15	1,597	3,509
Total	8,660		8,161		−198	8,443	17,285

[a] Springborn (2006)
[b] 1995–2000 net migration rates using the forward survival rate method
[c] 2000 Inyo County fertility rates multiplied by 5
[d] Infant and child survival rates of 0.99431 and 0.99539 for males and females, respectively
[e] Survived births (Ages 0–4)
2005 survived population + net migration (Ages 5–9 to 60–64)
(e.g., 2005 population aged 5–9 uses the survived population and net migration of the population aged 0–4 in 2000)
2005 survived population + net migration population aged 60–64 and 65+ (Ages 65+)
[f] Assumes that 0.51 of total births are male. Female births equal total births minus male births

References

Alho, J., & Spencer, B. (2005). *Statistical demography and forecasting*. Dordrecht: Springer.

Costner, H. (1965). Criteria for measures of association. *American Sociological Review, 30*, 341–353.

Hyndman, R., & Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility, and migration. *International Journal of Forecasting, 24*, 323–342.

Lutz, W., Sanderson, W., & Scherbov, S. (1999). Expert-based probabilistic population projections. In W. Lutz, J. Vaupel, and D. Ahlburg (Eds.), *Frontiers of population forecasting* (pp. 139–155). New York, NY: The Population Council. (A supplement to *Population and Development Review*, 24).

Rong, W., Anderson, R., Curtin, L., & Arias, E. (2012). *U.S. decennial life tables for 1999–2001: State life tables. National Vital Statistics Reports 60(9)*. Hyattsville: National Center for Health Statistics.

Rosenau, M. (1981). *Successful project management*. Belmont: Wadsworth.

Ševčíková, H., Raftery, A., & Gerland, P. (2013). Bayesian probabilistic population projections: Do it yourself. Working Paper 13.2. Joint Eurostat/UNECE Work Session on Demographic Projections, Rome, Italy, 29–31 October.

Smith, S. K., & Tayman, J. (2003). An evaluation of population projections by age. *Demography, 40*(4), 741–757.

Springborn, R. (2006). *Abridged life tables for California, 2004. Data summary No. DS06–1102*. Sacramento: California Department of Health Services, Center for Health Statistics.

Swanson, D., & Tayman, J. (1995). Between a rock and a hard place: The evaluation of demographic forecasts. *Population Research and Policy Review, 14*, 233–249.

Swanson, D., Burch, T., & Tedrow, L. (1996). What is applied demography? *Population Research and Policy Review, 15*(5), 403–418.

Tayman, J., & Swanson, D. (1996). On the utility of population forecasts. *Demography, 33*(4), 523–528.

Chapter 16
Concluding Remarks

16.1 Introduction

We hope that you now have a good idea of why we wrote a book about cohort change ratios (CCRs) and trust that we have demonstrated that CCRs have a wide range of uses and a high level of utility, features that are useful to applied demographers. Taking a cue from David Letterman, the recently retired late night TV personality, whose "top ten" lists became a mainstay of his show, we would like to conclude this exposition of the CCR method with our "top ten reasons" for using it (Swanson and Tedrow 2016). As did Letterman, we work through the list in reverse order, starting with reason number 10 and ending (although without a drum roll) with reason number 1. Following the initial list, we elaborate a bit on each reason.

Exhibit 16.1 The Top Ten Reasons to Use the CCR Method

10. You only need two census counts of population by age to generate of population forecast.
9. You can run it forward or backward, as a forecast or a backcast.
8. You can use it at virtually any level of geography.
7. You can use it for any population for which cohort data are available over time, including institutionally or administratively defined populations.
6. You can use it to estimate life expectancy.
5. It provides formal demography enthusiasts with an efficient numerical means for generating stable population, incorporating both sexes and migration.
4. It is a great method for doing multi-race population forecasts.
3. With lagged regression models, it can provide forecast intervals.
2. It is a re-expression of the fundamental demographic equation.
1. The number one reason is that it is easy to explain and implement.

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0_16

16.2 Top Ten Reasons to Use the CCR Method

Reason # 10. You Only Need Two Census Counts of Population by Age to Generate a Population Forecast As shown in Chapters 1, 2, 4, 8, and 9 you only need data from two census counts of population by age to generate a population forecast and locating data generally is easier than is the case for the cohort-component method (see, Chapter 3, for example).

Reason # 9. You Can Run It Forward or Backward, as a Forecast or a Backcast You can not only run the CCR method forward in time as a forecast, but as shown in Chapter 10 also in reverse, as a backcast. Recall that this approach was used to estimate the size of the Native Hawaiian population in 1778, the year of first European contact. Evaluations of this approach supported the idea the CCR method is capable of producing reasonably accurate historical estimates. Also, Swanson (2016) also used reverse CCRs, in conjunction with other methods, to reconstruct both the Hawaiian and part-Hawaiian populations by age of Hawaii from 1778 to 1990 and then applied CCRs to estimate them to 2010 and forecast them to 2030.

