Projection and Estimation of International Migration

Jonathan J. Azose

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Adrian E. Raftery, Chair

Adrian Dobra

Charles Hirschman

Program Authorized to Offer Degree:
Statistics

University of Washington

Abstract

Projection and Estimation of International Migration

Jonathan J. Azose

Chair of the Supervisory Committee:
Professor Adrian E. Raftery
Statistics

I propose techniques for improving both estimation and projection of international migration. By applying a Bayesian hierarchical modeling approach to net migration data, I produce projections of international migration that are global in scope and have well quantified uncertainty. My projections are of an appropriate form to be included as the migration component in probabilistic population projections, as has by done by Azose et al. (2016). (The current practice of the United Nations Population Division is to produce probabilistic population projections which include deterministic projections of migration.) The net migration model may be improved by incorporating a correlation matrix, but estimating such a matrix is difficult because the dimension of the matrix is large while the number of available data points is small. I demonstrate a method for estimating a correlation matrix which includes a prior belief that correlations which are large in magnitude are more likely among countries which are "close", either because of geographical or historical ties. Including correlations improves projections when net migration is aggregated across regions. I also propose a method for improving existing estimates of bilateral migration flows based on migrant stock data. A current state-of-the-art estimation method (Abel, 2013a) relies on an unrealistic assumption that the total number of migrants is as small as possible, resulting in estimates with many structural zeroes. By weakening that assumption, I produce estimates of migration flows between all pairs of countries that allow for substantial return migration flows.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

AR: autoregressive

ARIMA: autoregressive integrated moving average

BHM: Bayesian hierarchical model

CEPII: *Centre d'Etudes Prospectives et d'Informations Internationales*

CRPS: continuous ranked probability score

EU: European Union

IFE: independent forecast errors

JAGS: Just Another Gibbs Sampler

LDC: least-developed countries

LPOC: Laplace prior on correlations

MAE: mean absolute error

MAP: maximum *a posteriori*

MCEM: Monte Carlo expectation maximization

MCMC: Markov chain Monte Carlo

MIMOSA: Migration Modeling for Statistical Analysis

MM: minimum migration

MSE: mean squared error

OECD: Organization for Economic Co-Operation and Development

OLS: ordinary least squares

PB: pseudo-Bayes

RMSE: root mean squared error

SE: squared error loss

SEL: squared error loss in log-flows

SER: squared error loss in flow rates

UN: United Nations

UNPD: United Nations Population Division

UK: United Kingdom

US OR USA: United States of America

WPP: World Population Prospects

ACKNOWLEDGMENTS

First and foremost, I want to thank Adrian Raftery for his mentorship. His unflagging support and guidance made this research possible. Thank you also to Adrian Dobra and Charlie Hirschman, for both serving on my supervisory committee, and offering your advice and expertise.

I am grateful for financial support from the University of Washington Department of Statistics and Center for Statistics and the Social Sciences; from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (Grants R01 HD54511 and R01 HD70936); and from Science Foundation Ireland (ETS Walton visitor award, grant reference 11/W.1/I2079).

Finally, I am grateful to my wife, Ellen, and my daughter, Nora, who have been, respectively, an invaluable pillar of support and only a mild hindrance.

DEDICATION

There are always, in each of us,
these two: the one who stays,
the one who goes away —

But I am the one
who always goes away.

Sujata Bhatt

To the migrants.

Chapter 1

INTRODUCTION

1.1 Motivation

Migration looms large in today's political climate. Donald Trump, in announcing his candidacy in the 2016 presidential election, promised that "I will build a great, great wall on our southern border. And I will have Mexico pay for that wall" (Valverde, 2016). Just a week later, the British public narrowly voted for the United Kingdom to leave the European Union. Prime Minister David Cameron asserted that if the EU had allowed him to control migration to the UK, Britain's exit from the union could have been avoided (Parker et al., 2016). In 2015, Germany took in nearly two million migrants (Grobecker & Brückner, 2016), including 1.1 million asylum seekers (Al Jazeera America, 2016), prompting consternation from German right wing parties (Faiola, 2016). Meanwhile, Japan faces an aging population with a shrinking labor force, and cabinet ministers disagree on whether their historically strict immigration policy should be loosened to allow additional labor migration (Yoshida, 2015).

Despite the great political and demographic importance of migration, estimates and projections of migration continue to be elusive. In the countries with the best data collection infrastructure, estimates of migration are often available from population registers, census data, demographic and health surveys, or other administrative records. Reliable record-keeping along these lines is typically available only in the developed world. However, even in Europe, where many countries track migration via detailed population registers, migration estimates can be of poor quality (Abel, 2010; Nowok et al., 2006). In places where administrative data sources are unavailable or unreliable, migration can be estimated as the residual term in the population balancing equation; any change in population which is not at-

tributable to births or deaths must be due to migration. The practice of the United Nations Population Division (UNPD) is mostly to estimate migration using this residual method (United Nations, 2015b). A benefit of the residual approach is that population, births, and deaths tend to be easier to measure directly than migration, so residual estimates of migration are usually of good quality. A downside is that this method necessarily produces estimates of *net* migration—that is, the total number of in-migrants minus the total number of out-migrants. The use of net migration in population modeling is somewhat controversial (Rogers, 1990), and obscures interesting information about where migrants are moving to and from.

Much of the research in this dissertation is motivated by the goal of improving the United Nations Population Division's treatment of migration. The UNPD produces a biennially revised data set called the World Population Prospects (WPP). This data set contains estimates and projections of a variety of demographic quantities, including migration, but also variables such as total fertility rate, life expectancy at birth, and potential support ratio. Currently, the only migration statistics contained in the WPP are net migration counts and rates for all countries.

A desire to produce more detailed bilateral migration flow estimates, alleviating our current dependency on net migration, motivates the work in Chapter 4 of this dissertation. Detailed sets of flow estimates do exist in limited geographic contexts, typically for flows where at least one of the sending and receiving country is economically advanced (Organization for Economic Co-Operation and Development, 2015; Raymer et al., 2011). We provide a method which produces global flow estimates for all countries, including flows between developing countries.

As of the 2015 revision of the WPP (United Nations, 2015b), the UNPD practice is to produce probabilistic population projections by combining probabilistic projections of fertility and mortality with deterministic migration projections. Without probabilistic migration projections, the resulting population projections are not fully probabilistic. Consequently, the current population projection methodology understates uncertainty substantially, espe-

cially in Europe and North America (Azose et al., 2016). A key focus of this dissertation is producing well calibrated probabilistic migration projections which can be integrated into the WPP framework for probabilistic population projection. Details of this research can be found in Chapters 2 and 3 of this dissertation, as well articles published in *Demography* (Azose & Raftery, 2015) and the *Proceedings of the National Academy of Sciences* (Azose et al., 2016).

1.2 Background

We now provide some brief background on migration theory and population modeling. This section is based in part on the article by Azose et al. (2016). Additional background targeted to more specific subtopics can be found in the Introduction sections of each body chapter (Sections 2.1, 3.1, and 4.1.)

1.2.1 Theories of migration

Bijak (2010) gives a thorough overview of theories and models of international migration. He classifies theories of migration according to their most closely corresponding scientific discipline (i.e. sociological, economic, or geographic theories). As the work in this dissertation focuses on modeling migration rather than explaining its causes, I find it useful to instead categorize theories of migration according to the type of modeling approach that suits them. To that end, we might draw a distinction between micro-level theories of migration, which deal with decisions to migrate at the level of individuals, households, or small networks thereof, and macro-level theories of migration, which deal with aggregated indicators of migration (e.g. migration flows, net migration, or migrant stocks). Micro-level theories lend themselves well to agent-based modeling approaches while macro-level theories are better suited to population-level statistical inference methods like the Bayesian hierarchical models we will use for net migration.

The cornerstone of many micro-level migration theories is the idea of push and pull factors (Lee, 1966). In this framework, a decision to migrate is a rational choice based on

an individual's subjective assessment of the benefits available at the potential destination and the negatives associated with staying at the current location. Personal preferences and circumstances weigh into these subjective assessments as well; even with the same external factors, not all individuals will make the same decision about whether the expected benefits of migration outweigh the costs. Push factors are sometimes subdivided into "hard" push factors, dramatic events like war and famine that necessitate immediate action, and "soft" push factors, like poverty, which pose less of an immediate crisis (Öberg, 1996). Migrant characteristics may differ depending on whether migration is driven by hard or soft push factors. Hard push factors may produce a broader cross-section of migrants than soft ones, which tend to have a positive selection effect.

Other micro-level theories go into more depth on specific push and pull factors. A microeconomic approach says that individuals will migrate when the expected earnings at the destination exceed the costs of migration (Sjaastad, 1962; Harris & Todaro, 1970). It should be noted that migration costs include not only the money paid to migrate, but also opportunity costs associated with lost work, and psychological costs (which economists may nevertheless translate into monetary value.) Migration may also be undertaken as a way to mitigate economic risk at the household level (Stark & Bloom, 1985; Stark, 1991). Other factors which have been shown to be substantially associated with individual migration decisions are migration policy (Massey & Espinosa, 1997), geopolitical conflict (Castles & Miller, 2003), and quality of the natural environment (Myers, 2002; Reuveny & Moore, 2009).

Despite the extensive list of factors that cause individual desires to migrate, the high costs associated with migration are prohibitive for many individuals. A Gallup poll from 2011 found that although an estimated 630 million adults globally had a desire to make a permanent international move, only 19 million were preparing for such a move (Esipova et al., 2011).

On the macro-level, one of the strongest predictors of migration flows is the existence of an established migrant population. Networks of migrants provide a feedback mechanism such that migration flows tend to perpetuate themselves over time (Massey, 1990; Hatton &

Williamson, 1998). An existing population of compatriots at the destination of an international move serves as a pool of social capital, effectively lowering the costs associated with migration. Zipf (1946) provides the prototypical gravity model of migration, relating migration flows to the populations of origin region, destination region, and the distance between them with a relation akin to that of the Newtonian gravitational force.

A more macroeconomic argument is that migration occurs in response to differences in global supply and demand for labor (Wallerstein, 1974; Portes & Walton, 1981). An oversupply of labor may raise unemployment rates, making migration a more attractive option. Conversely, a country with a labor shortfall may institute policies to encourage new migration to replace the missing labor force. Billari & Dalla-Zuanna (2011) find empirical evidence that such replacement migration is currently taking place in Europe.

Despite their acknowledged role in driving migration, we will almost entirely eschew push and pull factors, economic or otherwise, as covariates in our models. Such factors are largely too difficult to predict to be of use in long-term projections. Instead we appeal to the inertia of self-perpetuating migration patterns. In the projection domain, we do so by modeling migration as an autoregressive process; in the estimation domain, we allow data about existing migrant stocks to influence flow estimates.

1.2.2 Modeling migration and population

Historically, most methods for projecting population have been deterministic. If the current population is known, broken down by age and sex, and future age- and sex-specific rates are projected for fertility, mortality and migration, then the cohort-component method produces population projections (Leslie, 1945). The UN Population Division now produces probabilistic projections of population, fertility, and mortality for all countries, but these projections still condition on deterministic migration projections (Raftery et al., 2012; United Nations, 2015b). The current methodology in the UN's World Population Prospects differs from country to country, but typically projects that net migration counts will remain constant until 2050 and drop deterministically to zero by 2150 (United Nations, 2015b). A deterministic

gravity model that assumes migration is proportional to population size raised to some power (Cohen et al., 2008; Kim & Cohen, 2010) is more flexible than the simpler WPP migration projections, but still lacks quantification of uncertainty. Probabilistic population projection models that account for migration uncertainty have been developed for a small number of countries, typically only those with good data (Wiśniowski et al., 2015). One of the key contributions of our work is a method for projecting migration that produces projections with uncertainty for all countries.

