UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Wood's Method -- a Method for Fitting Leslie Matrices from Age-Sex Population Data, with some Practical Applications

Permalink

https://escholarship.org/uc/item/47t278gq

Author

Sprague, William Webb

Publication Date

2013

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Wood's Method -- a Method for Fitting Leslie Matrices from Age-Sex Population Data, with some Practical Applications

By

William Webb Sprague

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Demography

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Kenneth W. Wachter, Chair
Professor Ronald D. Lee
Professor John Wilmoth
Professor William Hanks, Outside Member

Fall 2013

Abstract

Wood's Method -- a Method for Fitting Leslie Matrices from Age-Sex Population Data, with some Practical Applications

by

William Webb Sprague

Doctor of Philosophy in Demography

University of California, Berkeley

Professor Kenneth Wachter, Chair

This dissertation is dedicated to an exploration of “Wood's Method” -- a novel approach to fitting demographic transition matrices to age and sex population count data. Demographic transition matrices, otherwise known as “Leslie matrices,” are extensively used to forecast population by age, sex, and other characteristics. Our implementation of Wood's Method simplifies the creation of age and sex population forecasts greatly by reducing the amount of data necessary to create a demographic transition matrix. Furthermore, the method can be used to infer a demographic component of change (one of migration, fertility, or mortality) if the other two components are specified.

In Chapter One, we introduce Wood's Method, as well as showing some illustrative examples. In Chapter Two, we evaluate the accuracy of Wood's Method by cross-validating age and sex specific forecasts for 3,120 US counties. In Chapter Three, we present a simpler, alternative derivation of Wood's Method with an extensive example and show some extensions to the method made possible by this new formulation. In Chapter Four, we use the method to examine migration rates at the US County level and show important results regarding clustering of migration. Each chapters is independent of the others, but should be read in order.

To our knowledge, this is the first time Wood's Method has been used for forecasting human populations. We hope to show its viability as a forecasting and analysis method and sketch directions for further research.

Dedication

To Chrisanne, of course.

Contents

1 Fitting Leslie matrices with Wood's Method **1**

1.1 Leslie matrices and cohort component forecasting 2
1.2 Fitting Leslie matrices with age/ sex data 3
1.2.1 Optimization background 4
1.2.2 Population input format 5
1.2.3 Output Leslie matrix 5
1.2.4 Wood's method 8
1.2.5 Constraints on demographic rates 12
1.3 Optimization and demographic analysis 15
1.4 Testing Wood's method on US county populations 18
1.4.1 Data 18
1.4.2 Error structure of ten year forecasts 19
1.4.3 County examples 20
1.5 Extensions to Wood's method for applied settings 28
1.5.1 Forecasting to a control total 28
1.5.2 Forecasting vital counts 28
1.5.3 Choosing input data 28
1.6 Conclusion 29

2 Error analysis of Wood's Method **30**

2.1 Approaches to forecast evaluation 30
2.1.1 Challenges in scalar forecast evaluation 30
2.1.2 Step ahead distance for cross-validation 32
2.1.3 Challenges in age-sex specific forecast evaluation 32
2.1.4 Problem of changing social dynamics 33
2.1.5 Error metrics used in evaluating WM 34
2.2 Cross validation results 36
2.2.1 Basic error measures for 10 year cross validation forecast 36
2.2.2 Error relationship to population characteristics. 42

2.2.3 Comparison to other forecasts . . . 46
2.2.4 Measuring errors in proportional age distributions . . . 48
2.3 Prediction bands . . . 61
2.3.1 Empirical confidence intervals . . . 61
2.3.2 Prediction bands analogous to bootstrap calculations . . . 61
2.3.3 County examples of confidence intervals . . . 63
2.4 Conclusion . . . 68

3 An alternative formulation of Wood's Method 70
3.1 Mathematical preliminaries . . . 70
3.2 Rederivation of Wood's Method, unconstrained . . . 73
3.2.1 Derivation . . . 73
3.2.2 Example . . . 74
3.3 Adding constraints to matrix fits . . . 77
3.3.1 Single coefficient constraints . . . 78
3.3.2 Sum-of-coefficient constraints . . . 81
3.3.3 Spine constraints and supplementary variables . . . 82
3.3.4 Other constraint systems . . . 84
3.3.5 Discussion – black- to gray- to white-box models . . . 84
3.4 Exact solutions to the Sylvester equation $C = AXB$ . . . 85
3.5 Extensions to the Wood's Method framework . . . 87
3.5.1 Replacing the identity matrix, part 1 . . . 87
3.5.2 Replacing the identity matrix, part 2 . . . 87
3.5.3 Transforming input . . . 88
3.5.4 Regularization and bi-criterion objective functions . . . 89
3.5.5 The interpretation of the objective value . . . 90
3.6 Running time and memory requirements . . . 91
3.7 Conclusion . . . 92

4 Geographic structure of migration 93
4.1 Migration index – method and data . . . 94
4.1.1 Migration index . . . 94
4.1.2 Index calculation . . . 96
4.2 US county migration index examples . . . 98
4.2.1 Graph format . . . 98
4.2.2 Individual county discussion . . . 99
4.3 Migration distribution and geographic clustering . . . 105

4.3.1 Migration index histogram . . . 105
4.3.2 Migration index maps . . . 106
4.3.3 Ratio of variance . . . 119
4.3.4 Migration index variance . . . 120
4.3.5 Semiovariogram . . . 124
4.4 Conclusion . . . 127

5 Summary and further research **129**
5.1 Advantages to WM . . . 129
5.2 Problems to WM . . . 130
5.3 Moving forward with WM . . . 130

Bibliography **132**

List of Figures

1.1 Spokane County forecast cross validation 23
1.2 King County forecast cross validation 24
1.3 Clark County forecast cross validation 25
1.4 Clallam County forecast cross validation 26
1.5 Yakima County forecast cross validation 27
1.6 Whitman County forecast cross validation 27

2.1 Percentile error by age, counties above 10,000 38
2.2 Percentile error by age, counties below 10,000 40
2.3 Error boxplots for all counties 42
2.4 Proportional error by county population 43
2.5 Proportional error by average county growth rate 44
2.6 Proportional error by county percent growth rate 45
2.7 Proportional error by county standard deviation of growth rate 45
2.8 County HPM histogram 49
2.9 Forecast age structure of minimum HPM county 52
2.10 Three-decade age structure of minimum HPM county 52
2.11 Forecast age structure of 25%-tile HPM county 53
2.12 Three-decade age structure of 25%-tile HPM county 54
2.13 Forecast age structure of 50%-tile HPM county 54
2.14 Three-decade age structure of 50%-tile HPM county 55
2.15 Forecast age structure of 75%-tile HPM county 56
2.16 Three-decade age structure of 75%-tile HPM county 56
2.17 Forecast age structure of 100%-tile HPM county 57
2.18 Three-decade age structure of 100%-tile HPM county 58
2.19 HPM by county population 58
2.20 HPM by average county growth rate 59
2.21 HPM by county percent growth 59
2.22 HPM by standard deviation of growth rate 60
2.23 Spokane County forecast with prediction bands 65

2.24 King County forecast with prediction bands 65
2.25 Clark County forecast with prediction bands 66
2.26 Clallam County forecast with prediction bands 67
2.27 Yakima County forecast with prediction bands 67
2.28 Whitman County forecast with prediction bands 68

4.1 USA Census geographical hierarchy 95
4.2 Multnomah County (OR) age-specific migration index 100
4.3 King County age-specific migration index 101
4.4 Whitman County age-specific migration index 102
4.5 Clark County (OR) age-specific migration index 102
4.6 Clallam County age-specific migration index 103
4.7 Yakima County age-specific migration index 104
4.8 Migration index histogram 105
4.9 USA total migration county map 108
4.10 USA 60 plus migration county map 111
4.11 USA 55 and under migration county map 112
4.12 USA county migration average state map 113
4.13 USA county migration standard deviation state map 115
4.14 USA county migration average division map 117
4.15 USA county migration standard deviation division map 118
4.16 State migration average and standard deviation 120
4.17 Region migration index boxplot 124
4.18 Division migration index boxplot 125
4.19 State migration index boxplot 126
4.20 Migration index semivariogram 128

List of Tables

1.1 Fertility constraints 13
1.2 Survival constraints 14
1.3 Diagonal migration constraints 15
1.4 MAPE by age 20
1.5 Quantile error structure 21
1.6 Quantile error structure 22

2.1 Percentiles of Percentage Absolute Error 37
2.2 MAPE by percentile and sex, counties > 10,000 in 2010 39
2.3 MAPE by percentile and sex, counties < 10,000 in 2010 41
2.4 MAPE by population 46
2.5 MAPE by percentage growth 47
2.6 State and WM forecast error comparison 47
2.7 County HPM distribution 49

3.1 Error for unconstrained fit 77
3.2 Error for sparsity constrained fit 80
3.3 Error for all constraints 83

4.1 Standardized population for migration index 98
4.2 Migration summary, Southern region 121
4.3 Migration summary, Midwest region 121
4.4 Migration summary, Northeast region 122
4.5 Migration summary, West region 122

Acknowledgments

First of all, I would like to thank Prof. Ken Wachter, who guided the completion of this dissertation with just the right amount of patience, impatience, and finesse.

I would also like to thank Prof. Jenna Johnson-Hanks. Her comments on a working draft of Chapter One were crucial to the success of the subsequent dissertation. More importantly, her counsel and kindness over the last decade has made my time at the Demography Department possible and fulfilling.

Prof. Ron Lee read a draft of Chapter One as well and gave very important comments that made the project possible. He has been an important teacher and mentor for me as well. I have enjoyed and grown by my working relationship with Prof. Lee on LCFIT and the opportunity to work on large-scale programming of demographic analyses; without this experience, the programming supporting Wood's Method would have been impossible.

I would like to thank Prof. John Wilmoth for sitting on my dissertation committee and being an important teacher to me. I still fondly remember reading classic demographic papers with him and fellow graduate students during my first year at Demography. I would also like to thank the students in his Spring 2012 graduate seminar in demography for their comments on a working draft of Chapter One.

I would like to thank Prof. Bill Hanks, who agreed to read a dissertation quite distant from his interests with his consummate good nature. I also forgive him for cultivating in me a taste for theoretical linguistics and semiotics.

Professor Loo Botsford at UC Davis first introduced me to Leslie matrices and mentioned the potential of optimization and parameter fitting for population biology and demography when I worked in his lab in 2001-2003. I very much appreciate the time I spent working with him.

My colleagues at the State of Washington have been very supportive of this work and deserve special thanks. These include my supervisor Yi Zhao, the assistant director of OFM's Forecasting Division Dr. Marc Baldwin, Dr. Don Pittenger, Mike Mohrman, Dr. George Hough, Dr. Tom Kimpel, Dr. Erica Gardner, and Diana Brunink.

Prof. Dave Swanson has read various chapters at various times, and I thank him for his comments and encouragement.

Dr. Jack Baker of the University of New Mexico and Eddie Hunsinger of the State of Alaska commented on early drafts and have been excellent correspondents on all topics in applied demography. I thank them both for their time and their detailed emails.

Bill Schooling and Dr. Walter Schwarm at the California Demographic Research Unit provided cross-validation data for Chapter Two. Risa Proehl at the Oregon Population Research Center provided both comments and cross-validation data as well. I thank them for going out of their way to support this research.

Prof. Leonard Cohen at UC Berkeley made my survival in the Anthropology Department not just possible but often pleasant. Training in anthropology has sensitized me to the highly local and subtle social dynamics that can have a huge impact on demographic behavior at the applied scale. I thank him for his time and encouragement.

To the countless named and anonymous authors, testers, and users of open source software, I owe a great thanks and much of my working life as a programmer and data analyst. The software used to implement this dissertation was almost completely Open Source, including Octave, FreeBSD, Lyx, Latex, Git, PostgreSQL, PostGIS, and LibreOffice.

To my fellow students, co-workers, and friends in the Demography Department whose support and comraderie I will alway remember I owe a great thanks. I would especially like to mention Dr. Romesh Silva and his wife Professor Sarah Tom, whose hospitality made many visits to Berkeley possible.

A special thanks is due to Tadashi Tsuchida and Katrina Keating, good friends who allowed me to stay with them when I attended class and who have provided excellent fellowship as well.

I would like to send good wishes and thanks to my sister Caroline, her family Cole, Sasha and Doug, my mother Minka, my father Bill, my step-mother Michelle, and my parents-in-law Liz and Dave and Bill and Carol; Joel, Tony, John and the other Olympia sailors.

I would be remiss if I didn't mention the feline friends who kept me company and gave me joy during this period: Mustafa, Frank, and Nigel.

I would like to reiterate my thanks to my wife Chrisanne, who has been supportive of this and many endeavours for many years; I look forward to many more adventures with her.

1 Fitting Leslie matrices with Wood's Method

In this chapter, we introduce a new technique for generating demographic transition matrices ("Leslie matrices") from simple age and sex population counts, using an implementation of "Wood's Method" (Wood, 1997); these matrices can be used to forecast population by age and sex (the "cohort component" method) using simple matrix multiplication of a starting population vector. In other words, Wood's Method finds a best-fit demographic transition matrix from a set of age and sex population counts without requiring separately derived demographic rates for migration, fertility, and mortality. Our approach improves on previous methods for creating Leslie matrices in two respects: one, it eliminates the need to calculate input demographic rates from "raw" data, and, two, our new format for the Leslie matrix more elegantly reveals the population's demographic components of change (fertility, mortality, and migration). This chapter is organized around three main themes. First, we describe the underlying algorithm, "Wood's Method," which uses quadratic optimization to fit a transition matrix to age and sex population counts. Second, we use demographic theory to create constraint sets that make the algorithm useable for human populations. Finally, we use the method to forecast 3,120 US counties and show that it holds promise for automating cohort-component forecasts. This chapter describes the first published successful application of Wood's method to human populations; it also points to more general promise of constrained optimization techniques in demographic modeling. This chapter also provides an introduction to the rest of the dissertation.

1.1 Leslie matrices and cohort component forecasting

Computing age and sex specific forecasts (ASSFs) is a standard task of applied demographers, as these forecasts form a fundamental input for planning throughout industry and government. Examples include regular forecasts by the the U.S. Census Bureau, K-12 school forecasting, and caseload projection (hospitals, prisons, medicaid, etc). As one example, Washington State produces ASSFs as part of their statewide land use planning program; these forecasts are used to plan for innumerable downstream projects, from hospital construction to future transportation projects.

The cohort component method invented by (Whelpton, 1928) is a well established procedure for creating ASSFs. The Leslie matrix formulation of the cohort component method is particularly elegant for many reasons: it allows a forecasting problem to be formulated as simple matrix multiplication, the matrix provides a succinct characterization of the population dynamics, and the linear algebraic structure of the matrix yields interesting results. However, creating Leslie matrices requires demographic rates for the population under question, which must either be drawn from model demographic rates or created directly from vital event counts and age/sex-specific base population estimates. Even when these prerequisites are met (which can be surprisingly difficult in an applied demographic institutional environment), the Leslie matrix generated usually requires extensive tuning to yield plausible forecasts, because the results are highly sensitive to small changes in the matrix, whether due to error or mis-estimation. Using Wood's method, however, the Leslie matrix is guaranteed to generate at least plausible forecasts insofar as it is forced to interpolate between the input population vectors; in our experiments, it extrapolates populations well too. In this paper we describe a procedure such that an applied demographer can simply input historic age/sex population counts – no need for separate rate data – and automatically get a plausible Leslie transition matrix that can be used for cohort-component projection.

Most applied demographers depend either on spreadsheets or SAS "arrays" in the SAS "data step" to perform cohort component analysis and do not use matrix methods at all. We hope that this paper will make matix based demographic forecasting methods more attractive by showing their power and elegance.

To our knowledge, the underlying algorithm in our technique – Wood's method (WM) – has never been successfully applied to human demographic problems, even though it shows much promise for applied work. The lack of uptake may be due to several reasons, not least of which is that formulating the quadratic optimization problem in WM requires intermediate skills in computer programming and access to a good

matrix programming language.[1] More importantly, though, without well chosen constraints for the optimization model, the method gives completely implausible results, which might have led researchers to abandon the technique prematurely. In the only work on WM besides the original paper, Caswell's influential book (Caswell, 2001), Caswell recommends rather simplistically chosen constraints that force the cell and row sums in the output Leslie matrix to fall between zero and one; these constraints may be reasonable for a closed single sex population, but when used to derive a Leslie matrix for human populations with open migration flows, they yield a matrix which has no demographic interpretation and which gives very poor forecasts. However, if constraints are developed using basic human demographic theory, WM provides interpretable projection matrices which yield very reasonable forecasts.

We will not do extensive comparisons to other forecasts in this chapter, but we do compare a WM generated forecast to some official forecasts in Chapter Two.

1.2 Fitting Leslie matrices with age/ sex data

There are a number of moving parts to the method and we will take each in turn: to begin, we will give a short introduction to the idea of optimization under constraint, then we will describe the input format to Wood's method, followed by the transition matrix format, then the Woods inference algorithm, and finally a description of the constraints. It will be helpful to have a basic overview of the method before diving into the details: First, at least three sets of age/ sex population counts are assembled, with five year age widths and time periods at five year intervals, and with 85 years old as the open interval.[2] Second, a set of constraints is developed that sets bounds on the survival, migration, and fertility cells in the output matrix, and enforces any relationships between age-specific rates. Third, the population count data and this constraint set are

[1] We would not consider SAS or a spreadsheet program to fulfill this requirement, and these are the only environments one can count on seeing in applied demography shops. R is an improvement, but R's syntax is awkward for implementing complex matrix algorithms. The author prefers Octave, a Matlab compatible language; if Octave is not available, Matlab provides a reasonable second-best alternative.

[2] This open interval chosen because age specific demographic data in the United States is most commonly found with the open interval at 85 years old. This method is applicable to any regular age division; in the future we will use it with population count data using 100+ as the open interval.

input to a computer program which rearranges the population count data into a format suitable for quadratic optimization. Fourth, a computer program performs constrained quadratic optimization on a decision vector in which each optimized element is a non-zero cell in the desired Leslie matrix. Finally, the resulting vector is rearranged into a Leslie matrix and returned. Each of these steps will be described in detail below, but first we will try to give some background on optimization methods.

It is important to remember Wood's Method does not estimate age specific rates directly, but simply regresses each population age distribution on the distribution one step earlier. Most forecasting approaches first derive age specific rates and then use them to forecast population using demographic accounting identities.

1.2.1 Optimization background

As our presentation of WM highlights optimization techniques, it is worth sketching those for the reader; please note, however, that it would be impossible to do justice to this very well-developed topic – for an excellent introduction see (Boyd and Vandenberghe, 2005), but for a well written and classic work see (Dantzig, 1998). Optimization models try to find the best values for a set of variables under a set of constraints as measured by a summarizing "objective function" – for example, finding the best diet as measured by total nutrition and constrained by a budget, or the best allocation of retail products in a store as measured by total profit and constrained by shelf space, or the best fit of a parametrized function as measured by the minimum sum of squares and constrained by bounds on the parameters (the optimization problem used in WM). When the objective function is "convex" – meaning roughly that no line that connects two points on the objective surface crosses this surface – every local optimized value is guaranteed to also be globally optimized value over the (usually constrained) feasible solution space; linear functions are convex, and many second order polynomials are convex as well, including the function corresponding to sum of squares.

Most applied optimization problems are formulated in terms of strictly linear equations, including the objective function, which is usually a large summation involving the parameters and variables; this formulation gives us the "linear programming" that underpins modern operations research. In "quadratic programming," the objective function is a second order polynomial; quadratic programming is used in many variance minimization problems including Markowitz portfolio optimization. Optimization problems are ubiquitous in many fields, especially economics, operations research, automatic controls, and machine learning – even modern cognitive linguistics – and the algorithms and problem formulations are very well developed. See below for a

longer discussion of the potential power of constrained optimization approaches in demographic modeling beyond forecasting.

1.2.2 Population input format

The input is a set of age sex population counts stored in a 36 by T matrix, with the rows representing each of the 36 age and sex combinations and the columns representing each five year interval. Male and female population counts for the same year are concatenated vertically in a single vector, and these vectors are concatenated horizontally into a matrix; for an example see (1.1). At least three years of data are required, but as many years of data as are available can be used by the method. In (1.1), column 1 holds the population for 1990, column 2 for 1995, column 3 for 2000.

$$\begin{bmatrix} {}_5m_0^{1990} & {}_5m_0^{1995} & {}_5m_0^{2000} \\ {}_5m_5^{1990} & {}_5m_5^{1995} & {}_5m_5^{2000} \\ \vdots & \vdots & \vdots \\ {}_\infty m_{85}^{1990} & {}_\infty m_{85}^{1995} & {}_\infty m_{85}^{2000} \\ {}_5f_0^{1990} & {}_5f_0^{1995} & {}_5f_0^{2000} \\ {}_5f_5^{1990} & {}_5f_5^{1995} & {}_5f_5^{2000} \\ \vdots & \vdots & \vdots \\ {}_\infty f_{85}^{1990} & {}_\infty f_{85}^{1995} & {}_\infty f_{85}^{2000} \end{bmatrix} \tag{1.1}$$

1.2.3 Output Leslie matrix

Single sex Leslie matrix review

The Leslie matrix method should be familiar, at least in outline, to analysts with formal demographic training; if not, excellent introductory works include (Wachter, 2012). However, it is worth reviewing the essentials of the basic single sex Leslie matrix for a closed population. Leslie matrices project a population vector like (1.2) forward by a single time step of the same length as the age classification (e.g., if the population counts are reckoned in five year intervals, then each projection step is five years long as well). The matrix multiplies the population vector for time step 0 to yield a population vector at time step 1: $\mathbf{p}_1 = A\mathbf{p}_0$, where $\mathbf{p}_1$ is the projected age-specific population vector, $\mathbf{p}_0$ is the starting age-specific population vector, and A is the demographic transition matrix. This single-sex version of the matrix has the format

shown in (1.3), with ${}_5S_a$ representing five year survival at age a and ${}_5F_a$ representing the five year fertility rate at age a; zeros and elided entries are "structural zeros" – elements which must always be zero in a demographic model (for example, fertility to women in non-childbearing years and survival transitions that go backwards, as well as survival transitions that jump more than a single forecast step).

$$\begin{bmatrix} {}_5p_0^0 \\ {}_5p_5^0 \\ {}_5p_{10}^0 \\ {}_5p_{15}^0 \\ \vdots \\ {}_5p_{80}^0 \\ {}_\infty p_{85}^0 \end{bmatrix} \tag{1.2}$$

$$\begin{bmatrix} 0 & 0 & {}_5F_{10} & {}_5F_{15} & \dots & 0 & 0 & 0 \\ {}_5S_0 & & & & \dots & & & \\ & {}_5S_5 & & & \dots & & & \\ & & {}_5S_{10} & & \dots & & & \\ & & & {}_5S_{15} & \dots & & & \\ & & & & \ddots & & & \\ & & & & \dots & {}_5S_{80} & & \\ & & & & \dots & & {}_\infty S_{85} & 0 \end{bmatrix} \tag{1.3}$$

Two sex Leslie matrix with migration

The transition matrix generated by our version of WM follows the approach of the work of Rogers in two-sex transition matrices (Rogers, 1994), though it is a new formulation that uses the diagonal to account for net migration. Our matrix format is 36x36, which can be thought of as four 18x18 blocks. The upper left block projects males, with the sub-diagonal in this block representing male survival and the diagonal representing male net migration. The lower right block projects females, with the same structure as the upper left block, except that it also contains entries in the top row of the block which generate female babies. The upper right block's top row generates male babies; it is placed in this block because it represents a transition from female mothers to male children. This matrix does not model a population closed to migration, so the rows can sum to any value, including negative numbers. See (1.4) for a schematic, with "F"

representing fertility, "S" survival, "M" net migration, and the zeros standing for structural zeros in the matrix. "MF" represents "migratory fertility," babies born to mothers migrating in to the population. To project a population forward, one simply multiplies this matrix and a 36 x 1 population vector, as $\mathbf{n}(t+1) = A\mathbf{n}(t)$.

This form of the transition matrix is comprised of fairly standard pieces, but it differs from other Leslie matrix formulations in important ways. The upper left corner elements in the lower right and upper right blocks encode babies born in transit to mothers migrating into the population; note that we use the symbols "MF_m" and "MF_f," for these coefficients, respectively, to highlight their combined migration and fertility significance. The lower right corner in the upper left and lower right blocks encodes both migration and survival of the 85+ open interval. The rest of the diagonal cells describe net migration – population change unaccounted for by cohort survival, the residual after the population has aged forward. By placing migration in the diagonal and survival in the sub-diagonal, the transition matrix is more easily interpreted in terms of demographic components of change. For counties that have high college populations, both the sub-diagonal and the diagonal play a role in migration (see the discussion on the sub-diagonal constraint below for more detail). Placing migration in the diagonal assumes that the population "at risk for net migration" is the current age group, corresponding to the method for calculating migration in (Voss et al., 2004). However, in the literature, the population at risk for migration is more often considered to be the age class younger than the new migrants, and this net migration would be accounted for in the sub-diagonal as a cohort change ratio, in contrast to our formulation. Caswell uses Leslie matrices primarily in the context of stage classified models and closed populations; he uses the diagonal for those individuals remaining in a given stage, but requires that a given row sum to less than unity because, by definition, there is no migration in a closed population. None of these approaches uses the corner elements for migratory fertility, but we found their use to yield better fits through experimentation. Our matrix formulation, though slightly unusual, works well for its purpose.

$$\left[\begin{array}{ccccccc|ccccccc}
0 & 0 & & \dots & & & & MF_m & 0 & {}_5F_{10} & \dots & 0 & 0 \\
{}_5S_0 & M_5 & & \dots & & & & & & & \dots & & \\
 & {}_5S_5 & M_{10} & \dots & & & & & & & \dots & & \\
 & & & \ddots & & & & & & & \dots & & \\
 & & & \dots & {}_+S_{85} & {}_+M_{85} & & & & & \dots & & \\
\hline
 & & & & & & & MF_f & 0 & {}_5F_{10} & \dots & 0 & 0 \\
 & & & & & & & {}_5S_5 & M_5 & & \dots & & \\
 & & & 0 & & & & & {}_5S_5 & M_5 & \cdots & & \\
 & & & & & & & & & & \ddots & & \\
 & & & & & & & & & & \dots & {}_+S_{85} & {}_+M_{85}
\end{array}\right] \tag{1.4}$$

1.2.4 Wood's method

We will now describe the optimization algorithm at the core of the technique we are presenting. We closely follow pages 144-145 in (Caswell, 2001), which forms the basis for the computer implementation underpinning Chapters 1, 2, and 4 of this dissertation. We feel that Caswell's original derivation is unnecessarily obscure, and we present what we believe is a more straightforward formulation in Section 1.2.4, along with a easy to follow, small scale example.

Mathematical preliminaries

Quadratic optimization is formulated in general (without any specific application to Wood's Method) as follows:

$$\begin{aligned}
&\textit{minimize} && \frac{1}{2}\mathbf{p}^{\mathbf{T}}Q\mathbf{p} - \mathbf{c}^{\mathbf{T}}\mathbf{p} && (1.5)\\
&\textit{subject to} && G\mathbf{p} \leqq \mathbf{h} && (1.6)\\
& && A\mathbf{p} = \mathbf{b} && (1.7)
\end{aligned}$$

$\mathbf{Q}$ is a constant matrix, $\mathbf{c}$ is a constant vector, and $\mathbf{p}$ is the decision variable (a vector), the elements of which are varied to find the minimal value of (1.5). The matrix $\mathbf{G}$ and vector $\mathbf{h}$ encode the inequality constraints on $\mathbf{p}$, and the matrix $\mathbf{A}$ and vector $\mathbf{b}$ encode the equality constraints.

The Kronecker product $A \otimes B$ is defined on page 27 of (Golub and Loan, 2012) as multiplying the entire matrix B once for each element of A and concatenating the resulting block matrices. Note that it is not commutative.

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix} \tag{1.8}$$

The *vec* operator is defined on page 28 of the same work and examined in detail on pages 710-711. This operator takes a matrix and "stacks" it as a single column vector,

$$C = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{bmatrix} \tag{1.9}$$

$$vec(C) = \begin{bmatrix} c_{11} \\ \vdots \\ c_{m1} \\ \vdots \\ c_{1n} \\ \vdots \\ c_{mn} \end{bmatrix} \tag{1.10}$$

The *reshape* operator is also defined on page 28. This operator takes a vector v, an integer r for the number of rows, and an integer c for the number of columns, and creates a matrix by reshaping the input vector to have the specified number of rows and columns. This operator is an inverse of *vec*. $vec(C)$ can also be written as $reshape(C, m \cdot n, 1)$.

$$c = \begin{bmatrix} c_{11} \\ \vdots \\ c_{m1} \\ \vdots \\ c_{1n} \\ \vdots \\ c_{mn} \end{bmatrix} \tag{1.11}$$

$$C = reshape(c,m,n) = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{bmatrix} \tag{1.12}$$

Vector norms are examined on pages 68 through 70 of (Golub and Loan, 2012). Norms measure the distance, broadly defined, between two vectors. The 2-norm corresponds to the Euclidean distance between two points, and it is the most important vector norm in this exposition. It is defined as follows:

$$\|x\|_2 = (x_1^2 + \ldots + x_n^2)^{\frac{1}{2}} = (x^T x)^{\frac{1}{2}} \tag{1.13}$$

Wood's Method

Let $\mathbf{n}(t)$ be a k element vector representing the population at time = t. In our application, $k = 36$, reflecting our 36 age/ sex classes. t is a time step in the range $0 \ldots T$, with $t = 0$ for the first population measurement and $t = T$ the last population measurement.

Let A be the unknown demographic transition matrix, with elements $a_{i,j}$. $vec(A)$ is a column vectorized version of A after the *vec* operation is performed.

Let $\mathbf{p}$ be the vector after all elements of the corresponding demographic transition matrix A which are known to be zero due to demographic dynamics are removed. For example, the super diagonal in a demographic transition matrix is always zero because populations do not age backwards; we remove the corresponding row in $vec(A)$. The elements of $\mathbf{p}$ are the nonzero elements only of the unknown transition matrix, and they are allowed to vary as we try to fit this matrix.

Calculate $N(t) = \mathbf{n}(t) \otimes I$, where $\otimes$ is the Kronecker product, and $\mathbf{I}$ is a k by k identity matrix. In (1.14), we show an example of $N(t)$, in which ${}_cm_a$ signifies the counts of males of age a to $a+c$, and ${}_cf_a$ signifies females of age a to $a+c$.

$$N(t) = \begin{bmatrix} {}_5m_0 & 0 & & 0 & {}_5m_5 & 0 & & 0 & 0 & {}_\infty f_{85} & 0 & & 0 \\ 0 & {}_5m_0 & & 0 & 0 & {}_5m_5 & & 0 & 0 & 0 & {}_\infty f_{85} & & 0 \\ & & \ddots & & & & \ddots & & & & & \ddots & \\ 0 & 0 & & 0 & 0 & 0 & & {}_5f_{80} & 0 & 0 & 0 & & 0 \\ 0 & 0 & & {}_5m_0 & 0 & 0 & & 0 & {}_5f_{80} & 0 & 0 & & {}_\infty f_{85} \end{bmatrix} \tag{1.14}$$

We then remove columns of $N(t)$corresponding to the rows removed from $vec(A)$ in order to create $\mathbf{p}$ above. Let this new matrix be $M(t)$.

Now let $\mathbf{z}$ be the vertical concatenation of $n(t)$ for all $t = 1..T$:

$$\mathbf{z} = \begin{bmatrix} n(1) \\ n(2) \\ \vdots \\ n(T) \end{bmatrix} \tag{1.15}$$

Let M be the vertical concatenation of $M(t)$ for $t = 0..T-1$:

$$M = \begin{bmatrix} M(0) \\ M(1) \\ \vdots \\ M(T-1) \end{bmatrix} \tag{1.16}$$

Now we seek to minimize $\|\mathbf{z}-M\mathbf{p}\|^2$, the distance between the empirical populations stored in $\mathbf{z}$ and the projected populations formed by $M\mathbf{p}$.

$$\begin{aligned} \|\mathbf{z}-M\mathbf{p}\|_2 &= (\mathbf{z}-M\mathbf{p})^{\mathbf{T}}(\mathbf{z}-M\mathbf{p}) & (1.17) \\ &= \mathbf{z}^{\mathbf{T}}\mathbf{z} - \mathbf{z}^{\mathbf{T}}M^{\mathbf{T}}\mathbf{p} - \mathbf{p}^{\mathbf{T}}M^{\mathbf{T}}\mathbf{z} + \mathbf{p}^{\mathbf{T}}M^{\mathbf{T}}M\mathbf{p} & (1.18) \end{aligned}$$

We drop the $\mathbf{z}^{\mathbf{T}}\mathbf{z}$ term, since it is constant and thus won't play a role in the optimization, and we collect like terms to yield:

$$min\ \frac{\mathbf{p}^{\mathbf{T}}M^{\mathbf{T}}M\mathbf{p}}{2} - \mathbf{z}^{\mathbf{T}}M\mathbf{p} \tag{1.19}$$

Because is $M^{\mathbf{T}}M$ is positive definite, this objective function is convex. Since the constraints are all linear, the optimization problem is well defined, and we are guaranteed to find a unique global minimum as long as the problem is feasible (Boyd and Vandenberghe, 2005). The matrices $\frac{M^{\mathbf{T}}M}{2}$ and $\mathbf{z}^{\mathbf{T}}M$ are both passed into the software system as parameters.

Note that this method optimizes the transitions between adjacent input periods pairwise, but it does not optimize the transition from the first period of input to data to the last. Our approach has the advantage that it is a linear sum of squares fitting problem $\|\mathbf{n}(t+1) - A\mathbf{n}(t)\|_2$, so it is extremely "well behaved" analytically. It is also especially appropriate for the sort of short term forecasts that are regularly encountered in applied work at the state and local levels. The disadvantage is that the method may not be suitable for long term forecasting, especially when rates are changing as well as

population numbers; in these situations one would want to optimize $\|\mathbf{n}(t) - A^t\mathbf{n}(0)\|_2$, the best fit of a single matrix for the entire input. However this latter objective function yields nonlinearities and is likely not as appropriate to short term forecasts as Wood's Method.

1.2.5 Constraints on demographic rates

The values in the fitted Leslie transition matrix are enforced by constraints passed to the quadratic optimization routine `qp()`. In the discussion below $L(r,c)$ will refer to the cell in the Leslie matrix at row = r and column = c.

We use constraints to make entries in the Leslie matrix correspond to known demographic dynamics and rates. Many entries in a Leslie matrix are equal to zero, and all entries should be positive. Cells modeling survival must be between 0 and 1.0. The sum of the fertility cells should sum to a reasonable amount. Sometimes proportional relationships between cells must be enforced. Other constraints will be described in detail below.

Fertility (the cells labeled "F" in (1.4)) is constrained such that the sum of these cells falls between 1.0 and 6.0. The constraint determining the sum of the fertility cells is stored in G and $\mathbf{h}$ in (1.6).

Fertility is also constrained such that the relative proportions of all the fertility cells are constant, even though the total of these cells is determined by the optimization fitting routine. This constraint models the fact that human fertility is somewhat consistently distributed over fertile ages in an approximately log-normal shape. These percentages are derived from the Human Fertility Database (Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria), 2012) for women in 1980 USA, using 51.14% as the male sex ratio. This proportional constraint for fertility is described in Table 1.1. Each entry represents the percentage of the total fertility, between 1.0 and 6.0, that is allocated to the given age and sex entry in the fitted Leslie matrix. Fertility percentages for ages above 50 are left out of the table since their percentage is equal to zero; fertility percentages for age 45-50 is trace but we left it for completeness. The fertility percentage for male babies is slightly higher because of the higher male sex ratio at birth, for example 2.56% for males and 2.44% for females. The equality constraints determining this proportional fertility shape are stored in the matrix A and the vector $\mathbf{b}$ in (1.7).