Reason # 8. You Can Use It at Virtually Any Level of Geography The CCR method is well-suited for use in states and counties and as shown in Chapters 4 and 14, you can also do small area (i.e., subcounty) forecasts with this method. For example, in Chapter 4 forecasts were produced for the city of Bellingham in Washington and for the Pacific Beach and Mission Valley communities in San Diego, California, while in Chapter 14, they were produced for census tracts in Albuquerque, New Mexico.

Reason # 7. You Can Use It With Any Population for Which Cohort Data Are Available Over Time, Including Institutionally or Administratively Defined Populations You can also use the CCR method for populations such as school enrollment by grade. You can do this in two ways, directly and "embedded" within a CCR generated forecast by age, with the first method generally used for short-term forecasts and the second, for long-term forecasts. As an example of the first approach, Chapter 7 showed how the K-12 enrollment by grade of the Riverside (California) Unified School District is forecast for fall 2015 (using fall 2013 and fall 2014 enrollment data). In terms of the second approach, Chapter 7 showed how the embedded method generated a longer-term (10-year) public school enrollment forecast by grade for the Memphis, Tennessee School District.

Reason # 6. You Can Use It to Estimate Life Expectancy In Chapter 11, the CCR method is used to provide estimates of life expectancy at birth and other mortality-related indicators. When used with populations that have negligible migration the CCR approach can provide accurate estimates of these characteristics. As opposed to more data and analytically intensive methods (e.g., life tables), the CCR method has minimal data requirements in that it uses available census data and does not require a great deal of judgment or "data-fitting" techniques to implement.

Reason # 5. It Provides Formal Demography Enthusiasts with an Efficient Numerical Means of Generating Stable Populations, Incorporating Both Sexes and Migration In Chapter 12, the CCR method was used to generate stable populations for Greece and over 60 other countries. That chapter also demonstrated that the CCR results were consistent and in-line with results obtained using classical stable population mathematics.

Reason # 4. It Is a Great Method for Doing Multi-race Population Forecasts The 2000 census was the first to allow respondents to list themselves as belonging to one or more racial categories, as a result racial data in and after the 2000 census are inconsistent with racial data prior to 2000. In addition, racial classifications from the decennial census and America Community Survey are not completely consistent with the classification used for vital statistics data, making it difficult to develop reliable estimates of the components of change for racial and ethnic groups. Because it is based solely on data for two age distributions, the CCR method avoids these complications and provides a viable alternative to the full cohort-component models for forecasts of race, especially for forecasts of the multi-racial population (Swanson 2013). Given that the U.S. census only started counting multi-race people in 2000, it would be very difficult, for example, to construct a pre-2000 estimate of a given multi-race population in the absence of a CCR method backcast. Such an example is shown in Chapter 10 where we used the CCR method to estimate the 1990 multi-racial population in California's Riverside and San Bernardino counties (combined). Also, as mentioned under Reason # 9, Swanson (2016) used reverse CCRs and forward CCRs to reconstruct both the Hawaiian and part-Hawaiian populations by age of Hawaii from 1778 to 2010 and to provide a forecast of these populations by age to 2030.

Reason # 3. With Lagged Regression Models, It Can Provide Forecast Intervals In Chapter 6, the CCR method in conjunction with lagged regression models was used to generate formal measures of uncertainty (i.e., forecast intervals) for forecasts by age and for the total population for four states, Georgia, Minnesota, New Jersey, and Washington. The forecast intervals generated for every 10-year period from 1930 to 2010 were found to be both reasonable and informative.

Reason # 2. It Is a Re-Expression of the Fundamental Demographic Equation As noted by Land (1986), any quantitative approach to forecasting is constrained to satisfy various mathematical identities. Accordingly, a demographic approach should ideally satisfy demographic accounting identities, which are summarized in the identity known as the fundamental demographic equation: $P_t = P_0 + \text{Births} - \text{Deaths} + \text{In-migrants} - \text{Out-migrants}$. The Appendix shows that the CCR method does, in fact, satisfy the fundamental demographic equation, which provides the theoretical foundation that connects it to the Life Table (Chapter 11) and Stable Population Theory (Chapter 12). This foundation also facilitates the ability to decompose differences between CCRs into meaningful factors (Chapter 13) and provides a conceptual basis for the substantive interpretation of CCRs and their characteristics (Chapter 1).

Reason # 1. The Number One Reason Is That It Is Easy to Explain and Implement As discussed throughout this book, the CCR Method can be used in a variety of situations with minimal data requirements and, as such, comes with an inherently high level of utility (see Chapter 15). The minimum data needed is simply population data by age at two censuses. More detail is easily added to yield more detailed results as is the case with multi-race projections.