1.3 Outline of the Dissertation

The remainder of this document is organized as follows.

Chapter 2 presents a method for projecting net international migration rates for all countries using a Bayesian hierarchical model. This model is extended in Chapter 3 to include between-country correlations. Chapter 4 focuses on estimation of international migration, rather than projection, providing a method for improving estimates of bilateral migration flows between all pairs of countries. Finally, Chapter 5 summarizes my work and provides ideas for future directions of research.

Chapter 2

BAYESIAN PROBABILISTIC PROJECTION OF INTERNATIONAL MIGRATION

2.1 Introduction

In this chapter we propose a method for probabilistic projection of net international migration counts and rates. Our technique is a simple one that nonetheless overcomes some of the usual difficulties of migration projection. First, we produce both point and interval estimates, providing a natural quantification of uncertainty. Second, simulated trajectories from our model satisfy the common sense requirement that worldwide net migration sum to zero for each sex and age group. Third, our projected trajectories approximately replicate the observed frequency of countries switching between positive and negative net migration. Lastly, we sidestep the difficulty in projecting a complete large matrix of pairwise flows by instead working directly with net migration. Sample projections from our model for several countries are given in Figure 2.1.

In the remainder of the introduction, we provide background and describe global trends in migration. In the next section we describe our data and methods for producing probabilistic projections. This is followed by a summary of our main results, including an evaluation of our model's performance and what our projections predict about future global migration trends. Finally, we conclude with evaluative discussion.

This chapter is closely based on the article by Azose & Raftery (2015) published in *Demography*. Projected migration rates and counts for all countries are available as supplemental material to the published article.

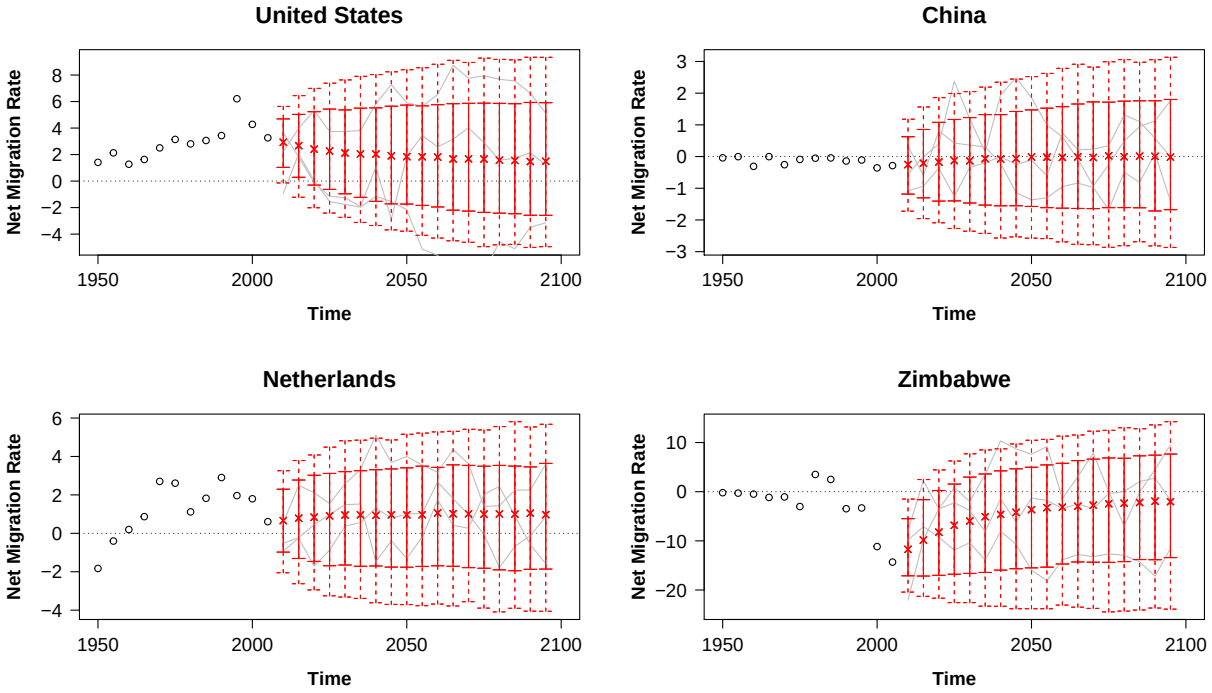

Figure 2.1: Probabilistic projections of net international migration rates: Predictive medians (indicated by "x"), 80% (solid vertical lines), and 95% (dashed vertical lines) prediction intervals for four countries, with example trajectories included in gray, and past observations shown as black circles. Rates are annualized and per thousand individuals in the specified country

2.1.1 Motivation and Background

There is a clear demand for migration projections. Organizations such as the United Nations, the UK Office for National Statistics, and the U.S. Social Security Administration have identified a necessity for migration forecasts (United Nations, 2011; U.S. Social Security Administration, 2013; Wright, 2010).

Our work is motivated by the needs of the UN Population Division in producing probabilistic population projections for all countries. The UN has recently adopted a Bayesian approach to projecting the populations of all countries as the basis for its official medium projection, and issued probabilistic population projections for all countries for the first time

in July 2014 (Raftery et al., 2012; United Nations Population Division, 2014). The underlying method can account for uncertainty about fertility and life expectancy through Bayesian hierarchical models (Alkema et al., 2011; Raftery et al., 2013). However, the approach does not yet take account of uncertainty about international migration. Instead, the UN probabilistic population projections are conditional on deterministic migration projections that essentially amount to assuming that current migration levels will continue into the medium term. To make the method fully probabilistic would require probabilistic projections of net international migration for all countries.

Lutz & Goldstein (2004), in answering the question of how to deal with uncertainty in population forecasting, pointed to the need for simple approaches to probabilistic forecasting of migration. Our article attempts to meet this need. Despite the demand, some experts have been pessimistic about the possibility of predicting migration at all. For example, ter Heide (1963) argued that the task of finding a usable model for migration is "virtually impossible." Bijak & Wiśniowski (2010:793–794) updated this opinion, drawing the similarly disheartening conclusions that "migration is barely predictable" and "forecasts with too long horizons are useless."

Nevertheless, there have been efforts to forecast international migration. These attempts have mostly been limited in geographic and/or chronological scope. Bijak & Wiśniowski (2010) produced migration projections for seven European countries to 2025 using Bayesian hierarchical models. Using another geographically focused method, Fertig & Schmidt (2000) projected migration flows from a set of 17 mostly European countries to Germany over the 1998–2017 period. One drawback of these two approaches in the context of population projections for *all* countries is that both require the use of data on migration flows between pairs of countries. Estimates of reasonable quality of these flows are now available for most pairs of European countries (Abel, 2010), making such techniques feasible for Europe and probably also for other developed regions. Estimates for global pairwise migration flows are also available (Abel, 2013a), but the quality of these estimates varies with the reliability of record keeping in the countries involved.

Hyndman & Booth (2008) provided another forecasting method: a stochastic model for indirect migration forecasting by forecasting fertility and mortality, with migration taken to be the appropriate quantity to satisfy the balancing equation. Their method provides estimates for individual countries for which reliable age- and sex-specific estimates of fertility, mortality, and migration are available. However, their method is not suitable for many of the world's countries, where such detailed breakdowns are either unavailable or unreliable. The 2012 revision of the United Nations *World Population Prospects* 2012 took a simpler approach by including point projections that generally project migration counts to persist at or near current levels for the next couple of decades and drop deterministically to zero in the long horizon. Cohen (2012) provided a method for point projections of migration counts for all countries using a gravity model. See Bijak (2006) for a review of other methods.

2.1.2 Theory of International Migration

There is a general consensus about the major causes of international migration. On the individual level, desire to migrate is caused largely by economic factors (Esipova et al., 2011; Massey et al., 1993). Refugee movements may be precipitated by political or social factors rather than economic ones (Richmond, 1988). However, both economic and political factors are unlikely to be predictable in the long run with any useful degree of certainty. For the purposes of projection, Kim & Cohen (2010) argued for the use of more predictable demographic variables in place of less predictable economic ones. They proposed a model for prediction of migration flows that incorporates life expectancy, infant mortality rate, and potential support ratio as predictor variables. Kim & Cohen found these variables to be significant predictors of migration flows. Furthermore, because demographic variables tend to change much more slowly than economic or political ones, it is often possible to project the values of demographic variables decades into the future with less uncertainty. Our model projects net migration on the basis of only past migration figures and an initial projection of populations for all countries, for which forecasts can be made with enough precision to be useful.

One additional demographic variable of interest in modeling migration is age structure, which is important to migration modeling in two different ways. First, projected age structures for all countries can potentially be used as predictor variables in projections of future migration. Because labor migration is common, the age structure of the sending and/or receiving countries can be used in making projections (Fertig & Schmidt, 2000; Hatton & Williamson, 2002, 2005). Kim & Cohen (2010), in a study of pairwise migration flows, found that a young age structure in the country of origin is associated with high migration flows, while a young age structure in the country of destination is associated with low flows.

Second, it may be of interest to project not only net migration counts but also *age-specific* net migration counts. Rogers & Castro (1981) provided a parametric multiexponential model migration schedule that can be used in converting from projected net migration counts to age-specific counts. Their model incorporates a principal migration peak among young adults, who often migrate for reasons of economics, marriage, or education, as well as a secondary childhood peak for the children of those young adult migrants. The model includes an additional option for waves of retirement and post-retirement migration, which are common patterns of regional migration but are less common internationally. Use of these model migration schedules can be particularly problematic when working with net migration rather than inflows and outflows (Rogers, 1990), but they may still provide a first-order approximation of age structures when no better data are available. Raymer & Rogers (2007) noted the complication that the age structure of a migrating population is dependent on direction of migration. For example, we would expect a labor migration and a subsequent return migration to have different age structures. This can be taken into account to some extent in a model like ours if data on the age structure of recent net migration are available.

For projection purposes, Bayesian modeling is well suited to modeling international migration. The difficulty in making accurate point projections emphasizes the need for an approach that produces estimates of uncertainty. Because our data set includes only 12 time points per country, non-Bayesian inference could be difficult; the Bayesian approach alleviates this by allowing us to borrow strength across countries and to incorporate prior

knowledge. Studies with limited geographical scope confirm this intuition. In a comparison of several methods for forecasting migration to Germany, Brücker & Siliverstovs (2006) found performance of a hierarchical Bayes estimator to be superior to that of simpler estimators based on ordinary least squares regression, fixed effects, or random effects. Well-calibrated results have come out of Bayesian forecasting efforts for fertility and mortality (Alkema et al., 2011; Raftery et al., 2012, 2013). In addition to forecasting, estimation of demographic variables also lends itself to Bayesian methodology (Abel, 2010; Congdon, 2010; Wheldon et al., 2013).

2.2 Methods

2.2.1 Data

We use data from the 2010 revision of the United Nations Population Division's biennial *World Population Prospects* (WPP) report (United Nations, 2011). WPP reports contain estimates of countries' past age- and sex-specific fertility, mortality, and net international migration counts and rates, as well as projections of future migration.