Age	Males	Females	Matrix cells
0	2.56%	2.44%	$a_{1,19}, a_{19,19}$
10	1.53%	1.47%	$a_{1,21}, a_{19,21}$
15	10.23%	9.77%	$a_{1,22}, a_{19,22}$
20	13.81%	13.19%	$a_{1,23}, a_{19,23}$
25	12.78%	12.21%	$a_{1,24}, a_{19,24}$
30	7.67%	7.33%	$a_{1,25}, a_{19,25}$
35	2.05%	1.95%	$a_{1,26}, a_{19,26}$
40	0.51%	0.49%	$a_{1,27}, a_{19,27}$
45	0.00%	0.00%	$a_{1,28}, a_{19,28}$

Table 1.1: Fertility constraints

The sum of the fertility cells is critical to the functioning of the method, but the distribution of the fertility over ages has very little effect on the suitability of the matrix for single period projections. We choose the distribution described because it is convenient and reflects an average distribution, but we are not arguing here that this particular fertility model is a particularly good choice or not. However, it does allow the matrix to be used for longer term projections, when new generations pass through these fertility cells.

Survival ("S" in (1.4)) is constrained to fit between a lower bound corresponding to 1970 USA males and an upper bound corresponding to 2008 USA females, computing survival as $S = L(x)/L(x-5)$, where $L(x)$ is the total number of person-years lived by the cohort from age x to $x+1$. Each cell in the subdiagonal is also constrained to be less than or equal to the next younger cell. This approach fails to capture relationships between ages such as a mortality shape, but when these relationships are better specified, they can be easily incorporated as constraints.

When the cohort ratio for 15 to 20 year olds in a given county is above 1.4 – i.e. when there are 1.4 times as many 20 year olds as 15 year olds between two periods – the sub-diagonal constraints for several young adult ages are relaxed and allowed to range more widely than would happen if the cell was modeling mortality alone. This change in constraints is made because counties with this attribute typically contain large university populations, and without relaxing the constraints on the sub-diagonal the population of 20-25 year olds will "age forward" instead of migrating out, and an inaccurate population wave will propagate through the county age structure (see Figure 1.6 for a plot of the age structure of a university-dominated county). These inequality constraints on survival cells are described in Table 1.2, and stored in matrix **G** described

above. Like all constraints these are linear, just wider than many other constraints due to the extreme behavior in university dominated counties.

Age	Cells	Male min	Male max	Female min	Female max
0	$a_{2,1}, a_{20,19}$	0.99614	0.99887	0.99709	0.99908
5	$a_{3,2}, a_{21,20}$	0.99769	0.99921	0.99854	0.99938
10	$a_{4,3}, a_{22,21}$	0.99523	0.99775	0.99784	0.99885
15	$a_{5,4}, a_{23,22}$	0.98988	0.9938	0.99645	0.99781
20	$a_{6,5}, a_{24,23}$	0.98952	0.99285	0.99605	0.99741
25	$a_{7,6}, a_{25,24}$	0.98955	0.99283	0.99503	0.9968
30	$a_{8,7}, a_{26,25}$	0.98676	0.99178	0.99258	0.99561
35	$a_{9,8}, a_{27,26}$	0.98066	0.98877	0.98875	0.99333
40	$a_{10,9}, a_{28,27}$	0.96986	0.98278	0.98285	0.98954
45	$a_{11,10}, a_{29,28}$	0.95364	0.97377	0.97465	0.98425
50	$a_{12,11}, a_{30,29}$	0.92778	0.96162	0.96300	0.97767
55	$a_{13,12}, a_{31,30}$	0.89199	0.94611	0.94672	0.96668
60	$a_{14,13}, a_{32,31}$	0.84307	0.92023	0.92108	0.94846
65	$a_{15,14}, a_{33,32}$	0.78197	0.88355	0.87881	0.92151
70	$a_{16,15}, a_{34,33}$	0.69812	0.82576	0.80976	0.87797
75	$a_{17,16}, a_{35,34}$	0.59611	0.73736	0.71033	0.80798
80	$a_{18,17}, a_{36,35}$	0.46707	0.61004	0.57199	0.6961

Table 1.2: Survival constraints

Migration ("M" in (1.4)) is constrained based on the age – it is allowed to range widely for people in their early twenties and retirement ages, but it is kept smaller for other ages. The corner cells ($L(1,19)$ and $L(19,19)$ – for 0-5 males and 0-5 females respectively) are considered part of the fertility row and are constrained to be part of the fertility proportion described above, because the solver otherwise puts all the fertility dynamics into these cells. These cells are labeled "MF" (1.4) to highlight their dual nature. The entry for 85+ year-olds ($L(18,18)$ and $L(36,36)$ for males and females, respectively) does "double duty" for mortality and migration, as well. These inequality constraints on the migration cells are listed in Table 1.3, and stored in matrix $\mathbf{G}$ described above.

In this approach, migration is calculated as a residual; the bounds on fertility and survival are fairly tight and the remaining dynamics are forced into the migration cells. However, if one constrained the migration and fertility based on known rates, the subdiagonal would give an estimate of survival.

Age	Cells	Value
5	$a_{2,2}, a_{20,20}$	+/-0.50
10	$a_{3,3}, a_{21,21}$	+/-0.50
15	$a_{4,4}, a_{22,22}$	+/-1.50
20	$a_{5,5}, a_{23,23}$	+/-1.50
25	$a_{6,6}, a_{24,24}$	+/-3.50
30	$a_{7,7}, a_{25,25}$	+/-1.50
35	$a_{8,8}, a_{26,26}$	+/-0.75
40	$a_{9,9}, a_{27,27}$	+/-0.50
45	$a_{10,10}, a_{28,28}$	+/-0.50
50	$a_{11,11}, a_{29,29}$	+/-0.50
55	$a_{12,12}, a_{30,30}$	+/-0.50
60	$a_{13,13}, a_{31,31}$	+/-0.75
65	$a_{14,14}, a_{32,32}$	+/-0.75
70	$a_{15,15}, a_{33,33}$	+/-0.75
75	$a_{16,16}, a_{34,34}$	+/-0.50
80	$a_{17,17}, a_{35,35}$	+/-0.50
85+	$a_{18,18}, a_{36,36}$	+/-0.65

Table 1.3: Diagonal migration constraints

These constraints are encoded in matrices and passed to the software along with the formulations in the above section. Due to the large number of constraints (over 100), we do not show their matrix formulations.

1.3 Optimization and demographic analysis

We believe that the approach of WM taken here – determining the elements of a transition matrix based on an optimal fit to data, encoding prior theoretical knowledge in constraints that restrict the matrix elements and impose (linear) relationships between them – may signal a new approach to demographic modeling which has the potential to yield many fruitful results beyond age/ sex forecasting. Constraints allow almost arbitrary prior knowledge to be easily and incrementally incorporated into the inference

stage of the forecast. Using a transition matrix to store the population dynamics gives us a succinct, interpretable representation of population change at the system level. Using optimization to find a best fit transition matrix allows us to go beyond deterministic modeling using fixed input rates. The constrained optimization approach draws on a wealth of prior research into constrained optimization and mathematical programming; working within this well established framework may enable fruitful collaborations. Finally, this basic approach is not limited to age/ sex forecasting, but is applicable to any situation that can be modeled with transition matrices. (A similar flexibility and power is possible in Bayesian network approaches, but these only answer questions about probability distributions; admittedly systems of probability distributions encompass a huge number of models, but they cannot model population processes directly.)

Given that in operations research it is well known that constraints are where "the magic happens," it is worth exploring our choice of constraints above in light of traditional demographic ideas.

First, we think of constraints as being "another level of indirection" above model demographic rates. The constraints used in this paper are based on model rates from large scale demographic data collection projects, specifically the Human Mortality Database (University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), 2012) and the Human Fertility Database (Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria), 2012), and from demographic theory regarding invariants in the age structure of mortality and fertility. However, rather than use a set of model rates deterministically, the constrained optimization procedure allows the population count data to have an influence on the fitted matrix within upper and lower bounds determined by the model rates. On the other hand, if an analyst wants to fix demographic rates deterministically, this is trivially accomplished by setting equality constraints rather than the inequality constraints presented here.

We also find it useful to think in terms of a few "constraint sets" rather than in terms of atomistic constraints – we use a lower-bound constraint set for survival, an upper-bound constraint set for survival, an equality constraint set to enforce fertility proportions, etc. Each constraint set is almost directly analogous to a model rate profile, except that it provides a bound or a relationship rather than a set of deterministic values. A constraint set may be derived empirically, analogously to "borrowing" a lifetable to model a population for which life tables have not been tabulated (often due to small sample problems). Constraint sets can also be derived using the spine-plus-parameter ideas of Brass (Wachter, 2012), from a purely analytic expression like the Gompertz mortality model, or using the quadratic model of (Wilmoth et al., 2012). A forecaster must still

decide on the values in a constraint set – a mortality constraint set for high mortality countries would be derived with a higher α in the Gompertz approach, for example – but the additional level of indirection allows for a combination of analyst judgment, information "borrowing," and data inference that is impossible with deterministic model rates alone. Finally, by using generic constraint sets with fairly wide bounds, the analyst can create reasonable forecasts with little population-specific knowledge (the approach taken in this paper). In fact, the analyst can get decent forecasts with the defaults presented here, even with no demographic theoretical background.

Additionally, the constraint approach allows us to gracefully and incrementally combine small- and large- scale data sources, by modifying constraint sets with "point constraints." Establishing rates for small populations is a chronic problem in demography, especially for mortality due to its rareness, so demographers typically "borrow" rates from large population lifetables when forecasting small populations. In the constraint approach, however, one can use local knowledge to adjust these general rates by adding point constraints at certain ages – for example, if homicide among 20 to 25 year old males is a known issue in a specific population, one can force survival at this age to be arbitrarily low without changing any other elements of the mortality constraint set, and then fit the transition matrix using these modified assumptions. Adjustments to the survival constraints for university dominated counties in this paper is another example of local, small-data information being included for specific populations at the constraint level.

Finally, the constraint approach is not limited to bounding single matrix elements. Constraints can be written as arbitrary linear combinations of the Leslie matrix elements, which allows for staggeringly complex interrelationships to be enforced. This technique is used with the fertility elements of the matrix described above: we enforce a constraint such that the sum of fertility elements is between 1.0 and 6.0, and that the fertility elements must have a fixed proportion to each other. As a more speculative example, we might constrain migration at ages 20 to 30 to be a proportion of retirement migration (as the former ages may work in service industries for the latter) and examine the resulting matrix for plausibility or theoretical interest.

More extensive analysis of this framework will be presented in the sequel. However, it is important to understand the power of this approach and its relationship to traditional demographic methods.

1.4 Testing Wood's method on US county populations

1.4.1 Data

The test data are an almost complete set of U.S.A. county populations from the US Census censal data and intercensal estimates, from 1970 through 2010, at 5 year intervals (US Census Bureau, 2011). The populations at each interval are divided into 5 year age widths, for males and females separately. The data was compiled from a variety of locations on the US Census website, often aggregating over race detail in the original data. Then the dataset was analyzed to determine which counties maintained a consistent FIPS code over the 40 years, and only those counties whose code was consistent were retained. This final dataset was stored as a 36 x 3120 x 9 three dimensional matrix, with 36 age/sex rows, 3120 county columns, and a third dimension storing the 40 years of data.

Below we present the results of this test. In Chapter 2 we explore comparisons between our forecasts and other official forecasts and examine the error structure somewhat exhaustively. Additionally, we examine demographic rates (births, deaths, and migration) implied by the Leslie matrix projection. We have chosen to separate the exhaustive empirical chapter from this chapter in order to keep the exposition clear.

To test the method, each county was run separately to compare forecasted population with actual population. The age specific population in 1980 through 2000 – four time steps, or $T = 4$ – was used as the training set. This set was processed by an Octave routine that reads in population data and default constraints and returns the transition matrix using the method outlined above. These resulting transition matrices were used to forecast 2000 population counts two steps forward, yielding an age specific forecast for 2010 using simple matrix multiplication $\mathbf{p}_{t+2} = \mathrm{A}^2\mathbf{p}_t$. The resulting population vector for 2010 was compared to the empirical data for 2010, and absolute percentage differences were computed for each age and sex cell. These percentage differences and other metadata were stored for all the counties as another set of 36 by 3120 matrices.

There are some caveats with this dataset, though we believe it provides a sufficient test of the method to prove its utility. Two counties failed to converge when their data was processed through the Woods method, and these were dropped from the error analysis below. Certain important counties were dropped because they had changes in their FIPS codes over the input period, including Dade County in Florida. Counties which changed their geographic boundaries but retained their FIPS code are assumed

to be homogeneous across time, even though their populations could have changed in ways that might bias the fitted Leslie matrix. We used intercensal estimates for the periods in between censuses (1975, 1985, 1995, 2005). Finally, data quality is at the mercy of the US Census Bureau.

Counties were chosen as a unit of analysis for several reasons. First, this research was initiated to support Growth Management Act (GMA) forecasting at Washington State, and counties are the basic unit as determined by GMA legislation. Second, institutions throughout the United States are regularly called upon to forecast population at the county level, so these techniques would frequently be exercised on county data if they were adopted. Third, county data is of excellent quality overall, as counties form both a basic enumeration and a basic tabulation geography for the U. S. Census Bureau. Fourth, counties provide a wide range of population sizes on which to test the method, ranging in 2010 population from 9,818,605 (Los Angeles County, California) to 82 (Loving County, Texas). Finally, counties also show a diversity of of population dynamics due to their sometimes very particular social contexts, with smaller counties often "specializing" in ways that influence age structure, such as providing retirement communities or housing university student populations.

This dataset should be considered as an extended example of the method but not an exhaustive analysis of the method's error properties.

1.4.2 Error structure of ten year forecasts

After all the counties were run, the 2010 forecasts were compared to actual 2010 census counts.

The Mean Absolute Percentage Error (MAPE) for Wood's Method over all cells is 10%; this is comparable to the MAPEs for 10 year county forecasts given in Smith and Tayman (2003). If we test the method on counties with populations larger than 50,000, the MAPE over all cells is only 7% for Wood's Method, and the MAPE for counties with less than 50,000 population is 11%.

The age specific MAPEs are slightly larger, see Table 1.4. Here we see that the age specific MAPEs for WM are comparable to the forecasts presented on in Smith and Tayman (2003) for several county forecasts in Florida. The first column is the age group. The second column gives the MAPE for Wood's Method (we summed the forecast and empirical populations to agree with the age categories in Smith and Tayman (2003)). The third column presents their forecasts from 1980 to 1990 for all counties. The fourth and fifth columns gives MAPEs for two different forecasts from 1990 to 2000 for all counties. They only use county forecasts for the state of Florida, while we produced

forecasts over almost all counties in the US; Florida experiences high migration, which may contribute to higher error due to more volatility in population change.

Unfortunately the US Census Bureau does not produce age/ sex population projections for counties, so it is difficult to assemble a large dataset against which to compare this method at the county level.

Age	WM 2010	S&T 1990	S&T 2000, #1	S&T 2000, #2
0	11.4%	9.6%	16.4%	10.3%
5	10.8%	9.1%	13.7%	8.0%
10				
15	11.9%	11.2%	13.1%	10.2%
20				
25	14.2%	13.3%	18.0%	12.7%
30				
35	12.9%	10.4%	14.7%	11.6%
40				
45	8.7%	9.5%	12.3%	11.2%
50				
55	6.9%	9.3%	15.5%	13.3%
60				
65	5.3%	11.0%	17.8%	10.1%
70				
75				
80				
85				

Table 1.4: MAPE by age

Also interesting is the quantile pattern of errors in Tables 1.6 and 1.5. Here we see that the more mobile ages (20-30 years old) are harder to forecast than the more staid working and childhood ages. Also we see higher MAPEs for 80 plus; this high MAPE is likely due to small sample and quantization problems.

1.4.3 County examples

As a further example of the utility and limitations of the method presented, we con-

Age	Sex	50% MAPE	80% MAPE	97.5% MAPE	Max
0	F	10%	18%	36%	135%
5	F	9%	17%	38%	202%
10	F	6%	13%	36%	164%
15	F	6%	13%	38%	298%
20	F	10%	20%	61%	416%
25	F	10%	23%	64%	217%
30	F	9%	19%	46%	127%
35	F	6%	15%	36%	112%
40	F	6%	13%	34%	111%
45	F	5%	11%	29%	113%
50	F	4%	10%	28%	103%
55	F	4%	9%	25%	146%
60	F	4%	9%	21%	145%
65	F	4%	9%	22%	162%
70	F	4%	9%	22%	162%
75	F	4%	10%	28%	226%
80	F	5%	12%	31%	258%
85	F	7%	17%	45%	342%

Table 1.5: Quantile error structure

sider several counties in Washington State by looking at plots of the forecasts. In the plots, the solid line shows the empirical population in 2010, the triangled line is the 2010 forecasted population, and the circled line is the jump off population in 2000. The x-axis is the bottom of the five year age class at which the population is counted, and the y-axis represents the population at that age class, whether actual or forecasted.

These plots allow us to see many of the basic population dynamics for each county, and show that our method forecasts the age/ sex composition well. If a peak or a trough in the circled 2000 line moves rightward, to the plain and triangled 2010 lines, this implies that part of the cohort is aging forward and staying in the county. If the peaks and troughs do not move to the right, but rather move up and down in the same age, it implies that people are moving into the area at a given age and back out again when they get older (see Whitman County for an example – people move into the county for college and move out when they graduate). We will use the terms "cohort age dynamic" and "migratory age dynamic" for these two population dynamics, respectively.

Age	Sex	50% MAPE	80% MAPE	97.5% MAPE	Max
0	M	10%	18%	37%	151%
5	M	8%	16%	36%	117%
10	M	6%	14%	36%	109%
15	M	6%	13%	34%	185%
20	M	10%	22%	61%	405%
25	M	11%	25%	65%	323%
30	M	10%	22%	55%	286%
35	M	8%	19%	50%	221%
40	M	7%	17%	48%	168%
45	M	5%	12%	37%	215%
50	M	5%	11%	34%	132%
55	M	5%	11%	30%	96%
60	M	5%	10%	25%	138%
65	M	7%	12%	25%	129%
70	M	9%	15%	29%	115%
75	M	9%	15%	30%	200%
80	M	11%	19%	36%	256%
85	M	12%	21%	45%	704%

Table 1.6: Quantile error structure

These counties were chosen to represent specializations in county attributes and corresponding ages of residents. They include counties with large regional centers (Spokane County), large internationally important cities (King County, which contains Seattle), suburban counties (Clark County, which provides housing for the Portland, OR metropolitan area), high retirement population counties (Clallam County), university oriented counties (Whitman County), and agricultural counties (Yakima County); each of these specializations have distinct age structures, all of which are forecasted well by Woods Method. Although these examples were chosen informally based on our knowledge of their social context, the idea of county classification owes much to (Pittenger, 1976; Parker, 2012).

Note that population growth in Washington state has slowed considerably below long term averages since 2008 due to the recession (in-migration was 13,000 in 2010 and about 4,000 in 2011, compared to a yearly average of about 45,000). This slowdown affects the accuracy of the forecast totals, since the training period had higher overall

growth than the forecast period. Even when the total forecast is lower than the empirical number, however, Wood's method often retains the basic shape of the age structure; this ability to keep structure is a benefit of the method.

Regional city center (Spokane County, FIPS 53063)

Spokane County is located on the Eastern border of Washington. It has a population of about 470,000, a diverse and vibrant economy, including manufacturing, farming, and multiple universities (Gonzaga, Eastern Washington University, and other smaller private colleges). Historically, Spokane County has had slow but steady growth. While the county contains a typical metropolitan center (the city of Spokane), the county is large enough to also include most of the family-oriented suburbs serving the central city. It provides an example of a "well-rounded" county, with a fairly complete range of ages. In Figure 1.1, we can see a cohort age dynamic as the baby boomers and their children both age forward in time. The forecast is quite good, except at ages 20-25, due possibly to increased college enrollment in the current recessionary environment.

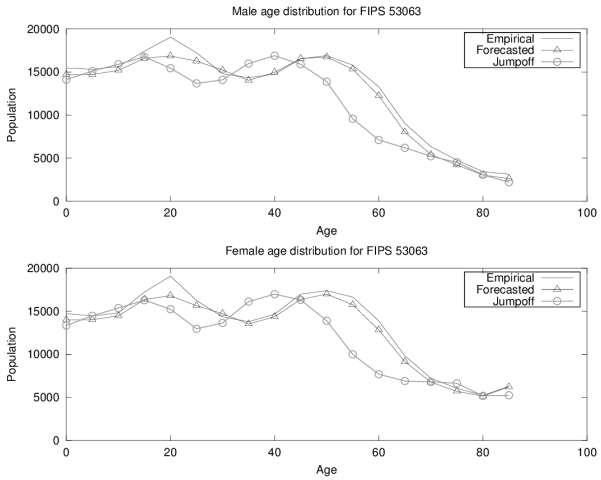


Figure 1.1: Spokane County forecast cross validation

Large metropolitan center (King County, FIPS 53033)

King County contains Seattle, a thriving high-tech industry, a large amount of urban-

ized area, and many universities, including the University of Washington (see Figure 1.2). Like many highly urbanized counties, there is a preponderance of people in their twenties and early thirties, who migrate into the county for employment and education but move out to the suburbs as they start families; this migratory age dynamic causes the steady bulge in the twenties and early thirties that we see in the Figure. We also see a fair number of baby boomers experiencing cohort age dynamics, as that bulge extends to older ages in 2010. Note the relative small numbers of children, a proportion of the population that is subject to migratory dynamics rather than cohort age dynamics due to the tendency of families to move to the suburbs when their children reach school age. Note also that the forecast under-predicts age 0-5, consonant with the anecdotal evidence that in the last decade more parents are staying in urban areas than in the 1990s and before. The forecast is good, except for the unpredicted peak at age 25-30, possibly due to the technology's sector strong performance even during the recessionary environment.

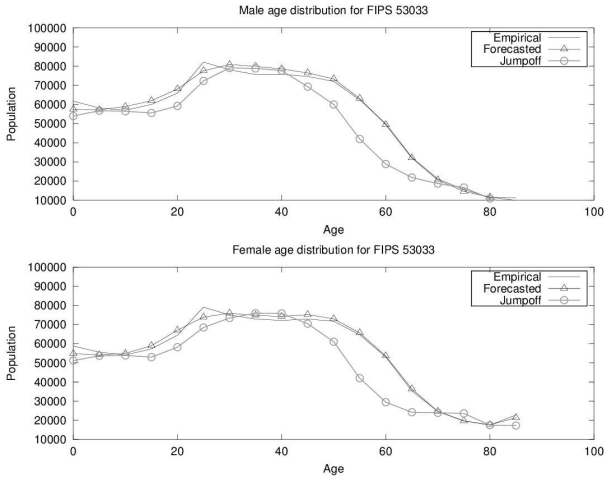


Figure 1.2: King County forecast cross validation

Suburban satellite (Clark County, FIPS 53011)

Clark County shows a typical suburban dynamic, in which young people move out in their teens and early twenties, and family-age people move in in their later twenties and

thirties (see Figure 1.3), which is what we expect given that Clark County is a bedroom community to nearby Portland and Washington County in Oregon. Note that we see the baby boom peak experiencing cohort age dynamics, but the younger ages evincing migratory age dynamics with no forward cohort movement. Note also that the forecast over-predicts total population due to the current recession, but it retains the shape of the population extremely well.

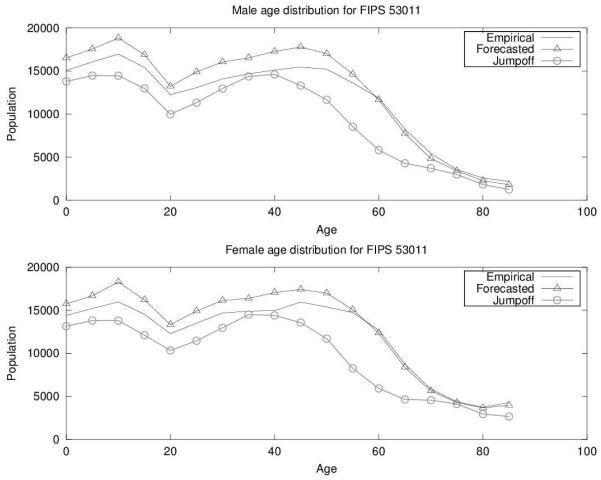


Figure 1.3: Clark County forecast cross validation

Retirement/ amenity (Clallam County, FIPS 53009)

Clallam County is a rural county on the Olympic Peninsula, with high recreational amenities (coast and mountains) but low employment opportunity. Like many similar counties, it shows typical older in-migration, with a huge bulge at about 55 years old. This age group displays cohort aging, but it also shows migratory age dynamics as the peak increases from migration (growing upward as well as moving to the right). The trough at 20-25 years old has migratory age dynamics, presumably as these young adults leave the rural county for opportunity and education. The forecast predicts the

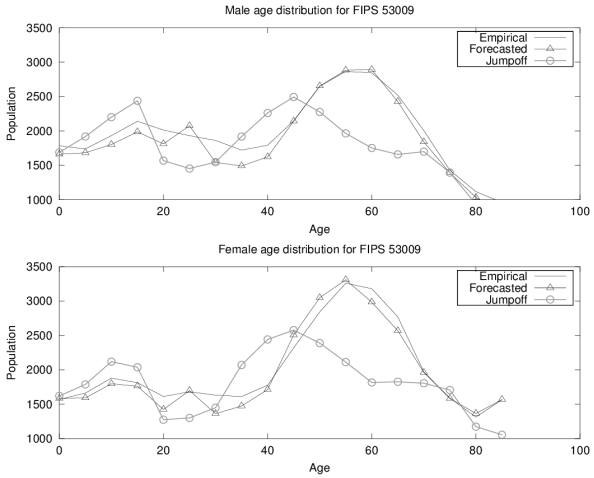


Figure 1.4: Clallam County forecast cross validation

older population very well, but the younger population is under-forecast.

Agricultural region (Yakima County, FIPS 53077)

Yakima County has an agricultural economic base, a large Hispanic population, and high fertility. There is large out-migration in the early twenties for non-Hispanics. See Figure 1.5. The forecast is quite good, except for over-predicting population in their twenties, which may be due to the current recessionary environment damping the force of Latino in-migration.

University specialization (Whitman County, FIPS 53075)

Whitman County contains Washington State University, a large public university. Besides the university, however, the next most important local industry is low-labor wheat farming, and there are few recreational attractions in nearby; this configuration is consistent with the large population spike at college ages (see Figure 1.6), which sharply decreases at the next age interval (25-30) with young people leaving shortly after graduation. Notice that the forecast is for decreasing population from 2000, while the empirical data show an increase in population. This misforecast is due to prior trends being incorporated into the transition matrix through the training data; from 1990 to 2000, Whitman saw an overall drop in population at these ages, but due to the recent recessionary environment, enrollment at WSU has probably seen higher enrollment

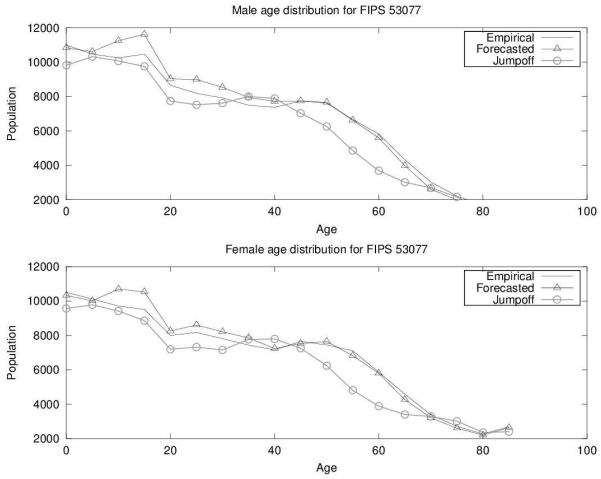


Figure 1.5: Yakima County forecast cross validation

in 2000-2010. Even with this error in forecast magnitude, the shape of the forecast population is correct.

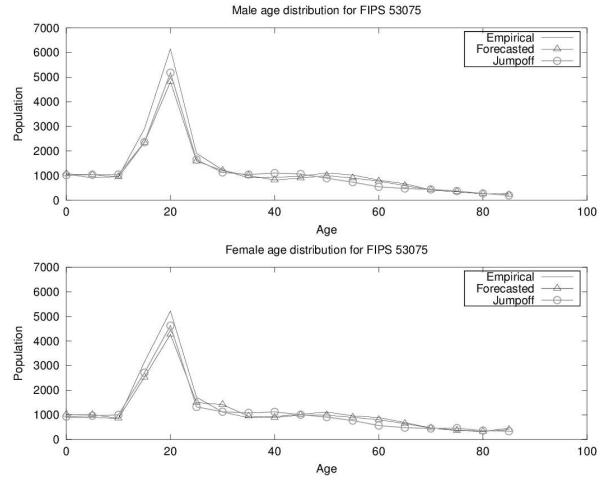


Figure 1.6: Whitman County forecast cross validation

1.5 Extensions to Wood's method for applied settings

There are a few extensions to this method that would make it more useful in applied settings. We sketch how these can be implemented or discuss plans for future research relating to them.

1.5.1 Forecasting to a control total

Often a population is forecast by projecting a total number, with the age/sex specific proportions "controlled" to that total. This approach is trivial to effect using Wood's Method as presented here. First, forecast the age/sex specific populations with WM, then derive age/sex proportion vectors for each forecast step by dividing each age/sex specific vector by its sum, and finally multiply the projected total forecast numbers by the corresponding age/sex specific proportions.

1.5.2 Forecasting vital counts

It is also important to be able to forecast numbers of vital events (births, deaths, net migration) along with population numbers. An estimate of these events should be easy to derive by using the various sections of the Leslie transition matrix corresponding to the vital event. To estimate births, set the migration diagonal and the survival sub-diagonal to zero, then multiply the age/sex population vector by the the remaining matrix. A directly analogous technique can be used for migration. For deaths, project the population using only the survival sub-diagonal, then find the difference of the starting vector and the finishing "cohort-wise." These techniques all depend on the choice of intelligent constraints in the original fitting routine, as otherwise the matrix components cannot be assumed to have demographic interpretations. Unfortunately, this approach is impossible with college populations, as the sub-diagonal no longer corresponds to survival; refining the model for college counties is ongoing. We have not tested this technique in practice.

1.5.3 Choosing input data

It is also important to consider how to choose input data to derive the Leslie transition matrices. Since Wood's Method incorporates an average of all transitions, if past trends do not reflect future trends, the forecasts will be incorrect. There are no hard

and fast rules for choosing training data, but there are two basic rules of thumb, which, unfortunately, can contradict each other in any given application. One, generally forecasters recommend going as far back for input as the forecast goes forward in time. Two, examine past data and look for articulation points where trends change direction, and avoid using data from before the most recent "elbow" in the data.

1.6 Conclusion

We believe that the forecasting method outlined above shows great promise for use in applied settings to forecast age and sex because it both streamlines the creation of Leslie matrices, and it flexibly incorporates previous demographic knowledge in the constraints.

Additionally, the method should be generally applicable to all populations using the above approach of combining theoretically based constraints with empirical data. While different populations will require different matrices and constraint sets, the basic applicability of the approach has been shown.

There are three main directions we see for exploring the application of this method. The first is extending it to more complex matrix models, including race, multi-regional models, and educational state models. The second direction will be in creating and refining constraints, as both limits on rates and subtle population interactions are better approximated and modeled; sensitivity analyses, partly based on the wealth of optimization theory and the approaches of Caswell will be extremely important in this line of investigation. Finally, the mathematical properties of the optimization method are not well explored.

2 Error analysis of Wood's Method

In this chapter, we examine the error dynamics for age-sex population forecasts generated using the version of Wood's Method presented in Chapter 1. Our primary datasets are counts by age-sex for most US county populations from 1970 to 2010; we also examine selected county projections from various state offices. We examine forecasts using 1990 to 2000 data for training, and 2010 as the cross-validation year, describing the error characteristics when 2010 forecasts are compared to empirical 2010 population counts. We also look at a new measure, at least for applied demography, of how well the age-sex "shape" of the forecasted population fits the shape of an empirical population. Finally, we examine techniques for generating prediction intervals. We conclude that Wood's Method is a suitable age-sex forecasting method, especially when one considers the ease of its application.

2.1 Approaches to forecast evaluation

It is quite difficult to find a "best practices" approach to evaluating a forecast methodology. We examine some of the challenges to forecast evaluation and present our approaches to evaluating WM forecasts. We do not use a single approach for evaluation, but use several methods that elucidate forecast performance in a number of different ways.

2.1.1 Challenges in scalar forecast evaluation

There are several challenges to evaluating forecasts of any domain, even when we restrict our attention to single dimension time series.

Formulas for confidence intervals derived from sampling theory do not apply since we have the full set of measurements, so one cannot rely on the familiar p-value as an evaluative metric.

ARIMA approaches such as described in (Chatfield, 2003) require stationary time series. However, for county age-sex populations in the US, the available time series

are fairly short, extending only from the 1970s for intercensal five-year age-sex data; once the time series is differenced enough to make it stationary there is very little data left with which to estimate parameters. This problem is exacerbated if the analyst uses less than the full time series in order to avoid polluting current forecasts with historical social dynamics that no longer apply. These issues prevent ARIMA from being widely used in applied demography.

Within-sample explanation of variance statistics, like R-squared and similar metrics, can be useful. However, within-sample error statistics will always overstate the accuracy of the forecasts, because the residuals are calculated using an interpolated series rather than a true forecast. In the case of WM, within-sample variation can only describe how well the linear model fits a nonlinear process. We use within-sample residuals to derive prediction bands via bootstrap populations (see below), but not as an evaluative tool.

The best approach is to examine compare forecasts to historical data for the same year, such as (Lee and Miller, 2001). However, this approach is limited because demography is so dependent on changing social conditions, and errors derived for one historical period may not apply to subsequent periods, even for the same population. (Remember demography's failure to predict the Baby Boom!) If, however, there is some consistency in social conditions over time for a population, this type of cross validation can be very useful. In any case, it is the approach best we have.

It would also be useful to compare our errors to other errors for age-sex population forecasts, but effecting such comparisons is surprisingly difficult as there is no equivalent of the Human Mortality Database (University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), 2012) for population forecasts. Current county age-sex forecasts are available for many states, but older forecasts for cross-validation purposes must be obtained by special request. The US Census Bureau has discontinued their state forecasts and has never generated county forecasts. We could only find one vendor with age and sex forecasts by county, Woods and Poole (http://www.woodsandpoole.com/), but their product costs approximately $1,000.00 for full county coverage over the USA, their older forecasts for cross-validation may not be available, and it might violate contractual agreements to use their forecasts from other customers. According to one state demographer, Woods and Poole forecasts are also generally too high, which she thought was due to their overuse of labor demand as a driver in their model.

In order to at least simulate comparison, it would also be useful to compare age-sex forecasts to a naive model, such as (Hyndman and Koehler, 2006) do for scalar time series. With scalar time series, the naive model is usually either a random walk

with drift, linear extrapolation, or some other simplistic automatic method. As WM is really the only automatic cohort-component method available, and thus the only "naive model" for age-specific forecasting, there is no way to run such comparisons.