One of the authors of this book (Swanson) was engaged as an expert witness in a court case that involved population and enrollment projections for which he used the CCR method. It made an economist serving as the opposition's expert witness grumble that the method was so simple that one of his children could understand and operate this technique. The US Federal Judge hearing the case understood the method and how it operated as well. The side the economist was representing lost the case (Thomas 2012). We would like to think that Hamilton and Perry (1962) and Hardy and Wyatt (1911) would be pleased at the Judge's decision.

References

Hamilton, C. H., & Perry, J. (1962). A short method for projecting population by age from one decennial census to another. *Social Forces, 41*, 163–170.

Hardy, G. F., & F. B. Wyatt. (1911). Report of the actuaries in relation to the scheme of insurance against sickness, disablement, etc. *Journal of the Institute of Actuaries XLV*, 406–443.

Land, K. (1986). Methods for national population forecasts: A review. *Journal of the American Statistical Association, 81*, 888–901.

Swanson, D. (2013). *People of the Inland Empire: Changes in age, ethnicity, and race*. Paper presented at the Center of sustainable suburban development. Riverside: University of California Riverside.

Swanson, D. (2016). *A history of identity: The native and part-Hawaiian population of Hawai'i, 1780–2030*. Paper presented at the 2016 Conference of the American Sociological Association, Seattle, Washington, 20–23, August.

Swanson, D., & Tedrow, L. (2016). *The top ten reasons to use the cohort change ratio method*. Paper presented at the 2016 Conference of the Population Association of America, Washington, DC, 31 March – 2 April.

Thomas, W. (2012). *Municipal schools fight makes shared vision on public education virtually impossible*. Retrieved from http://www.commercialappeal.com/news/wendi-c-thomas-municipal-schools-fight-makes-shared-vision-on-public-education-virtually-impossib-ep-329136811.html

Appendix

Cohort Change Ratios and the Fundamental Demographic Equation

It is important that a demographic technique satisfy various mathematical identities and, in particular, the demographic accounting identity known as the fundamental demographic equation:

$$P_{t+k} = \mathrm{P_t} + \text{Births} - \text{Deaths} + \text{In-migrants} - \text{Out-migrants}. \tag{A.1}$$

This equation states that the population at a given point in time, P_{t+k}, must be equal to the population at an earlier time, P_t, plus the births and in-migrants and minus the deaths and out-migrants that occur between time = t and time = $t + k$.

The Cohort Change Ratio method moves a population by age from time t to time $t + k$ using cohort-change ratios (CCRs) computed from data in the two most recent censuses. It consists of two steps. The first step uses existing data to develop CCRs and the second step applies the CCRs to the cohorts of the launch year population to move them into the future. The formula for the first step, the development of a CCR is:

$${}_{\mathrm{n}}\mathrm{CCR}_{\mathrm{x,t}} = {}_{\mathrm{n}}\mathrm{P}_{\mathrm{x,t}} / {}_{\mathrm{n}}\mathrm{P}_{\mathrm{x-k,t-k}} \tag{A.2}$$

where,

${}_nP_{x,t}$ is the population aged x to $x + n$ at the most recent census (t),
${}_nP_{x\text{-}k,t\text{-}k}$ is the population aged x–k to x–$k + n$ at the 2nd most recent census (t–k), and
k is the number of years between the most recent census at time t and the one preceding it at time t–k.

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0

The basic formula for the second step, moving the cohorts of a population into the future is:

$$_{n}P_{x+k,t+k} = {}_{n}CCR_{x,t} \times {}_{n}P_{x,t} \tag{A.3}$$

where,

$_{n}P_{x+k,t+k}$ is the population aged $x + k$ to $x + k + n$ at time $t + k$, and
$_{n}CCR_{x,t}$ and $_{n}P_{x,t}$ are as defined in Eq. (A.2).
In terms of the CCR Method satisfying the fundamental demographic equation, let

$$_{n}CCR_{x,t} = ({}_{n}P_{x-k,t-k} + B - D + I - O)/({}_{n}P_{x-k,t-k}) \tag{A.4}$$

where,

$_{n}P_{x\text{-}k,t\text{-}k}$ is the population aged $x–k$ to $x–k + n$ at the 2nd most recent census (t-k),
B = Births between time $t–k$ and t
D = Deaths between time $t–k$ and t
I = In-migrants between time $t–k$ and t, and
O = Out-migrants between time $t–k$ and t. Since,

$$_{n}P_{x+k,t+k} = (({}_{n}P_{x-k,t-k} + B - D + I - O)/({}_{n}P_{x-k,t-k})) \times ({}_{n}P_{x,t}). \tag{A.5}$$

then,

$$_{n}CCR_{x,t} = ({}_{n}P_{x-k,t-k} - D + I - O)/({}_{n}P_{x+k,t+k}), \tag{A.6}$$

where, $x + k >= 10$.

Thus, we see that the CCR method expresses the individual components of change—births, deaths, and migration—in terms of Cohort Change Ratios. As such, it satisfies the fundamental demographic equation.

Index

J. Baker et al., *Cohort Change Ratios and their Applications*,
DOI 10.1007/978-3-319-53745-0