Our work is motivated by a desire to incorporate probabilistic migration projections into probabilistic population projections. Thus, the quantity we are interested in forecasting is $y_{c,t}$, the net number of migrants in country c in time period t. Because net migration is sufficient to determine population change due to migration, we need not consider inflows and outflows separately. We condition on known population projections, $\tilde{n}_{c,t}$, taken from the WPP 2010 revision. So long as projected populations are known, we can freely convert between net migration counts, $y_{c,t}$, and net migration rates, $r_{c,t}$. In the WPP data, rates are reported in units of migrants per thousand individuals in the specified country.[1]

[1]Strictly speaking, this is not a rate. A rate should divide counts of some event by the population exposed to risk of that event. Here, if a country is a net receiver, the real exposed population is that of the rest of the world rather than the population of country c, so our "rate" doesn't have the correct exposed population in the denominator. Nevertheless, we follow convention and continue to call this a net migration rate, even though the terminology is controversial. This convention is fairly widely used, including in the WPP.

2.2.2 Probabilistic Projection Method

Our technique is to fit a Bayesian hierarchical first-order autoregressive, or AR(1), model to net migration *rate* data for all countries. Recall that our motivation is to obtain probabilistic migration projections for incorporation into population projections for all countries—an application that requires projected net migration *counts* rather than *rates*. Nevertheless, it is advantageous to model on the rate scale and convert the output to counts rather than modeling counts directly. The primary disadvantage to modeling net migration counts is that variability in count data grows roughly in proportion to population size. This suggests dividing counts by population sizes as a way of stabilizing the variance, resulting in a model on migration rates.

We model the migration rate, $r_{c,t}$, in country c and time period t as

$$(r_{c,t} - \mu_c) = \phi_c(r_{c,t-1} - \mu_c) + \varepsilon_{c,t},$$

where $\varepsilon_{c,t}$ is a normally distributed random deviation with a mean of zero and a variance of σ_c^2. We put normal priors on each country's theoretical long-term average migration rate μ_c, and a uniform prior on the autoregressive parameter ϕ_c. Under this model, simulation of trajectories requires us to estimate or specify values of μ_c, ϕ_c, and σ_c^2 for all countries; thus, the complete parameter vector is given by $\boldsymbol{\theta} = (\mu_1, \ldots, \mu_C, \phi_1, \ldots, \phi_C, \sigma_1^2, \ldots, \sigma_C^2)$, where C is the number of countries.

The full specification of the model, including prior distributions, is as follows:[2]

$$\text{Level 1} \begin{cases} (r_{c,t} - \mu_c) = \phi_c(r_{c,t-1} - \mu_c) + \varepsilon_{c,t} \\ \varepsilon_{c,t} \overset{\text{ind}}{\sim} N(0, \sigma_c^2) \end{cases}$$

[2]Other sensible choices of prior yield very similar results. For example, fixing $\lambda = 0$ and taking $\sigma_c^2 \sim IG(0.001, 0.001)$ both produce only small changes in predictions. We incorporate an extra level of hyperpriors in part to encourage more shrinkage of parameter values toward a global mean. Additionally, more informative priors would be possible if one wished to incorporate knowledge from other sources, such as region-specific knowledge of means or variances in migration.

$$\text{Level 2} \begin{cases} \phi_c \overset{\text{iid}}{\sim} U(0,1) \\ \mu_c \overset{\text{iid}}{\sim} N(\lambda, \tau^2) \\ \sigma_c^2 \overset{\text{iid}}{\sim} IG(a,b) \end{cases}$$

$$\text{Level 3} \begin{cases} a \sim U(1,10) \\ b|a \sim U(0, 100(a-1)) \\ \lambda \sim U(-100, 100) \\ \tau \sim U(0, 100), \end{cases}$$

where $X \sim N(\mu, \sigma^2)$ indicates that the random variable X has a normal distribution with a mean of μ and a variance of σ^2 (and hence a standard deviation of σ), $U(c,d)$ denotes a uniform distribution between the limits c and d, and $IG(a,b)$ denotes an inverse gamma distribution with probability density function (as a function of x) proportional to $x^{-a-1}e^{-b/x}$.

We obtain draws from the posterior distributions of all parameters using Markov chain Monte Carlo methods. In our implementation, we use the Just Another Gibbs Sampler (JAGS) software package for Markov chain Monte Carlo simulations (Plummer, 2003).

Having obtained a sample $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N)$ of draws from the joint distribution of the parameters, we use these draws to obtain a sample from the joint posterior predictive distribution. For each sampled value $\boldsymbol{\theta}_k$ from the joint posterior distribution of the parameters, we first simulate a set of joint trajectories $\tilde{r}_{c,t}^{(k)}$ for net migration rates at time points until 2100, where k indexes the trajectory. However, this procedure generally produces trajectories that are impossible in that they give nonzero global net migration counts. We therefore create corrected trajectories for net migration counts and rates using the following method:

1. On the basis of the parameter vector $\boldsymbol{\theta}_k$, project net migration rates for all countries a single time point into the future. Denoting the next time period in the future by t', this allows us to obtain a collection of (uncorrected) projected values $\tilde{r}_{c,t'}^{(k)}$ for all countries c.

2. Convert net migration rate projections $\tilde{r}_{c,t'}^{(k)}$ to net migration count projections $\tilde{y}_{c,t'}^{(k)}$. To

convert from rates to counts, we multiply the rate $r^{(k)}_{c,t'}$ by the projected average population. Projected average populations are taken from the deterministic population projections in WPP 2010 (United Nations, 2011).

3. Further break down migration counts by age a and sex s to obtain estimates of net male and female migration counts for all countries and age groups, $\tilde{y}^{(k)}_{c,t',a,s}$. This is done by applying projected migration schedules to all countries. For the projections in this article, we take each country's projected age- and sex-specific migration schedule to be the same as the distribution of migration by age and sex in the most recent time point for which detailed data are available for that country.

4. For each simulated trajectory, within each age and sex category, apply a correction to ensure zero worldwide net migration. The correction we apply redistributes any overflow migrants to all countries, in proportion to their projected populations. Specifically, we take the corrected migration count projection $\tilde{y}^{*(k)}_{c,t',a,s}$ to be

$$\tilde{y}^{*(k)}_{c,t',a,s} = \tilde{y}^{(k)}_{c,t',a,s} - \frac{\tilde{n}_{c,t'}}{\sum_{j=1}^{C} \tilde{n}_{j,t'}} \sum_{j=1}^{C} \tilde{y}^{(k)}_{j,t',a,s}.$$

5. Convert the corrected age- and sex-specific net migration counts $\tilde{y}^{*(k)}_{c,t',a,s}$ back to corrected net migration rates $\tilde{r}^{*(k)}_{c,t'}$ by aggregating and converting counts to rates. In practice, the corrections from the previous step are typically small on the net rate scale. In more than 95% of cases, the resulting change in countries' projected net migration rates $\tilde{r}^{*(k)}_{c,t'}$ is less than 0.2 net annual migrants per thousand.

6. Continue projecting trajectories one time step at a time into the future by repeating steps 1–5.

Although the uncorrected net migration rates $\tilde{r}_{c,t'}$ come from the desired marginal posterior predictive distributions, the correction in step 4 changes those distributions by project-

ing them onto a lower-dimensional space. Sensitivity analysis suggests that the correction introduces only minor changes between the marginal distributions with and without the correction.

Also worth noting is that the projected net migration rates from our method are not very sensitive to changes in the population projections $\tilde{n}_{c,t'}$, justifying the use of fixed WPP 2010 population projections that include migration. It would be possible instead to project all components of population change simultaneously, including migration.

Probabilistic projections of net migration rates and counts for all countries for the time periods from 2010 to 2100 are included as Online Resources 1 and 2 to the article by Azose & Raftery (2015) published in *Demography*.

2.3 Results

2.3.1 Evaluation

We evaluate projections in the form of net migration rates. In the modeling stage, the choice to use rates rather than counts was mathematically motivated. A model on net migration counts would have required variance proportional to population size, a complication that is not necessary on the rate scale. The choice to evaluate on rates rather than counts is motivated by the application we have in mind. The goal is to produce migration projections for all countries in response to the needs of the UN Population Division. Evaluation on the count scale would effectively give heavy weight to our model's performance on a small number of high-migration countries. To better assess performance on all countries, we work instead on the rate scale.

We do not know of any other model that produces *probabilistic* projections of net migration for all countries. However, we can take our model's median projections to be point projections and compare them with models that produce point projections only. First, as a baseline for comparison, we evaluate them against simple *persistence models,* which project either net migration rates or net migration counts to continue at the most recently observed

levels indefinitely into the future. For up to 35 years into the future, the model that projects persistence of net migration counts is similar to the expert knowledge-based projections in the WPP (United Nations, 2011).

Second, we compare against point projections produced separately for all countries using Cohen's 2012 gravity model based method. The gravity model produces projected migration counts, which we convert to rates for evaluation. For each country c, the gravity model makes projections as follows: let $L(t)$ be the population of country c at time t, and let $M(t)$ be the population of the rest of the world at time t. Then expected in-migration to country c is given by $a \times L(t)^{\alpha} M(t)^{\beta}$, where a is a country-specific proportionality constant. The exponents α and β are constant across countries, with values estimated by Kim & Cohen (2010). Similarly, expected out-migration from country c has the form $b \times L(t)^{\gamma} M(t)^{\delta}$, where b is to be estimated, and γ and δ come from Kim & Cohen (2010). The constants of proportionality a and b for each country are chosen to minimize the sum of squared deviations between estimates of net migration produced by the gravity model and historical values of net migration from the WPP 2010 revision (United Nations, 2011). Having estimated a and b for a particular country, we calculate net migration projections by $a \times L(t)^{\alpha} M(t)^{\beta} - b \times L(t)^{\gamma} M(t)^{\delta}$, where $L(t)$ and $M(t)$ are now projected populations. Implementation details are given in Appendix A.2.

Our historical data consist of a series of migration rates $r_{c,t}$ for 197 countries at 12 time points in five-year time intervals, spanning the period from 1950 to 2010. We perform an out-of-sample evaluation by holding out the data from the m most recent time points for all countries and producing posterior predictive distributions on the basis of the remaining $(12 - m)$ time points. As point forecasts, we used the median of the posterior predictive distribution. We report out-of-sample mean absolute error as a measure of the quality of point forecasts, and interval coverage as a measure of quality of our interval predictions.

Table 2.1 contains these evaluation metrics for our Bayesian hierarchical model and the mean absolute errors for the persistence and gravity models. Our point projections outperformed the gravity model and both persistence models at all forecast lead times, and our

interval projections achieved reasonably good calibration. Appendix A.3 contains additional tables with evaluation metrics broken down by region. Our Bayesian hierarchical model outperformed the gravity model in all regions and the persistence models in most regions.

Table 2.1: Predictive performance of different methods: Mean absolute errors (MAE) and prediction interval coverage for our Bayesian hierarchical model, the gravity model, and the persistence models

Validation Time Period	Model	MAE	80% Cov. (%)	95% Cov. (%)
5 Years	Bayesian	3.24	91.4	96.4
	Gravity	4.70	—	—
	Persistence (of rates)	3.57	—	—
	Persistence (of counts)	3.58	—	—
15 Years	Bayesian	4.76	84.9	93.4
	Gravity	6.57	—	—
	Persistence (of rates)	6.74	—	—
	Persistence (of counts)	6.30	—	—
30 Years	Bayesian	5.12	77.2	89.3
	Gravity	12.32	—	—
	Persistence (of rates)	7.17	—	—
	Persistence (of counts)	5.82	—	—

2.3.2 Migration Trends

The primary goal of our model is to produce point and interval projections. However, it is also desirable for our model to replicate current trends in the migration data.