2.1.2 Step ahead distance for cross-validation

We then must determine the best step-ahead distance for temporal cross validation. While it might seem obvious to use a five year step-ahead for our comparisons, because a five year interval forms a single iteration of the Leslie matrix projection process, we use ten years (two projection steps) for a two reasons. First, population estimates produced by the US Census Bureau for years ending in "5" are "intercensal estimates" derived by interpolating between censuses, with concomitant inaccuracies; population datasets for years ending in "0" are enumerated in the decennial census, so using two iterations provides as accurate a comparison as possible. Second, doing a ten year forecast is a typical task in an applied demography office, so an error analysis derived from a ten year step ahead distance should prove more useful for evaluation than an error analysis based on a single five year step.

2.1.3 Challenges in age-sex specific forecast evaluation

There are unique challenges to evaluating an age-sex population forecast due to its multidimensionality, beyond the challenges outlined above for scalar time series.

Traditional ARIMA time series theory is best developed for scalar time series, while our task is to forecast age and sex vectors; either we need to apply a dimension reducing technique like the singular value decomposition used in the Lee-Carter method (Lee and Carter, 1992), or draw on less well-developed theory in multi-dimensional time series analysis.

We can reduce dimension and evaluate forecasts by merely comparing the total populations summed over age and sex; while reductionist, it does give a basic measure of the utility of the method. We do not perform these comparisons as they do not apply to the particular strength of WM, which is easy age and sex detailed forecasting.

One can also derive errors cell-by-cell, computing the errors of all (age, sex, county) triples, and then summarizing these calculated errors using a typical measure like the MAPE or by a non-parametric percentile distribution (a histogram of errors). The problem with this approach is that different groups of cells may have different error structures; for example, the middle ages have less error in their forecasts than people in their early twenties. We can either stratify the groups we are analyzing (by age, sex,

and county population), or we can use a single strata for ease of comparison and lose some of the structure of the error. We use both approaches, but it is worth noting that determining appropriate strata is still more art than science.

Additionally, there is no canonical approach in the applied demography literature to computing errors describing the coherent "shape" of forecasts (its form irrespective of individual cell error or a constant shift of the population vector up or down). Computing such measures of "shape error" may seem silly, but WM could be used as part of a multistep projection algorithm to calculate age and proportions that are "controlled" to a total population forecasted by some other method, and in this case its ability to accurately forecast proportions would be very important. In this paper we use the Hilbert projective metric, which is used extensively in theoretical demography and population biology (Golubitsky et al., 1975; Wachter, 1986) to examine differences in the shape between forecasts and empirical age distributions.

2.1.4 Problem of changing social dynamics

It is important to remember that social dynamics can change in a population dramatically, with projections inevitably suffering due to forces exogenous to demographic processes. Unfortunately, there is no way to use purely demographic data like county age-sex populations to predict exactly how overall social conditions will change, even though causality in the other direction (from social conditions to demographic behavior) is somewhat better understood. In other words, the demographic data is "Y" data, and we do not necessarily have information on the "X" data.

As an example of changing social conditions influencing demographic patterns in a way that defies forecasting, consider Multnomah County, Oregon, which experienced very low migration in the 1980s due to the collapse of Oregon's timber economy and Portland's reputation as a rather unpleasant post-industrial city; since 1990, however, Multnomah county has had very high migration due to high-tech economic development and Portland's recent emergence as a nationally important cultural center. In Multnomah County's case, evaluations derived by comparing a 2000 forecast (with a 1990 jump-off) to 2000 empirical data would be less useful than comparing a 2010 forecast (2000 jump off) because the social changes of the 1990s would not be part of the training data in the former.

Interestingly, in Portland's case, there were hints in the 1960s and 1970s of Portland's future as one of the most desirable cities in the United States; these include an innovative and enlightened planning community at both the city and the state levels, the high cost of nearby California real estate leading to migration in order to buy homes,

the strength of Nike as a shoe and apparel company, the basis for a technology industry in Tektronix, and a slow migration of "creative class" types described in (Florida, 2002). Additionally, while there was some discussion locally of the decreasing timber inventory that Oregon depended on before the 1980s, the collapse of the industry due to globalization, de-industrialization, increased efficiency, and more stringent regularization was a surprise as well. However, none of these trends would have been strong enough to stand up to scientifically rigorous testing, even though they turned out to be defining for the region's demography; this sensitivity to social conditions and nonlinearity points at the difficulty of forecasting social dynamics using scientific methods.

Additionally, the presence of large group quarters in a county, like universities, military installations, and prisons, makes population dynamics especially volatile because the populations of these institutions depend on government policies and budgets, which are far more variable than the demographics of the conventional household population. For example, Newport County in Rhode Island housed the Cruiser and Destroyer Fleet for the U.S. Navy until 1973; the population age profile for 1970 shows a typical military bulge for males in their early twenties, which utterly disappears in 1975. Such a change is impossible to predict using demographic input alone (and thus by using WM), and any cross-validation tests based on 1960 to 1970 training data would show a large error in that county.

Somewhat ironically, when forecasters try to compensate for these social dynamics by incorporating "expert opinion," they may introduce errors in population forecasts by decreasing the "discipline of the model" and applying their intuition incorrectly. We do not know of any way to overcome the inherent difficulty of social dynamics in forecasting, but it is important to note it when considering error in population forecasts.

2.1.5 Error metrics used in evaluating WM

Considering the difficulties outlined above, we use several different error metrics in attempt to understand the error characteristics of WM forecasts. The most important metrics are the Mean Absolute Percentage Error (MAPE) and the Median Absolute Percentage Error (MedAPE) statistics. The MAPE is the industry standard measure, and the MedAPE provides a a robust version of the MAPE. Both the MAPE and MedAPE also provide a simple ranking measure that is easy to understand and use in making choices between forecast methodologies.

We have stressed the absolute error rather than the signed error throughout this article as it gives a measure of "wrongness" which is easy to interpret and visualize; for the evaluation of a forecasting method, absolute error provides a simple metric – closer to

zero is better, farther away is worse. Signed errors give a more nuanced insight into the interaction of a forecasting method with the underlying social dynamics, but they are less useful for simply evaluating the method. We report the Mean Algebraic Percent Error (MALPE) in some of our tables, but we do not examine it in detail. We plot the signed errors, present boxplots, and discuss their distribution briefly below.

We also examine the quantile distributions and histograms of the absolute proportional error in order to understand the error characteristics in a richer way than is possible with single measures of central tendency. In applied environments, the worst-case error and the 95% error are often as important as the mean or median error when evaluating a forecasting method or when applying population forecasts to real world tasks like developing budgets. (Note that there is a possible source of confusion when we talk about percentiles of percent error below – remember that "95% error is 150%" means that 95% of the forecasts have an error less than 150%.)

We also measure the fit of the age sex shape, irrespective of its magnitude, using the Hilbert projective metric (HPM). This measure is hard to interpret directly, so we show some examples of low or high HPMs along with their percentile values and a plot of forecasted versus empirical proportions. We present a detailed discussion of this metric below, as it may be unfamiliar to many applied demographers.

There is a recent literature on scaled variations of the MAPE, like MAPE-R and SMAPE (for example, (Campbell, 2002; Hyndman and Koehler, 2006)). We do not use these measures for a number of reasons. Most importantly, MAPE-R and similar error measures are opaque to interpretation. Furthermore, while advocates of MAPE-R claim that it corrects overly pessimistic evaluations of forecasts, we would rather be conservative in our evaluation of our forecasts. Additionally, if we apply MedAPE (or MAPE, for that matter) to all the "contestants" in a comparison, they serve as fair measures because of their relative values. Finally, as MAPE is a purely formal computation, it cannot be more or less "accurate" than another measure, so supposed improvements are not as easy to validate as their proponents claim.

We also do not use unscaled measures of error like the Root Mean Square Error (RMSE), as they would be completely nonsensical in comparing errors between county populations which can vary by almost five orders of magnitude (Loving County, Texas, had a population of 82 in 2010, while Los Angeles County CA had a population of 9,818,605).

Finally we would like to note that in order to best compare forecast methodologies, the cost of inaccuracy should be calculated using an age-sex specific "loss function" such as described in (Bryan, 2000; Hough and Swanson, 2006), and a forecast methodology should be chosen to minimize the total loss due to inaccuracy. A loss function

measures the economic implications of errors in a forecast; for some ages, forecast error has higher economic implications than other ages, and the forecast methodology that best accounts for these important segments of the population should be chosen. As an example of different loss functions, consider state versus federal government in the US. States fund educational and prison services, which are used mostly by populations in the younger ages, so a misforecast of these ages would have larger consequences than a misforecast of other age groups, and this age-specific weight would be reflected in the state level loss function. Retirees, in contrast, are supported by the federal programs of Social Security and Medicare, so misforecasts of aging populations would have more of an impact on the Federal government's loss function. The MAPE or the MedAPE imply a degenerate loss function in which all proportional errors are weighted the same. We do not explore loss functions any more in this paper, but it is important to note them as generalizations of the simple error measures used here.

2.2 Cross validation results

In this section we present details of our error analysis. Note that we use the inter-censal data for July 1, not Census Day April 1, to keep consistent with middle of the decade estimates. These data disagree with a lot of easy to find numbers like Wikipedia.

2.2.1 Basic error measures for 10 year cross validation forecast

In Table 2.1, we tabulate a summary of overall errors in the cross validation forecast when compared to actual population numbers in 2010. With a median error of 7.8% the forecast is quite reasonable. The worst case (100%) and "almost-worst-case" (90%) percentage errors are more extreme. Our 90% errors are comparable to the 90% errors in (Rayer, 2008) for counties with a 10 year horizon, even though their errors describe total population projection error rather than age-sex-specific forecasts.

We also compared our forecast "internally" to see how much difference in forecast error at each age/ sex cell is due to using the demographic transition matrix approach as opposed to using a naive multiplicative projection. First we calculated the percent change between 2000 and 2010 for each county. Then, for each county, we multiplied all age/ sex cells by this single percentage growth number and stored this naive county age/ sex forecast. Then we calculated the absolute percentage error for each age/ sex category by subtracting these cells from empirical data for 2010. The summaries of

Percentile	% Error (pooled)	% Error (pooled & naive)	% Error >10k	% Error <10k
0%	~0.0%	~0.0%	~0.0%	~0.0%
5%	0.7%	1.3%	0.6%	1.1%
50%	7.8%	13.9%	6.9%	12.5%
90%	25.4%	33.5%	21.4%	38.9%
95%	34.5%	40.2%	28.4%	52.6%
100%	1324.5%	538.5%	41.6%	1324.5%

Table 2.1: Percentiles of Percentage Absolute Error for Age-Specific Forecasts

these errors are found in the third column of Table 2.1. We see a marked reduction in age/ sex errors by using WM generated transition matrices instead of simply projecting each age/ sex category with the total percentage growth.

In Figure 2.1, we graph the median, 90%, and maximum percentage errors, discriminated by age and sex, for US counties above a population of 10,000 in 2000. This graph packs a lot of meaning into a small space, so we will explain it in some detail. On the x-axis, we show each 5 year age group of the population for which we are calculating errors. Each point on the graph has two features – a shape (triangles for males, circles for females) and a color (red, blue, and green for whether the point represents the value for 50th percentile of MAPE, the 90th percentile, or the maximum MAPE for that age group). On the y-axis, we have the MAPE for each percentile group represented by the points, on a logarithmic (base 10) scale. For example, consider the blue circle for twenty year olds; this means that the median MAPE across for the 3120 counties is about 11%, with 1560 counties having a MAPE less than 11%. The green circle for twenty year olds indicates the 90th percentile of MAPE for 20-25 year old males is about 25%; i.e. 90% of the counties have a MAPE less than 25% for the 20-25 year old male group. The maximum MAPE (the 100th percentile) is about %1100 for this group. Color coding is necessary in order to in order to put three different levels of information on the graph (sex, age, and error level).

The graphic is based on the ideas of Jacques Bertin (Bertin, 1998) and his work on the semiology of graphics, using two "semiological axes" for percentile (color) and sex (shape) to complement the two mathematical axes of x and y. Bertin theorizes that we can perceive simultaneous differences in the following six categories of marks in an image: shape, size, color hue, color value, color intensity, and texture. Only shape, color hue, and texture should be used to distinguish between qualitative categories because value, size, and intensity all imply an ordered quantitative scale that confuses the viewer if it is not appropriate; we chose to use color instead of texture due to the

available software for generating graphs. Bertin is a cornerstone of cartographic theory, occupying a place like William Cleveland does for statistical graphics. For those without access to color displays or printers, perceiving the gestalt line for each each error level shouldn't be too difficult as each line is plotted higher as the percentile error increases; for example the top line with triangles in Figure 2.1 refers to male error at 100th percentile, the next line with triangles downward represents the 90th percentile error, etc. Percentiles are well ordered by definition, and in these graphs none of the percentile error lines cross other percentiles; for example, no 90th percentile lines cross 50th percentile lines.

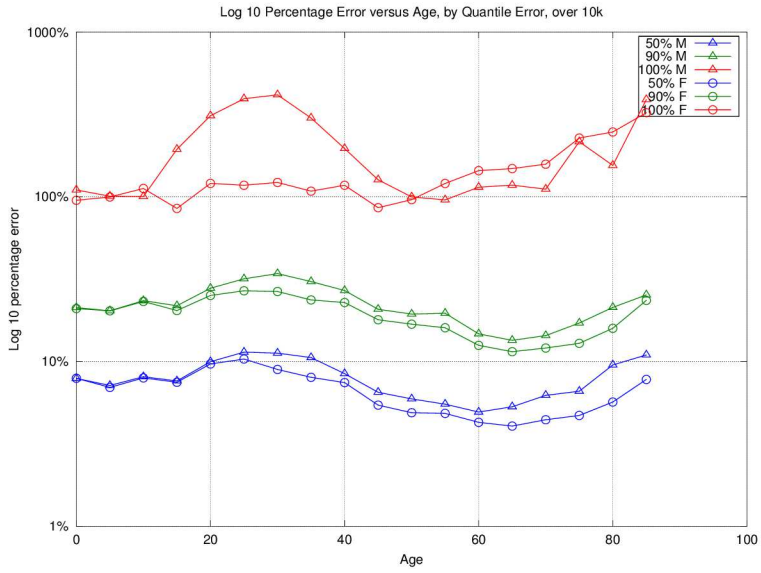


Figure 2.1: Percentile error by age, counties above 10,000

We can see a distinct pattern, with forecasts of people in their early 20s and 30s and above 80 showing more error than other ages. Forecasts for people in their 40s and 50s have the least error. Forecasts of men have consistently higher error than forecasts of women, though the shapes of each gender's error are similar. The exception is that during their twenties, the maximum error is much higher for men, possibly due to the disproportionate impact of men on counties with prisons and military bases, causing the group quarter volatility discussed above. The high percentage error for the aged may be due to small numbers biasing the error calculation. For people in their 20s in of both genders, however, populations are far more mobile than at other ages, more likely to be in college group quarters, and are more difficult to track using administrative data

like driver licenses due to sloppy behavior in updating addresses. Table 2.2 contains the numbers plotted in Figure 2.1.

	MAPE by percentile, counties over 10K					
	Males			Females		
Age	50%	90%	max	50%	90%	max
0-5	7.9%	21.2%	110.2%	7.9%	21.0%	95.3%
5-10	7.2%	20.3%	100.9%	7.0%	20.3%	99.8%
10-15	8.1%	23.4%	100.4%	8.0%	23.2%	112.5%
15-20	7.6%	21.8%	194.5%	7.5%	20.4%	85.0%
20-25	10.0%	27.9%	310.8%	9.7%	25.2%	120.7%
25-30	11.4%	31.8%	394.4%	10.4%	26.9%	117.6%
30-35	11.2%	34.2%	416.2%	9.0%	26.6%	122.2%
35-40	10.6%	30.6%	301.0%	8.0%	23.7%	108.2%
40-45	8.4%	27.0%	196.5%	7.5%	22.8%	117.6%
45-50	6.5%	20.8%	126.9%	5.4%	17.9%	85.9%
50-55	5.9%	19.4%	99.8%	4.9%	16.9%	96.2%
55-60	5.5%	19.7%	95.8%	4.8%	16.0%	120.6%
60-65	4.9%	14.7%	114.5%	4.3%	12.5%	144.3%
65-70	5.3%	13.4%	117.8%	4.1%	11.5%	148.5%
70-75	6.2%	14.4%	111.4%	4.4%	12.1%	157.9%
75-80	6.6%	17.1%	216.0%	4.7%	12.9%	227.4%
80-85	9.5%	21.3%	155.2%	5.7%	15.9%	247.7%
85+	11.0%	25.4%	389.4%	7.8%	23.5%	324.0%

Table 2.2: MAPE by percentile and sex, counties > 10,000 in 2010

In Figure 2.2 we graph the same metrics for counties smaller than 10,000, showing the tabulated numbers in Table 2.3. Note that the scales are different in these two graphs in order to show the age specific error shapes clearly. Overall, errors are higher than for counties with populations less than 10,000. The maximum errors for both males and females shows a less coherent shape than for larger counties, likely due to small sample size problems.

It is critical to remember that small populations have more volatile populations characteristics and higher forecasting errors due to both purely formal considerations and to social dynamics "on the ground." Formally, small denominators magnify errors when percentages are calculated. Additionally, and still formally, percentages fluctuate more

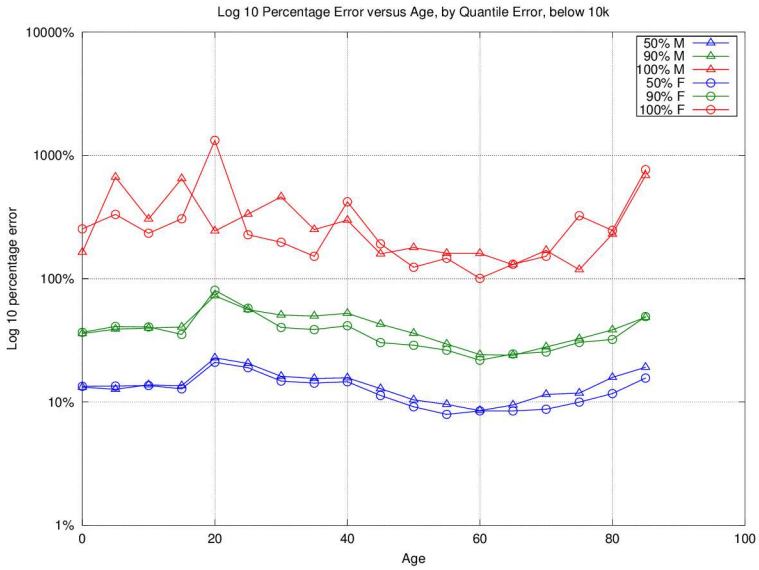


Figure 2.2: Percentile error by age, counties below 10,000

widely under the influence of "quantization," as forecasts are rounded to make whole numbers before being compared. For example, if a forecast for a small population group is 4.6 and the actual population is 4, the absolute error will be 1-5/4 = 20%, which seems quite extreme but is only due to the fact that demography behavior occurs at whole numbers.

To understand further why small population measurements are more volatile due to social reasons, consider that households – not individuals – are often the unit of demographic choice and behavior for migration; families move together and make migration decisions together, thus increasing the impact of a single decision in terms of individual population numbers. For example, in a county of 200 people, if four households move out in a year (as "net" migration), that may be a change of 6% in the total population from only four decisions, and an even higher percentage when the scope of the ages is considered – all those moves will likely be in ages zero through 15 and twenty to fifty year olds, the family ages.

It is useful to compare these graphs to boxplots of the proportional errors in Figure 2.3. Notches represent a robust estimate of the uncertainty about the median. Points that lie outside 3 times the interquartile range (IQR) are represented by circles, and points between 1.5 and 3 times the IQR are represented by pluses. A few outliers have been left off the plots. We see that low population counties have a similar error pattern

	MAPE by percentile, counties under 10K					
	Males			Females		
Age	50%	90%	max	50%	90%	max
0-5	13.1%	35.4%	163.6%	13.3%	35.5%	152.7%
5-10	12.6%	36.5%	180.1%	13.3%	39.2%	177.5%
10-15	13.7%	38.6%	185.6%	13.6%	40.5%	233.5%
15-20	13.3%	38.2%	485.3%	12.6%	33.8%	305.8%
20-25	22.6%	68.4%	243.8%	20.7%	75.6%	1324.5%
25-30	20.4%	55.2%	333.3%	18.8%	55.2%	176.3%
30-35	16.0%	47.9%	327.1%	14.4%	39.0%	118.7%
35-40	15.4%	49.2%	250.4%	13.9%	38.3%	151.9%
40-45	15.3%	50.2%	297.9%	14.6%	41.0%	420.3%
45-50	12.7%	40.4%	158.8%	11.2%	29.4%	138.8%
50-55	10.3%	35.5%	178.7%	9.0%	27.1%	123.8%
55-60	9.5%	28.7%	160.5%	7.9%	26.4%	146.5%
60-65	8.4%	23.9%	160.3%	8.4%	21.6%	99.7%
65-70	9.4%	23.6%	129.7%	8.3%	23.5%	131.0%
70-75	11.3%	26.7%	169.6%	8.5%	24.9%	151.9%
75-80	11.6%	31.6%	117.7%	9.8%	30.3%	324.1%
80-85	15.8%	37.9%	230.1%	11.5%	31.9%	246.3%
85+	18.9%	46.5%	685.2%	15.4%	48.2%	372.8%

Table 2.3: MAPE by percentile and sex, counties < 10,000 in 2010

by age, with greater spread than higher population counties. We also see a consistent bias toward overforecasts, except at the more elderly ages. Underforecasting older people may be due to nonlinearly increasing lifespan, or to more rural counties providing retirement destinations but being weighted disproportionally in county numbers when not normalized by population.

Using two population size pools for county forecast evaluation is justified for a number of reasons. Counties over 10,000 comprise over 98% of the population of the US, even though they are 22% of the total number of county jurisdiction (percentages calculated using our slightly incomplete version of county data, which is described in Chapter 1). More importantly, the loss function for accurate forecast is much larger for larger population counties because larger counties have large impacts on social programs and economics. By examining the entire USA, we look at many small counties,

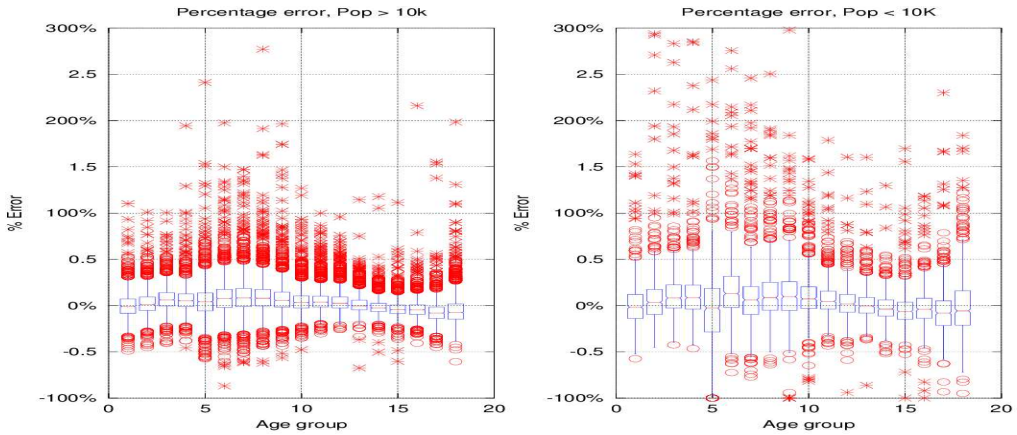


Figure 2.3: Error boxplots for all counties

and even smaller age classes, which is why we do most of our analysis for populations greater than 10,000.

2.2.2 Error relationship to population characteristics.

We graphed the percentage error for 40-44 year old females with respect to several population characteristics in an attempt to see if the error bears a relationship to easily calculated population attributes. For independent variables, we used four different numbers: the total population in 2010, the average of the growth rate ("R") from 1970 to 2010, the standard deviation of R over the same period, and the percent growth from 1990 to 2000. We include a LOWESS line (Cleveland, 1979, 1981; Cleveland and Devlin, 1988) plotted using a public ally available LOWESS package (Burkey, 2013) found on the Matlab file exchange, with a smoothing factor of 0.5. We chose to plot only 40-44 year old females as a means to "scalarize" the error data because middle-aged women are generally the least volatile age/ sex group and will thus highlight trends more clearly. The plots (Figures 2.4, 2.5, 2.7, and 2.6 respectively) do not show a dramatic relationship between the percent error and these population characteristics, but by examining the LOWESS line in each, we can see some relationships. We did not run any statistical tests because of the apparent (and surprising to the author) weak relationship. Interestingly, when we examine plots like these for the Hilbert projective

metric below, we see a much stronger relationship.

We see a slight slope toward zero in the error trend with respect to county population in Figure 2.4. Note that we have graphed this as percentage error, without taking absolute values, showing the variance of the error bettor. This relationship is consonant with accepted wisdom that larger populations are easier to forecast with less error. The LOWESS line has the most consistent trend between 10,000 and 100,000; at lower and higher populations, there are fewer sample points, which might affect the LOWESS line. The LOWESS line is at about +15% error at 10,000 decreasing to almost zero at 100,000. Most of the errors are above zero, showing a bias toward over-forecasting.

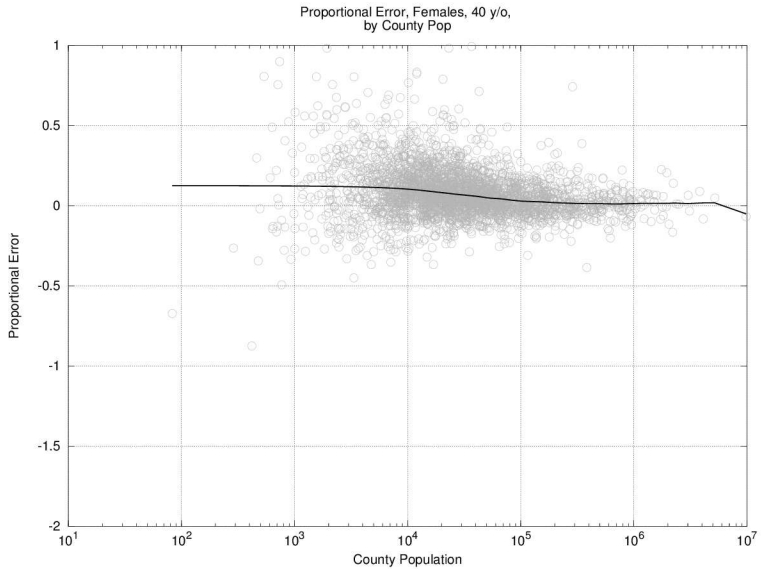


Figure 2.4: Proportional error by county population

We see a mild U-shaped curve with respect to average R (Figure 2.5), with the least error when R is close to zero. Faster changing populations, whether growing or shrinking, can be more dynamic and harder to forecast. Growth is more likely due to migration in the USA than to natural increase; migration fluctuates more than natural increase from year to year, especially with respect to employment demand which follows economic cycles. Note that these are average values for R, reflecting the dynamics of the county since 1970. The errors are still mostly above zero, reflecting a tendency to over-forecast in our run, which, as we discussed earlier, may be due to the higher growth in the 1990-2000 decade in the USA biasing the test forecasts for 2000-2010.

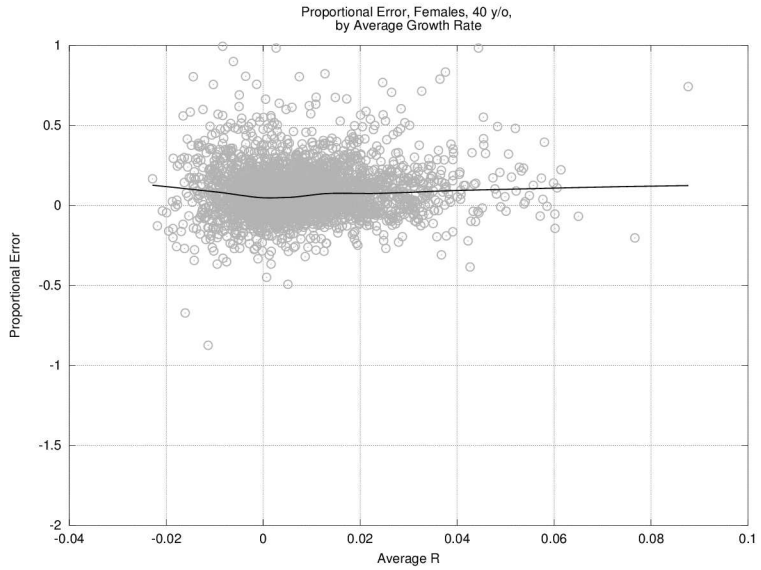


Figure 2.5: Proportional error by average county growth rate

With respect to growth in 1990-2000 (Figure 2.6) we see a fairly distinct downward slope as growth from negative values (population loss) to zeros, then a much slower decrease as growth rises. There are a couple of reasons likely for this dynamic – many of the rural counties have only experienced population loss recently, so forecasts would tend to be too high when based on historical data, reflected in the line above zero. Also, depopulating counties also tend to be smaller and more rural, which makes forecasting more difficult for reasons described above. Interestingly, there is a distinct articulation point at zero, where growth goes positive and after which the proportional error decreases much more slowly.

We also see a rise in proportional error corresponding to an increase in the standard deviation of R calculated from 1970 to 2010 (Figure 2.7); it is quite reasonable that greater variability in population change would yield greater forecast error. The downward slope to the right of the data cloud is likely due to the sparsity of data for those high values and the one high standard deviation outlier on the right with an average R of about 0.1 but a very low proportional error.

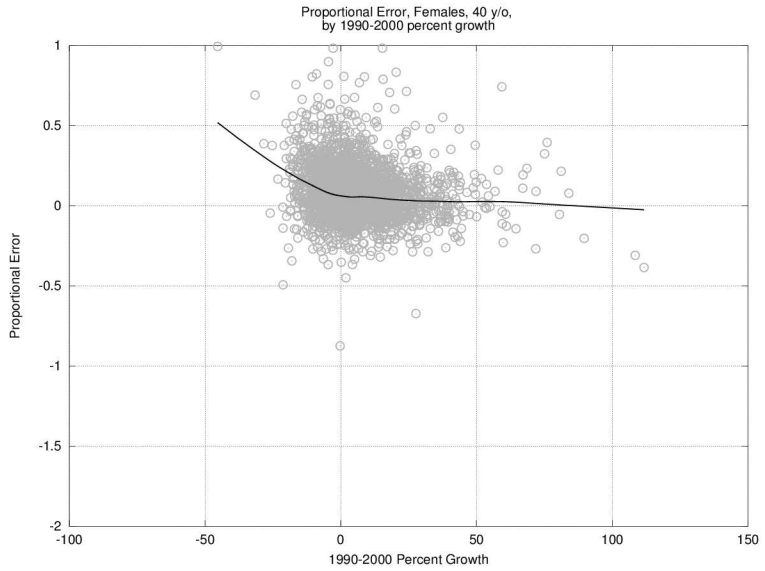


Figure 2.6: Proportional error by county percent growth rate

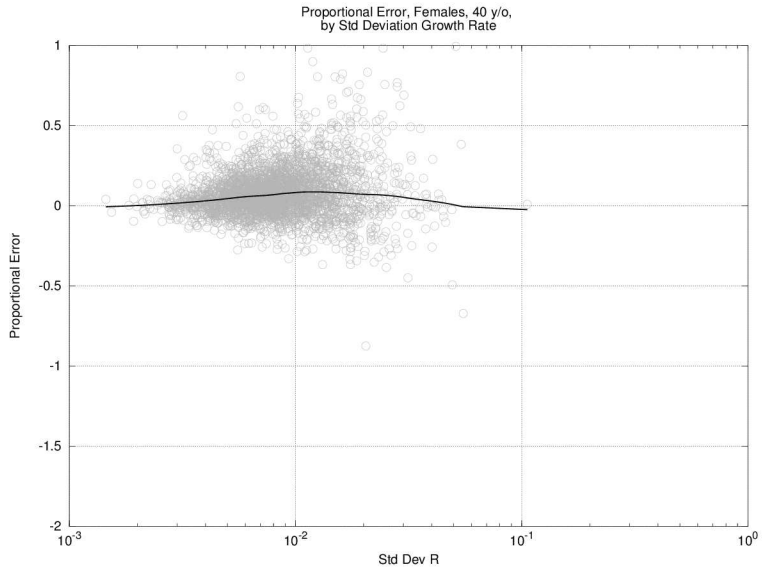


Figure 2.7: Proportional error by county standard deviation of growth rate

2.2.3 Comparison to other forecasts

Smith et al.

When MAPEs and MedAPEs are tabulated similarly to (Smith et al., 2001) in Table 2.4, we see the same error patterns as describe in that work, with average and median errors decreasing monotonically with the size of the population. Note that we used two single age-sex classes – males age 20-24 and females age 40-44 – in order to have a single number for comparison. These two age classes also present two extremes of population dynamics, in that young males are much less demographically predictable than middle aged females, yet they both show the same error dynamics with respect to population size with a difference only in magnitude.

Population in 2000	N	MAPE (m20)	MAPE (f40)
0 - 4,999	275	44.8%	27.6%
5,000 - 14,999	746	20.6%	15.5%
15,000 - 24,999	532	15.9%	13.0%
25,000 - 49,000	639	13.6%	9.9%
50,000 - 99,000	390	10.7%	8.6%
100,000+	525	10.2%	6.7%

Table 2.4: MAPE by population

Like the errors presented in (Smith et al., 2001), our errors follow a "U" shape with respect to percent growth over the previous decade, see Table 2.5. Note that we added a -10.0% to -5.0% bin in order to clarify the pattern, even though this category was not present in (Smith et al., 2001). The LOWESS line in Figure 2.6 reflects this bathtub curve as well.

Other population characteristics may yield more clear relationships, especially if the characteristics are economically or socially significant (e.g. the presence of group quarters, recent employment history, large concentrations of certain races, ethnicities, etc). There may also be a more clear relationship if different time intervals are used over which to calculate the population characteristics. However, such an investigation is beyond the scope of this dissertation as it is an empirical sociological problem, and may be highly specific to each county.

There is more variation in proportional error for smaller counties as can be seen in figures 2.1 and 2.2, along with more variability at the problematic ages of young adulthood and the aged. The data for these graphs has both males and females pooled.

Percent growth 1990-2000	N	MAPE (m20)	MAPE (f40)
>-10%	194	37.3%	24.9%
-10% - 5%	289	20.5%	14.8%
-5% - 0%	607	15.9%	12.1%
0% - 10%	1229	15.1%	11.0%
10% - 25%	576	16.4%	11.8%
25% - 50%	182	16.2%	11.4%
50% - 100%	28	18.6%	15.7%
<= 100%	2	26.3%	34.7%

Table 2.5: MAPE by percentage growth

Oregon, Washington, and California state forecasts

In Table 2.6, we show summary statistics for our forecasts and forecasts performed by demographers in California, Oregon, and Washington, obtained by personal communication. California's forecast was released in 2003, Oregon's forecast in 2004, and Washington's forecast in 2002; our WM forecasts are performed with data that would have been available by 2003. All of the forecasts include age and sex detail. California's forecast has age detail up to 100+ years old as the open interval, so we summed all ages from 85 and above to correspond to our open interval of 85+. We did not examine the methodology of these forecasts as part of our comparison. Our results compare favorably to all three forecasts, especially when the ease of performing a WM forecast is considered.

Measure	CA	WM (CA)	OR	WM (OR)	WA	WM (WA)
0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5%	0.8%	0.5%	0.6%	0.8%	0.8%	0.6%
50%	7.7%	5.9%	7.0%	8.3%	7.8%	8.2%
95%	41.6%	29.0%	31.2%	37.8%	28.1%	27.6%
100%	2400.0%	765.0%	135.0%	161.8%	100.0%	84%
MAPE	14.0%	9.7%	10.1%	12.2%	10.3%	10.7%
MALPE	-7.5%	-3.0%	-2.2%	6.3%	-1.4%	-4.4%

Table 2.6: State and WM forecast error comparison

2.2.4 Measuring errors in proportional age distributions

The Hilbert projective metric (HPM) has been used extensively to prove theorems in stable population theory because it is able to compare the shape of two vectors irrespective of their magnitude; see (Golubitsky et al., 1975; Wachter, 1986) for its application to convergence in stable population processes. We use it to measure how well a forecast retains the relative age-sex proportions of the population controlled to total population. The proportional shape of the population forecast is important for two main reasons. First, proportional relationships between age classes are often crucial for social dynamics, especially in the calculation of dependency ratios and marriage markets. Secondly, age-sex specific forecasts are often calculated in two steps – first, the total population is forecasted using time series or econometric methods, then the ages are allocated from that total; determining the age distribution for allocation would be a very likely application of WM.

We apply the formula in (2.1), with p_i as the true empirical data, and q_i as the forecasted age sex vector. We calculate the HPM of each forecast with the true age sex distribution. Note that we are measuring the divergence between the age-sex vectors for both sexes, and not breaking them up into male and female; since males and females might have very different dynamics in counties with large group quarters such as prisons or military bases, measuring the sexes separately would be worthwhile exploring in the future.

$$D_H(p,q) = log\left(\frac{max(p_i/q_i)}{min(p_i/q_i)}\right) \tag{2.1}$$

Representative counties for given HPM

Table 2.7 shows quantiles of the distribution of all counties cross-validated in 2010. This table is difficult to interpret – intuitively, we want to say the best match on age proportions is about 40 times better than the worst (4.75759/0.11485 = 41.4), but it is difficult to know what "better" means. This difficulty is the reason we investigate some typical counties for each percentile of the HPM below.

Figure 2.8 shows a histogram describing the distribution of the HPM between forecasted population and actual population for all the counties analyzed. The x-axis contains the HPM, ranging from zero to five. Each bar represents the proportion of counties having a given HPM between actual and forecasted population, with that proportion shown on the y-axis. We see a distinct skewed shape, quite likely reasonably fitted by

Percentile	HPM
0% (minimum)	0.11485
5%	0.22270
20%	0.30976
50% (median)	0.44368
80%	0.69541
95%	1.18006
100% (maximum)	4.75759

Table 2.7: County HPM distribution

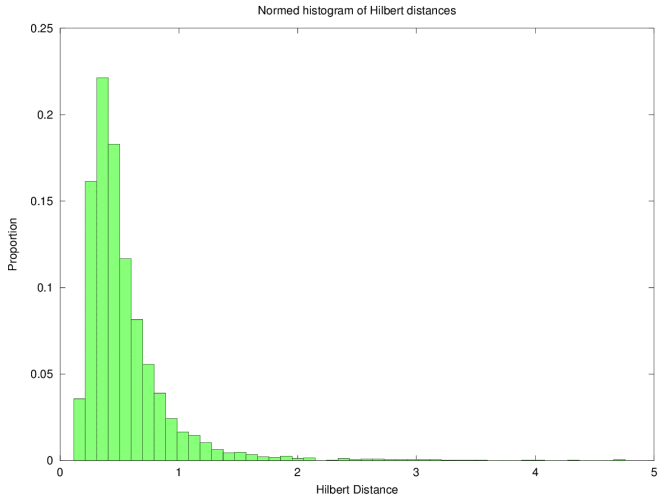


Figure 2.8: County HPM histogram

a lognormal distribution.

It is worth, therefore, examining the difference in age structure between actual and forecasted populations for a range of HPM values. Figures 2.9 through 2.17 show county age distributions and forecasts from a range of HPM percentiles, with males on the top graphic and females on the bottom; we will examine each in turn. Remember that we are comparing the difference in the shapes of the demographic forecast, not the total difference.

Reading three period age-dynamics graphs

In the discussion below, we present graphs showing age profiles from three decades

(1990, 2000, and 2010) on the same graph in an attempt to understand the dynamics of the county and how it might affect the Hilbert distance and forecast error, for example Figure 2.9. Young demographers may not be familiar with plots like these, so we want to spend a moment to explain how to interpret them.

These graphs are basically age pyramids turned horizontally, with colored lines and line markers used to distinguish between periods on the same age axis. The x-axis represents the bottom cutoff for the age class at which the population is being graphed. The blue lines with triangles at each age represent forecasted age specific populations in 2010 from a 2000 jumpoff year. The green lines with circles represent the actual population in 2010. The gray dashed line represents the jumpoff population in 2000; this age profile is important to see in order to visualize the dynamics as it moves forward to the forecasted and actual populations.

Relationships between aging and migration are particularly easy to pick out in these graphs. We see "aging in place" – when residents do not move out so their cohort population moves through the age classes – by watching humps in the age shape move horizontally at ten year distances (on the x-axis) with each new decade. Net migration into the county can be inferred by picking a point on the age-shape graph, moving 10 years right (reflecting the aging of the cohort), and determining whether the population is higher for in-migration or lower for out-migration. Sometimes we see "standing waves" at certain ages, when the hump or trough remains at the same age for different decades, showing that local migration is heavily age dependent and consistent; counties with large colleges have large consistent humps at the 15-25 ages, with young people moving in and then leaving 5 to 10 years later. Rural and suburban counties have similar dynamics in reverse, with a "standing trough" during these ages as young people move out for education, military service, and employment, but more middle age people move in to raise families. In certain counties we see both "standing waves" and "aging in place" in the same population; this can happen, for example, if there is a college locally that draws lots of young people who move in and out again shortly but also a large baby-boom (born in 1945 to 1960) population that ages in place. We can also watch the shape of a cohort change as it moves forward in time, usually spreading out and becoming less peaked as the effect of a particular demographic event becomes dissipated.

Intergenerational dynamics may be visible as well. Fertility echoes can sometimes be seen by comparing peaks or troughs and similar features about 25 years earlier (a generation). Populations which show a fairly flat age distribution show a "well rounded" social dynamic, with jobs, education, and housing appropriate to the multiple age groups; smaller counties often show a more peaked age structure because their social context is better suited only to a specific age group.

While these graphics reflect important social dynamics, like all of demography, aggregate numbers may hide lots of activity. This aggregate masking is especially true for migration, in which high levels of gross migration may cancel out to show moderate levels of net migration.

Example HPM counties

Hillsborough County, FL (FIPS 12057). Figure 2.9 shows a plot of the forecast and empirical population for Hillbsborough County, Florida, the best HPM fit in the 3120 counties. In this example, the forecast undershot the actual population in 2010 by quite a bit, but retained its shape quite well. Hillsborough County had a population of 1,229,226 in 2010 Census, and a 4.8% MedAPE over all ages. This county has a "well rounded" population, probably due to the presence of the major city of Tampa, with a Baby-Boomer hump at ages 45 to 55, their children forming an echo in the late teens. Otherwise we see a fairly flat distribution until the higher mortality ages above 60. However, even though the magnitude of the difference at any given age is high, the peaks and valleys in the 2010 forecasted population correspond very closely to the peaks and valleys in the empirical population, with the difference in population between empirical and forecast falling more or less evenly over all the ages from 0 to 45-50. For ages above 50 the forecast population matches the empirical population very closely. Better matching at older ages makes sense, as ages 0 to 50 typically have higher migration than older ages, hence are more sensitive to volatility in economic conditions and are less likely to show consistent aging in place.

By examining population change between 1990 and 2010 in Figure 2.10, we can see that Hillsborough County has active population dynamics. The baby-boomers – the hump in the blue line at ages 25-35 in 1990, moving and growing to the right in 2000 (green) and 2010 (red) – experience both aging-in-place and high migration. Their children can be seen by examining the red line for population in 2010, with peaks at age 45 and at 20, about a generation apart. The peakedness of the population distribution attenuates as it moves from 1990 to 2010. The low HPM for this forecast is not an artifact of a static population, but rather of consistent age dynamics. In this case, WM reflects the social dynamics closely for this county, even though it does not predict the higher overall growth seen in 2000-2010. This county also has a high population, which helps with the stability and the low HPM.

Payne County OK (FIPS 40119). Figure 2.11 shows a plot of the Payne County, Oklahoma forecast, which had a population of 77,350 in the 2010 Census. The forecast in the 25% best HPM of the 3120 counties, and has a 4.9% MedAPE over all ages. Payne County is about 50 miles north of Oklahoma City and 50 miles west of Tulsa.

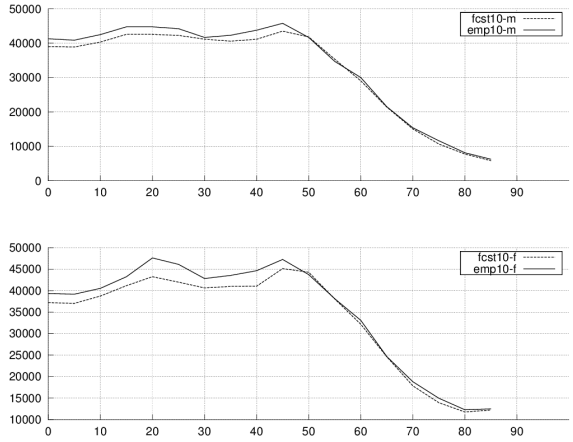


Figure 2.9: Forecast age structure of minimum HPM county

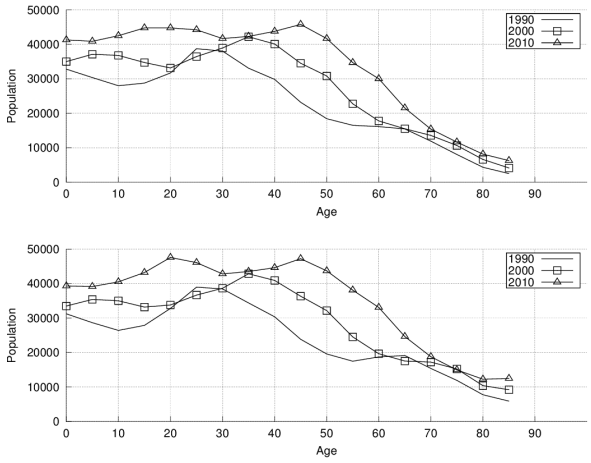


Figure 2.10: Three-decade age structure of minimum HPM county

Stillwater is the biggest city in the county, with Oklahoma State University, and the large peak of population at age 20-25 reflects the college specialization of the county. The proportional age structure of the county is well forecasted by WM, except that certain ages are missed (such as age 20 females) while some are very accurate (such as age 10 males), leading to a less well-formed forecast profile and a worse HPM than

if the errors were consistent across ages. The ability of HPM to measure proportional shape is shown in the fact that Hillsborough County (above) has a markedly better HPM than Payne County, even while the overall MedAPE is only slightly better.

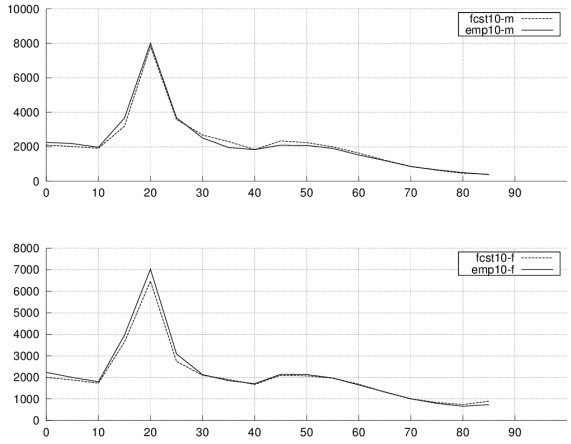


Figure 2.11: Forecast age structure of 25%-tile HPM county

The three decade graph for Payne County, OK, in Figure 2.12 shows that the population in their early twenties has been rising, but these young people leave the county and do not age in place. There is a slight Baby Boom hump that becomes evident in 2000 and 2010 but that is overshadowed by the spike in college aged population in 1990, presumably when the last of the boomers were attending college.

Stutsman County, ND (FIPS 38093). Figure 2.13 shows a plot of the Stutsman County, North Dakota forecast. Stutsman County had a population of 21,100 in 2010, and the forecast is the median HPM over our 3120 counties, with a 4.9% MedAPE over all ages. This county is located in the middle of North Dakota, containing the county seat of Jamestown, which also has most of the county population (about 15,000) and manufacturing jobs as well as supporting local agriculture. We can see that the Baby Boomer peak is well forecasted but that the dynamics of the younger ages is less well captured; this dynamic is consonant with lower migration of older people who are more likely to age in place facilitating easier forecasting. (Anecdotally, fewer young people stay in rural counties than they used to, as good headquarter jobs move to metropolitan areas with the increased consolidation of the 20th century.)

In Figure 2.14, we see that the Baby Boomer wave progresses clearly in Stutsman County, experiencing some out-migration but retaining its shape as it ages in place,

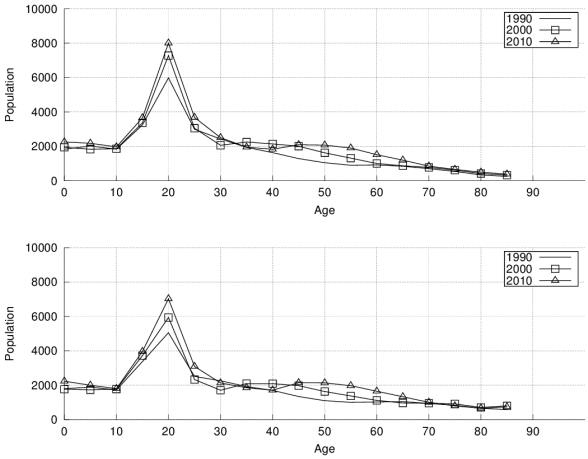


Figure 2.12: Three-decade age structure of 25%-tile HPM county

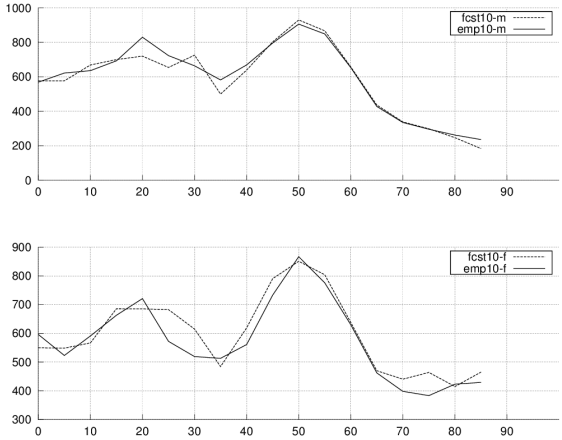


Figure 2.13: Forecast age structure of 50%-tile HPM county

which explains its good forecast fit at these ages (45-60). The echo is much less well defined, however, due to the substantial outmigration of young people, especially females. Note that the MedAPE for Stutsman County is similar to the MedAPE for Hillsborough County, while the HPM is noticeably worse; this difference is due to the bad shape fit from 25 to 35 years old, and again, shows the importance of the HPM in

fully understanding forecast error.

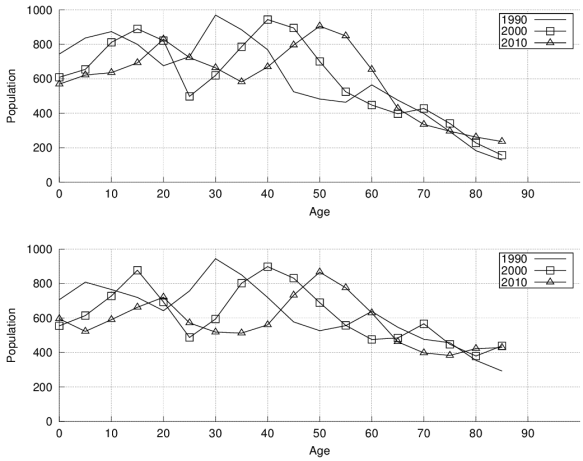


Figure 2.14: Three-decade age structure of 50%-tile HPM county

Hildago County, NM (FIPS 35023). Figure 2.16 shows a plot of the forecast for Hildago County, New Mexico, with a population of 4,894. Hildago County is the 75% worst HPM county match, with a 13.8% MedAPE. Hildago County has a high poverty rate and an low-value agricultural economic base of ranching. It has been declining in population for the last half century except for 1970-1980, but the decline was especially precipitous in the last decade, concentrated in working age people and their children. Because of this sudden – and age-localized – decline, we see large errors in the forecast of under-twenty ages for 2010, even while the rest of the forecast is fairly good. This age localized error impacts the HPM more than if the error were distributed evenly.

In the three decade age dynamics graph in Figure 2.16, we see a steep decrease in young people by examining each line during the younger years and seeing the decrease as we go from 1990 (blue), 2000 (green), to 2010 (red). Interestingly, Hildago County seems to retain attractiveness for people over 45. This precipitous decline in the young age but consistent population in the older ages contributes to Hildago County having shapes with high HPM.

Hemphill County, TX (48211). (Figure 2.17), with a population of 3,807 in 2010, and a MedAPE of 18.7% in the forecast, presents an interesting contrast to Hildago County, while still being a small western county. It is located in the Texas panhandle, historically dependent on ranching, but recently oil and gas extraction has become

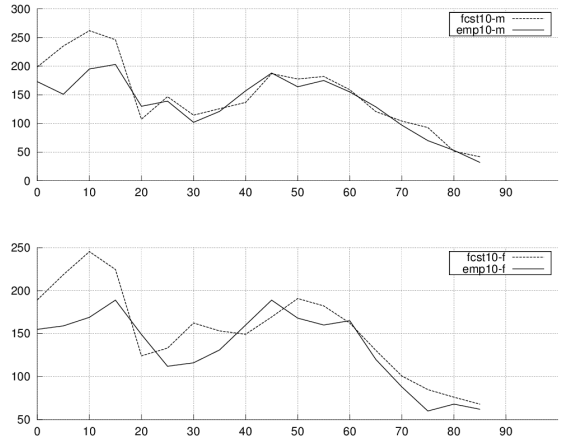


Figure 2.15: Forecast age structure of 75%-tile HPM county

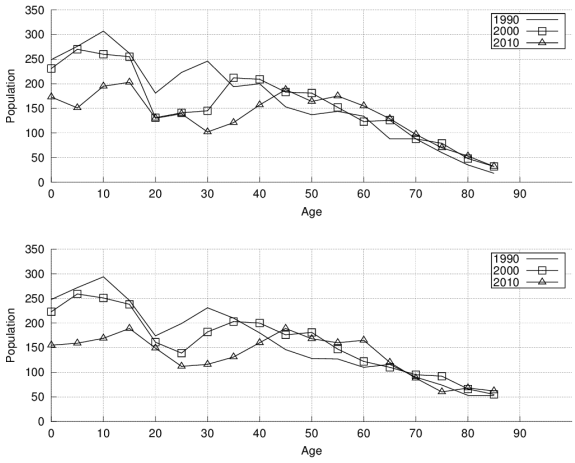


Figure 2.16: Three-decade age structure of 75%-tile HPM county

dominant. Here we see a decrease in population during the training period which predates the energy boom, 1990 to 2000, but there is an influx from 2000 to 2010 which would not have been available to the training routine. Note that forecast for ages over 50 is relatively good, due to the decreased influence of potentially volatile migration at the older ages, with all the error happening in the family and working ages. The

misforecast of younger people is the determining factor in the high HPM.

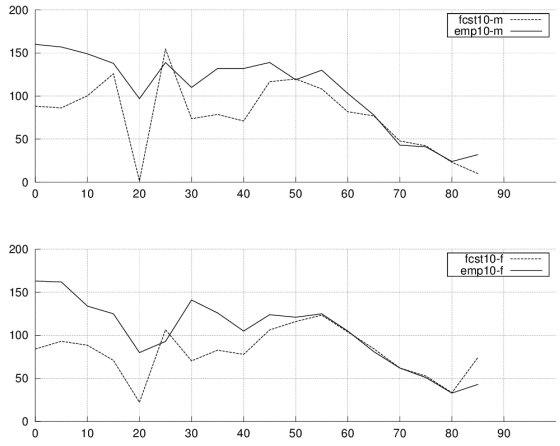


Figure 2.17: Forecast age structure of 100%-tile HPM county

In Figure 2.18, we see a rise across all ages. The rise in the higher ages was forecast correctly as discussed above, primarily as aging in place with moderate out-migration – watch the peak that starts with 1990 at ages 25-35 and moves rightward with the green and red lines. This aging in place is especially clear in the female population, which would be less affected by energy industry labor migration. However, the rise in the younger ages, we see, was not captured in the forecast.

HPM as a function of county demographic characteristics.

We now consider how HPM varies with total population, growth rate, recent growth, and growth variability, in parallel with our description of proportional error summarized in figures 2.19 through 2.22. In the case of HPM, we do not need to summarize using only a single (and somewhat arbitrary) age and sex classification since HPM is a scalar, holistic measure of the age shape.

Figure 2.19 shows the clear relationship of lower HPM with larger populations, similar to Figure 2.4, except with a much clearer trend for HPM.

Figure 2.20 shows the HPM depending on the average growth rate of the county from 1970 to 2010, similar to Figure 2.5. With HPM, we see a much more pronounced effect

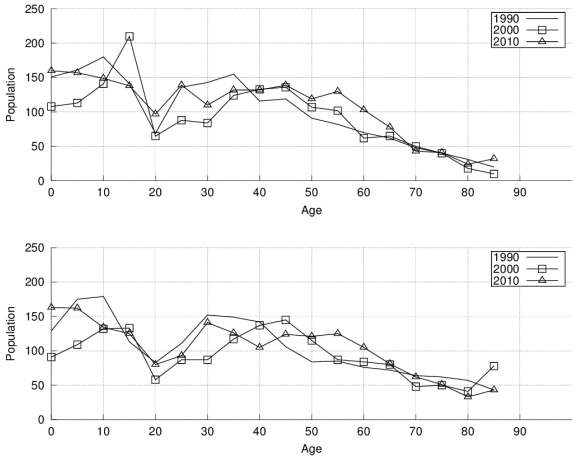


Figure 2.18: Three-decade age structure of 100%-tile HPM county

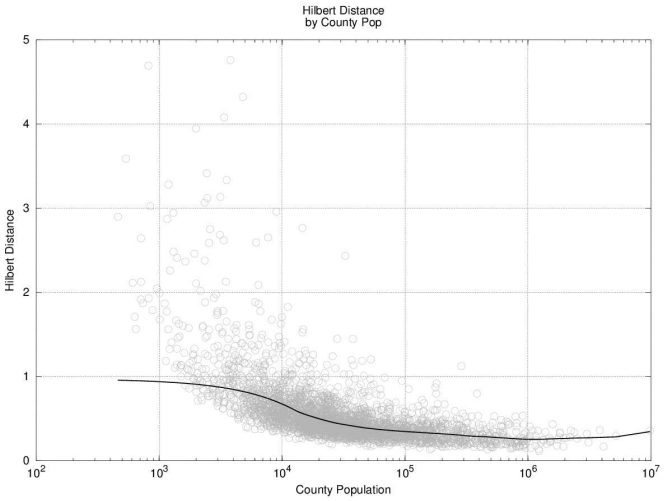


Figure 2.19: HPM by county population

on counties experiencing depopulation (with negative average growth); this trend is attenuated at zero growth, but then HPM rises slowly as counties experience higher and higher growth. This is similar to the bathtub curve for error dependent on population growth shown in Figure 2.5 and Table 2.5; in other words, more dynamic populations are harder to forecast, especially if they are shrinking.

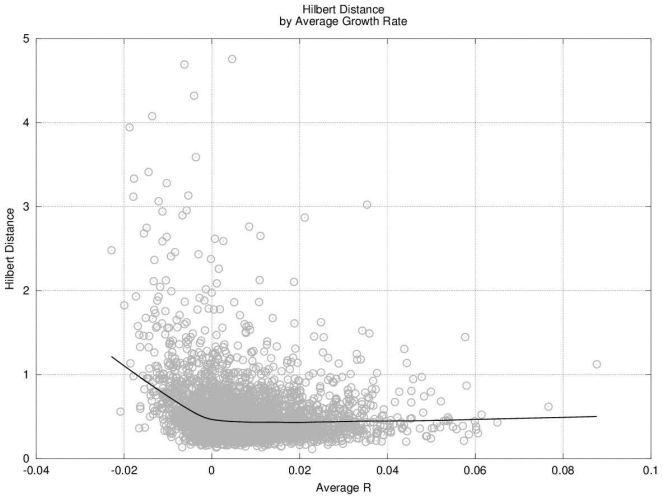


Figure 2.20: HPM by average county growth rate

Plots of recent growth and its relationship to HPM can be seen in Figure 2.21. This graph is very similar to Figure 2.6, which describes proportional error in terms of recent growth. Just like with average growth, we see that depopulation leads to less accurate forecasts, as does high growth, with a sort of bathtub curve and much more of an effect at negative growth.

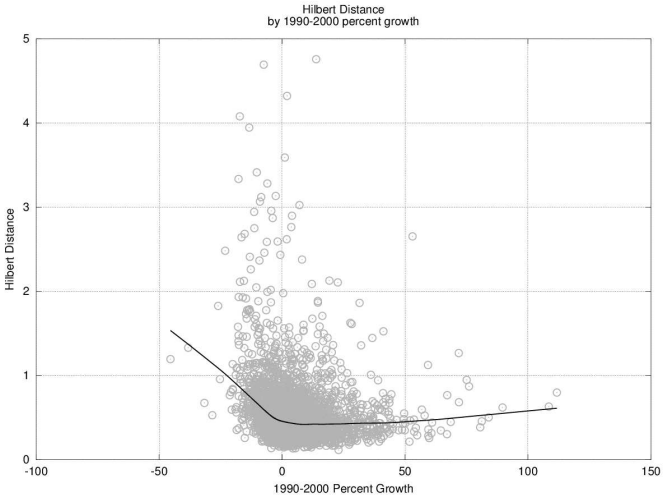


Figure 2.21: HPM by county percent growth

Finally, we see a clear increase in HPM with variability in county growth as measured by standard deviation of R from 1970 to 2010, in Figure 2.22 consonant with Figure 2.7.

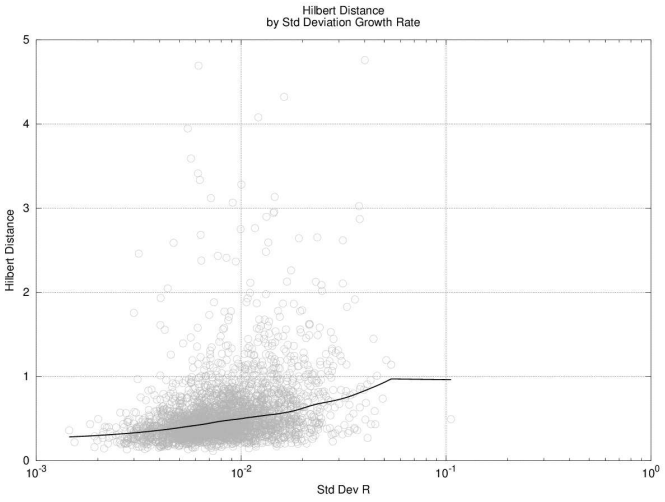


Figure 2.22: HPM by standard deviation of growth rate

Interestingly, the relationship of demographic characteristics is the same for both HPM and proportional error at a given sex and age classification, except there is a much more pronounced increase in error with HPM than with MAPE's for a single age sex classification. This additional clarity in the graphs might be due to the fact HPM is able to capture the dynamics of the entire population.

HPM Conclusion

As the HPM is a new measure for applied demography, if not for population biology and formal demography, we consider these results to be exploratory. However, accurately forecasting the age and sex shape is useful in its own right, as often age and sex distributions are controlled to total forecasts that are derived using econometric or other models. Additionally, the age/ sex distribution reflects local social dynamics in a way that total population does not. We find that HPM usefully describes the accuracy of proportional forecasts and is consonant with other forecast error measures.

2.3 Prediction bands

It is important in any forecast to provide reasonable bounds on projection numbers. Unfortunately, the usual armature of sampling theory fails us here again, even while it whets our appetite for a comfortingly familiar confidence interval. There are two main approaches that seem appropriate – deriving empirical confidence intervals and simulating forecasts from bootstrap perturbations to training data. We will discuss each in turn.

2.3.1 Empirical confidence intervals

The first approach we considered was to develop empirical confidence intervals as advocated by (Rayer et al., 2009). While an elegant idea, it is not obvious how to pool populations before calculating confidence intervals for each age-sex category. One could use base-10 logarithmic cuts, for example, but that categorization is arbitrary. One might also use a single pool, which is straightforward, but such an approach is overly pessimistic as it assumes homogeneity of errors across counties large and small, even though large counties have much less error when pooled without any smaller counties, as we saw in the above error analyses, especially figures 2.19, 2.4, and Table 2.4. Using a single pool is trivial, however, and we have almost done it already as part of our evaluation above – just look up the appropriate percentage error in tables 2.2 or 2.3, then apply it to get a count by which to adjust the age sex forecast. Percentages for intervals besides 10 years would be trivial to develop as well.

2.3.2 Prediction bands analogous to bootstrap calculations

An alternative approach was developed to generate simulated populations that have similar error patterns to within-sample fits, and then run bootstrap simulations from these. This procedure is somewhat ad hoc and only tentatively justified, but it seems to perform decently well.

In order to avoid the problems described with pooling above, we must rely only on a single population's data, which creates problems generating a range of possible dynamics. Our solution is to assume that the WM is the best linear approximation to population dynamics for a population, and that any additional variability for that population is due to effects that are captured by the distribution of the residuals. This residual distribution is used to generate simulated populations, which are then used to create new projection matrices. These projection matrices are used to create multiple

forecasting traces, which are summarized by quantiles of 5% and 95%, forming the confidence bands. This approach is loosely based on Freedman and Bickel's work on bootstrapping linear regression (Bickel and Freedman, 1981; Freedman, 1981).

More formally, let $w(P_{.,.})$ be a function that generates the WM projection matrix A from an M by N population matrix $P_{.,.}$; this "function" is merely the Woods Method algorithm described in Chapter 1, given mathematical notation for ease of exposition. The dots in the subscripts of P signify all possible M age-sex positions and all N year positions. For example, to represent all age-sex cells for a single year t, we would write $P_{.,t}$; For example, to represent all years for male 20 year olds only (found on row 4), we would write $P_{4,.}$. With no subscripts, P represents the entire population.

$$A = w(P) \tag{2.2}$$

Then calculate a fitted population $P^{'}$:

$$P^{'}_{.,1} = P_{.,1} \tag{2.3}$$

$$P^{'}_{.,2} = A \cdot P_{.,1} \tag{2.4}$$

$$P^{'}_{.,3} = A^2 \cdot P_{.,1} \tag{2.5}$$

Then calculate a matrix of residuals R by elementwise subtraction:

$$R = P - P^{'} \tag{2.6}$$

Then calculate the mean and standard deviation of this residual matrix for each age row-wise. These means and standard deviations will be vectors of dimension 36 because there are 36 age-sex cells in a single time step of population.

Let $v(\mu, \sigma, t)$ be a function that generates a random matrix of size 36 by t, with row-wise mean μ and column-wise standard deviation σ. This function, like $w(P_{.,.})$, represents an algorithm coded in Octave using more succinct mathematical notation. The function (a variant of Matlab's `mvnrnd()`) inputs the variance/ covariance matrix of the residuals (call this covariance matrix S) and their age-specific row-wise means (μ), takes the Cholesky decomposition of S into $L * L' = S$, and calculates a random matrix of residuals as $M = R * S + \mu$, where R is a random normal matrix with mean = 0 and standard deviation = 1. This matrix M has the same covariance and mean as the original matrix of residuals R, at least to the degree that those residuals can be modeled as normally distributed.

Let $P''_{m,t,i}$ be a population matrix with the third dimension representing number of iterations, m the number of age-sex categories, and t the number of time steps in the population. Generate this simulation matrix by adding the simulated residuals to the fitted population for each iteration i:

$$P''_{.,.,i} = P'_{.,.} + \alpha v(\mu, \sigma, t) \tag{2.7}$$

Finally, calculate i projections, one for each iteration of the simulated population, and take the 5%, 50%, and 95% quantiles across each age to give a low, median, and high forecast.

To widen the confidence bands, we experimented with introducing a variance multiplier, α, to make the simulated bootstrap inputs more variable than they would be based on the above procedure. This approach may work in practice, but it should be applied on a county-by-county basis, and the approach needs more analysis to be justified. Note that this procedure assumes that the residuals are distributed normally about the modeled population. No statistical tests have been performed to substantiate this supposed normality, but the histograms look sharply peaked and symmetric so we assume they are "close enough" to normal to use.

We would include an example calculation, but the mass of subscripts would only cloud the algorithm.

2.3.3 County examples of confidence intervals

Below we show several counties in Washington state, paralleling the case study counties in Chapter 1, with the empirical population in 2010, a cross validation forecast based on 1990-2000 input data, and prediction bands calculated using the algorithm outlined above. As usual, the x-axis represents age classes, and the y-axis the age-specific populations, whether forecasted or empirical. The lines with triangle markers represent the forecasted population and the lines with circles represent the actual population at these age classes. The dashed lines represent the prediction bands calculated using the method described above.

While the prediction bands cannot capture variation due to change of social dynamics within a county, they can capture variation in counties which retain their social dynamics over both the training and the forecast periods. We see that the prediction bands encompass the forecast shape well, but they do not reflect the population magnitude as well as they do the overall age/ sex shape of the population. This is reasonable in that to change age/ sex specific dynamics in a county would require a change in county

social characteristics – for example, the building of a new university or highway, or the closing of a military base – but a change in magnitude may be due to a socially more transitory effect like lower employment in the same array of job types present in the county. We see such a complex social change in Spokane County, below, with likely higher enrollment in the universities and an increased proportion of people in their young twenties. We see an example of a simple magnitude change in Clark County, which experience much higher unemployment during 2000-2010 than in 1990-2000 due to the 2007 recession, but which experienced no structural changes in the social environment.

The prediction band approach might be especially useful in a several stage forecast process. Since the band approach presented here only accounts for linear trends, complex and age-specific as they may be, these bands should be used as a first cut at confidence bands, then expanded at selective ages by consulting county experts for their opinions on possible new local trends.[1] As these bands are only linear, they should be considered a lower bound on forecast inaccuracy.

Regional city center (Spokane County, FIPS 53063)

The prediction bands capture the empirical population to some degree. There is clear peakedness at age 20 in 2010 which is not captured, probably because it is due to an influx of young people, a new trend for Spokane County. (This new migration pattern is likely due to increased college enrollment, at universities like Gonzaga, Whitworth, and Eastern Washington University.) This change in direction surprised our office, but it might have been easier to see coming if we were more familiar with local dynamics.

Large metropolitan center (King County, FIPS 53033)

The bands capture King County's actual population in 2010 very well, probably due to there being very few structural changes in 1990-2000 compared to 2000-2010; the King County economy has had the same economic and housing drivers for two decades.

Suburban satellite (Clark County, FIPS 53011)

Clark County's 2010 population is not captured in the bands at all, due to the fact that it was very hard hit in the housing bubble collapse and experienced subsequent high unemployment. The county basically provided a release valve for affordable housing

[1] Unfortunately, population forecasting happens in a politically and economically charged context, so "experts" often see new trends – or the lack thereof – based only on wishful thinking.