One prominent feature of the historical migration data to consider is the frequency with which countries switch between being net senders and net receivers of migrants. Such switches have been relatively common over the past 50 years. In fact, in the 2005–2010 period, 46% of countries had different migration parity than they had in 1955–1960 (i.e., they switched either from net senders to net receivers or vice versa). In contrast, the current United Nations methodology (United Nations, 2012) projects *no* crossovers between now

and 2100. Our model projects crossover behavior that is more in line with historical trends. Further analysis of projected parity changes is given in the case study on Denmark later in the article.

A second question is what our projections say about the *magnitude* of migration. Because we have only directly modeled net migration counts $y_{c,t}$ and the associated rates $r_{c,t}$, looking at the associated magnitudes $|y_{c,t}|$, or equivalently $|r_{c,t}|$, can serve as a model validity check. For example, a model could produce reasonable marginal migration projections for all countries despite being consistently biased toward projecting too much migration. We think it is worth confirming that our model does not have such a fault.

Furthermore, the analysis in the Evaluation section was concerned only with *marginal* projections for each country. However, because our projections actually take the form of joint trajectories for all countries simultaneously, we should confirm that the joint projections look reasonable. We do so by condensing high-dimensional joint projections of absolute migration into a single dimension using two different averages of net absolute migration.

One meaningful average of absolute net migration rates is

$$u(t) = \frac{\sum_{c=1}^{C} |r_{c,t}|}{C},$$

the *unweighted* mean absolute net migration rate across all countries. Because net migration represents the contribution of migration to population change, $u(t)$ can be interpreted as a heuristic measure of whether it is typical for countries to experience a lot of population change from the effects of migration.

Weighting absolute migration rates in proportion to population size rather than uniformly produces a measure of what the typical individual experiences, rather than the typical country. Such a weighted average is given by

$$w(t) = \sum_{c=1}^{C} |r_{c,t}| \frac{n_{c,t}}{\sum_{j=1}^{C} n_{j,t}}.$$

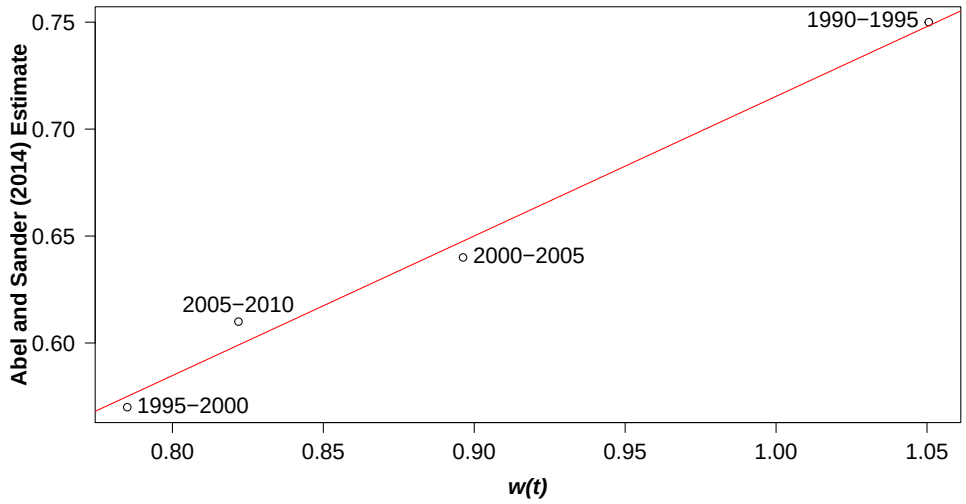

Figure 2.2: Estimated percentage of the world population migrating compared with a population-weighted average of absolute net migration rates, $w(t)$, for five-year periods from 1990 to 2010. Estimates of percentage of world population migrating are taken from Abel and Sander (2014). For this figure, we convert rates from the usual "net annual migrants per thousand" to "net five-year migrants per hundred" to put them on a comparable scale with percentage of world population migrating. There is a strong and significant correlation between the two quantities ($R^2 = .989$, $p = .006$)

If it were true that countries with net outflows had no inflows and vice versa, then $\frac{1}{2}w(t)$ would give the total proportion of the world population migrating. Of course, substantial cross-flows are common, so in reality, $\frac{1}{2}w(t)$ substantially underestimates the total proportion of the world population migrating. Nevertheless, comparison with flow estimates for 1990–2010 from Abel & Sander (2014) shows that $w(t)$ is strongly correlated with the total proportion of the world population migrating. Figure 2.2 compares Abel & Sander's estimates with $w(t)$. The correlation between the two measures is strong and significant ($R^2 = .989$, $p = .006$).

Figure 2.3 shows the historical values of $u(t)$ and $w(t)$ as well as our projections into the future. Our forecast shows no clear growth or shrinkage in $u(t)$, which is consistent with its historical trend. Meanwhile, we predict that $w(t)$ will continue to grow, leveling off in the

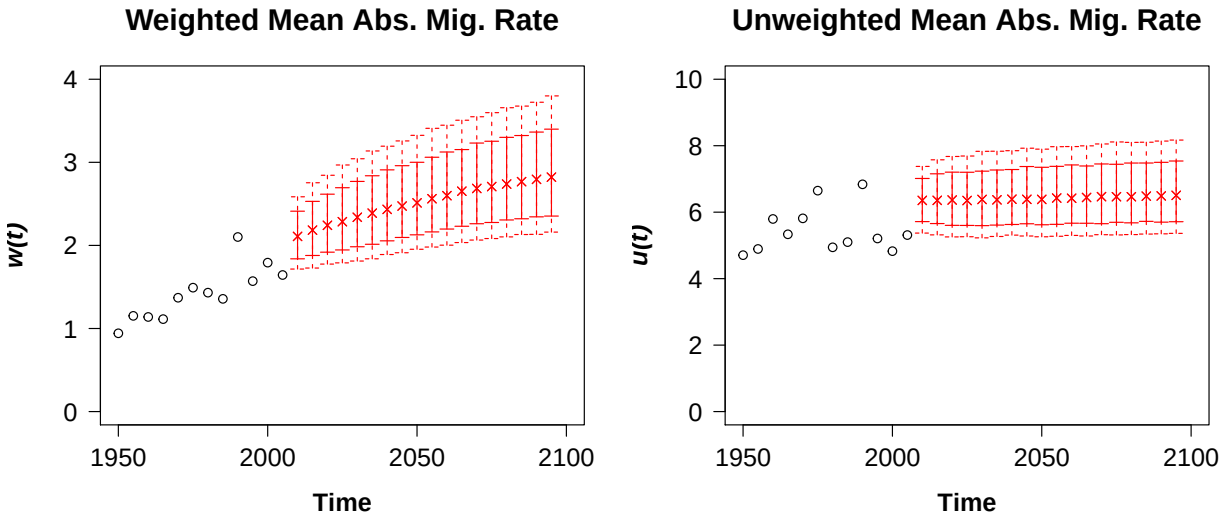

Figure 2.3: Observed historical data on population-weighted (left) and unweighted (right) averages of absolute annual migration rates per thousand for five-year periods from 1950 to 2010 (indicated by circles). Median estimates (indicated by "x") and 80% and 95% prediction intervals (indicated with vertical lines) from our model for periods out to 2100

long horizon. Despite the apparent contradiction, there is no inconsistency in the fact that $w(t)$ has grown quite substantially over time while $u(t)$ has not. This discrepancy is largely explained by the facts that (1) the largest countries have experienced mild increases in their absolute migration rates over time and (2) net migration rates and counts in the Gulf States grew enormously over this period. This first observation can be viewed as evidence of a form of globalization in international migration, in which net migration rates for large countries, once very low, are becoming more similar to those for other countries.

2.3.3 Case Studies

We now examine projected migration rates for a selection of four countries: Denmark, Nicaragua, India, and Rwanda. These four countries were selected both to provide geographic diversity and a variety of observed net migration trends since the 1950s. Denmark has experienced a shift from being a net sender of migrants to a net receiver, a pattern

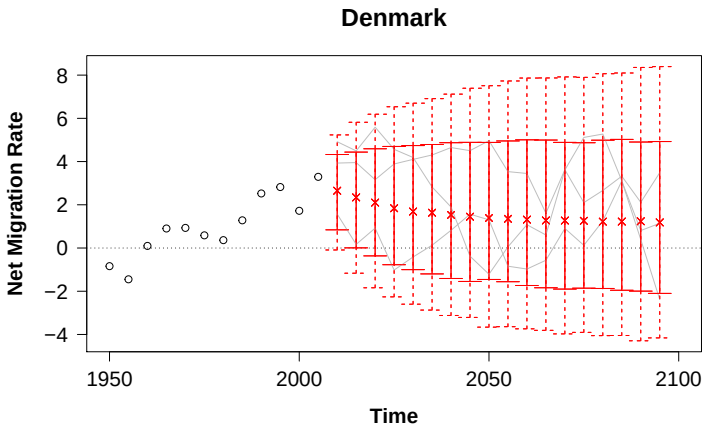

Figure 2.4: Probabilistic projections of net international migration rates: Predictive medians and 80% and 95% prediction intervals for Denmark, with example trajectories included in gray.

common in European countries. Nicaragua has had relatively stable and consistently negative net migration since the 1950s. India has had migration rates close to zero, which is common among the largest countries. Finally, Rwanda provides an example of a country that has experienced a large spike in absolute migration rate. This is not intended to be an exhaustive catalog of observed trends in net migration rates, although many countries have followed patterns similar to one of these four example countries.

Following these four case studies, we also present projections for the least-developed countries versus all other countries.

Denmark

Denmark experienced net emigration through the 1950s but has consistently received net immigration since the 1960s. This pattern of changing from a net sender to a net receiver within the last 60 years is common to many European countries, including Norway, Finland, the UK, and Spain. These countries' net out-migration in the middle of the twentieth century serves as a reminder that the global migration to northern and western Europe, which now

seems so firmly established, is a relatively recent phenomenon.

Our median predictions for Denmark have the country continuing to be a net receiver of migrants for as far out into the future as we care to project. However, we also see that the probability of Denmark switching over to a net sender increases over time. Based on the history of the twentieth century, it seems realistic to include the possibility of changeovers in Denmark and other European countries in probabilistic migration projections. Correspondingly, projections that do not take account of this possibility seem unrealistic.

The European countries are not alone in having oscillated between being net senders and net receivers of migrants. As mentioned in the discussion of migration trends, 46% of countries had different migration parity in the 1955–1960 period than they had in 2005–2010. Thus, they switched either from net senders to net receivers, or vice versa, during the past 55 years. Our Bayesian hierarchical model projects that 46% of countries will have different migration parity in 55 years (i.e. in 2055–2060) than they do now.[3] This projection is in line with the number of historical parity changes. In contrast, the gravity model (Cohen, 2012) projects that only 29% of countries will change parity by 2055–2060. Both persistence models and the WPP migration projections (United Nations, 2012) project no parity changes.

Nicaragua

Migration rates in Nicaragua have increased steadily in magnitude over the last six decades. Nevertheless, although our model projects a small probability of continued growth in the magnitude of the net migration rate, it gives higher probability to scenarios in which migration rates move back toward zero. In general, our model favors trajectories in which net migration rates move toward zero rather than continuing current trends of growth in magnitude where such trends exist.

Statistically, this tendency for migration rates on average to reverse course and tend back toward zero reflects past trends through from the hierarchical nature of the model.

[3]This figure is robust to incorporating a threshold. If we require that absolute migration rate be at least 1 net annual migrant per thousand either before or after the change, the figure is still 46%.