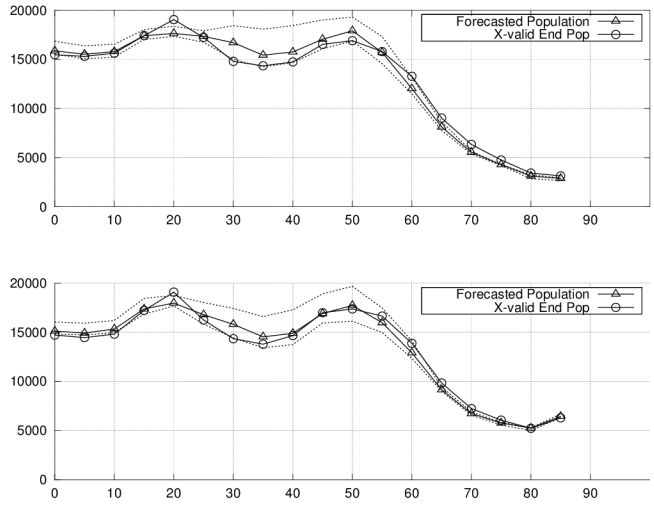


Figure 2.23: Spokane County forecast with prediction bands

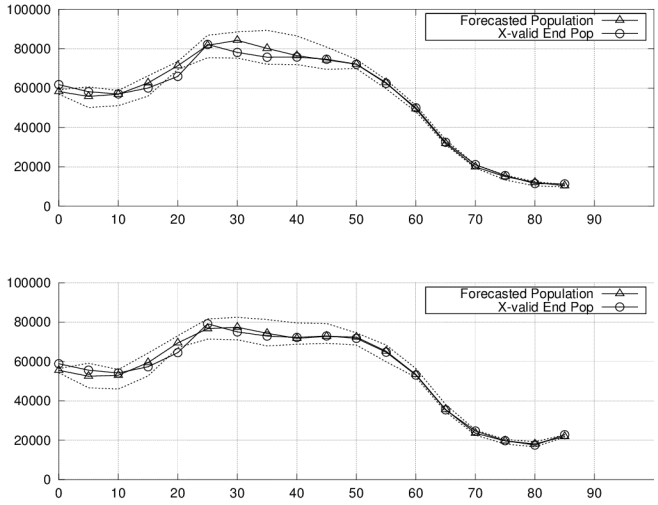


Figure 2.24: King County forecast with prediction bands

for the Portland Metro region, a release which was unnecessary after the bubble burst.

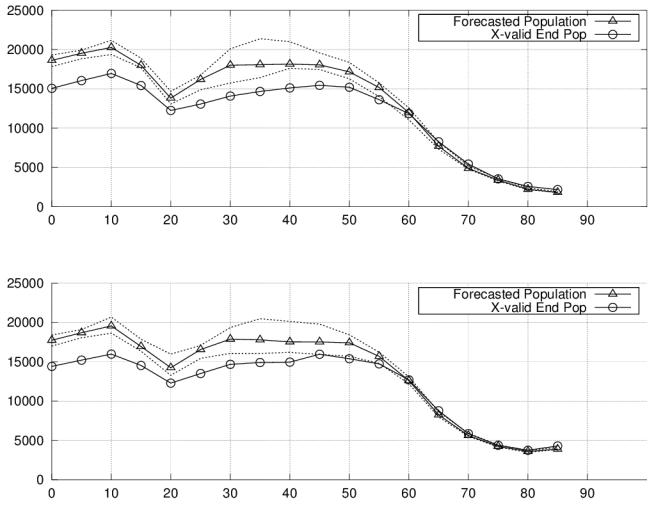


Figure 2.25: Clark County forecast with prediction bands

However the forecast bands retain the shape of suburban demographics very well, in which young adults move out for jobs and education, with a concomitant population decrease at ages 20-25, but people move in to raise families and show increased population over 30 and in the grade school ages.

Retirement/ amenity (Clallam County, FIPS 53009)

The bands capture the shape well, and even cover the empirical population in the volatile young ages. The retirement in-migration has curbed somewhat more than trend, possibly due to the recession.

Agricultural region (Yakima County, FIPS 53077)

We see a slowdown in population growth, probably due to a slower economy and the recession beginning in 2007, so the empirical forecast is not covered by the bands. However, the bands conform to the shape of the forecasted well, consonant with the lack of structural changes in Yakima County.

University specialization (Whitman County, FIPS 53075)

Whitman County provides an interesting case, in that enrollment trends drive the

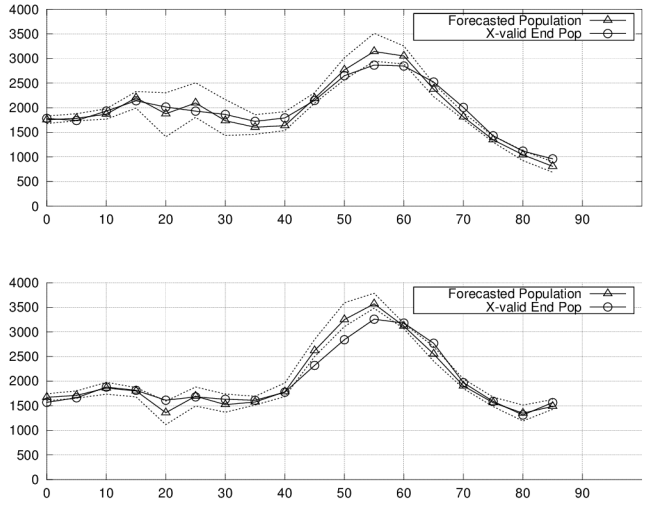


Figure 2.26: Clallam County forecast with prediction bands

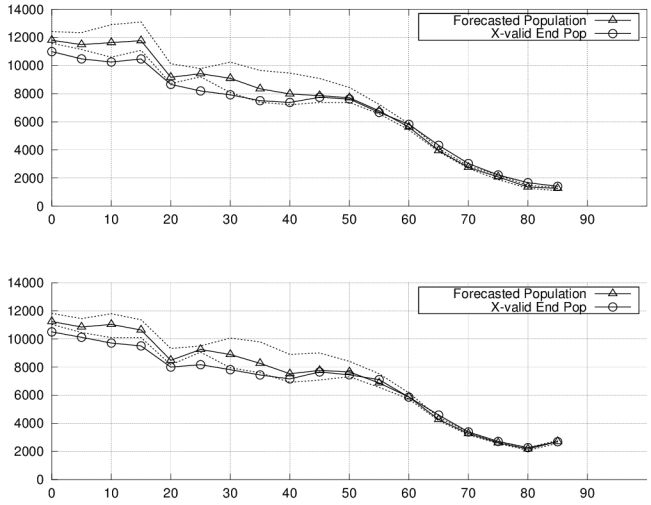


Figure 2.27: Yakima County forecast with prediction bands

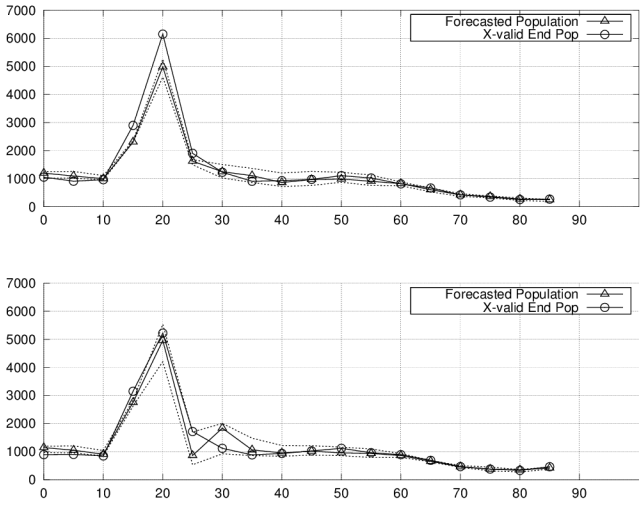


Figure 2.28: Whitman County forecast with prediction bands

forecast, and the out of trend increase in enrollment for males makes the forecast inaccurate. Happily, the forecast bands for females age 25 - 30 cover the spike in 30 year olds that was incorrectly forecast.

2.4 Conclusion

We conclude that the The errors for WM are quite acceptable when analyzed in several different ways, and that WM is a useful approach to forecasting by age and sex.

When the error is analyzed using MAPEs and MedAPEs by age and sex, these metrics are comparable with other forecasting methodologies. When analyzed using the Hilbert Projective Metric, the median HPM error shows that WM forecasts reflect actual age sex proportions well; there are no comparable studies to which to compare the our results for the Hilbert Projective Metric.

Interestingly, we see that errors can be due to an inaccurate forecast of either shape or quantity. Shape errors are measured best by HPM, but quantity errors are best measured by MAPE. Shape errors may reflect a qualitative change in the county social structure that is not transitory, but purely quantitative errors in the forecast probably reflect more transitory dynamics.

We also see that useful linear forecast bands can be generated using Wood's Method, though they tend to underestimate error in WM forecasts and should be used in conjunction with more expert knowledge.

Finally, we have begun to examine the influence of social conditions on Wood's Method errors. We see a higher errors due to negative or high positive growth and time series variability on both the HPM and the female 40-45 scalar measure. Due to local knowledge we were able to speculate on the causes of error in some typical Washington counties, but further work to elicit the causes of specific errors with respect to social drivers is promising.

3 An alternative formulation of Wood's Method

In this chapter, we present an alternative derivation of Wood's Method, in contrast to the formulation presented in subsection 1.2.4 above. Our new formulation is both easier to follow and more generally applicable than the formulation presented in Chapter One. This new formulation also allows us to better explain the use of constraints in fitting a model to demographic population data, and we show examples of the most important constraints for WM in Section 3.3

However, in the other chapters we use the implementation based on the original formulation; furthermore, the alternative method as presented in this chapter does not currently scale with more than a few age categories and will require significant computer programming changes to make it viable for the 36 by 36 matrices that form the basis of our application.

3.1 Mathematical preliminaries

Our new formulation will use several results and operators that are not part of the standard undergraduate linear algebra curriculum, but which have become almost classic in the last 25 years. These include the Kronecker product, the *vec* and the *reshape* operators, the use of vector norms, and the solution to the Sylvester equation $Y = AXB$. We sketch them here for reference in our reformulation of Wood's Method below, but the interested reader is encouraged to consult the excellent work (Golub and Loan, 2012) for further detail. We also use a somewhat nonstandard subscripting approach which we explain in detail in this section.

The Kronecker product $A \otimes B$ is defined on page 27 of (Golub and Loan, 2012) as multiplying the entire matrix B once for each element of A. Note that it is not commutative.

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix} \tag{3.1}$$

The *vec* operator is defined on page 28 of the same work and examined in detail on pages 710-711. This operator takes a matrix and "stacks" it as a single column vector,

$$C = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{bmatrix} \tag{3.2}$$

$$vec(C) = \begin{bmatrix} c_{11} \\ \vdots \\ c_{m1} \\ \vdots \\ c_{1n} \\ \vdots \\ c_{mn} \end{bmatrix} \tag{3.3}$$

The *reshape* operator is also defined on page 28. This operator takes a vector v, an integer r for the number of rows, and an integer c for the number of columns, and creates a matrix by reshaping the input vector to have the specified number of rows and columns. This operator is an inverse of *vec*. *vec* can also be written as $reshape(C, mn, 1)$.[1]

$$c = \begin{bmatrix} c_{11} \\ \vdots \\ c_{m1} \\ \vdots \\ c_{1n} \\ \vdots \\ c_{mn} \end{bmatrix} \tag{3.4}$$

[1] Using a term like "$reshape(c,m,n)$" may feel more like computer programming than mathematics, but if the authors of Golub and Loan (2012) – a canonical work in the field of matrix analysis – see fit to use it, we believe it can be used in this dissertation as well.

$$C = reshape(c,m,n) = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{bmatrix} \tag{3.5}$$

Vector norms are examined on pages 68 through 70 of (Golub and Loan, 2012). Norms measure the distance, broadly defined, between two vectors. The 2-norm corresponds to the Euclidean distance between two points, and it is the most important vector norm in this exposition. It is defined as follows:

$$\|x\|_2 = (x_1^2 + \ldots + x_n^2)^{\frac{1}{2}} = (x^T x)^{\frac{1}{2}} \tag{3.6}$$

where x^T is the transpose of the vector x. The 1-norm is defined in terms of absolute value, as follows:

$$\|x\|_1 = |x_1| + \ldots + |x_n| \tag{3.7}$$

We also depend crucially on a result (12.3.9) on page 711 regarding solutions to the Sylvester equation, where B^T signifies the transpose of B:

$$vec(AXB) = (B^T \otimes A)vec(X) \tag{3.8}$$

This identity translates a three term matrix expression AXB into a product of a matrix $B^T \otimes A$ and a vector X.

Finally, note that we will use a somewhat specialized subscript notation to refer to submatrices of a matrix. As is standard, indices separated by a comma indicated the given elements by row and column, so for example, $Z_{2,3}$ denotes the element of Z in the second row and third column. A dot (".") indicates that all elements for that that dimension are included, so for example, $Z_{.,3}$ denotes the entire third column of Z, a vector. An ellipsis with two dots separates a beginning and ending index (like "2..4"), indicates that all rows or columns between the begin and end are included. For example $Z_{2..4,1}$ refers to the second through fourth rows of Z in the first column only, yielding a vector.

This subscript notation applied to a matrix can yield submatrices ($Z_{1..3,2..4}$), vectors ($Z_{.,2}$), or scalars ($Z_{3,3}$). Variables can also be used in a subscript, as well; for example $Z_{k-1,.}$ refers to the second to last row in Z if Z has k rows. This notation lets us describe Wood's method much more succinctly.

3.2 Rederivation of Wood's Method, unconstrained

We now present the reformulation of Wood's Method. This derivation does not account for constraints, which are treated in the sequel.

3.2.1 Derivation

There are k states, corresponding to age-sex groups and there are observations on m time periods, presented in a k by m dimensional data matrix Z.

Let A be a k by k dimensional demographic forecasting matrix, to be estimated by the method.

Using dot notation, a single age specific projection step can be written as $Z_{.,t+1} = A \cdot Z_{.,t}$, where $1 \leq t < m$. All projection steps with the given data in Z and the projection matrix A can be written as the following, using the approximately equal sign $\cong$because generally there is no solution for A:

$$Z_{.,1..m-1} \cong A \cdot Z_{.,2..m} \tag{3.9}$$

In order to prove this, one merely needs to perform the multiplication using matrix elements with subscripts in Z. It is worth noting that the all the columns in $Z_{.,1..m-1}$ will be equal to $Z_{.,2..m}$, except for the first and last columns; the others will be "shifted" to the left by one column index.

Strict equality, as opposed to approximate equality, will hold for (3.9) in certain conditions described in 3.4, but for our application we are only interested in finding a best approximation to A, even if sometimes this best approximation is also an exact fit.

(3.9) is of the same form as the typical matrix equation $b \cong A \cdot x$ except that $Z_{.,1..m-1}$ and $Z_{.,2..m}$ are matrices rather than vectors, so we lack a way to approximate or solve this equation as presented. However, by introducing the k by k identity matrix I, we can rewrite it as:

$$Z_{.,1..m-1} \cong I \cdot A \cdot Z_{.,2..m} \tag{3.10}$$

We can then apply the *vec* operator to both sides of this equation and use the identity (3.8) to rewrite this as follows:

$$\begin{aligned} Z_{.,1..m-1} &= I \cdot A \cdot Z_{.,2..m} \\ vec(Z_{.,1..m-1}) &= vec(I \cdot A \cdot Z_{.,2..m}) \\ &= (Z^T_{.,2..m} \otimes I) vec(A) \end{aligned} \tag{3.11}$$

In (3.11), we now have a standard matrix equation with a vector on the left side ($vec(Z_{.,1..m-1})$) equaling a matrix (($Z^T_{.,2..m} \otimes I$)) times a vector ($vec(A)$), thus transforming the unknown demographic transition matrix A into a vector equation that can be solved using standard matrix inversion.

(3.11) rarely has a single solution, but rather has no solution or an infinite space of solutions, so we find the best fit, letting $vec(A)$ vary but holding the other terms constant. We use the 2-norm (3.6) as the criterion and minimize as follows:

$$\min_{vec(A)} \left\| vec(Z_{.,1..m-1}) - (Z^T_{.,2..m} \otimes I) vec(A) \right\|_2 \tag{3.12}$$

This 2-norm formulation leads to the same quadratic minimization approach as presented in (1.5), where we square the term inside the norm:

$$\min_{vec(A)} \sqrt{(vec(Z_{.,1..m-1}) - (Z^T_{.,2..m} \otimes I) vec(A))^T \cdot (vec(Z_{.,1..m-1}) - (Z^T_{.,2..m} \otimes I) vec(A))} \tag{3.13}$$

After some algebra, dropping of constant terms, and dropping the square root (because the form without the square root is an equivalent optimization problem, see ((Boyd and Vandenberghe, 2005)for justification), we have the following:

$$\min_{vec(A)} \frac{vec(A)^T \cdot (Z^T_{.,2..m} \otimes I)^T \cdot (Z^T_{.,2..m} \otimes I) \cdot vec(A)}{2} - vec(Z_{.,1..m-1})^T \cdot (Z^T_{.,2..m} \otimes I) \cdot vec(A) \tag{3.14}$$

After performing the optimization, we find $vec(A)$ that provides the best coefficients of the best fit demographic transition matrix. We use the *reshape* operator to get our best fit transition matrix from this vector:

$$\hat{A} = reshape(vec(A), k, m) \tag{3.15}$$

3.2.2 Example

We now present an example fitted demographic transition matrix based on the derivation above and a simulated population. We will see that the fitted matrix yields only

small forecasting errors but is implausible demographically. In the sequel, we will add constraints to get a more demographically plausible fit.

Example population and transition matrix

First we need an example population with which to test demographic transition matrices. We create a simulated population in this section which we will use throughout the chapter.

Consider an imaginary three age-class population. This example population will be closed so all the dynamics of the population will be due to mortality and fertility. N measurements of this population can be represented by a 3 by N matrix, call it P. Three separate measurements of the population might be represented thus:

$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} \\ p_{2,1} & p_{2,2} & p_{2,3} \\ p_{3,1} & p_{3,3} & p_{3,3} \end{bmatrix}$$

We use a survival vector where each element s_i is the proportion of the population surviving from age class i to age class $i+1$, with a somewhat arbitrary set of survival coefficients. These will fill the subdiagonal of a Leslie matrix used to the population.

$$s = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.5 \end{bmatrix} \tag{3.16}$$

We need fertility multipliers in the top row of the matrix as well. They are as follows:

$$f = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.8 \\ 0.6 \end{bmatrix} \tag{3.17}$$

These elements yields the following demographic transition matrix T:

$$T = \begin{bmatrix} 0.2 & 0.8 & 0.6 \\ 0.8 & 0 & 0 \\ 0 & 0.5 & 0 \end{bmatrix} \tag{3.18}$$

In order to generate an example population, we calculate a starting population vector having stable proportions based on T, derived by multiplying a vector of ones by T^{100},

$$P_{start} = T^{100} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

We then use that population vector to determine the proportions in each age class,

$$P_{proportion} = P_{start} / sum(P_{start})$$

We multiply this vector of proportions by 100. We calculate the population series by iteratively perturbing the matrix T to yield T_i for each time step forward i. For each time step i, we generate a normally distributed matrix with mean zero and standard deviation 0.025, and multiply this matrix elementwise (Hadamard) by the matrix T. This multiplication scales the perturbation by the size of the cell in T and forces the necessary structural zeros to appear in the perturbation matrix. Finally, we add this perturbation matrix to T to yield T_i. Formally:

$$T_i = T + T \circ N(0, 0.025)$$

With each step, we force also all negative entries to be zero, and all entries greater than one to be one. We thus generate a 10 element population which will serve as input. In the code, we seed the random number generator with zero so that all experiments can be duplicated. This process yields a data matrix P:

$$P = \begin{bmatrix} 465 & 484 & 506 & 509 & 561 & 569 & 590 & 616 & 638 & 689 & 714 \\ 358 & 372 & 389 & 419 & 413 & 440 & 473 & 465 & 504 & 531 & 564 \\ 173 & 181 & 179 & 189 & 209 & 214 & 210 & 243 & 232 & 237 & 266 \end{bmatrix}$$

Example fit with no constraints

If we fit a transition matrix M from P using the method presented above, the result is:

$$M = \begin{bmatrix} 0.0657 & 1.0073 & 0.5387 \\ 0.7240 & 0.1041 & 0.0219 \\ 0.1323 & 0.3774 & -0.1122 \end{bmatrix} \tag{3.19}$$

This matrix was returned after only two iterations of the quadratic programming algorithm. The value of the objective function returned is -38.9611.

The one-step projected population is found by multiplying $P_{+1} = MP$ is:

$$P_{+1} = \begin{bmatrix} 484 & 504 & 522 & 557 & 565 & 596 & 628 & 640 & 675 & 708 \\ 378 & 393 & 411 & 416 & 454 & 462 & 481 & 500 & 519 & 559 \\ 177 & 184 & 194 & 204 & 217 & 217 & 233 & 230 & 249 & 265 \end{bmatrix}$$

If we calculate the cell-by-cell difference from the actual population to the projected population, we see the following distributions of difference in Table 3.1

	Count error	Abs count error	Abs percent error
0% (min)	-15	0	0%
25%	-5	3	0%
50% (median)	1	5	1%
75%	4	10	2%
100% (max)	14	15	4%

Table 3.1: Error for unconstrained fit

We see that the matrix does a good job of fitting the data. Unfortunately, the negative numbers and the lack of sparsity in M make it demographically implausible.

When we run the same example with only three columns of input data, the solution is underdetermined because there are more variables (age class transitions) than there are data columns. The algorithm searches for a solution up to a maximum number of iterations (200 is the default), and then exits with an error code indicating that it does not converge. However, when the fitting procedure is run with five columns, it does converge, though returning a similarly demographically implausible transition matrix.

3.3 Adding constraints to matrix fits

Well chosen constraints are necessary for the suitability of WM for generating transition matrices, as we can see by the example above, in which matrix (3.19) does not meet the form of a Leslie matrix even while it predicts one-step ahead population change very well. On one hand, overdetermined input data with unconstrained fits can generate overfitted transition matrices with implausible entries, like negative fertility, which

nevertheless minimize the objective function. On the other hand, underdetermined input data with unconstrained fits will create objective surfaces that are "flat," with a large space of objectives being equal and no single minimum, and in this case, constraints are necessary to ensure convergence. Additionally, we may want to avoid potentially implausible fitted transition matrices by constraining the matrix coefficients to be demographically reasonable, either setting certain transitions to zero (e.g., disallowing people to age backwards through a nonzero entry on the superdiagonal), or by forcing sets of cells to be demographically plausible (e.g., keeping subdiagonal survival between 0 and 1). We also use constraints to add demographic prior information of more complexity.

In this section we will examine several types of important constraint systems for use with Wood's Method, with examples. We confine ourselves to linear constraints, which can be written in the form

$$G \cdot \mathbf{a} \leqq \mathbf{g} \tag{3.20}$$

$$H \cdot \mathbf{a} = \mathbf{h} \tag{3.21}$$

Where G and H are matrices, $\mathbf{a}$ is a vector containing $vec(A)$ (the transition matrix is A as above) and possible "supplementary variables" appended at the bottom (see below for more discussion of supplementary variables), and $\mathbf{g}$ and $\mathbf{h}$ are constant vectors.

These conditions are passed into computer optimization routines in various ways which we will not specify here.

3.3.1 Single coefficient constraints

The first constraint method to be discussed is fixing single entries in the transition matrix to equal given constants. The most important use of this constraint type is used to force an appropriately sparse demographic transition matrix. In a Leslie matrix, most of the entries are structural zeros (such a matrix is said to be "sparse"). Without setting constraints to enforce this pattern, the fitted matrix might be be filled with non-zero entries which have no demographic interpretation. In the derivation we present in this chapter, structural zeros in the fitted matrix are created by using equality constraints passed to the quadratic optimization routine. In Wood's original formulation, however,

structural zeros are enforced by deleting the zero elements of the transition matrix before the optimization is performed.

We enforce sparsity constraints on for our example matrix M by setting the constraint matrix (1.7) as follows:

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{a} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.22}$$

Note that since $\mathbf{a} = vec(M)$, each row in the constraint matrix corresponds to an "unwound" version of A. So the elements in the demographic transition matrix on the left are mapped to the elements in a single constraint row on the right:

$$\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{bmatrix} \tag{3.23}$$

Thus, to force $M_{1,3}$ to be zero, we put a 1 in the 4th element of the first row of the constraint matrix.

The transition matrix fitted with the above data and these new constraints is found in (3.24) Notice how the subdiagonal more clearly follows the overall pattern for mortality given in (3.16) than the matrix generated without sparsity constraints. The fitted matrix does not follow the pattern for fertility given in (3.17) as closely, but it is a better approximation than (3.19):

$$M = \begin{bmatrix} 0.0657 & 1.0073 & 0.5387 \\ 0.8127 & 0 & 0 \\ 0 & 0.4948 & 0 \end{bmatrix} \tag{3.24}$$

The objective value for a fit with these sparsity constraints is -15.7604, less negative (and thus worse) than when the same fit is performed with no constraints. This change for the worse in the objective is to be expected, since the added constraints prevent the search from finding the very lowest parts of the configuration space. On the other hand, the difference in forecast error from the non-constrained fit is inconsequential, as can be seen by the tabulation of the cross-validated forecast error in Table (3.2);.

With these constraints, the minimization routine will converge with more limited data, even if the fit is not as good. For example, if we fit the matrix using only four columns of population data ($Z_{.,1..4}$) as input data, the matrix is distinctly worse (objective equals -4.5414), and it has implausible negative entries:

	Count error	Abs count error	Abs percent error
0% (min)	-15	0	0%
25%	-4	3	0%
50% (median)	0	5	1%
75%	4	10	2%
100% (max)	14	15	5%

Table 3.2: Error for sparsity constrained fit

$$M = \begin{bmatrix} 0.8157 & -0.6498 & 1.9500 \\ 0.8114 & 0 & 0 \\ 0 & 0.4904 & 0 \end{bmatrix} \tag{3.25}$$

One can also use the same constraint matrix approach to fix matrix entries to known values to non zero entries. For example, if we had a measurement for the fertility in the youngest age of 0.2, we could use the following condition to force $M_{1,1}$to equal this known value while keeping the sparsity constraints above:

$$H \cdot \mathbf{a} = \mathbf{h} \tag{3.26}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \mathbf{a} = \begin{bmatrix} 0.2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.27}$$

This can be useful if one already knows some of the applicable rates in the population and wants the demographic transition matrix to reflect that prior information.

One can use a directly parallel approach to enforce inequality constraints for single coefficients, using the G matrix and $\mathbf{g}$ vector in (1.6). To enforce a greater-than inequality, one uses a negative number in the appropriate cell in G. For example, to enforce a constraint such that $M_{1,1}$ be bounded between 0 and 2.0, one would set up an inequality constraint system like the following:

$$G \cdot \mathbf{a} \leq \mathbf{g} \tag{3.28}$$

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \mathbf{a} \leq \begin{bmatrix} 0 \\ 2.0 \end{bmatrix} \tag{3.29}$$

To ensure that all non-sparse entries of M are positive we would set up the following constraint matrix:

$$G \cdot \mathbf{a} \leq \mathbf{g} \tag{3.30}$$

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \end{bmatrix} \cdot \mathbf{a} \leq \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.31}$$

The matrix fitted with both sparsity and positivity constraints presented above is the same at our display precision as (3.24), with a slightly worse (higher) objective of -15.7603.

3.3.2 Sum-of-coefficient constraints

We can specify how sets of variables combine in sums, either as equality constraints or inequality constraints.

For example, if we know that the total fertility of a population is a certain amount, we can force the fertility entries to sum to that number. In our example population, we could force the fertility row to sum to 1.6, while still keeping sparsity constraints, by setting up an equality constraint with an additional row as follows:

$$H \cdot \mathbf{a} = \mathbf{h} \tag{3.32}$$

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \mathbf{a} = \begin{bmatrix} 1.6 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.33}$$

If we only wanted to ensure that total fertility was between 1.0 and 2.0, but retain the positivity constraints in (3.30),we would set up a inequality constraint matrix with two additional rows in G, like the following:

$$G \cdot \mathbf{a} \quad \leq \quad \mathbf{g} \tag{3.34}$$

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \end{bmatrix} \mathbf{a} \leq \begin{bmatrix} 2.0 \\ 1.0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.35}$$

3.3.3 Spine constraints and supplementary variables

Finally, we can use supplementary variables to "tie together" a number of transition matrix coefficients in a way that parallels the use of parametrized "spines" in many demographic models, including the Brass logit model (Wachter, 2012). For example, we might use a fertility spine derived from empirical data to force relative proportions among the fertility coefficients, but keep the sum of these coefficients within a bound set by inequality constraints.

In order to accomplish a multiple element constraint, we introduce a supplementary variable s. We modify the objective function (1.5) to add as many zeros to the lower right corner of $\mathbf{Q}$ and the bottom of $\mathbf{p}$ as there are supplementary variables; we use zeros in the objective because we do not want the supplementary variables to affect the value of fit criterion, but merely to facilitate more complicated constraint specifications. For each supplementary variable, we add a column to the constraint matrices G and H and a row to the variable vector$\mathbf{a}$ and the constant vectors $\mathbf{g}$ and $\mathbf{h}$. We then adjust the constraint matrices to reflect the structure of the spine.

For example, if one wanted to constrain the matrix such that fertility has the same proportions given in (3.17), but with a sum that is allowed to vary between 1.0 and 2.0, we would create the following equality and inequality constraint specifications (matching (1.6) and (1.7) respectively) while keeping the sparsity and positivity constraints from above. Note that the supplementary columns and rows are set off by lines in the matrices.

$$G \cdot \mathbf{a} \leq \mathbf{g} \tag{3.36}$$

$$\left[\begin{array}{ccccccccc|c} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \left[\frac{\mathbf{a}}{s}\right] \leq \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \hline 1 \\ 2 \end{array}\right] \tag{3.37}$$

$$H \cdot \mathbf{a} = \mathbf{h} \tag{3.38}$$

$$\left[\begin{array}{ccccccccc|c} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0.125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.375 & 0 & 0 & -1 \end{array}\right] \left[\frac{\mathbf{a}}{s}\right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ \hline 0 \\ 0 \\ 0 \end{array}\right] \tag{3.39}$$

Fitting the simulated population with these constraints yields the following transition matrix:

$$M = \begin{bmatrix} 0.2008 & 0.8031 & 0.6023 \\ 0.8127 & 0 & 0 \\ 0 & 0.4948 & 0 \end{bmatrix} \tag{3.40}$$

This fit has an objective of -15.7601 and a summary of forecast error as in Table (3.3)

	Count error	Abs count error	Abs percent error
0% (min)	-15	0	0%
25%	-5	3	0%
50% (median)	0	6	1%
75%	6	9	2%
100% (max)	16	16	5%

Table 3.3: Error for all constraints

3.3.4 Other constraint systems

Innumerable other constraint systems are possible, but the above constraint systems seem to cover most forecasting situations encountered by an applied demographer.

The most obvious extension is to constrain sets of matrix elements to be linear combinations of more than one set of spines. Let v_1 and v_2 be two independently varying multipliers and $\mathbf{s}_1$ and $\mathbf{s}_2$ be two spines, then any set of cells in the matrix can be constrained to be in the space defined by $S = v_1\mathbf{s}_1 + v_2\mathbf{s}_2$ with $c_{1,min} \leq v_1 \leq c_{1,max}$ and $c_{2,min} \leq v_2 \leq c_{2,max}$. These spines can be derived from rate data directly using the singular value decomposition. However, in our experiments, no forecasting accuracy was gained by including this extension.

The constraint system also allows one to model age/ sex relationship beyond mere survival and exogenous migration. For example, one could use equality constraints to make migration of 0-5 year olds a multiple of migration of 20-40 year olds, modeling children's migration associated with their families. Such highly specified models require more data than is commonly available in applied demography, however.

3.3.5 Discussion – black- to gray- to white-box models

The beauty of this system is that model specifications derived from prior knowledge can be added incrementally via constraints. Adding constraints moves the model, in a sense, from a "black box" in which no prior knowledge is assumed regarding the transition matrix, to a "white box" in which all elements of the transition matrix are specified beforehand from prior knowledge. In the intermediate realms of "gray," we can specify those elements exactly if we know their values, or specify their ranges if we only know reasonable bounds. Additionally, by using equality constraints, we can specify arbitrary linear interrelationships between elements if we have data to support those models.

Consider an analyst tasked with a forecasting project. If her data only include age and sex counts, she would use a lightly specified Wood's model, with sparsity and positivity constraints only; such a model would be a "dark gray." If she knows something about the mortality and fertility rates, she can include ranges on those elements or some specification of their shape through spines, creating a model of lighter gray. If she knows some of the rates exactly or decides to borrow them from a source like the Human Mortality Database, she can specify those as equality constraints and let the other elements "float," leading to an even more more transparent ("white") model. Finally, the most transparent model of all would be for her to specify all matrix elements except for a

single one-dimensional set, in which case WM serves basically as a simple equation solver; in this approach WM can be used as a method to estimate rates, as we do in Chapter 4 on migration below.

In Wood's original formulation, this flexibility of the constraint framework is much harder to understand and implement, necessitating our new, more abstract formulation. Wood's original method does not treat structural zeros as constraints explicitly and thus clouds the theoretical connection of matrix fitting and constrained optimization techniques. The original method also requires more computer programming to implement other constraints, hindering experimentation.

However, Wood's original approach of deleting non-active coefficients from the fitting before starting the fit has two advantages. The first advantage with Wood's formulation is that the Leslie model is a sparse model from the outset. In our new formulation, we must add sparsity as a constraint. The second advantage to Wood's original formulation is due to sparsity as well, but more related to computational power than to model specifications. With large matrices, our formulation does not currently run to completion quickly enough to be feasible, but Wood's original ignores the zero coefficients from the very beginning, and so can finish in a reasonable amour of time for large matrices, including the 36 by 36 matrices we use for county age and sex projections in other chapters.

3.4 Exact solutions to the Sylvester equation $C = AXB$

We now include some information about the character of exact solutions to (3.10). We relegate this discussion to a tangential section because the character of exact solutions, if they exist, to $C = AXB$ are not important for Wood's Method.

There are three mutually exhaustive possibilities for solutions to the Sylvester equation $C = AXB$: no solution, a unique solution, or an infinite linear space of solutions of some dimension. In our formulation in C corresponds $Z_{.,2..m}$, A to I, X to A, and B to $Z_{.,1..m-1}$ in (3.10), but we will use the less cluttered notation of $C = AXB$ in this section.

The domains of the known matrices are as follows: $A \in R^{m \times n}$, $B \in R^{p \times q}$, and $C \in R^{m \times q}$

According to Laub (2005), a solution exists if and only if

$$AA^{+}CB^{+}B = C \tag{3.41}$$

Where A^{+}signifies any pseudoinverse of matrix A.

The solution is unique if

$$BB^{+} \otimes A^{+}A = I \tag{3.42}$$

Finally, the solution, if it exists, is (BB^{+} is I if and only if B is full rank and square):

$$X = A^{+}CB^{+} + Y - A^{+}AYBB^{+} \tag{3.43}$$

with $Y \in R^{n \times p}$ arbitrary.

In human demographic applications with five year age categories, B and C are usually full rank, since it is highly unlikely that any given sample year age specific population is a linear combination of the other sample years. There may be occasional exceptions, but only in very small populations; these exceptional populations often also cause errors in the algorithm due to numeric ill-conditioning. Both B and C typically have the same number of columns since one is a time shift of the other, while the measured age classes are the same. A is typically the identity matrix since both $col(C) = col(B)$. When these three conditions hold, then a solution (or a space of solutions) to the non-constrained equation exists as long as the number of columns is less than or equal to the number of rows. In this case the system is either underdetermined or has a single solution.

If the number of columns (samples) is greater than the number of rows (variables), then $CB^{+}B \neq C$ due to properties of the pseudoinverse, there is no solution, and the system is overdetermined.

The solution is unique only when the number of rows for both B and C are equal to each other and to the number of columns, since BB^{+} only equals the identity when B is full rank, and it is only full rank when the rows and columns are the same size; this situation would occur very rarely in applications.

In our county forecasting application presented in Chapter 1, there are an infinite number of solutions (an underdetermined equation) for each unconstrained county fit because the number of rows is 36 but the number of columns is between three (for data spanning a decade only) and nine (for the full series from 1970 to 2010). In non-human applications such as presented by Caswell, there are typically no solutions because the number of columns (measurements) is greater than the number of rows (age or developmental categories); these systems are overdetermined.

Exact solutions are never used in the WM application, so we do not examine this facet of the derivation in any more detail.

3.5 Extensions to the Wood's Method framework

In this section we consider some extensions to the simple formulation presented above. Without an example matrix they will difficult to understand, but they should nevertheless be considered for completeness and to point toward further research.

3.5.1 Replacing the identity matrix, part 1

Interestingly, the matrix I in (3.10) need not be an identity matrix, though it is implicitly presented as such in both Wood and Caswell. Using the identity matrix makes sense for many demographic applications, where the before- and after- populations are measured with the same categories (36 categories for age/ sex in our case) and are isomorphic between domain and range. However, one can imagine a more general situation in which the output measurement (the "after," or time = t+1) has a different format than the input. For example the output might be measured as total population, but the input might be measured in thirty-six age/ sex categories; in this case one could use a 1×36 matrix for A consisting of all ones to transform an age vector to an total count and find the best transition matrix with limited data. We have not implemented this approach, so it remains to be seen how effective it is in practical applications.

For an example using the simulated population above, consider the following:

$$Q_{t+1} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} f_1 & f_2 & f_3 \\ s_1 & 0 & 0 \\ 0 & s_2 & 0 \end{bmatrix} P_t \tag{3.44}$$

Here Q_{t+1} represents the one-step ahead total population, with the transition matrix and age classified population described above. The transition matrix can be fitted by minimizing

$$\left\| Q_{t+1} - \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} f_1 & f_2 & f_3 \\ s_1 & 0 & 0 \\ 0 & s_2 & 0 \end{bmatrix} P_t \right\|_2$$

using the techniques described above.

3.5.2 Replacing the identity matrix, part 2

The identity matrix I in (3.10) can also be used to incorporate variables describing independent inputs into the fitting procedure. Consider the following version of

(3.10)with I appended with a column of coefficients:

$$P_{t+1} = \begin{bmatrix} 1 & 0 & 0 & c_1 \\ 0 & 1 & 0 & c_2 \\ 0 & 0 & 1 & c_3 \end{bmatrix} \begin{bmatrix} f_1 & f_2 & f_3 & 0 \\ s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ e \end{bmatrix} \tag{3.45}$$

Here we have a non-population variable e which increases or decreases the output population by the same proportion depending on the variables c_i in the A matrix. This non-population variable might, for example, represent unemployment rates, and the proportional coefficients c_i might each represent how much of an effect unemployment has on that specific age's population change. This non-population variable is not regressed on itself, addressing Wood's concern with his method modeling autoregressive dynamics only. We have not implemented this approach, so it remains to be seen how effective it is in practical applications. We would minimize

$$\left\| P_{t+1} - \begin{bmatrix} 1 & 0 & 0 & c_1 \\ 0 & 1 & 0 & c_2 \\ 0 & 0 & 1 & c_3 \end{bmatrix} \begin{bmatrix} f_1 & f_2 & f_3 & 0 \\ s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} P_t \\ E \end{bmatrix} \right\|_2$$

as above. In this case, we would need to specify an element constraint for the 1 in $A_{4,4}$.

3.5.3 Transforming input

Transforming the data before fitting the transition matrix may yield better forecasting results.

The easiest approach is to "denoise" the data into smaller rank data with the singular value decomposition (SVD). In the case of county population, 90% of the variation is described by the first three singular values and corresponding singular vectors, based on numeric experiments. A sketch of the procedure is as follows: take the SVD of the data matrix, drop small singular values, reconstruct a dataset with this smaller rank data, and finally find the transition matrix T by applying algorithm to this new data. We have not experimented with this approach yet.

One also apply "median adjusting" to the input by dividing the data input matrices by their global median (across all rows and columns). This transformation retains the multiplicative structure of the data while reducing the conditioning number of the associated matrices, leading to increased numeric stability. Median adjusting is now used

by default in all Wood's Method code.

3.5.4 Regularization and bi-criterion objective functions

We can add terms to (3.12) to effect "regularization," in which the norm of the demographic transition matrix enters into the objective criterion as well as the fit of the transition matrix. The idea behind regularization is that a fitted matrix with smaller norm is "better" that one with a larger norm. This is intuitively reasonable, as a lower norm matrix is more likely to be sparse and less like to have both positive and negative entries "overcompensating" for each other. In a physical context, a lower norm matrix corresponds to a lower energy system, and may have interesting connections to variational mechanics.

The most promising approach to regularization is l_1-norm regularization, as presented in Boyd and Vandenberghe (2005) on pages 308-310. This approach retains convexity in the objective function and may encourage sparsity in the demographic transition matrix without the use of constraints. The parameter δ is adjusted to effect a higher or lower stress on the regularization criteria on the right in comparison to the best-fit criterion on the left.

$$\min_{vec(A)} \delta \left\| vec(Z_{.,1..m-1}) - (Z^T_{.,2..m} \otimes I) vec(A) \right\|_2 + (1-\delta) \left\| vec(A) \right\|_1 \tag{3.46}$$

Another interesting use of regularization is to smooth the coefficients of the demographic transition matrix using an approximation to the first derivative given by Δ:

$$\min_{vec(A)} \delta \left\| vec(Z_{.,1..m-1}) - (Z^T_{.,2..m} \otimes I) vec(A) \right\|_2 + (1-\delta) \left\| \Delta \cdot vec(A) \right\|_2 \tag{3.47}$$

with

$$\Delta = n^2 \begin{bmatrix} 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & -2 & 1 \end{bmatrix}$$

This technique may address some of the ideas of Girosi and King (2008) and their

advocacy of smoothing demographic rates as a source of prior knowledge. However, one must be careful with the arrangement of Δ for two reasons.

From a purely technical standpoint, $vec(A)$ can be arranged in a surprising order because of the way a matrix "unravels" into a vector; for example when A is a 3 by 3 matrix, the $A_{1,2}$ element represents the fertility for the middle age group. To smooth this matrix coefficient, we need to balance it with $A_{1,1}$ and $A_{1,3}$, as these entries are "adjacent" in terms of fertility rates. However, $A_{1,1}$, $A_{1,2}$, and $A_{1,3}$ are in $vec(A)_1$, $vec(A)_4$, and $vec(A)_7$ (respectively) when the matrix is unraveled. In this case the row of Δ smoothing this coefficients would have to be written as:

$$\Delta = 81 \cdot \begin{bmatrix} 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

More importantly, smoothing may not apply when large age classes are used. For example, there is a dramatic jump in migration between ages 10-15 and 15-20 for college counties, and smoothing this discontinuity might lead to inaccurate projections in these cases. If the age classes are small enough that the coefficients approach a continuous function, then discontinuity would be less of a problem.

3.5.5 The interpretation of the objective value

Unfortunately we still find the objective function (3.12) opaque. The objective function is not the norm being minimized but rather an "equivalent problem" in the sense of (Boyd and Vandenberghe, 2005), page 130. In other words, when we minimize the objective, we also minimize the norm, but the objective may or may not have a useful substantive connection to the norm. It should be possible to interpret the norm (3.13) in terms of goodness-of-fit for each coefficient in the transition matrix, but I have not determined the mathematical formula connecting them.

Related to the objective value, the "lambdas" returned by the optimization routine represent the linear (tangent) unit costs of each constraint in terms of the objective. Concretely (if hypothetically), by relaxing a constraint by an unit amount (for example, letting $a_{1,1} + a_{2,1} \leqq 1$+dx instead of $a_{1,1} + a_{2,1} \leqq 1$) we should see a corresponding decrease in the objective measured by the lambda corresponding to the constraint. We would also see an improvement in a goodness-of-fit metric, even though the demographic realism of the transition matrix may suffer. Interpreting the lambdas requires an interpretation of the objective function.

3.6 Running time and memory requirements

It is important to consider the amount of computing power necessary to run a Wood's Method analysis, because if the number of age/ sex classes becomes large, the above approach can quickly exhaust the processing power of even today's computers. We first consider the size of memory necessary to run Wood's Method, as we found that memory is the barrier to running analyses on larger systems.

First consider space requirements. Let $size(X)$ be the number of elements in a matrix M, $rows(X)$ and $cols(X)$ the number of rows and columns, and $X = (Z^T_{.,2..m} \otimes I)^T$ from (3.12). Let c the number of age categories and s be the number of samples in the data. A "Big O" analysis of memory shows that the size of X grows more than cubically with the number of categories:

$$\begin{aligned} size(X) &= rows(X)cols(X) \\ &= size(B^T)size(A) \\ &= (rows(B^T)cols(B^T))(rows(A)cols(A)) \\ &= c \cdot s \cdot c \cdot c = c^3 s \end{aligned}$$

If there are 3 age categories and 6 samples (as presented in Caswell as an example), the size of X is greater than 162 elements, or more than 1296 bytes of storage – matrices of these sizes are easily stored in computer memory. However, if there are 36 categories and 30 samples, the matrix has greater than 139,968 elements, most of which are zero due to the Kronecker product of the identity matrix. This requires 11,197,440 bytes for a single matrix (at 8 bytes per cell); while a single matrix of this size is easily accommodated on modern computers, a few of these matrices stored as intermediate values in a numeric process easily exhausts computer memory.

The time complexity is N^3, dominated by eigenvalue computations in the quadratic programming implementation. There are typically about 20 iterations for each call to qp().

Sparse matrix functionality within Octave can address the space issues during the set-up of the problem, especially by using sparse identity matrices when taking Kronecker products. However, the quadratic programming routine provided in Octave does not yet take advantage of sparsity. Additionally, Octave's quadratic programming implementation calculates singular values and eigenvalues as intermediate values during its iterative algorithm; matrices of dimension 1,000 by 1,000 and larger – which occur due to the use of Kronecker products in the method – can make this implementation

unusable due to the length of time necessary to calculate eigenvalues for large matrices.

In the original application of Wood's Method to demographic forecasting counties, 36 by 36 matrices are used, which prevents us from re-implementing county forecasting using the above more classical derivation of WM and the associated software framework. Lanczos and Arnoldi methods, which only calculate a few eigenvalues more quickly and take advantage of sparsity, may mitigate this problem, but Octave's "qp()" function does not use these techniques even though they are included in the Octave environment.

Somewhat surprisingly, the computing memory and time required to effect Wood's Method both increase as we add constraints, even though by constraining the solution space further we make it mathematically smaller. The increased memory and time requirements are due to the larger matrices required to hold the constraints, matrices which are used in multiple eigenvalue computations throughout the search algorithm. Each row in the constraint matrices is as long as the number of variables in the optimization, further increasing the time and space requirements of adding age categories.

3.7 Conclusion

We see that Wood's Method can be derived much more elegantly using classic approaches than with the approach presented in Wood's original paper and our Chapter One. Our approach also clarifies the use of constraints in WM. We see that the output of the transition matrix need not be in same form of the input matrix if the basic fitting equation is appropriately formulated. We can also use our form to include independent variables by adding columns to the matrices appropriately, expanding the WM formulation beyond just a regression to itself. Our formulation allows us to see how we can add supplementary variables which structure the elements of the matrix using linear combinations of vectors; this connects matrix formulations and classic demographic approaches to characterizing rates by using spines. We briefly showed how one can optimize not just the fit of the transition matrix, but the character of the matrix itself using regularization techniques. We speculated on possible interpretations of the objective and the lambdas, but unfortunately have not been able to fully characterize their meanings demographically. Finally, using our new approach allowed us to analyze the computer resources and complexity of WM, which is especially important as it points the way to improving the algorithm for larger and more frequent population analyses.

4 Geographic structure of migration

In this chapter we show the broader utility of Wood's Method to demographic questions. One demographic task that is hard to perform with readily available data is characterizing migration behavior on a local scale. In the USA, we collect excellent birth and data through the public health and vital statistics infrastructure, but, for the most part we have no direct source of migration data. This dearth of migration data is not consistent worldwide. In countries and states where citizens have a strong incentive to update the government with their current address it is much easier to acquire migration data by examining address change records. Positive incentives include participation in health care programs or income transfer programs (Social Security or the Alaskan Permanent Fund), while negative incentives include fines and possibly jail sentences for not updating address information. Many states in the USA require an address update for driver licenses, but the regulation is not universal. Even in states that do technically require documentation for each address change, licensees or residents often forget or do not bother; administrative record data for migration collected through these channels are well known to be bad, especially for young people and people in university.

However, by generating a Leslie transition matrix with mortality and migration in different cell positions, Wood's Method allows us to isolate net migration dynamics even when we lack actual migration data. To do this, we constrain the subdiagonals in the matrix to correspond to well known ranges of mortality, but let the diagonal be fitted to account for the leftover change, which is necessarily due to migration from the basic demographic accounting equation. Note that this technique only allows us to measure *net* migration rates, in which the in-migration is measured relative to the receiving county, rather than relative to the sending place ("gross" migration).

We explain this technique in more detail below, but first we describe how it can be used to answer a broader demographic question about migration, namely the clustering of migration behavior. One important aspect of any demographic behavior – fertility, mortality, migration – is its geographic patterning, often termed "clustering." An unexpected discovery of the Princeton Fertility Project (Coale and Watkins, 1986) was

that the onset of low-fertility in Europe was locally patterned and localized, essentially coinciding with sub-national speech dialect regions. We seek to extend that line of investigation to the possible patterning of migration.

4.1 Migration index – method and data

In order to characterize the migration of a county, we first derive a scalar migration index for each county by applying the migration components of the fitted transition matrix to a standard population. We then "smoke test" this index with some counties to see if the migration indexed is plausible given what we know about these counties' migration patterns. We map this migration index at the county level and investigate clusters that appear visually on these maps. We examine clustering at multiple geographic scales in the USA, corresponding to counties, states, and census divisions (see Figure 4.1) to determine whether there is a hierarchy in migration patterns. We then examine the spatial autocorrelation of the migration index to discover if any clustering is discernible with a pure numeric analysis. Finally, we attempt to use our fitted matrices to model net migratory redistribution within the USA.

4.1.1 Migration index

The first step is to calculate a migration index for each county.

One of the difficulties in establishing a migration index has been the lack of good data on migration, as discussed above, especially when compared to the widely available high quality fertility and mortality data. Typically, net migration is calculated as a residual from all other known population change, and thus requires fertility, mortality, and age specific population at two separate time points. Gross migration does not require these supplementary data, but instead depends on individual level data from surveys or (nearly) full-coverage address registers describing the source and destination of each individual mover. The American Community Survey calculates gross migration by using a question about residence one year previous to the survey, but these data cannot be used for many small counties due to small sample sizes and privacy concerns. As described above, most states in the US lack complete migration registers, except for older residents receiving social security and medicare; this lack of good migration rules out an important data source for gross migration calculation.

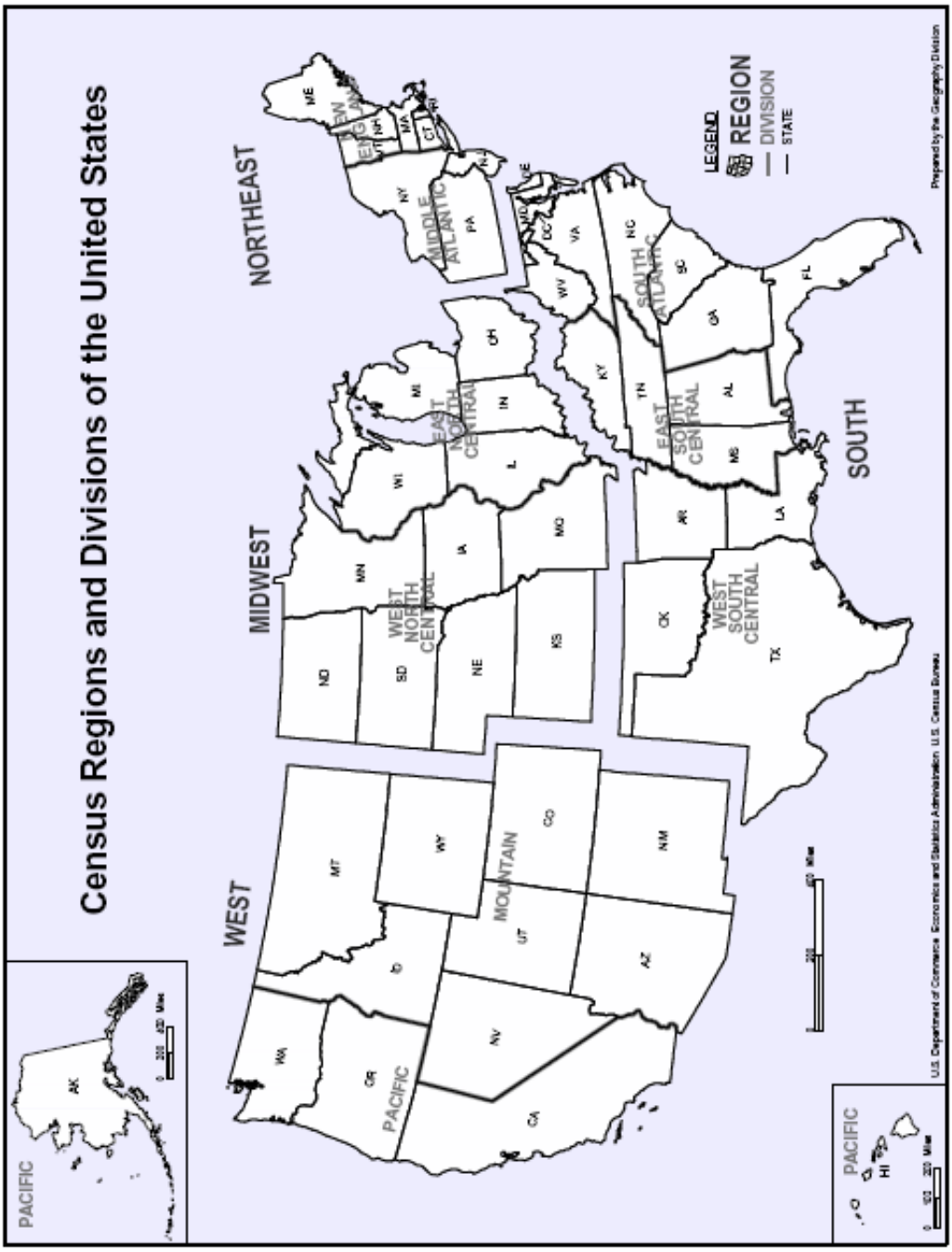


Figure 4.1: USA Census geographical hierarchy

The applied demography group at Wisconsin, Madison has calculated net county migration for all US Counties each decade since 1960 – see (Winkler et al., 2013) for a description of the most recent analysis and dataset – but this dataset is not very widely known outside the rural sociology subdiscipline; it is an excellent contribution, but it is hugely labor intensive and must always lag the Decennial Census by at least three years because of upstream data release schedules. While we could generate migration indices from this dataset without the use of WM, we want to show a useful non-forecasting application of WM.

4.1.2 Index calculation

The index we compute is a single scalar value that can be assigned to each county, facilitating comparison between counties. We use an approach known as called "direct standardization," in which a set of rates for various populations are all applied to a standard population, with these new populations then compared to each other to see the effect of the differing rates. More detail can be found in, for example, (Preston et al., 2001).

In order to calculate something like net migration rates, we first generate transition matrices using Wood's Method for each county, as described in Chapter 1. Then we delete the fertility and mortality components of the matrices, leaving only migration on the diagonal cells. We set the migration component at age 0-5 to zero, because the population at this age is a combined result of childbirth and migration, and hence its contribution to a migration index is ambiguous. We set migration at the open interval (85+) to zero as well because high age data seem to contribute to the migration index anomalously, and for relative comparisons between counties, these ages are less important than the middle ages in any case. Referencing the matrix presented in Chapter 1, repeated here in (4.1), we set ${}_5S_a$ and ${}_5F_a$ to zero for all a; we set MF_m, MF_f, and ${}_+M_{85}$ to zero as well:

$$\left[\begin{array}{ccccc|ccccc}
0 & 0 & & \dots & & MF_m & 0 & {}_5F_{10} & \dots & 0 & 0 \\
{}_5S_0 & M_5 & & \dots & & & & & \dots & & \\
 & {}_5S_5 & M_{10} & \dots & & & & & \dots & & \\
 & & & \ddots & & & & & \dots & & \\
 & & & \dots & {}_+S_{85} & {}_+M_{85} & & & \dots & & \\
\hline
 & & & & & MF_f & 0 & {}_5F_{10} & \dots & 0 & 0 \\
 & & & & & {}_5S_5 & M_5 & & \dots & & \\
 & & & 0 & & & {}_5S_5 & M_5 & \cdots & & \\
 & & & & & & & & \ddots & & \\
 & & & & & & & & \dots & {}_+S_{85} & {}_+M_{85}
\end{array}\right] \tag{4.1}$$

We derive a standardized population by taking l_x from HMD for 2010 all-USA male and female life tables. The l_x for this population can be found in Table 4.1 (the vector is presented with separate a column for each sex for readability, while it is actually a stacked vector when calculated). We then and normalize it by dividing each sex by its total, yielding an age- and sex-specific standardized population, presented in the same table.

For each county, we project forward this standard population by the county's specific migration-only matrix to yield a vector of age/ sex specific migration, in which entries may be either positive or negative. We project one step forward only, in order to avoid biasing our index with fertility and mortality dynamics being left out. To calculate the total migration index, we find the sum over all the ages and sexes of the projected population.

Our migration indices suffer somewhat from problems with specifications in mortality and fertility rates for the current use of WM. By allowing the mortality and fertility encoded in the horizontal and subdiagonals to vary, even slightly, these parts of the transition matrix absorb some of the dynamics which may belong in the migration components and vice versa. The amount of allowed variability is small, so for the purposes of generating a migration index, our method should be sufficient. When mortality and fertility rates are available for a county, they should be specified exactly via cell constraints, which would then make the migration components more accurate.

Additionally, if the standard l_x above does not reflect a county's survival characteristics, the migration index might be slightly off target as well, but given the relative homogeneity of mortality in the developed world, this is only a slight drawback; the

Age	Male l_x	Female l_x	Male proportion	Female proportion
0	100,000	100,000	0.0645	0.0614
5	99,211	99,349	0.0639	0.0610
10	99,147	99,299	0.0639	0.0610
15	99,066	99,239	0.0638	0.0609
20	98,724	99,100	0.0636	0.0609
25	98,103	98,878	0.0632	0.0607
30	97,442	98,603	0.0628	0.0606
35	96,729	98,247	0.0623	0.0603
40	95,882	97,743	0.0618	0.0600
45	94,700	96,993	0.0610	0.0596
50	92,813	95,788	0.0598	0.0588
55	90,005	94,010	0.0554	0.0577
60	85,999	91,579	0.0520	0.0562
65	80,728	88,079	0.0520	0.0541
70	73,494	82,843	0.0474	0.0509
75	63,770	75,226	0.0411	0.0462
80	50,845	64,196	0.0328	0.0394
85	35,026	49,060	0.0226	0.0301

Table 4.1: Standardized population for migration index

method could be applied with county specific l_x's when data are available.

4.2 US county migration index examples

As a simple check on the validity of the age-specific migration index, we graph several counties in the Pacific Northwest with well known migration behavior. The migration index has a similar shape to these counties' population profiles and to their demographic dynamics, showing that the index reflects demographic reality in these test cases. We will examine these specific counties below.

4.2.1 Graph format

Before looking at each specific county in detail, it is worthwhile examining the

graphs in Figures 4.2 through 4.7. Like the other age profile graphs above, they are basically age-sex pyramids turned on their side. The x-axis is the age at which point the migration index is calculated, i.e. at five year intervals, but not including age 0-5 or the open interval as described above. Each triangle or circle is a data point, for males and females respectively. These points are connected by lines in order to convey the gestalt of an overall age distribution; with only points on the graph, it would be difficult to distinguish the male or the female subpopulation age profiles as a whole.

The y-axis shows the migration calculated at each age; the sum of these numbers is the total migration index used below in maps. There is a darker line through zero on the y-axis; points above this line show positive migration – people moving into the county at the given age – while points below show negative migration and the fact of net out-migration at the given age. For example, in Figure 4.2, we see that there is negative migration at ages 5-10, as these children are often leaving with their families and moving to suburban counties like nearby Washington County, Oregon.

In all the counties we have examined there is a high index value in the 75-80 and 80-85 age ranges. We believe these high numbers are anomalous, but we have not determined why they are present. Conceivably, as life expectancy increases, there may be unexpectedly large numbers of people in the higher ages, which would show up as migration if survival was not appropriately adjusted.

All the graphs share the same vertical scale, between -0.05 and +0.05, to allow for easy visual comparison.

4.2.2 Individual county discussion

Multnomah County, Oregon is known to be a cultural center and migration destination for young adults interested in the arts and recreation, and this migration dynamic is reflected in Figure 4.2. At ages 20-25 and 25-30, we see an extremely high migration index, which falls rapidly on both sides. Some of the positive migration at 15-20 is likely due to college migration, with Portland State, Reed, Lewis and Clark, and other smaller colleges as destinations. We see a distinct out-migration at ages 35-50, with the migration index becoming negative; many of these migrants are likely looking for a more family oriented living situation with bigger houses and less urban school districts than they can find in Multnomah County. These families will bring their small children when they move out of Multnomah County, and we see a similar negative migration indices for age 5-10. We see out-migration at age 60-65 as well, with only small in-migration in the ages between. Importantly to local migration dynamics, the major high tech and sportswear employers of the Portland area – Intel, Nike, Columbia Sportswear,

and their suppliers – are all located in suburban Washington County, Oregon, making the regional pull of the suburban counties away from Multnomah County even stronger than if residents were only looking for more suburban housing.

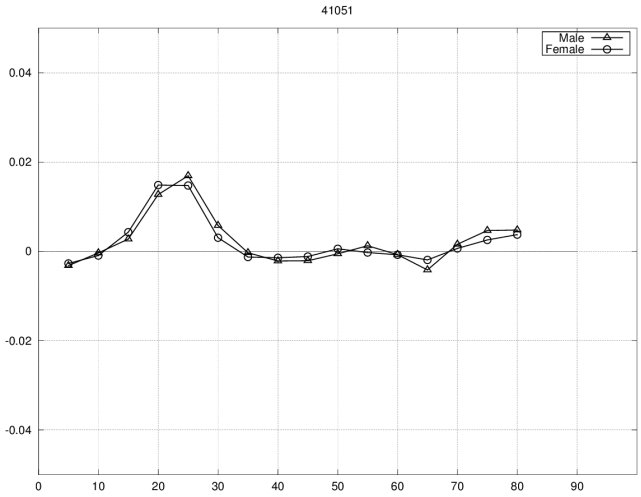


Figure 4.2: Multnomah County (OR) age-specific migration index

King County, Washington is similar to Multnomah in that is known to attract large populations in their twenties. In contrast to Multnomah, however, there is a much wider economic base in King County, with numerous corporate headquarters including Microsoft, Starbucks, Amazon, and Costco; there are many Boeing suppliers and plants throughout King County as well. King County Washington is also much larger geographically than Multnomah Oregon, at 2,307 square miles versus 466 square miles; with more geographic area and more employment, King County can encompass a wider variety of socioeconomic configurations which are reflected in different migration-index patterns. Specifically, while Multnomah County is densely urbanized throughout, King County includes traditional suburbs like Redmond and Bellevue; King county even has some mountainous settlements to the West, which likely have rural amenity/ retirement migration patterns like Clallam County, Washington, which we describe below. Correspondingly, the migration index for 35-50 year olds and their 5-10 children is about zero, unlike the negative amounts in Multnomah County, Oregon, due to this employment diversity attenuating the aggregate index.

Whitman County (Figure 4.4) contains Washington State University, but little other economic opportunity besides low labor wheat farming; the landscape – flat and dry farmland - would not be considered high amenity either. The enrollment at WSU in

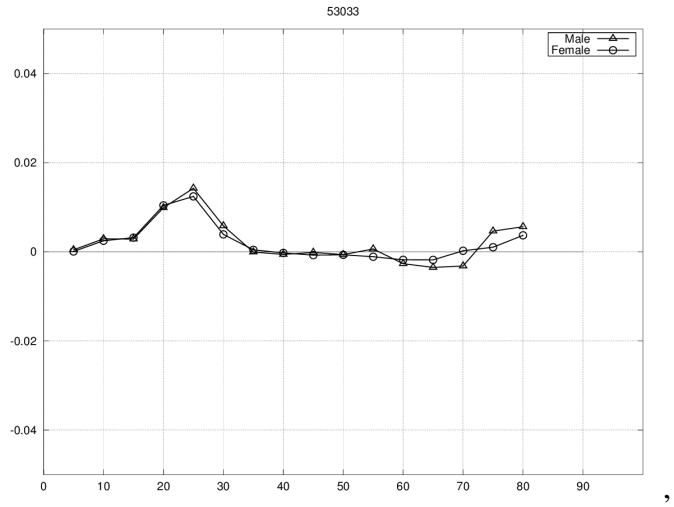


Figure 4.3: King County age-specific migration index

2012 is about 19,000 and the county total population is about 46,000, so the county population dynamics are dominated by students at the university. Whitman shows an extremely high migration index at the college ages, but either negative or flat at other ages. Note that in Multnomah and King Counties, which both have important population dynamics due to young people in their early twenties and thirties but less important undergraduate populations (at least proportionally to the total population), the large peak in these urban counties is shifted one age class to the right, from 15-25 to 20-30 in comparison to Whitman. Whitman also has large out-migration in the age-class 25-35, with the index going strongly negative, as these people leave the county due to the lack of employment outside the university. Multnomah and King Counties are flat at these ages, with far less out-migration than Whitman, due their more diverse employment and housing opportunities.

Clark County (Figure 4.5) is located to the north of Multnomah County, providing a suburban, commuter oriented environment. While there is a community college and a branch campus of Washington State University in the county, there are no large universities in the county to draw young people. There is distinct out-migration index at 15 and 20 years old, in direct contrast to in-migration at these ages in college counties like Whitman (Figure 4.4); there is similarly contrasting high in-migration image at 25-35. This pattern is consonant with the intuition that young people leave Clark to pursue educational and work opportunities and return to raise families. The population forty years and older shows consistent in-migration as well, in contrast to Multnomah,

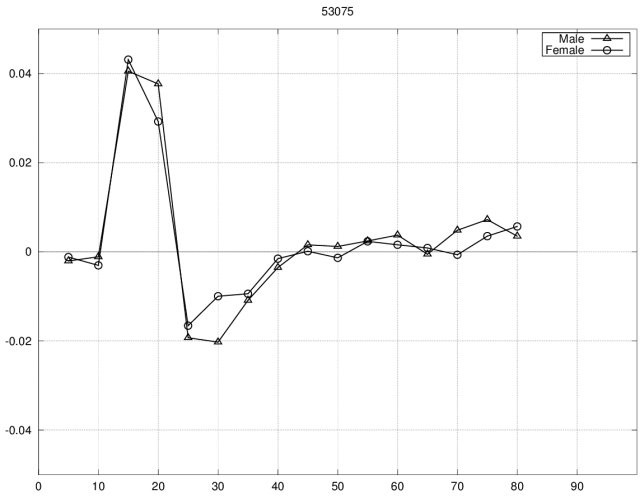
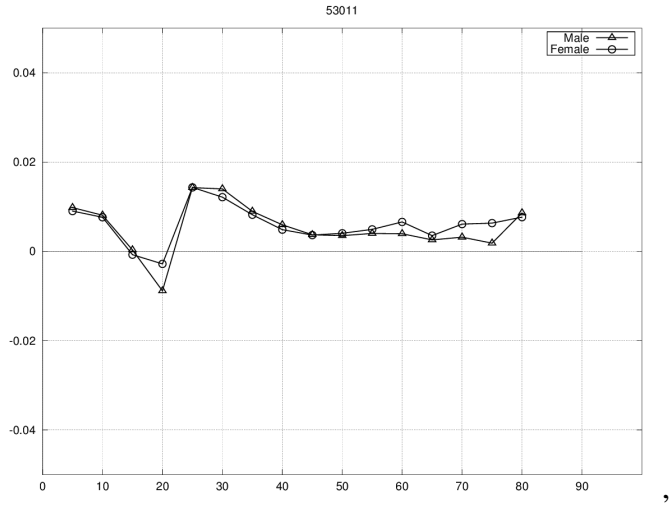


Figure 4.4: Whitman County age-specific migration index

King, and Whitman counties, all of which provide services and social environments for younger people. In contrast to Clallam County (Figure 4.6), however, there is no "retiree bulge" starting at age 50; likely, while Clark is a reasonable retirement area and shows no particular out-migration, it is not a retirement destination for non-locals.



,

Figure 4.5: Clark County (OR) age-specific migration index

Clallam County (Figure 4.6) is a recreational amenity county on the Olympic Peninsula. It contains the Olympic National Park with excellent outdoor recreation opportunities, close proximity to Seattle, numerous opportunities for water recreation on both rivers and the ocean, and overall mild weather. On the northeastern corner of the county there even is a "rain shadow," forcing precipitation to be about 30% of most of the nearby Puget Sound area, a major attraction for retirement movers to Sequim and its surrounding areas. We see particularly high migration indices of about 0.015 and 0.01 for males and females at the 60-65 year old retirement ages. Like many rural counties, we see strong out-migration by young people; this migration is due to lack of employment and higher education opportunities. We see these dynamics reflected in the highly negative index values at age 20-25 for females and males.

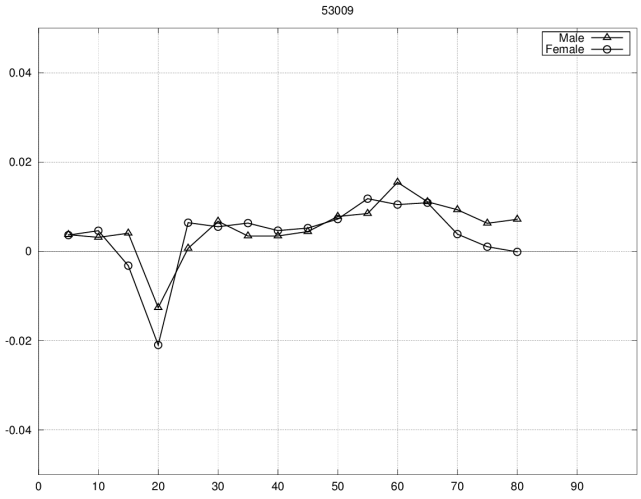


Figure 4.6: Clallam County age-specific migration index

Finally, we show the migration index for Yakima County, Washington. Yakima is a rural county, with a thriving agricultural base, but no higher education opportunities and fewer higher paying jobs than more urban counties; it also has a reputation as being somewhat less desirable than other Washington Counties, with higher crime and poverty levels than other counties in Washington and a concomittant lack of retirement migration. Yakima's young age specific migration index has a similar shape to Clallam, with the same exodus of 15-25 year olds, except less extreme than Clallam at these ages, probably due to a more diversified economy. Yakima also has an attenuated retirement bulge and lower overall 35-55 age in-migration in comparison to Clallam County.