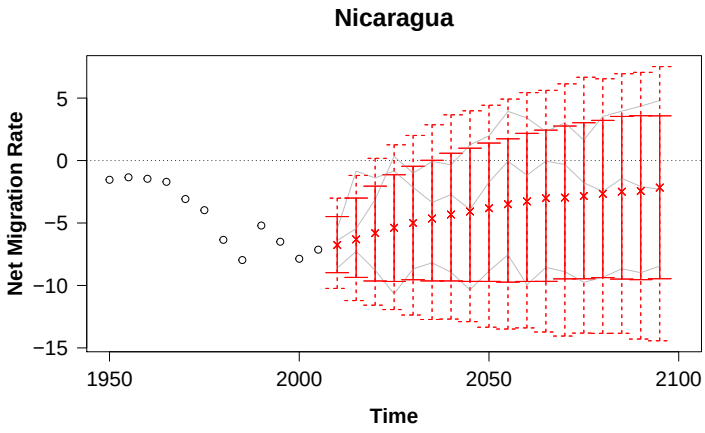

Figure 2.5: Probabilistic projections of net international migration rates: Predictive medians and 80% and 95% prediction intervals for Nicaragua, with example trajectories included in gray

Specifically, all of the μ_c values, which we can think of as the long-horizon median migration rates for each country, are assumed to come from a common $N(\lambda, \tau^2)$ distribution. As a result, the hierarchical "borrowing of strength" has a tendency to pull all the μ_c values toward a common center, λ, which has a posterior distribution with a mode close to zero. Although our model's median projections tend to predict reversal in growth trends, the predictive probability distributions give substantial probability to continuation and also to growth of rates.

India

Historically, India has had relatively low net migration rates, on the order of less than 1 per thousand. The 95% prediction intervals from our model are quite a bit wider than the range of India's historical data, expanding out to roughly ±3 per thousand.

Statistically, the width of a country's prediction intervals from our model is primarily controlled by the error variance σ_c^2. (The autoregressive parameters, ϕ_c, also influence the width of prediction intervals, but to a lesser extent.) The excess width of India's predic-

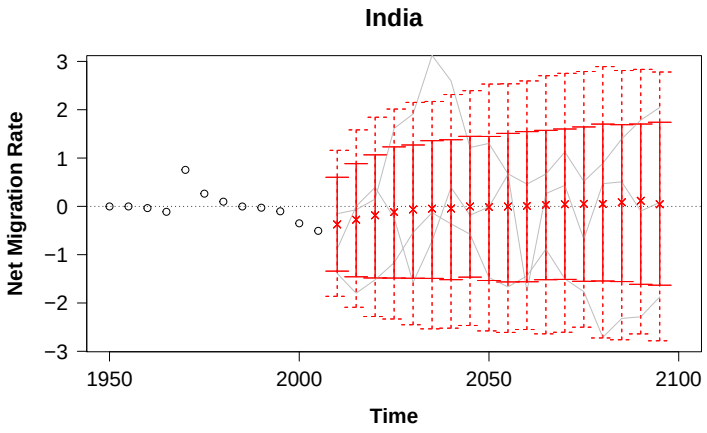

Figure 2.6: Probabilistic projections of net international migration rates: Predictive medians and 80% and 95% prediction intervals for India, with example trajectories included in gray.

tion intervals above its range of observed migration history is statistically a result of the hierarchical "borrowing of strength." Given that most other countries have larger ranges of migration rates, the posterior distribution of σ_c^2 for India gets inflated somewhat to values more in line with the rest of the world. The same inflation of σ_c^2 occurs in China, which has also experienced uncommonly low migration rates in the past.

Substantively, this seems realistic given the increasing globalization we have documented. As the largest countries become more like other countries in terms of migration patterns, it seems reasonable to expect that the variability of their migration rates in the future would also increase to become more like the levels of other countries.

Rwanda

In the early 1990s, Rwanda experienced high net out-migration, followed by high net in-migration in the late 1990s. These migration spikes were a result of emigration during the Rwandan genocide in 1994 and subsequent return migration. Outside of the 1990s, Rwanda had quite small and stable migration rates. This pattern of stability punctuated

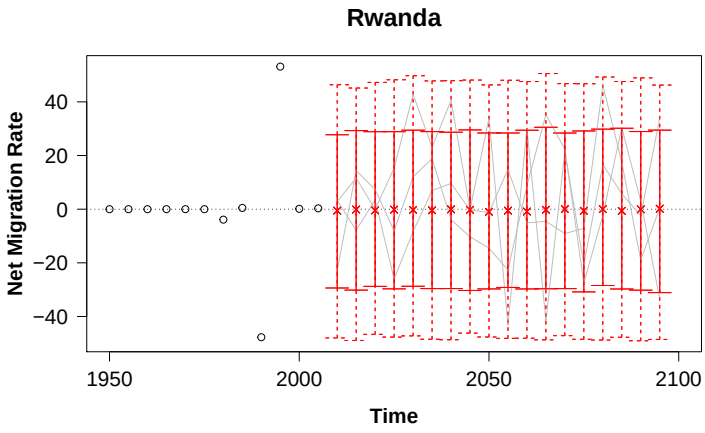

Figure 2.7: Probabilistic projections of net international migration rates: Predictive medians and 80% and 95% prediction intervals for Rwanda, with example trajectories included in gray.

by large shocks poses a problem for probabilistic projections: Do we get better performance with wide prediction intervals that encompass the high migration rates during the shock, or narrow prediction intervals that reflect the decades of stability around it?

Our model opts for wide prediction intervals in cases like Rwanda. A model that puts a heavy-tailed t distribution on the random error terms $\varepsilon_{c,t}$ rather than a normal distribution would produce narrower prediction intervals. However, we found that the normal model achieved better calibration of the main prediction intervals of interest—namely those of probability 95% and lower. The concluding discussion section contains a brief further discussion of a model with t-distributed errors.

The Least-Developed Countries

The United Nations publishes a list of the least-developed countries, with countries classified as least-developed based on assessments of their economic vulnerability, human capital, and gross national income (Committee for Development Policy and United Nations Department of Economic and Social Affairs, 2008). A total of 46 countries in our data fall into the least-

developed category. We now consider briefly the projections that our model makes for these least-developed countries in comparison to all other countries.

In the 2005–2010 period, only 26% of the least-developed countries were net receivers of migration, as compared with 43% of all other countries. The least-developed countries had an average net migration rate of -0.97 per thousand, compared with an average of 2.64 per thousand in all other countries. However, our model projects that this gap in migration between currently least-developed and all other countries will narrow over time. Key findings are summarized in Table 2.2. Over the coming decades, on average, we project mild growth in net migration rates among the least-developed countries and decline in net migration rate across all other countries.

Table 2.2: Mean projected change in migration rates (per thousand) among least-developed countries (LDC) versus all other countries (other), with 95% prediction intervals in parentheses

	LDC	Other
By 2020	$+0.02$ $(-3.09, +2.99)$	-1.50 $(-3.24, +0.33)$
By 2040	$+0.23$ $(-2.89, +3.38)$	-2.12 $(-4.23, +0.06)$
By 2060	$+0.35$ $(-2.78, +3.55)$	-2.29 $(-4.54, +0.14)$

2.4 Discussion

We have presented a method for projecting net international migration rates. Our method is novel in that it provides probabilistic projections of net migration for all countries. Furthermore, it satisfies the requirement that simulated trajectories have zero global net migration for each sex and age group.

Additionally, we observe a paradoxical trend in the evolution of global migration rates. Although there appears to be more migration than in the past as a proportion of the world population, countries' absolute migration rates, on average, have not been increasing. Our method successfully reproduces this pattern, which seems desirable for migration projection methods in general.

Our model includes the assumption that the random error terms $\varepsilon_{c,t}$ are independent across countries and time. That assumption is mathematically convenient, but for many pairs of countries, we expect to see nonzero correlations. For example, it is reasonable to expect that if Mexico undergoes particularly high net emigration during a quinquennium, then the United States will experience higher than usual net immigration during the same period. Thus, we might expect to observe negative correlation between the random errors for Mexico and the United States. At the same time, it is not unreasonable to expect positive correlation between error terms in neighboring pairs of countries whose economic fortunes tend to move together. Such a pattern is observed, for example, among the Baltic states. We attempted to find an optimal nontrivial covariance structure by constructing a variance-covariance matrix as a linear combination of matrices whose off-diagonal elements are pairwise, time-invariant covariates. However, this method offered no significant improvement over the assumption of independent residuals.

Migration data characteristically have outliers. Wars and refugee movements, for example, produce migration rates that are on a much larger scale than are typical during times of stability. This suggests that a model with a long-tailed error distribution, such as a t distribution, might be more appropriate than a model with normal errors. Furthermore, a model with t-distributed errors with degrees of freedom allowed to vary across countries is a natural way of handling the fact that some regions have quite stable migration rates over time (e.g., Western Europe) while others have quite a lot of volatility (e.g., Central Africa). However, in practice, we found that models with normally distributed errors tended to outperform models with t errors in out-of-sample evaluation of the resulting prediction intervals. Models with t errors often produce 80% and 95% prediction intervals that are so tight that they do not come close to covering the range of observed historical migration rates.

Statistically, the root of the problem is that in models with t errors, large outliers often do not have a large effect on the inferred scale parameter. Although using t errors often results in models with a high likelihood of the observed data, high likelihood does not necessarily correspond to good calibration of prediction intervals or qualitatively realistic migration

rates. For the migration forecasting problem, we believe that there is more value in forecast distributions with reasonable prediction intervals than in distributions that are likely to assign high probability density to future observations, if the choice has to be made. Thus, we used the normal model thoughout.

Note that by selecting the AR(1) model in advance, we are necessarily not accounting for variance due to model uncertainty. In the short term, this approach can be empirically justified by the fact that recent data are fit relatively well by the AR(1) model. If, in the long term, this ceases to be true, we expect our model to understate variances of posterior predictive distributions. Abel et al. (2013) demonstrated a Bayesian approach of averaging population forecasts across several plausible time series models that could be suitable to our data. We did not take this approach here because of empirical findings that higher-order autoregressive models don't offer significant increase in predictive power on this data set and because the AR(1) model offers qualitatively plausible long-term prediction intervals. However, it is quite possible that expanding our current method to take account of model uncertainty would improve the quality of the predictions.

Our model treats the United Nations estimates of net migration as the true migration counts without making any concession to the variability inherent in those estimates. Net migration is difficult to estimate directly, and ideally we would prefer to have estimates of net migration along with a measure of uncertainty in those estimates. Unfortunately, we know of no global source of estimates of migration with associated uncertainty. Appendix A.1 contains a breakdown of the sources of net migration data in the World Population Prospects along with discussion of the main causes and effects of uncertainty in migration estimates.

An additional source of variance which is unaccounted for in this article is the uncertainty in projected populations. The results presented here are conditional on the deterministic population projections in WPP 2010. Our methodology could be straightforwardly modified to allow instead for probabilistic population projections as inputs, including population projections that are updated at each time step to incorporate the migration projections output

by our model, as is done by Azose et al. (2016). The analysis presented here, however, is focused on producing reasonable migration projections taking known population projections as a given.

Chapter 3

ESTIMATING LARGE CORRELATION MATRICES FOR INTERNATIONAL MIGRATION

The United Nations is the major organization producing and regularly updating probabilistic population projections for all countries. International migration is a critical component of such projections, and between-country correlations are important for forecasts of regional aggregates. However, there are 200 countries and only 12 data points, each one corresponding to a five-year time period. Thus a 200×200 correlation matrix must be estimated on the basis of 12 data points. Using Pearson correlations produces many spurious correlations. In this chapter we propose a maximum *a posteriori* estimator for the correlation matrix with an interpretable informative prior distribution. The prior serves to regularize the correlation matrix, shrinking *a priori* untrustworthy elements towards zero. Our estimated correlation structure improves projections of net migration for regional aggregates, producing narrower projections of migration for Africa as a whole and wider projections for Europe. A simulation study confirms that our estimator outperforms both the Pearson correlation matrix and a simple shrinkage estimator when estimating a sparse correlation matrix.

3.1 Introduction

International migration is a major contributor to population change, but is hard to project, making proper quantification of uncertainty especially important. Existing global models for migration are well-calibrated marginally, i.e. for individual countries (Azose & Raftery, 2015), but typically rely on an unrealistic modeling assumption that forecast errors are uncorrelated across countries. If correlations exist, but are not modeled, the resulting projections may

still be well calibrated for countries individually, but can under- or overestimate variance in projections of migration for regions that span multiple countries. We present a method for estimating a correlation matrix from a small number of data points that uses informative priors, shrinking elements of the correlation matrix which we expect *a priori* to be small. In applying this method to migration, we choose priors based on empirical evidence of non-zero correlations among classes of countries which are "close" to one another according to a variety of distance covariates. Our method improves projections of migration for regional aggregates while mitigating the issue of spurious correlations that arises from trying to estimate a large correlation matrix based on many short time series.

3.1.1 Illustrative example

In this section we focus on six selected countries—Estonia, Latvia, Lithuania, South Africa, Zimbabwe, and Zambia—to highlight the need for regularization of the correlation matrix.

Migration rates in Estonia, Latvia, and Lithuania over the period from 1950 to 2010 look quite similar (top row of Figure 3.1.) All three countries share a spike in out-migration during the 1990–1995 time period, which appears as a large negative forecast error in a first-order autoregressive (AR(1)) model. This sudden jump in out-migration among the Baltic states shares a common cause, namely the fall of the Soviet Union, which both induced westward migration and prompted many ethnic Russians to return to Russia (Fassmann & Munz, 1994; Okólski, 1998)

Meanwhile, several countries in Southern Africa also experienced big shifts in migration rates during the 1990–1995 time period (bottom row of Figure 3.1.) From 1990 to 1995, South Africa received substantially more in-migration than it had in previous decades, while Zimbabwe and Zambia both switched from being net receivers of migrants to net senders. For these three countries, at least some of the change in migration was due to political shifts related to the end of South Africa's apartheid policy. For example, the number of legal entrants to South Africa who overstayed their visas grew dramatically during the 1990s, with many such entrants coming from other countries of the Southern African Development

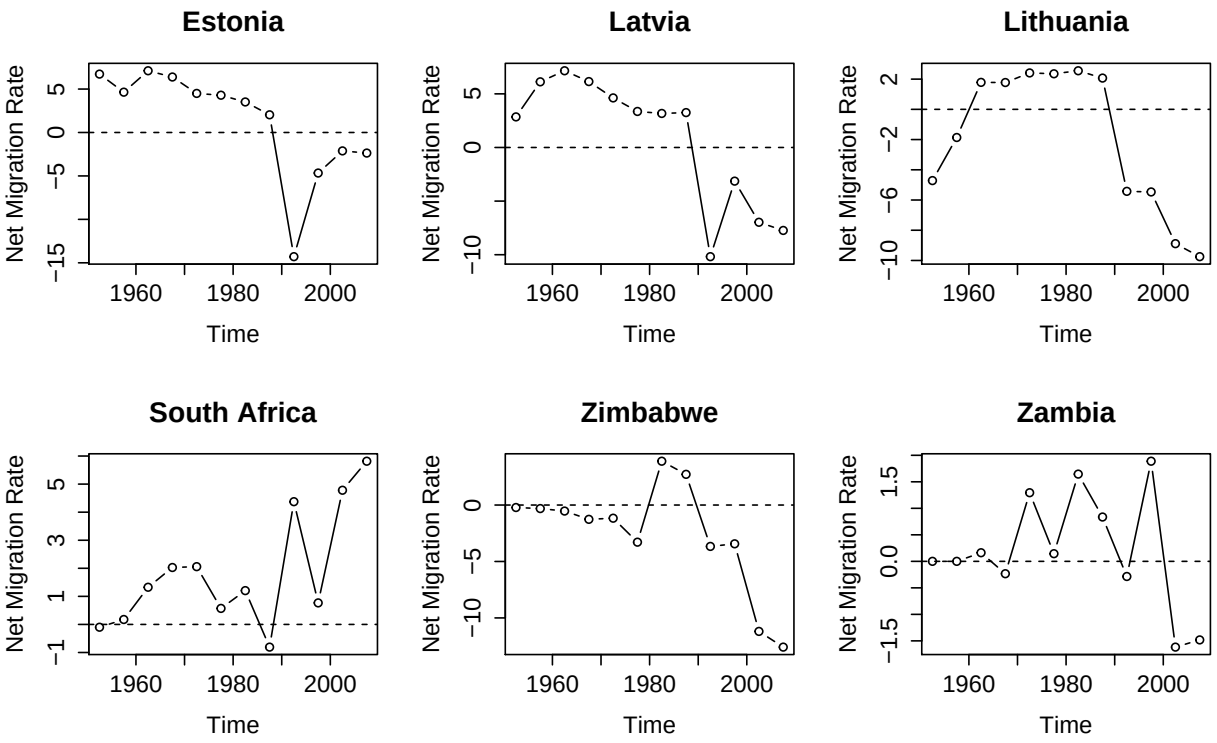

Figure 3.1: Net migration rates (net annual migrants per thousand individuals) for six countries.

Community (Crush, 1999).

Because all six countries experienced pronounced changes in migration rates during the same time period, the usual Pearson estimates of the correlation in forecast errors are relatively large for these six countries (left panel of Figure 3.2.) Knowledge of world affairs, however, suggests that some of these correlations may be spurious. There are plausible explanations for the correlations within the three Baltic nations and within the Southern African nations, but the cross-regional correlations are suspect. In fact, the cross-regional correlations seem to have arisen largely from a coincidental synchrony in the timing of disparate geopolitical events, and do not represent correlations that we would expect to continue to exist in future migration data. Our method is designed to shrink these seemingly spurious cross-regional correlations, producing the estimated correlation matrix shown in the right

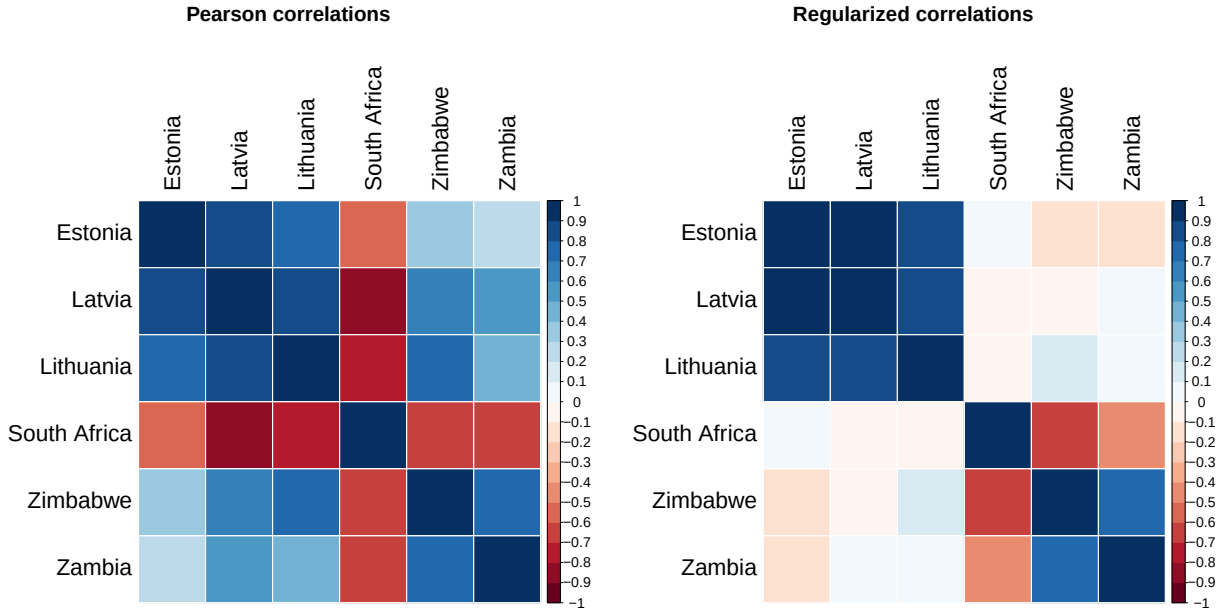

Figure 3.2: Estimated correlations among forecast errors for migration. Left panel shows Pearson correlation estimates. Right panel shows our regularized estimates.

panel of Figure 3.2. Cross-regional correlations decrease substantially in magnitude, while correlations within regions remain largely unchanged.

3.1.2 Background

Country-specific projections of international migration are an important input in policy-making decisions (Bijak et al., 2007; Brown & Bean, 2012). Projected migration figures are commonly used in long-term planning of social welfare programs (U.S. Social Security Administration, 2013; Wright, 2010). However, projection of migration is difficult—Bijak & Wiśniowski (2010) describe migration as "barely predictable"—and global modeling of migration remains somewhat rudimentary. The United Nations Population Division produces global projections of fertility, mortality, and migration for all countries (United Nations, 2012). For most countries, the 2012 revision of the World Population Prospects (WPP) deterministically projects net migration to persist at current levels until 2050 and decline

linearly thereafter.

To produce fully probabilistic population projections, one must incorporate probabilistic projections of fertility and mortality with a global probabilistic model of migration. It follows from the demographic balancing equation that the contribution of migration to population change is given by *net* migration (that is, in-migration minus out-migration.) Probabilistic models exist for both net migration (Azose & Raftery, 2015; Azose et al., 2016) and in- and out-migration separately (Wiśniowski et al., 2015). Both of these models are Bayesian hierarchical autoregressive models which treat forecast errors in migration as independent across countries, conditional on model parameters. This leads to projections that are well calibrated for individual countries, but may not be for multi-country aggregates. Our method aims to relax this independence assumption.

It is worth noting that a strong correlation in migration rates themselves need not translate to a strong correlation in forecast errors. For example, from 1960 through 2000, Mexico was consistently either the largest or second-largest source of migration flows to the US, with a conservative estimate of nearly 5 million Mexicans migrating to the US during the 1990's (Abel, 2013a). While we estimate that net migration rates for the USA and Mexico have a correlation of -0.56 based on quinquennial WPP data from 1950-2010, we estimate a correlation in forecast errors of only -0.07. That is, most of the relationship between the USA and Mexico is already captured by the autoregressive model parameters, and the "random" components of migration rates for the two countries are nearly independent conditional on the AR(1) model.

In this high-dimensional setting with short time series, the empirical correlation matrix is a poor estimator, in that it can include many spuriously large estimated correlations. Our goal is to use regularization to improve an empirical correlation matrix for forecast errors in migration. There is a large body of literature on regularized estimation of covariance matrices, with applications in genomics, image processing, and finance, among other fields (Fan et al., 2014). The novelty of our method is that it allows the incorporation of available prior information in an easily interpretable way.

Existing covariance estimators based on penalized likelihood maximization are typically maximum *a posteriori* (MAP) estimates under some prior belief about covariance, but these formulations are not well suited to specifying beliefs directly about elements of the correlation matrix. Perhaps the most similar method to ours is that of Bien & Tibshirani (2011), which allows informative priors on elements of the covariance matrix rather than the correlation matrix. Their method is not directly applicable to our setting, as our goal is to augment existing marginal variances with a suitable correlation structure. Other proposed MAP estimators include the graphical lasso (Friedman et al., 2008), which can be used to place an informative prior on the inverse covariance, and the method of Chi & Lange (2014), which penalizes covariance estimates that have very large or very small eigenvalues. An extreme example is given by Chaudhuri et al. (2007), who provide a method for covariance estimation in the presence of known zeroes. Zhang & Zou (2012) propose a variant on penalized likelihood maximization that replaces the negative log-likelihood with a simpler loss function.