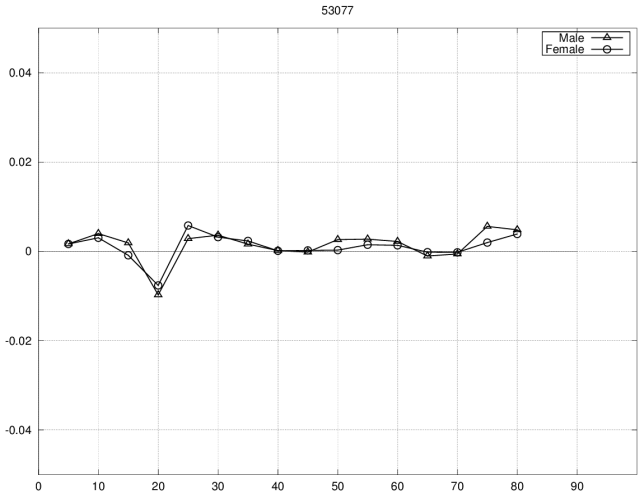


Figure 4.7: Yakima County age-specific migration index

4.3 Migration distribution and geographic clustering

4.3.1 Migration index histogram

Figure 4.8 shows the nonparametric distribution of the migration index over our entire dataset. The x-axis contains the migration index calculated for each county using the method described in Section 4.1.2. Each bar represents the proportion of counties with an index in a range that can be read of the x-axis; there are ten bars within each gridline section, representing a range of about 0.02 each in the migration index.

The median is 0.0239 and the interquartile range is 0.100798. With a median above zero, we can infer that most counties in the USA have positive migration. There is a noticeably longer tail extending to 0.7259 on the right, but only reaching -0.2968 on the left; this asymmetry is to be expected if migration is largely positive across all counties and thus more rightward in the graph. Remember that a county index of migration is the sum of all the age-specific migration indices within the county, so different ages may balance to bring the total closer to zero even if some ages have extreme values (for example 15-25 in a college county, like in Figure 4.4).

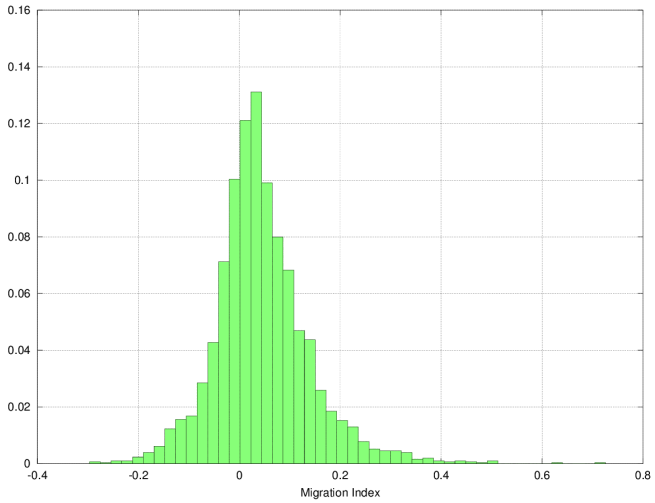


Figure 4.8: Migration index histogram

4.3.2 Migration index maps

County total migration distribution

In Figure 4.9, we map the migration index across 3108 counties in the lower 48 states of the USA. We do not treat Hawaii because it is too far from the mainland to have its migration be analyzed in terms of distances from other states. Alaska suffers from the same situation; additionally, many of Alaska's sub-state boundaries were created in the last 20 years, and thus it is difficult to calculate a migration index based on their historical behavior. There are seven other counties in the other 48 states which cause difficulties in the migration index calculation and are ignored in this analysis as well.

The counties are outlined in light black lines, except with counties with high out-migration. Because counties with high out-migration are indicated with dark colors, their boundaries are indicated with white, thin lines. The states are outlined in heavy dark lines. State boundaries often extend into water, which can be seen if one looks closely at the map along the shorelines, but this cartographic artifact is more obvious near the upper peninsula of Michigan and the Southern tip of Florida. Labels were not placed on states or counties in this map so as to avoid visual clutter, hiding any visual impression of migration dynamics. Counties in the Western states are often much bigger than counties in the East, biasing visual interpretations slightly, but this is unavoidable.

We classify migration into high in-migration, high out-migration, and moderate migration. We use only three categories because more categories would be difficult to distinguish visually, especially when using grayscale graphics and attempting to discern large scale patterns. (A rule of thumb for cartography is to use four to six categories in chloropleth maps, with five or more categories generally only being used in two-color thematic schemes, in which one color is used for negative values and another color for positive values. Examples of two color maps include elevation maps, with blue signifying below sea level elevation and a reddish brown signifying above sea level elevation.)

A county is considered to have high out-migration if the migration is in the 25th percentile over all migration indices, whether positive or negative. We classify negative total migration indexes close to zero as "moderate" in order to retain a consistent non-parametric approach to classification. A county is considered to have moderate migration if it is between the 25th and the 75th percentiles of the migration index. This interval incorporates some counties with out-migration, but it seems like the least arbitrary approach. Finally a county is considered to have high in-migration if the migration index in the 75th percentile.

We calculate the 25th percentile migration index to be -0.0097 and the 75th percentile to be 0.0911. The lowest migration index is -0.296812, in Blaine County Nebraska, located in the middle of the state and in the high plains, a region with well documented high out-migration. The county with the highest migration index of 0.7259 is Pinal County, Arizona, which is located between Pima County to the south and Maricopa County to the north which have experienced documented high growth in the last decade from retirement migration.

Because most counties have positive migration indices, this measure is not symmetrical. There are 920 counties with migration indices less than or equal to zero. The median value for all the county indices is 0.0338, consonant with the slight rightward skewness in the histogram in Figure 4.8.

A basic question that can be answered by visual inspection is whether the migration indices are uniformly speckled or clump into larger consistent areas, with the former pattern reflecting variance at small scales and the latter reflecting variance at larger scales. By inspecting the map, we see that most migration seems to be clustered at about 200-250 mile level, a scale which encompasses multiple small counties in the eastern USA while sometimes incorporating fewer larger counties in the West. There are numerous exceptions to this clustering, with isolated counties of high in-migration or out-migration scattered about. Interestingly, 200 miles or so intuitively "feels" like a region, sometimes defined as an area that encompasses the widest extent of commuting and industrial deliveries. The US Bureau of Economic Analysis (BEA) aggregates counties into "economic areas" using commuting as their main criteria (Johnson and Kort, 2004), and these often are about 100 miles across. There are some notable exceptions to this clustering pattern, for both in-migration and out-migration, which we will treat in turn below.

For negative indices, the only cluster bigger than a typical commuting region is the large expanse of out-migration in the middle of the country, starting in the south in Western Texas, spreading out as it goes north to cover most of the Dakotas and Montana. The prolonged population decline of the Great Plains is well documented (for example (Rathge and Highman, 1998)), due to lack of urban centers, increased farm automation, and the lack of outdoor amenities. In fact, the map of population decline matches the distribution of amenity landscape mapped in (McGranahan, 1999).

All the other clusters are more or less regional in size. These include both the Mississippi Delta region and Southwest Alabama (the "Black Belt") – two of the poorest regions in the United States, with very high proportions of African Americans. The coal region of southwest West Virginia and western Kentucky has a smaller cluster of out-migration, as does Northwestern Ohio plus Northeastern Indiana, the region sur-

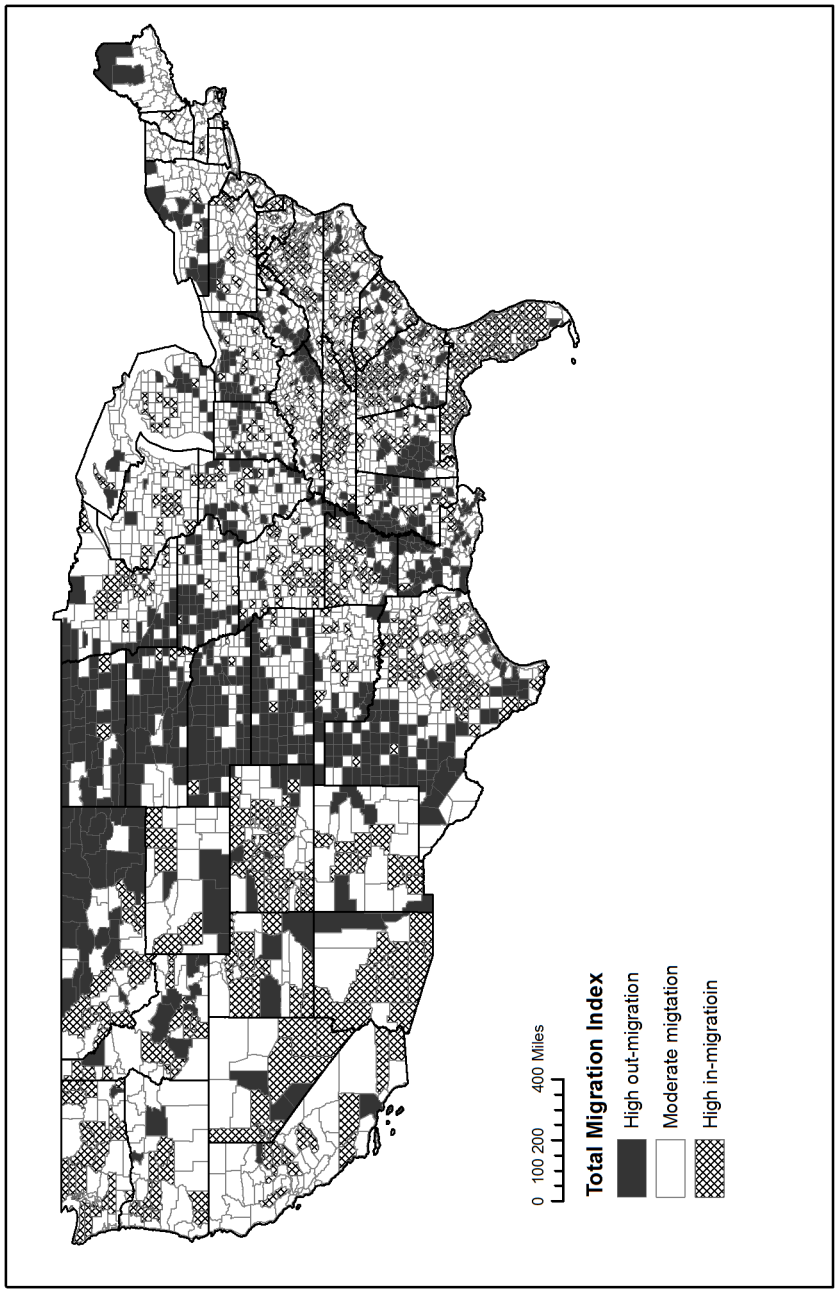


Figure 4.9: USA total migration county map

rounding Buffalo New York, and upstate New York on the Canadian border. There appears to be an out-migration cluster in Northern Maine around Aroostook County, but due to the low population density and large county sizes in this area it may be just a large county with high out-migration.

Interestingly, the in-migration clusters seem to be slightly larger than the out-migration clusters, and there are more in-migration clusters in total. The larger in-migration clusters include about half of Southern Arizona, most of Florida, and much of central Colorado. The smaller clusters seem to be either urban centers with employment in high tech, financial services, and (in the Southeast) manufacturing, or clusters due to an area providing outdoor amenities. The former group includes the Salt Lake City area in Utah, the Sacramento region in California, the two clusters in Eastern Texas, Northern Virginia, the Portland region in Oregon, the Minneapolis region in Minnesota, the Rayleigh area in central North Carolina, Central Colorado around Denver, Northern Alabama near Decatur, various smaller clusters in South Carolina, and the Nashville region in central Tennessee. The latter group of in-migration clusters, with rural amenities dominating, include the Northwestern part of Washington, the Ozarks on the boundary of Missouri and Arkansas, and Rocky Mountain counties in various Western states.

Age specific migration indices

In Figure 4.10, we map the migration index calculated for people over 60 plus only. We renormalized this map using the 25% and 75% cutoffs for 60 plus only – less than 0.01685 for high out-migration and greater than 0.0314 for high in-migration.

Most of the clusters are similar to the total migration index map in Figure 4.9, with one very interesting exception. For people over 60, the High Plains – Kansas, Nebraska, South and North Dakota, Iowa, and Eastern Montana – do not seem to be experiencing the great exodus that dominates discussions regarding high plains demography. In contrast, the West Texas plains do seem to be subject to high out-migration with people over 60, unlike the states to the North. Besides the northern plains, other rural counties in the Eastern US – the Mississippi Delta, the Black Belt, etc – seem to have the same out-migration characteristics regardless of age. In the western half, however, New Mexico, Utah, and Idaho all have fewer out-migration counties when only 60 plus migrants are considered. These counties likely provide high quality of outdoor recreation and low cost of living and thus attract retirees.

There are also smaller clusters of out-migration counties for these older ages in more urban and developed areas that show less out-migration when all ages are considered in total. These include the "Inner Midwest" – the states of Indiana and Ohio, along with much of nearby Northern Kentucky and West Virginia. Without gross migration

data we do not know where these people are moving, but is plausible that they may be retirees leaving their urbanized, work-oriented counties for higher amenity counties after retirement.

There are also several counties near the urban centers of the Northeast, Charlotte North Carolina, and a cluster in Eastern Alabama near Auburn with high out-migration for people 60 and over. Maine and New Hampshire have no out-migration counties for 60 and over people, though they have many counties with all-ages out-migration in the northern parts of the states.

In Figure 4.11 we show the migration index with people under 55 only. We renormalized the cutoffs for this age group as well – less than -0.0125 for high out-migration and greater than 0.0365 for high in-migration. It presents the same overall impression as the total migration index map in Figure 4.9. The main difference seems to be only in the High Plains with slightly more out-migration counties than in either the total migration or the 60-plus migration maps.

County to state relationship

Figure 4.12 shows the average migration index over each state, calculated by taking the mean total migration index over all the counties within each state; these statistics are calculated using aggregation functions in the SQL database language in our county database. The same approach is used for standard deviations by state in Figure 4.13. Unlike counties, for states and divisions, the thematic colors were chosen by our Geographical Information System[1] (GIS) routines, assigning shading based on the percentile of the particular average migration index across all states; the specific cutpoints are in the map legend. Each state label has the US Postal Service two letter code (for example "WA" for Washington). Some of the labels in the Northeast were offset to keep the map readable.

The state migration level maps reflect the overall impression of the county map in terms of average migration. Kansas, Nebraska, the Dakotas, and the Deep South show high county levels of out-migration, and the aggregated out-migration in these states is consonant with this fact, as can be seen in Figure 4.9. Louisiana shows overall low migration in the state map, and is filled with out-migration or moderate migration parishes in the county map as well. Interestingly, Louisiana's out-migration does not seem to be caused by Hurricane Katrina, since the out-migration parishes are in the Northeast and Central Western Louisiana; the Northeastern area is part of the Mississippi

[1] We used a combination of Geographic Information Systems for our analysis. We did the final rendering in ArcGIS, initial mapping in QGIS (an open source desktop GIS), and the data management in a PostgreSQL database with geographic extensions from PostGIS.

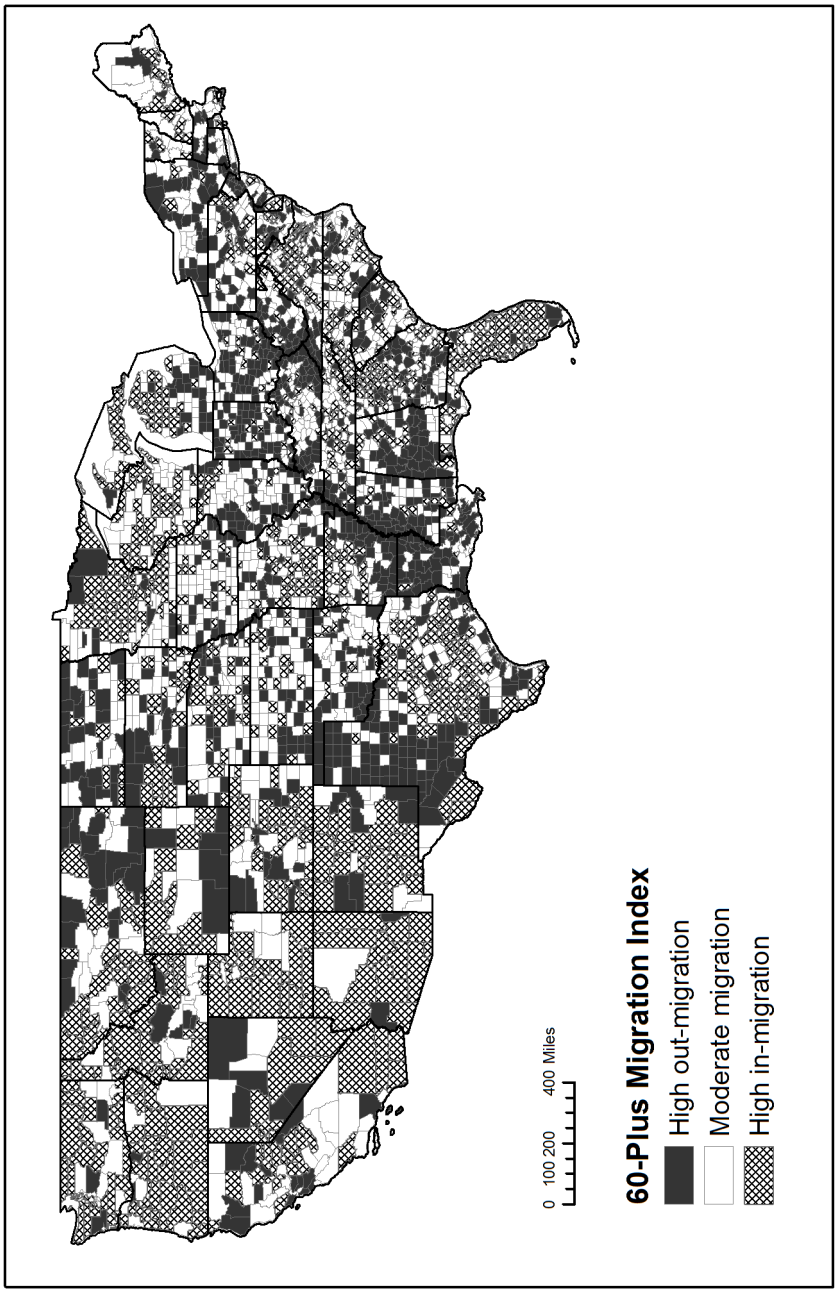


Figure 4.10: USA 60 plus migration county map

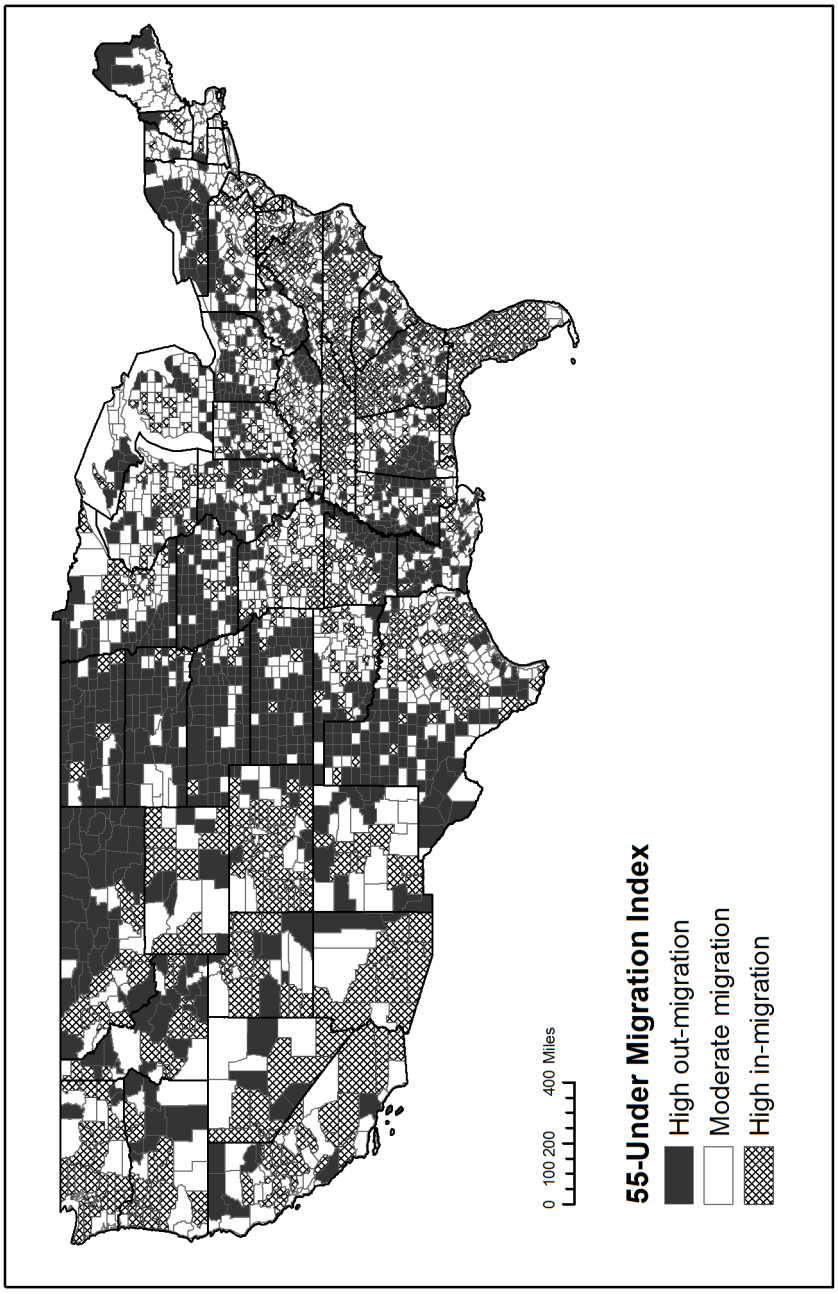


Figure 4.11: USA 55 and under migration county map

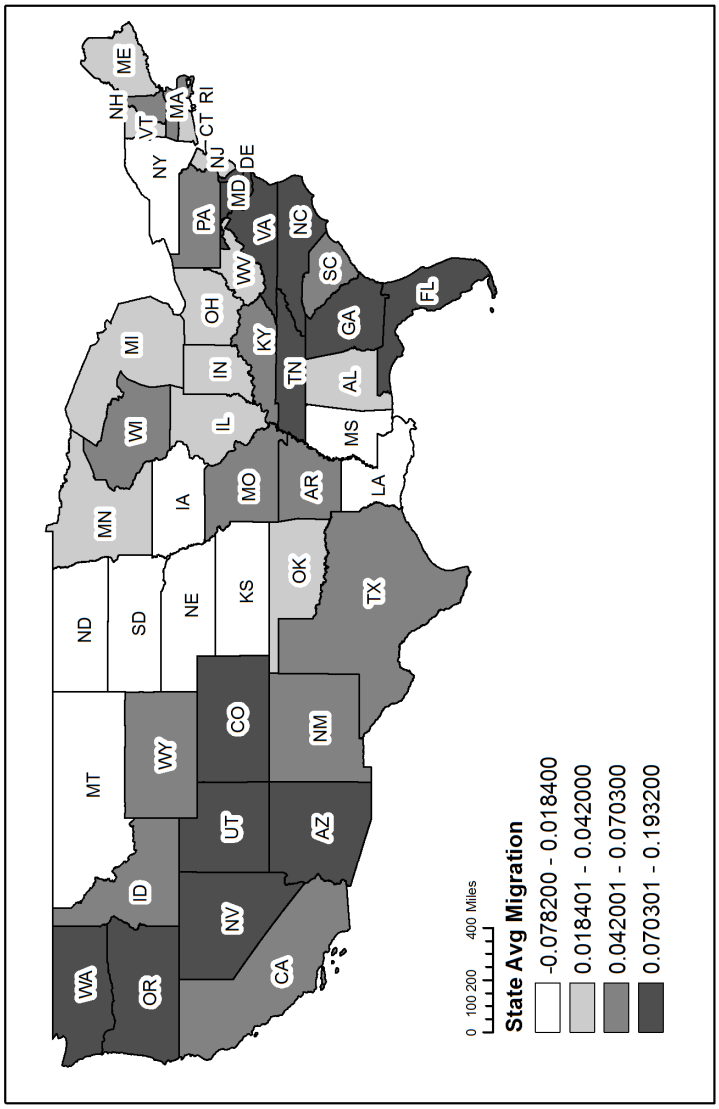


Figure 4.12: USA county migration average state map

River Delta system, whose migration dynamics were described in Section 4.1. Alabama, Minnesota, Iowa, Illinois, Ohio, Indiana, New York, and Maine all have many out-migration counties, and accordingly fall into the moderate and low migration classifications in the state map due to averaging with their moderate and high in-migration

counties. The rest of the Southeast is experiencing strong in-migration when one looks at the number of in-migration counties, and this high in-migration is reflected in the state maps. Montana is interesting, with most of its Eastern counties being located in high plains environments and having high out-migration at the state level like the Dakotas and Nebraska, while the Western counties are experiencing high in-migration, presumably for their mountain recreational amenities. Thus, while Montana's average migration is quite low, it's migration variation is quite high.

Figure 4.13 shows the standard deviation of the migration index in thea state by state, calculated over all the counties in each state. Florida has very high variance; this is surprising based on visual inspection of the county map in Figure 4.9. Interestingly, Montana and South Dakota also have high migration variance, while they show overall high out-migration at the state level. In the case of Montana, this variance is probably due to the two contrasting regions of migration described in Section 4.1, in reference to the smaller scale county maps – high in-migration on the Western side in the Rocky mountains, and high out-migration on the eastern side in the Great Plains. In South Dakota there are a few in-migration counties on the Eastern border, and several moderate migration counties on western border with Wyoming; presumably these create the high variance due to contrast with the many out-migration counties in the rest of the state. The Mountain and Great Basin states share the high variance of Montana even as they show overall high average in-migration; the county map for these states has a large of number of non-moderate counties with both in- and out-migration; we see the this contrast in in-migration and out-migration counties especially in Nevada, Utah, New Mexico, and Colorado. The Northeast shows the lowest variance in the country, reflecting the more or less consistent picture of moderate county migration for this region in Figure 4.9. The balance of the USA shows moderate variance, along with most states having a few out-migration counties, but mostly in-migration and moderate migration counties.

Census divisions

In Figure 4.14 we see a labeled chloropleth map of migration indices at the Census Division level for the USA, and Figure 4.15 has a map of the standard deviations for each division. We will consider each division in turn, examining average migration and variance together.

The Pacific Division ("P") has the second largest average in-migration index after the

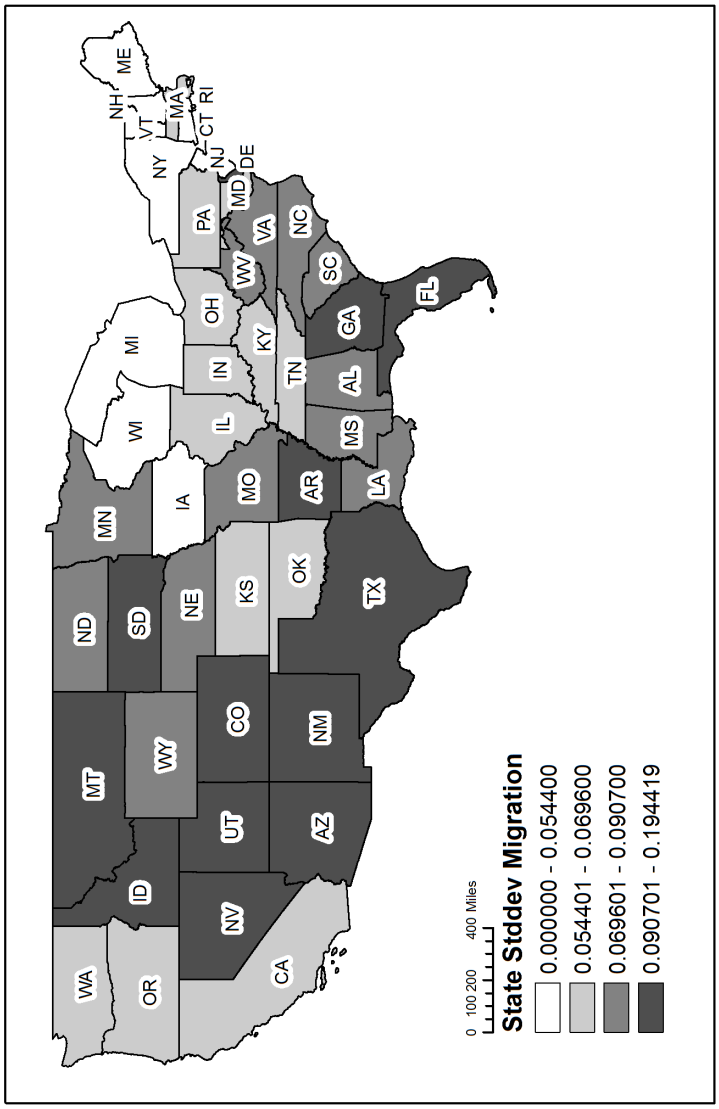


Figure 4.13: USA county migration standard deviation state map

South Atlantic, but a fairly low standard deviation. Encompassing California, Oregon, and Washington, the West has historically had high migration almost everywhere, and very few counties of out-migration as can be seen in Figure 4.9.

The Mountain Division ("MNT") has a moderately high in-migration index, but the

highest standard deviation in the country. Many parts of this region have experienced amenity-based in-migration, especially in the extensive scenic areas, but other counties are either remote desert or high plains and these regions have experienced out-migration similarly to the Great Plains. This Division includes Nevada and Arizona with high retiree and "snow-bird" migrants as well.

The West-North-Central ("WNC") division has lowest migration index of the USA, and is the only Division in which the average migration index is nearly negative. This low average is consonant with the the county map in Figure 4.9, which visually has the largest swath of out-migration counties in the USA as well. The standard deviation in this Division is fairly high at 0.08, probably due to isolated counties in the Dakotas, the attractive Minneapolis region, and the Southwestern Missouri Ozark retirement region contrasting with generally high out-migration in the other counties.

The West South Central ("WSC") division has moderate in-migration, probably because the high out-migration counties in West Texas, Louisiana, and Arkansas balance with the high in-migration counties in Eastern, more urban Texas, and the Ozark amenity region in Arkansas. Along with these contrasts, we see the second highest standard deviation in the USA.

The East North Central ("ENC") division has both a low average migration index and a low standard deviation of the migration index. From the county map, it is full of mostly moderate migration counties, with a few exceptions, leaning toward out-migration, similar but attenuated compared to the Plains states.

The East South Central ("ESC") division is small geographically but is one of the most dynamic migration regions in the USA, with high average migration even though it contains the Mississippi Delta, the Black Belt, and the coal region of Eastern Kentucky, all areas of high out-migration due to poverty and isolation. High in-migration is seen in much of Alabama and Kentucky and almost all of Tennessee raising both the average and the standard deviation of the division migration index.

The New England ("NE") division has low to moderate average in-migration but a fairly homogeneous and small geography, with concomitantly low standard deviation.

The Middle Atlantic ("MA") division has slightly lower average in-migration than New England, but a noticeably higher standard deviation, due to the contrast seen between the high in-migration urban counties near the Atlantic and the rural out-migration counties in Western New York and Western Pennsylvania.

The South Atlantic ("SA") division is probably the most dynamic migration division in the country. High in-migration can be seen in almost all Virginia and Florida counties, the Northern half of Georgia, and the Rayleigh area in North Carolina. In-migration is scattered along the coast and in isolated spots throughout the rest of the

region as well. However, this division also includes West Virginia, with many out-migration counties and a balance of moderate migration counties. With the many out-migration counties dotting the Appalachian spine contrasting to high in-migration closer to the coast, the standard deviation is the second largest in the USA.

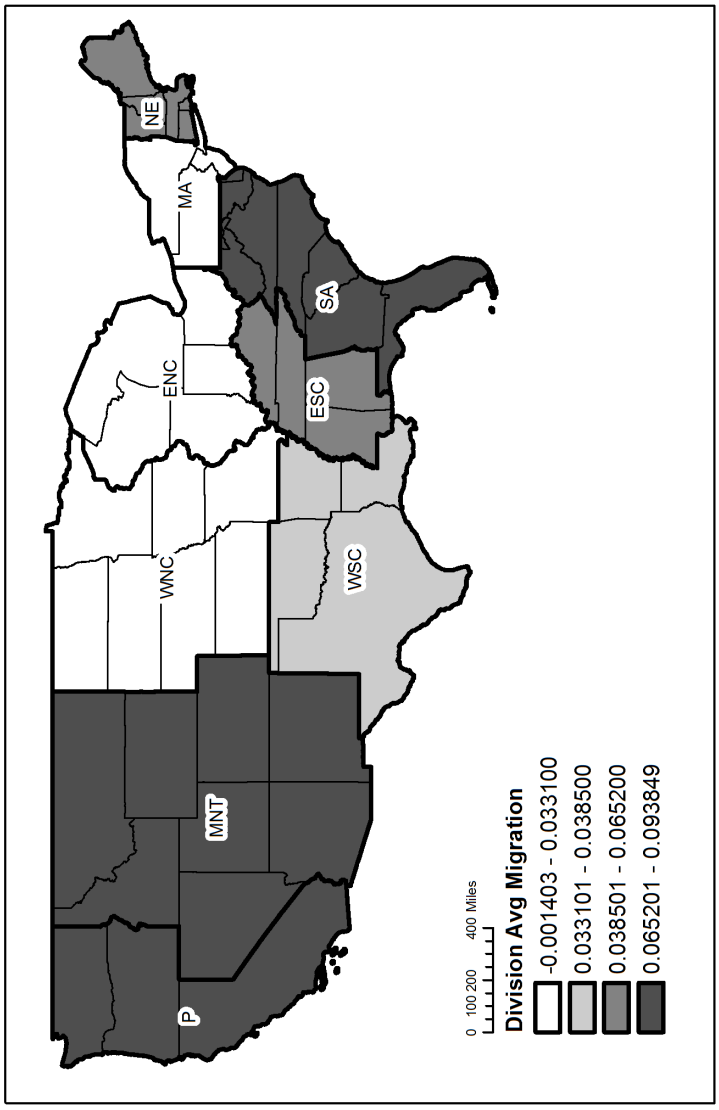


Figure 4.14: USA county migration average division map

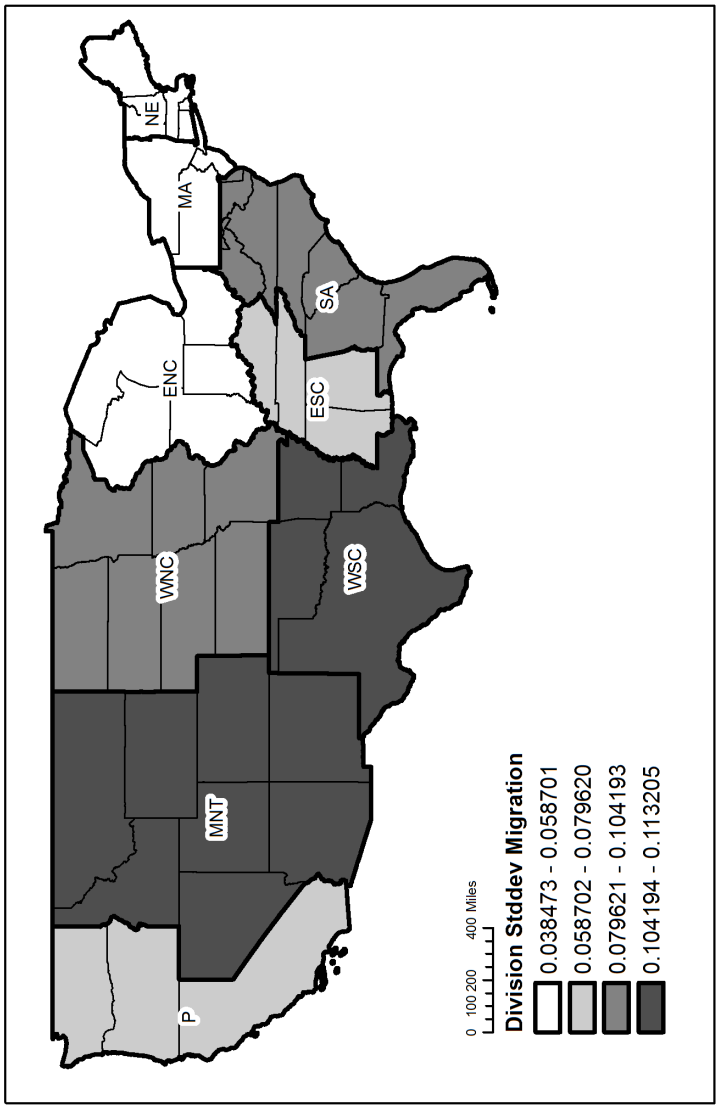


Figure 4.15: USA county migration standard deviation division map

Regional dynamics

When considered at the regional level, the West and South have high migration, the Northeast has moderate migration, and the Midwest has really low migration, as can be

seen from the Tables 4.3 through 4.5. The contributions to the West's high migration are almost all of Arizona as a high migration state, then many counties throughout the region with high local migration. We see a similar pattern in the South, with Florida contributing huge number of migrants and most of the other states with moderately high migration as well. The Northeast has moderate migration, largely due to no particularly strong out-migration or in-migration clusters. With a map this coarse, it is hard to distinguish any non-trivial trends, so we have not included it.

4.3.3 Ratio of variance

We tabulated the standard deviation and average of the total migration index by county for three nested levels of geography in order to determine whether there is any clustering of a propensity to migrate. We used Census Regions, Census Divisions, and States as our hierarchical geographic levels.

Figure 4.16 shows points representing the averages and standard deviations of all 48 states. The x-axis contains the average total migration index for a given state. The y-axis shows the standard deviation of total migration indices for the same state. Each point is drawn as a square, circle, triangle, or cross in order to distinguish whether the point belongs to one of the four Census Regions: Midwest, Northeast, South, or West, respectively. The ellipses encompass each set of points defined by a given region at two standard deviations, thus incorporating about 95% of the data. Each ellipse is labeled by its regional code: "MW" for Midwest, "NE" for Northeast, "S" for South, and "W" for West, and the numbers in parentheses in these labels denote the number of states within a region. The dashed ellipse represents the pattern of all the state data with no regional groupings. These ellipses are derived by calculating the eigenvectors of the covariance matrix of the total migration average and standard deviation for all states in a given regional grouping.

We see that for the Midwest region, the standard deviation shows a negative correlation with total migration, i.e. there is more variability of total migration within a state if the migration is low or negative. This negative correlation is surprising in that we expect faster growing geographies to have more variation. For all other regions, and for the all counties as a single group, variability is positively correlated with total migration.

We refrained from applying this technique to Census Division groupings because of the small numbers of data points in each Division would yield very busy graphics with

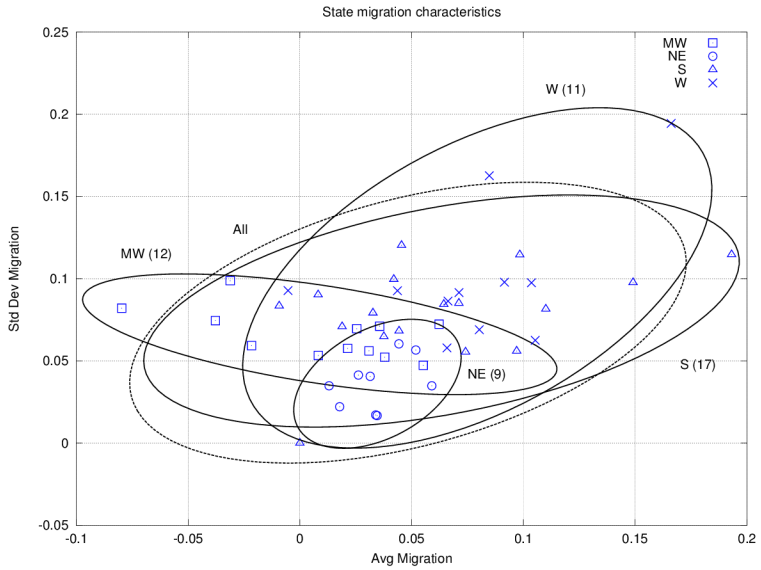


Figure 4.16: State migration average and standard deviation

tenuously derived ellipses.

We also show the tables below of state average total migration and standard deviation of total migration, sorted first by ascending region total average migration index, then by division total average migration index, and then by ascending state total average migration index. Tables 4.3 through 4.5 show these summary measures, sorted lexicographically by average migration index. The trends seem to correspond with the well known high growth in the South and West, low growth in the Midwest (along with the depopulation of rural counties in the plains states), and the moderate growth of the Northeast.

4.3.4 Migration index variance

It would be overreaching to apply ANOVA-type tests to this sort of data, but is well worth examining some exploratory data analysis graphics. Figures 4.17 through 4.19 show migration index boxplots for the same geographies discussed above – regions, divisions, and states.

Region	Avg	Stdev	Division	Avg	Stdev	State	Avg	Stdev
S	0.0630	0.1026	WSC	0.0361	0.1063	LA	-0.0092	0.0835
S	”	”	WSC	”	”	OK	0.0376	0.0648
S	”	”	WSC	”	”	AR	0.042	0.0996
S	”	”	WSC	”	”	TX	0.0455	0.1203
S	”	”	ESC	0.0478	0.0795	MS	0.0082	0.0902
S	”	”	ESC	”	”	AL	0.0327	0.0792
S	”	”	ESC	”	”	KY	0.0444	0.0682
S	”	”	ESC	”	”	TN	0.0970	0.0560
S	”	”	SA	0.0938	0.1042	WV	0.0189	0.0708
S	”	”	SA	”	”	SC	0.0644	0.0844
S	”	”	SA	”	”	VA	0.0711	0.0851
S	”	”	SA	”	”	MD	0.0741	0.0554
S	”	”	SA	”	”	GA	0.0984	0.1146
S	”	”	SA	”	”	NC	0.1100	0.0815
S	”	”	SA	”	”	DE	0.1491	0.0976
S	”	”	SA	”	”	FL	0.1932	0.1146

Table 4.2: Migration summary, Southern region

Region	Avg	Stdev	Division	Avg	Stdev	State	Avg	Stdev
MW	0.0129	0.0710	WNC	0.0045	0.0823	ND	-0.0797	0.0821
MW	”	”	WNC	”	”	NE	-0.0378	0.0745
MW	”	”	WNC	”	”	SD	-0.0312	0.0988
MW	”	”	WNC	”	”	KS	-0.0216	0.0593
MW	”	”	WNC	”	”	IA	0.0082	0.0533
MW	”	”	WNC	”	”	MN	0.0357	0.0711
MW	”	”	WNC	”	”	WI	0.0552	0.0473
MW	”	”	WNC	”	”	MO	0.0623	0.0723
MW	”	”	ENC	0.0287	0.0598	OH	0.0214	0.0576
MW	”	”	ENC	”	”	IL	0.0255	0.0696
MW	”	”	ENC	”	”	IN	0.0309	0.0561
MW	”	”	ENC	”	”	MI	0.0380	0.0523

Table 4.3: Migration summary, Midwest region

Region	Avg	Stdev	Division	Avg	Stdev	State	Avg	Stdev
NE	0.0324	0.0472	MA	0.0296	0.0505	NY	0.0131	0.0349
NE	”	”	MA	”	”	NJ	0.0315	0.0406
NE	”	”	MA	”	”	PA	0.0443	0.0604
NE	”	”	NE	0.0385	0.0385	RI	0.0178	0.0221
NE	”	”	NE	”	”	ME	0.0262	0.0414
NE	”	”	NE	”	”	CT	0.0339	0.0171
NE	”	”	NE	”	”	VT	0.0346	0.0168
NE	”	”	NE	”	”	MA	0.0519	0.0567
NE	”	”	NE	”	”	NH	0.059	0.0349

Table 4.4: Migration summary, Northeast region

Region	Avg	Stdev	Division	Avg	Stdev	State	Avg	Stdev
W	0.071	0.1005	MNT	0.066	0.1136	MT	-0.0053	0.0928
W	”	”	MNT	”	”	NM	0.0437	0.0927
W	”	”	MNT	”	”	WY	0.0663	0.0864
W	”	”	MNT	”	”	ID	0.0712	0.0916
W	”	”	MNT	”	”	NV	0.0848	0.1626
W	”	”	MNT	”	”	CO	0.0915	0.0979
W	”	”	MNT	”	”	UT	0.1036	0.0975
W	”	”	MNT	”	”	AZ	0.1661	0.1944
W	”	”	PAC	0.0814	0.0642	CA	0.0658	0.0580
W	”	”	PAC	”	”	OR	0.0803	0.0690
W	”	”	PAC	”	”	WA	0.1053	0.0625

Table 4.5: Migration summary, West region

Boxplots are a classic graphical technique invented by John Tukey in the 1970s (Tukey, 1977). As described in the Octave documentation, “The box plot is a graphical display that simultaneously describes several important features of a data set, such as center, spread, departure from symmetry, and helps identify observations that lie unusually far from the bulk of the data” (Eaton et al., 2012). We use the plots to examine these features at various levels of count total migration summaries.

The x-axis on our boxplots contains the codes for the various geographically groupings whose distribution is represented graphically. On the regional boxplot in Figure 4.17, the labels are “M” for the Midwestern region, “NE” for the northeast, “S” for the

south, and "W" for the West. It is useful to reference Figure 4.1 for a map of these geographies. The division labels are found on x-axis of Figure 4.18 as well, and here they are abbreviations of the names divisions shown on Figure 4.1 – for example "WNC" (leftmost label) corresponds to West-North-Central, the division including Kansas, Nebraska, the Dakotas, Missouri, Iowa, and Minnesota. Finally, the x-axis labels in Figure 4.19 correspond to the standard two character US Postal Service codes, "ND" for North Dakota etc.

In each box, the center line is the median, and the upper and lower edges are the 25th and 75th percentiles respectively; by examining whether the median line is closer to the top or bottom we can get a feel for the skewness of the distribution in that group. For our county migration index, each grouping – regional, divisional, and state – shows a fair amount of symmetry with the median bar pretty close to the median.

The boxplots are also ordered by median total migration index, with the median for the county grouping (region, division, or state) increasing left to right. For example, in Figure 4.17, we see that the central mark in the box climbs monotonically as we go from the the Midwest region with the lowest median total migration rightward to the West region with the highest median migration.

Each circle or cross-mark above a given label corresponds to a specific migration index in the grouping read off the y-axis. Points within the 25th and 75th percentile (points within the interquartile range – "IQR") are not shown directly but determine the height of the box, the width of the boxes represents how many of these points would be found, with wider boxes implying more points inside the IQR. Note how much wider the Midwest and South boxplots are in Figure 4.17, with geographically larger counties leading to smaller county counts in the West, and a smaller region altogether decreasing the counts in the Northeast. At the regional level the interquartile range seems to be pretty similar (IQR), though the median is quite different. At the divisional and state levels, the IQR is noticeably different between groups, however.

The notch on the boxplot gives a sort of confidence interval around the median. To paraphrase the Matlab documentation, the high and low corners of the notch correspond to $q2 \pm 1.57(q3\text{–}q1)/\sqrt{n}$, where $q2$ is the median (50th percentile), and $q1$ and $q3$ are the 25th and 75th percentiles. It is interesting to look at the various boxplots and consider where the confidence intervals would overlap to form groupings. At the regional level, the Midwest and the Northeast form a pair of low migration regions, with confidence intervals almost overlapping, while the South and the West form a pair of high migration regions with almost overlapping confidence intervals as well. At the Divisional level, we see that the West-North-Central is a zero migration region, due to the out-migration in the Great Plains. The East North Central and the Middle Atlantic form

one group of moderate in-migration divisions. The West South Central, Northeast, and East South Central form a slightly higher in-migration group. Finally, the Mountain, Pacific, and South Atlantic divisions form a group of very high migration. At the State level, we see that there are definite differences in median migration indexes over the counties, though it is harder to pick out any obvious groupings than in the other plots.

The boxplot whiskers extend to 1.5 times the IQR, and outliers are plotted as crosses or circles if they are past that range. The symbol for points that lie outside 3 times the interquartile range is 'o', and symbol for points between 1.5 and 3 times the interquartile range is '+'. In Figures 4.17 through 4.19, the only pattern we can distinguish is that most of the groupings have their outliers falling in high in-migration; this pattern would be consonant with the skewness toward positive migration present in the US seen above, especially in the histogram in Figure 4.8.

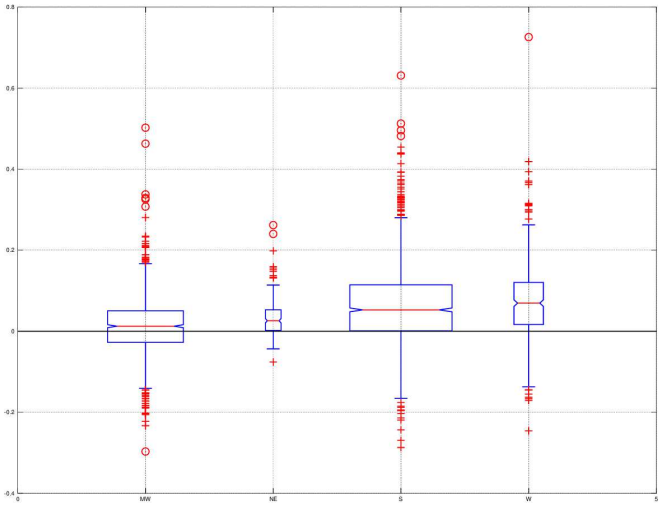


Figure 4.17: Region migration index boxplot

4.3.5 Semiovariogram

As another approach to understanding the spatial clustering of migration, we calculate a "semivariogram" of the total migration index.

The semivariogram describes the migration index variance as a function of distance between points, with the the distances binned into groups whose pair-wise distances fall

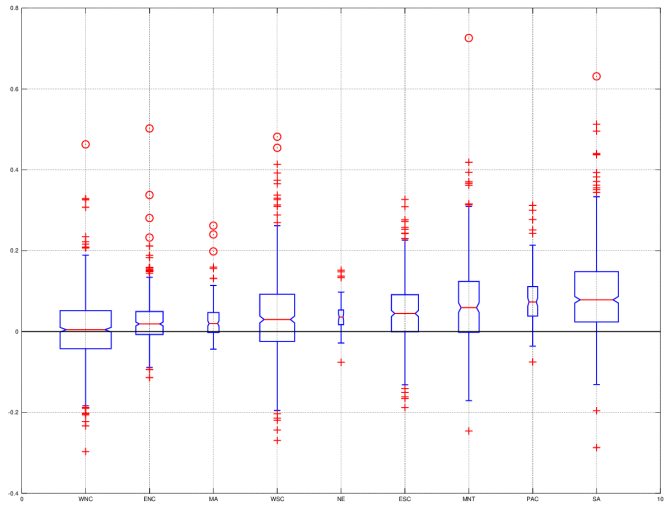


Figure 4.18: Division migration index boxplot

within one of a set of increasing intervals; for example, all counties within 0 to 25 miles of each other, all counties within 25 to 50 miles of each other, etc. This semivariance – $\gamma(h)$ in (4.2) – is plotted on the y-axis, and the distance h is plotted on the x-axis. Standard works on geostatistics give far more detail, such as (Isaaks and Srivastava, 1989). For reference, here is the equation for the semivariance at a give distance h:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [z(\mathbf{u}_\alpha + h) - z(\mathbf{u}_\alpha)]^2 \tag{4.2}$$

$z(u)$ is the "z-value" for an (x, y) pair represented by $\mathbf{u}$; in our case the z-value is the total migration index. h is the lag distance (25 miles in our case). $N(h)$ is the number of points at the lag h. $\gamma(h)$ is a vector with the semivariance computed for each lag h; h is also the distance between ticks on the x-axis on the semivariogram plot.

The semivariogram was developed in the context of minerals prospecting, where sampling cores are drilled to infer the shape of an "ore body below ground. The semivariogram attempts to discern the articulation point at which the point-to-point measurements transition from being clustered (indicating the presence of an ore body) to non-clustered (at the boundary of the ore body).

For the semivariogram, we need to calculate the pairwise distance between counties. To do this, we first load county boundaries from US Census TIGER geographic data into a PostGIS enabled database. We then compute centroids for each county, and

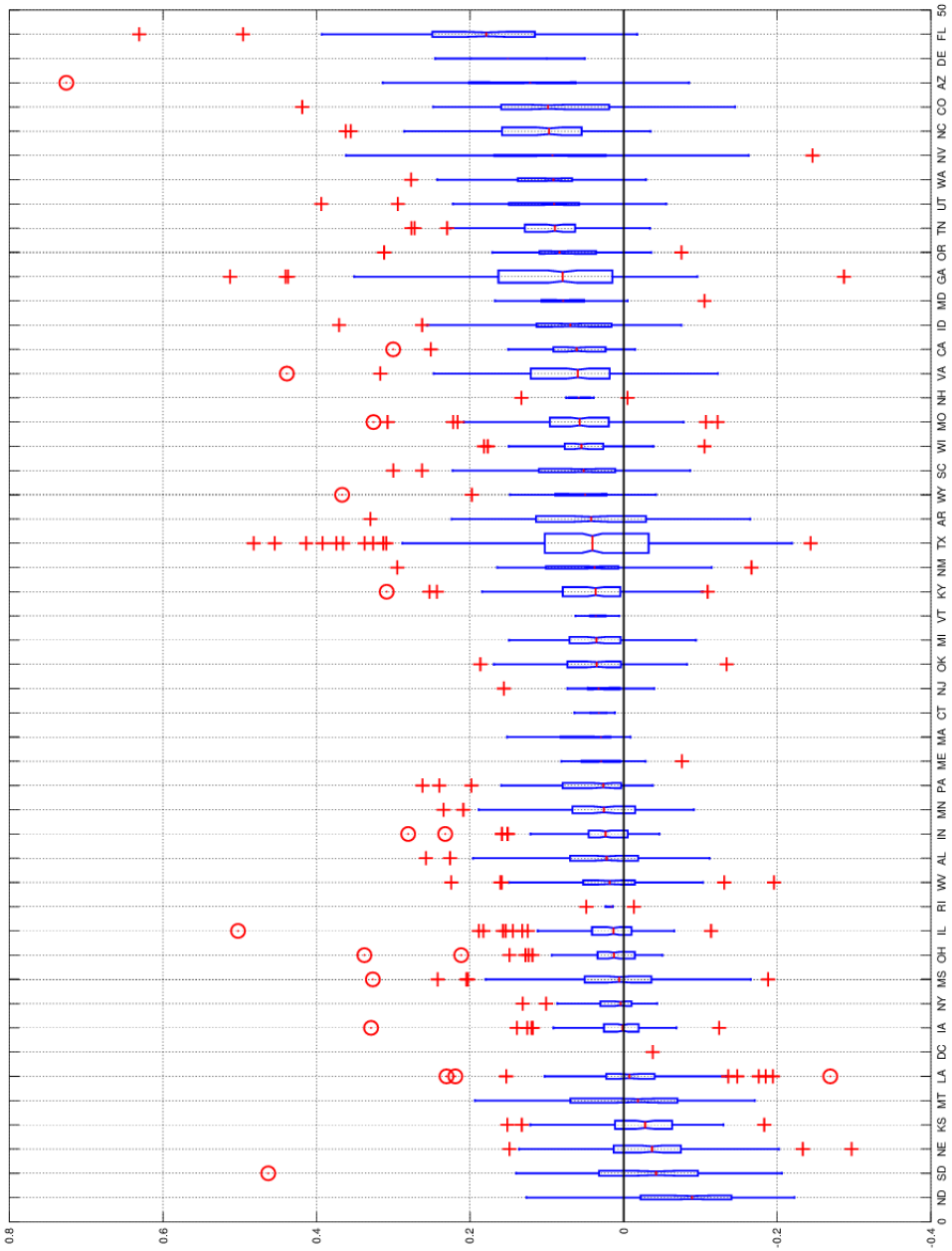


Figure 4.19: State migration index boxplot

calculate "geodesic" distances (distances which account for the curvature of the earth) between these centroids in miles, storing that result in a big distance matrix. We then import these distances into Octave in order to compute a semivariogram. We were unable to load all the county distance pairs due to memory limitations, so we restricted our dataset to pairwise distances less than 1,200 miles; since the USA (less Hawaii and Alaska) is about 3,000 miles across, any subnational clustering will be accounted for by this approach.

We present the calculated semivariogram in Figure 4.20. Each circle represents a set of counties a given distance apart. The x-axis represents the top cutoff for the distance defining a set of counties. The y-axis contains the semivariance of each county set. So the first circle, at 25 miles on the x-axis, represents the semivariance (about 0.0088) of all counties 0 to 25 miles distance from each other, the second circle shows the semivariance of all counties within 25 to 50 miles of each other, etc. For example, the first circle starting from the left on Figure 4.20 is the variance for the set of all migration indices which are between zero and 25 miles apart, the second circle is the variance of all migration indices 25 to 50 miles apart, etc

We see that semivariance climbs until 200 or so miles before it flattens, indicating clustering at distances below 200 miles. 200 miles is consonant with our discussion of region size in Section 4.1, in which 200 miles is about the greatest distance possible for a coherent labor and housing market.

After 400 miles, variance starts to decrease again, dropping to the end of the dataset at 1200 miles, with a slight flattening and uptick at about 600 miles. This decrease in variance after a peak is not typical behavior for a semivariogram; if autocorrelation and clustering were the only relationship between county migration indices, we would see a flattening with no subsequent drop. Perhaps the semivariogram presents evidence of a wider systems relationship in which county migration correlates due to long distance relationships rather than merely adjacent movers. In fact, the flattening at 600 might reflect the homogeneity or autocorrelation of out-migration in the midwestern Plains counties. Such a geographic relationship would stand in contrast to the purely adjacency-driven clustering we see in fertility behavior.

4.4 Conclusion

The migration index calculated above provides a new approach to understanding population flows. It yields reasonable results that are consonant with previous research,

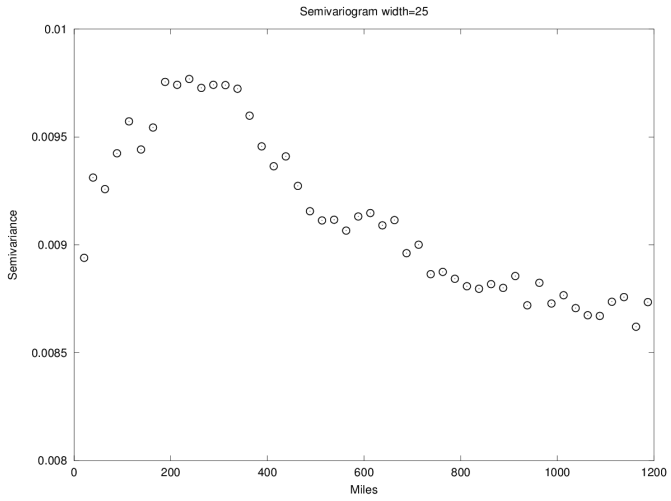


Figure 4.20: Migration index semivariogram

which gives us confidence in its utility as a method for future investigation. It is also far less labor intensive than other approaches, which require extensive tabulation in order to determine migration, such as performed by (Johnson et al., 2005), yet our technique yields similar results. Ours is also the first study that uses a standardized population rather than age-standardizing by calculating age specific rates.

The migration index also highlights the difference between the Midwest region and the rest of the US. In the Midwest, 60-and-over residents are not moving out the area, while 55-and-under residents seem to be fleeing in droves. Furthermore, in the Midwest variance is correlated negatively with growth within states; this odd situation may be related to the articulation points we see in error measures graphed in figures 2.4 through 2.6 and 2.19 through 2.21. In these graphs, negative growth counties have much higher error than positive growth counties, and the (negative) slope of that error is much higher than the slope of positive growth counties.

It is interesting that patterns of net migration flow correspond to conventional wisdom regarding the patterns of gross migration flow. While we lack data, we expect that net flows in any given direction are due to the balance of much larger gross flows in both directions, as formulated in Ravenstein's seventh law of migration, quoted in (Tobler, 1995): "Each main current of migration produces a compensating counter current." In the case of rural counties, especially in the Great Plains, it might be that the inflow and outflows are highly age specific, but without gross migration data we are unable to pursue this line of investigation any further.

5 Summary and further research

It is worth considering Wood's method as a whole, summarizing it's strengths and weaknesses as well as pointing toward further research.

5.1 Advantages to WM

Wood's Method, as presented here, has many advantages as a new demographic forecasting and modeling approach.

Wood's Method allows for far more flexible data inputs than typical age/ sex population forecasting systems by using parameterized constraints to set known rates, relationships between rates, or bounds on rates. Because of this approach, incremental changes to a model are possible, either as new data is available or in order to experiment with inputs. This system allows for prior knowledge to be easily incorporated into fitted models as well.

Wood's Method also avoids the "fiddly" assembly problems that can occur in less flexible forecast frameworks when inputs are changed. At Washington State, we have been successfully using a model to perform age allocation at the county level, but seemingly trivial changes to inputs can result in bizarre outputs of the model, and any input change has to be tested very carefully. Wood's method does not suffer from this brittleness, in our experience, since the optimization approach smooths out everything by searching for a good fit within the constraints. This robustness is very important in applied settings for practical reasons, and it also reflects a systems orientation that may be important theoretically.

The fitted matrix in Wood's Method can model arbitrary linear relationships, not just the renewal process used in this disseration. This flexibility may allow WM to tease out more complex population dynamics like feedback relationships due to density dependence.

Finally, Wood's Method has been shown to be feasible in investigating demographic attributes outside forecasting, including the migration dynamics examined in Chapter Four.

5.2 Problems to WM

Wood's Method suffers from overfitting when unconstrained because of the number of fitted coefficients – there are N^2 coefficients in every matrix used to project N state vectors. Relatedly, the fitted matrix may have a large norm due to large coefficients cancelling each other out with negative values.

Wood's Method is computationally challenging for large numbers of states due to large matrices, many states, and the use of iterative methods.

Wood's Method requires that an analyst have a basic understanding of optimization approaches not current in the social sciences, and an implementation still requires a fair amount of programming to effect.

We still do not understand the relationship between a fitted linear approximation and the underlying non-linear reality. At least this lack of understanding is true for all linear regression approaches in the social sciences, which puts us in good – or at least wide – company.

5.3 Moving forward with WM

We believe that the biggest potential for extending Wood's Method lies in incorporating the bicriterion optimization approaches described in Subsection 3.5.4 to reward parsimony and sparsity in the matrix fits. Rewarding parsimony is especially important as constraints on the fitted matrix are relaxed in order to model more complex population dynamics. Norm minimization is used with vector support machines to good effect; it is theoretically attractive in analogy to the energy minimization approaches underpinning variational mechanics.

We also hope to experiment with other models, especially feedback systems, using supplementary variables to model exogenous inputs and (finite) historical dynamics using elements of the matrix besides the diagonal, subdiagonal, and fertility row. The

Easterlin hypothesis of density dependence will be the first application of these techniques.

We also hope to improve the implementation to leverage sparsity and Lancosz methods to address problems in computational time and memory size.

In all, we believe that Wood's Method provides both an excellent forecasting methodology today and a promising avenue of research for the future.

Bibliography

Armstrong, J. S. and Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8(1):69–80.

Bertin, J. (1998). *Semiology of Graphics: Diagrams, Networks, Maps*. ESRI Press, Redlands, CA.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9:1196–1217.

Boyd, S. and Vandenberghe, L. (2005). *Convex Optimization*. Cambridge Univerity Press, Cambridge.

Brewer, J. W. (1978). Kronecker products in matrix calculus and systems theory. *IEEE Transactions on Circuits and Systems*, 25:772–781.

Bryan, T. (2000). US Census Bureau population estimates and evaluation with loss functions. *Statistics in Transition*, 4:537–549.

Burkey, J. (2013). LOWESS, locally weighted scatterplot smoothing for linear and nonlinear data (enhanced). `http://www.mathworks.com/matlabcentral/fileexchange/22470` (accessed February 2013).

Campbell, P. (2002). Evaluating forecast error in state population projections using census 2000 counts. *Population division working paper series*, 57.

Caswell, H. (2001). *Matrix Population Models: Construction, Analysis, and Interpretation, Second Edition*. Sinauer, Sunderland, Mass.

Chatfield, C. (2003). *The Analysis of Time Series: an Introduction*. Chapman and Hall/CRC, Boca Raton.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74:829–836.

Cleveland, W. S. (1981). Lowess: A program for smoothing scatterplots by robust locally weighted regressions. *The American Statistician*, 35:54.

Cleveland, W. S. and Devlin, S. J. (1988). Locally-weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83:596–610.

Coale, A. J. and Watkins, S. C. (1986). *The Decline of Fertility in Europe: the Revised Proceedings of a Conference on the Princeton European Fertility Project*. Princeton University Press, Princeton.

Dantzig, G. B. (1963/ 1998). *Linear Programming and Extensions*. Princeton University Press, Princeton.

Eaton, J. W. et al. (2012). Octave programming language. `http://www.octave.org`.

Florida, R. (2002). *The Rise of the Creative Class*. Perseus Book Group, New York.

Freedman, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics*, 9:1218–1228.

Girosi, F. and King, G. (2008). *Demographic Forecasting*. Princeton University Press, Princeton.

Golub, G. H. and Loan, C. F. V. (2012). *Matrix computations*. Johns Hopkins University Press, Baltimore.

Golubitsky, M., Keeler, E. B., and Rothschild, M. (1975). Convergence of the age structure: applications of the projective metric. *Theoretical Population Biology*, 7:94–93.

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

Hastings, A. (1997). *Population Biology: Concepts and Models*. Springer-Verlag, New York.

Hough, G. and Swanson, D. (2006). An evaluation of the american community survey: results from the Oregon test site. *Population Research and Policy Review*, 25:257–273.

Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22:679–688.

Isaaks, E. H. and Srivastava, R. M. (1989). *Applied Geostatistics*. Oxford University Press, Oxford.

Johnson, K. M., Voss, P. R., Hammer, R. B., Fuguitt, G. V., and McNiven, S. (2005). Temporal and spatial variation in age-specific net migration in the united states. *Demography*, 42:791–812.

Johnson, K. P. and Kort, J. R. (2004). 2004 redefinition of the BEA economic areas. *Survey of Current Business*, 84:68–75.

Judge, G. G. and Takayama., T. (1966). Inequality restrictions in regression analysis. *Journal of the American Statistical Association*, 61:166–181.

Keyfitz, N. and Caswell, H. (2005). *Applied Mathematical Demography*. Springer, New York.

Laub, A. J. (2005). *Matrix Analysis for Scientists and Engineers*. SIAM: Society for Industrial and Applied Mathematics, Philadelphia.

Lee, R. D. and Carter, L. (1992). Modeling and forecasting the time series of US mortality. *Journal of the American Statistical Association*, 87:659–671.

Lee, R. D. and Miller, T. (2001). Evaluating the performance of the Lee-Carter approach to modeling and forecasting mortality. *Demography*, 38:537–549.

Loan, C. V. and Pitsianis, N. (1992). Approximation with kronecker products. In Moonen, M. S. and Golub, G. H., editors, *Linear Algebra for Large Scale and Real Time Applications*, pages 293–314. Kluwer Publications.

Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria) (2012). Human fertility database. `http://www.humanfertility.org` (accessed July 2012).

McGranahan, D. A. (1999). Natural amenities drive rural population change. *USDA Agricultural Economic Report series*, 781.

Openshaw, S. (1983). The modifiable areal unit problem. *Concepts and Techniques in Modern Geography*, 38.

Parker, T. (2004, accessed on January 1, 2012). USDA county typology codes. `http://www.ers.usda.gov/data-products/county-typology-codes.aspx`.

Pittenger, D. (1976). *Projecting State and Local Populations*. Ballinger, Cambridge, Mass.

Preston, S., Hueveline, P., and Guillot, M. (2001). *Demography*. Blackwell, Malden, Mass.

Rathge, R. and Highman, P. (1998). Population change in the great plains. *Rural Development Perspectives*, 13:19–26.

Rayer, S. (2008). Population forecast errors: a primer for planners. *Journal of Planning Education and Research*, 27:417–430.

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

Rogers, A. (1994). *Introduction to Multiregional Mathematical Demography*. Wiley, New York.

Smith, S. K. (1987). Tests of forecast accuracy and bias for county population projections. *Journal of the American Statistical Association*, 82:991–1003.

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

Smith, S. K., Tayman, J., and Swanson, D. A. (2001). *State and Local Projections: Methodology and Analysis*. Kluwer Academic, New York.

Smola, A. J. and Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14:199–222.

Tobler, W. (1995). Migration: Ravenstein, thornthwaite, and beyond. *Urban Geography*, 16:327–343.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, New York.

University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany) (2012). Human mortality database. `http://www.mortality.org` (accessed July 2012).

US Census Bureau (2011). Census intercensal age and sex estimates. `http://www.census.gov/popest/intercensal/index.html` (accessed July 2011).

Voss, P. R., McNiven, S., Hammer, R. B., Johnson, K. M., and Fuguitt, G. V. (2004). County-specific net migration by five-year age groups, Hispanic origin, race and sex 1990-2000. *Center for Demography and Ecology, University of Wisconsin-Madison, Working Paper Series*, 2004-24.

Wachter, K. (1986). Ergodicity and inverse projection. *Population Studies*, 40:275–287.

Wachter, K. (2012). *Essential Demographic Methods*. Copy Central, Berkeley.

Whelpton, P. (1928). Population of the united states, 1925 to 1975. *American Journal of Sociology*, 34:253–270.

Wilmoth, J., Zureick, S., Canudas-Romo, V., Inoue, M., and Sawyer, C. (2012). A flexible two-dimensional mortality model for use in indirect estimation. *Population Studies*, 66:1–28.

Winkler, R. L., Johnson, K. M., Cheng, C., Voss, P. R., and Curtis, K. J. (2013). County-specific net migration by five-year age groups, Hispanic origin, race and sex 2000-2010. *Center for Demography and Ecology, University of Wisconsin-Madison, Working Paper Series*, 2013-4.

Wolfe, P. (1959). The simplex method for quadratic programming. *Econometrica*, 27:382–398.

Wood, S. N. (1997). Inverse problems and structured-population dynamics. In Tuljapurkar, S. and Caswell, H., editors, *Structured Population Models in Marine, Terrestrial, and Freshwater Systems*, pages 555–586. Chapman and Hall.

Wood, S. N. (2000). Modeling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society, Series B*, 62:413–428.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57:348–368.