A related class of covariance estimators relies on shrinkage of an empirical covariance matrix towards a simpler estimator, typically trading some bias for lower mean squared error (Ledoit & Wolf, 2003, 2004, 2012). A strength of these methods is that so long as the empirical covariance matrix is positive semi-definite and the shrinkage target is positive definite, a linear combination of the two will naturally be positive definite. Applying a shrinkage method to the migration setting would be difficult, as the elements we would like to penalize do not define a positive definite shrinkage target.

A form of regularization that is straightforward to implement is applying thresholding directly to elements of a covariance or correlation matrix (Bickel & Levina, 2008a; El Karoui, 2008); these authors show that a hard-thresholded covariance matrix is consistent in operator norm. Generalized thresholding (Antoniadis & Fan, 2001), developed in the context of wavelet applications, provides a class of related regularized estimators. A key difficulty with such estimators is that care must be taken to ensure that the resulting estimator is positive definite. In some problems, this can be handled by selecting a thresholding constant from an

appropriate range (Fan et al., 2013). Unfortunately, such an approach is not easily adapted to our problem. The structure of the elements we wish to penalize is such that we can tolerate only a small amount of shrinkage of all penalized elements before our estimated correlation matrix loses positive definiteness.

One fully Bayesian treatment is proposed by Liechty et al. (2004), who include substantive prior information by specifying clusters of correlations which they expect to be similar. This is unfortunately unsuitable to our setting, since geographical and cultural proximity can give rise to either positive or negative correlations. A. Huang & Wand (2013) describe a computationally attractive *non-informative* prior on covariances which does not easily extend to the informative priors we would like to include. Other fully Bayesian treatments are given by Barnard et al. (2000), who propose a prior on the correlation matrix which is either marginally or jointly uniform, and Leonard & Hsu (1992) and Deng & Tsui (2013), who propose Bayesian estimation of the logarithm of the covariance matrix, which is unfortunately hard to interpret.

In scenarios where there is a natural ordering to the variables, it is often reasonable to make the assumption that large values of $|i - j|$ imply near independence or conditional independence. When this is the case, one can regularize by banding or tapering of the covariance or inverse covariance matrix (Bickel & Levina, 2008b; Fan et al., 2007; Furrer & Bengtsson, 2007; Chen et al., 2013; Levina et al., 2008). These approaches are not suitable to our problem, as there is no natural ordering of countries.

Good overviews of other methods in covariance estimation are given by Fan et al. (2015) and Pourahmadi (2011).

3.2 Methods

We start with an established, well-calibrated autoregressive model on net migration rates for all countries (Azose & Raftery, 2015). This model has the form:

$$\boldsymbol{g}_t - \boldsymbol{\mu} = \text{diag}(\boldsymbol{\phi})(\boldsymbol{g}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t, \tag{3.1}$$

$$\boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathcal{N}_C\left(\mathbf{0}, \text{diag}(\boldsymbol{\sigma}) \cdot I_C \cdot \text{diag}(\boldsymbol{\sigma})\right), \tag{3.2}$$

$$\phi_c \overset{\text{iid}}{\sim} U(0,1), \tag{3.3}$$

$$\mu_c \overset{\text{iid}}{\sim} N(\lambda, \tau^2), \tag{3.4}$$

$$\sigma_c^2 \overset{\text{iid}}{\sim} IG(a,b). \tag{3.5}$$

(This is the same model introduced in Chapter 2, but with slight changes to notation. We have also omitted here the specifics of hyperpriors on a, b, λ, and τ, which are not important to this analysis.) Notationally, $\boldsymbol{g}_t$ is a length-C vector of net migration rates for all countries during the time period from t to $t+1$, where C is the number of countries analyzed. The quantities $\boldsymbol{\mu}$, $\boldsymbol{\phi}$, and $\boldsymbol{\sigma}$ are vectors of model parameters, and $\mathbf{0}$ is a length-C vector of zeroes. Notably, forecast errors in their model are treated as independent, conditional on the model's other parameters. Our method augments this model with an estimated correlation structure. Although this paper focuses on the migration context, the same technique could be applied to probabilistic models of other demographic indicators.

From this point forward, we refer to the model introduced in Chapter 2 as the Bayesian Hierarchical Model with Independent Forecast Errors (BHM+IFE). In principle, the methodology we describe here provides a means of estimating a correlation matrix to be adjoined to any probabilistic model with conditionally independent forecast errors.

The outline of our procedure for estimating a correlation matrix is as follows:

1. From the BHM+IFE model, draw a posterior sample of m realizations of model parameters, $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\phi}^{(1)}$, $\boldsymbol{\sigma}^{(1)}$, $\ldots$, $\boldsymbol{\mu}^{(m)}$, $\boldsymbol{\phi}^{(m)}$, $\boldsymbol{\sigma}^{(m)}$.

2. Convert the estimated forecast errors from the posterior sample of model parameters to a single empirical correlation matrix, $\tilde{R}$.

3. Combine the empirical correlation matrix with informative priors on correlations to obtain a maximum *a posteriori* (MAP) correlation estimate, $\hat{R}$.

This procedure can be viewed as performing a single step of the Monte Carlo EM (MCEM) algorithm (Wei & Tanner, 1990).

The posterior sampling in stage 1 can be performed using any reasonable sampling procedure. In practice, we performed our posterior sampling with a combination of Gibbs sampling and Metropolis-Hastings steps.

In the following sections, we first discuss the details of obtaining an MAP estimator (Section 3.2.1) and then the question of what to use for an empirical correlation matrix (Section 3.2.2). This is followed by an algorithm for computing the MAP estimator (Section 3.2.3), and finally discussion of a criterion for selecting a regularization parameter (Section 3.2.4).

3.2.1 MAP correlation estimate

Our goal is to estimate the correlation structure, R, of forecast errors, $\boldsymbol{\varepsilon}_t$. We assume a model of the form

$$\boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathcal{N}_C\left(\mathbf{0}, \Sigma\right), \tag{3.6}$$

where the variance matrix, Σ, decomposes into standard deviations, $\boldsymbol{\sigma}$, and a correlation matrix, R, as $\Sigma = \text{diag}(\boldsymbol{\sigma}) \cdot R \cdot \text{diag}(\boldsymbol{\sigma})$. To determine a MAP estimator for R, we express the posterior distribution for R as a product of likelihood and prior.

Data Likelihood

Equation (3.6) implies a likelihood function for R of the form

$$p(\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{T-1}|R, \boldsymbol{\sigma}) \propto_R \det(R)^{-(T-1)/2} \exp\left(-\frac{1}{2}\sum_{t=1}^{T-1} \boldsymbol{\varepsilon}_t' \mathrm{diag}(\boldsymbol{\sigma})^{-1} R^{-1} \mathrm{diag}(\boldsymbol{\sigma})^{-1} \boldsymbol{\varepsilon}_t\right), \quad (3.7)$$

restricted to the space Ω of valid correlation matrices (i.e. positive semi-definite matrices with ones on the diagonal.) Matrix trace identities simplify this likelihood to

$$p(\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{T-1}|R, \boldsymbol{\sigma}) \propto_R \det(R)^{-(T-1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}(R^{-1}\tilde{R})\right), \quad (3.8)$$

where

$$\tilde{R} := \frac{1}{T-1}\sum_{t=1}^{T-1} \mathrm{diag}(\boldsymbol{\sigma})^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mathrm{diag}(\boldsymbol{\sigma})^{-1}. \quad (3.9)$$

The evidence from the data is encapsulated in $\tilde{R}$, which is something akin to an empirical correlation matrix. Note that $\tilde{R}$ would be a sufficient statistic for R if the $\boldsymbol{\varepsilon}_t$'s and $\boldsymbol{\sigma}$ were known. In fact neither the $\boldsymbol{\varepsilon}_t$'s nor $\boldsymbol{\sigma}$ are known, and $\tilde{R}$ must be replaced with a sensible estimate in order to proceed. Details of the estimation of $\tilde{R}$ are given in Section 3.2.2.

Prior

Our choice of prior distribution on R is motivated by a desire to incorporate informative prior beliefs about which country pairs are likely to be nearly uncorrelated. As such, we choose a prior of the form

$$\pi(R) \propto_R \prod_{0 \le i < j \le C} \exp(-\lambda P_{ij}|R_{ij}|), \quad (3.10)$$

again restricted to Ω. The matrix P with entries P_{ij} is a penalty matrix that encodes the extent to which we believe that countries i and j may be correlated. In our application to migration, we constrain all the entries in P to be equal to 0 or 1, although in general

P may be allowed to have arbitrary non-negative entries. The parameter λ is an overall regularization parameter that encodes how strongly we want to penalize correlations.

The key benefit of this prior is its ease of interpretability. Setting $P_{ij} = 1$ expresses a belief that R_{ij} should be close to zero, with the strength of that belief controlled by λ. Setting $P_{ij} = 0$ implies that all values of R_{ij} are equally believable, *a priori*. Other penalized likelihood estimators have been proposed, corresponding to MAP estimators under implied priors on precision (Friedman et al., 2008), covariance (Bien & Tibshirani, 2011), or eigenvalues of the covariance matrix (Chi & Lange, 2014). None of these allow one to specify prior beliefs about correlations directly.

Note that under this specification, the prior distribution of the correlation R_{ij} is either uniform or truncated Laplace conditional on the rest of the correlation matrix, but marginal distributions will not be uniform or double exponential. Although it is possible to specify a marginally uniform prior on all elements of the correlation matrix (Barnard et al., 2000), we know of no way to specify a distribution that is marginally uniform for some elements and marginally peaked at zero for others.

Because the prior density is a product of Laplace densities on correlations, we will refer to our eventual correlation estimator as the LPoC (Lapalace Prior on Correlations) estimator. Augmenting the BHM+IFE with the LPoC correlation estimate produces the BHM+LPoC model.

Posterior

Combining the likelihood and prior, we obtain the log posterior distribution for R, equal to

$$\log p(R|\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{T-1}, \boldsymbol{\sigma}) = -\frac{T-1}{2}\log\det(R) - \frac{T-1}{2}\mathrm{tr}(R^{-1}\tilde{R}) - \frac{\lambda}{2}\|P * R\|_1 + c(\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{T-1}, \boldsymbol{\sigma}), \tag{3.11}$$

where $*$ denotes elementwise matrix multiplication, and $\|\cdot\|_1$ gives the sum of the absolute value of the elements of a matrix.

Thus, finding the MAP estimator for R is equivalent to solving the minimization problem

$$\text{Minimize}_{R\in\Omega}\left\{\log\det(R)+\text{tr}(R^{-1}\tilde{R})+\frac{1}{T-1}\lambda\cdot\|P*R\|_1.\right\} \tag{3.12}$$

Algorithmic details of a numerical solution are given in Section 3.2.3.

Note that if the penalty parameter, λ, is zero, then this minimization problem yields the maximum likelihood estimator (MLE) of R conditional on $\boldsymbol{\sigma}$. So long as $\tilde{R}$ is itself positive definite, this MLE is just $\tilde{R}$, the empirical correlation matrix. Similarly, if λ is held fixed as T grows, the penalty term in (3.12) goes to zero and the LPoC estimator converges to $\tilde{R}$. Since $\tilde{R}$ is consistent for R, the LPoC estimator is also consistent.

3.2.2 Estimating $\tilde{R}$

Since the forecast errors and model parameters of the BHM+IFE model are unknown, we do not have access to the true value of $\tilde{R}$. Instead we use an estimate of $\tilde{R}$. For practical reasons, we would prefer to have $\tilde{R}$ itself be a valid correlation matrix so that (3.12) will have a known analytic solution in the limiting scenarios where T grows or λ goes to zero. Accordingly, we might choose an estimator $\tilde{R}^{basic}$ with elements defined by

$$\tilde{R}^{basic}_{ij}:=\frac{\sum_{t=1}^{T-1}\hat{\varepsilon}_{i,t}\hat{\varepsilon}_{j,t}}{\sqrt{\sum_{t=1}^{T-1}\hat{\varepsilon}^2_{i,t}}\sqrt{\sum_{t=1}^{T-1}\hat{\varepsilon}^2_{j,t}}}, \tag{3.13}$$

where $\hat{\boldsymbol{\varepsilon}}_t$ is the posterior mean of $\boldsymbol{\varepsilon}_t$ from the BHM+IFE model. This estimate, $\tilde{R}^{basic}$, is the MLE for estimating the correlation matrix of a multivariate normal random variable with mean known to be zero and unknown marginal variance terms. By construction, $\tilde{R}^{basic}$ is guaranteed to be positive semi-definite and to have ones on the diagonal.

However, in our application, $\tilde{R}^{basic}$ is low rank, since T is small relative to the dimension of the matrix. For computational reasons, we would prefer to have a strictly positive definite

matrix, so we estimate $\tilde{R}$ by

$$\tilde{R}^{PD} = 0.99 \cdot \tilde{R}^{basic} + 0.01 \cdot I_C. \tag{3.14}$$

This change can be viewed as augmenting our estimates of $\boldsymbol{\varepsilon}_t$ with a small amount of additional uncorrelated data.

3.2.3 Solving the minimization problem

We apply a majorize-minimize algorithm similar to that used by Bien & Tibshirani (2011) to the minimization problem in (3.12). The function being minimized over is the sum of a convex and a concave component. The majorize-minimize algorithm repeatedly iterates through the following steps:

1. Replace the concave component with its tangent plane to obtain a fully convex function.

2. Find the global minimum of the convex function from Step 1.

3. Update the estimate of the tangent plane.

Notationally, we label our starting point for this algorithm as R_0 and subsequent iterations of this majorize-minimize algorithm are denoted with subscripts $R_1, R_2, \ldots$.

In (3.12), the concave component is $\log\det(R)$, which we replace with the tangent plane $\log\det R_i + \mathrm{tr}(R_i^{-1}(R - R_i))$. After simplifying and removing terms which are constant in R, the convex minimization problem in the ith iteration of the algorithm is

$$\text{Minimize}_{R\in\Omega} \left\{ \mathrm{tr}(R_i^{-1}R) + \mathrm{tr}(R^{-1}\tilde{R}) + \lambda\|P * R\|_1 \right\}. \tag{3.15}$$

Now all of the terms the objective function in (3.15) are convex, and all but $\lambda\|P*R\|_1$ are differentiable, so we can apply the generalized gradient descent algorithm (Beck & Teboulle,

2009). Each generalized gradient descent step takes the form

$$R_{new} = \text{argmin}_{\omega \in \Omega}\{(2t)^{-1}\|\omega - (R_{current} - t(R_i^{-1} - R_{current}^{-1}\tilde{R}R_{current}^{-1}))\|_F^2 + \frac{\lambda}{T-1}\|P * \omega\|_1\}. \tag{3.16}$$

If the restriction to Ω were not present, this problem would have a simple analytic solution, given by

$$R_{new} = \mathcal{S}\left(R_{current} - t(R_i^{-1} - R_{current}^{-1}\tilde{R}R_{current}^{-1}), \frac{\lambda}{T-1}tP\right), \tag{3.17}$$

where $\mathcal{S}$ is the element-wise soft-thresholding operator defined by

$$\mathcal{S}(X, \alpha)_{ij} = \text{sign}(X_{ij}) \cdot (|X_{ij}| - \alpha_{ij}) \cdot \mathbb{1}(|X_{ij}| > \alpha_{ij}). \tag{3.18}$$

(This move is actually restricted to the off-diagonal elements only, as the diagonal elements of a correlation matrix are constrained to equal 1.) Thus, if there were no positive definiteness constraint, each update step would consist of a gradient descent step according to the gradient of the differentiable component followed by soft-thresholding the result.

Although we do have to satisfy a positive definiteness constraint, we can start by trying the update step in (3.17). If this update results in a valid correlation matrix, then that matrix is our solution to (3.16), and we replace $R_{current}$ with R_{new}. However, sometimes the soft-thresholded gradient step results in a matrix that is not positive definite. In that case, it is possible to appeal to a slower, iterative solution to (3.16). One such solution is given by Cui et al. (2016). In practice, as long as we are looking for a solution in the interior of Ω, it is good enough to simply reduce step size rather than appealing to the relatively costly iterative algorithm whenever the generalized gradient descent suggestion lies outside of Ω.

Step size selection has a large impact on performance and convergence of this algorithm. Details of step size selection are discussed in Appendix B.1.

3.2.4 Selecting the regularization parameter λ

Although the penalty matrix P can be selected on the basis of world knowledge, we are less likely to have genuine prior beliefs about the value of the regularization parameter λ. Accordingly, we need some procedure for selecting a value for λ. In regularization problems, it is common to select the regularization parameter via cross-validation (Bien & Tibshirani, 2011; Chi & Lange, 2014; J. Z. Huang et al., 2006). This approach is too computationally intensive to be feasible for our application. Among shrinkage estimators, it is common to choose the amount of shrinkage in order to minimize an expected loss function (James & Stein, 1961; Ledoit & Wolf, 2003). However, no suitable analytic result exists that allows us to approximately minimize expected loss in our scenario.

Consequently, we developed a heuristic criterion that selects λ in a way that aligns with the goal of our regularization process. Our method's intent is to shrink the magnitude of penalized elements of the correlation matrix while leaving unpenalized elements more or less unchanged. In practice, although we succeed at bringing penalized elements towards zero, this shrinkage usually comes at the cost of inflating other elements. We have observed that this inflation tends to grow more pronounced as λ grows. For very large values of λ, our estimated correlation matrix may shrink nearly all penalized entries to zero at the expense of inflating a few elements (both penalized and unpenalized) to nearly ± 1. This is not a desirable outcome.

Although it may seem counterintuitive at first, the observed inflation is not an artifact of a coding error or poor convergence of our algorithm. A simple reproducible example of inflation in a 3×3 matrix is provided in Appendix B.2. In this low-dimensional setting, standard numerical optimization routines agree with the results from our code and both display inflation of unpenalized elements.

Our criterion for selecting λ compares the off-diagonal elements of $\tilde{R}$ and $\hat{R}(\lambda)$. We choose the value of λ which maximizes the difference between average shrinkage and average

inflation. Formally, our criterion is defined by

$$k(\tilde{R}, \lambda) = \underset{i,j \text{ s.t. } |\hat{R}(\lambda)_{ij}|<|\tilde{R}_{ij}|}{\text{mean}} \left(|\tilde{R}_{ij}| - |\hat{R}(\lambda)_{ij}| \right) - \underset{i,j \text{ s.t. } |\hat{R}(\lambda)_{ij}|>|\tilde{R}_{ij}|}{\text{mean}} \left(|\hat{R}(\lambda)_{ij}| - |\tilde{R}_{ij}| \right). \quad (3.19)$$

Large positive values of k are desirable, as they correspond to values of λ for which we induce a lot of shrinkage and not much inflation.

3.3 Results

In this section, we first report results from applying our method to global migration data in Section 3.3.1. Section 3.3.2 then provides a simulation study which demonstrates that our method outperforms Pearson correlations and the Ledoit-Wolf shrinkage estimator (Ledoit & Wolf, 2003) in the scenario where the penalty matrix P is appropriate to the true correlation structure.

3.3.1 Application to migration

Data

We use data on net migration from the 2012 revision of the World Population Prospects (WPP) United Nations (2012). The WPP contains estimates of net migration for all countries in five-year time periods from 1950 until 2010, a total of 12 time periods. We compute the net migration rate $g_{c,t}$ as the net number of migrants in country c over the five year period starting at time t, divided by thousands of individuals in country c at time t.

Because we want to express prior beliefs as a function of distance covariates, we restrict the set of modeled countries to the 191-country overlap between the WPP 2012 and the set of countries included in CEPII's GeoDist database, a database of bilateral distance covariates defined on pairs of countries (Mayer & Zignago, 2011).

Selection of P

Our estimation technique requires that we choose a penalty matrix, P, that reflects our prior beliefs about which country pairs are likely to be correlated. Although it would be possible to elicit expert opinion about each of the roughly 18,000 country pairs, we instead choose a P that can be characterized in terms of just a few covariates. Our matrix P penalizes a pair of countries if *none* of the following conditions is met:

1. The two countries are contiguous.

2. The two countries' most important cities are located less than 3000 km apart.

3. The two countries are in the same region according to the United Nations Population Division's division of the world into 22 regions, based on both geographical contiguity and cultural affinity (United Nations, 2012).

4. The two countries are currently in a colonial relationship.

This definition of P is in line with migration theory, which suggests that migrant flows are more likely when monetary and social costs of movement are low (Harris & Todaro, 1970; Lee, 1966; Sjaastad, 1962; Stark & Bloom, 1985), as will be the case with countries which are geographically proximate or share administrative ties. This definition penalizes 85% of country pairs, leaving 15% unpenalized. The average country is considered to be "close" to 29 other countries, and "distant" from the remaining 161.

In selecting these conditions, we examined nine candidate distance covariates. The first eight such covariates come from CEPII's GeoDist database (Mayer & Zignago, 2011), while the ninth is derived from the United Nations division into 22 regions. The left column of Table 3.1 gives the complete list of covariates considered. As an empirical basis for determining which criteria to include in defining our penalty matrix, we examined the elements of the sample correlation matrix for all pairs of countries meeting each criterion. Using a Kolmogorov-Smirnov test, we tested whether the distribution of these sample correlations

Table 3.1: Results of Kolmogorov-Smirnov test that empirical correlations are significantly different from the distribution of elements of a sample correlation matrix when the true error structure is uncorrelated. p-values lower than 0.05 are bolded.

Covariate	p-value
Contiguous	**0.019**
Common language (official)	0.23
Common language (spoken by 9% of pop.)	0.58
Geodesic distance less than 3000 km	**0.0003**
Colonial relationship after 1945	0.57
Common colonizer after 1945	0.11
Current colonial relationship	**0.035**
Ever had a colonial link	0.36
Same UN Region	**0.036**

was different from the distribution of elements of the sample correlation matrix under a null hypothesis of uncorrelated errors. The right column of Table 3.1 shows the p-values from these Kolmogorov-Smirnov tests. Our definition of the penalty matrix P includes all covariates with a p-value less than 0.05.

Selection of the regularization parameter, λ

We computed values of $\hat{R}(\lambda)$ for all values of λ from 0 to 3 in increments of 0.1. Figure 3.3 shows the value of $k(\tilde{R}, \lambda)$ over a range of λ values. We found that $k(\tilde{R}, \lambda)$ peaked at $\lambda = 0.6$, where we find average shrinkage of 0.13 compared with average inflation of 0.07. Increasing λ from 0.6 to 0.7 induces additional shrinkage, but at the cost of greatly inflating some correlations. Accordingly, we choose $\hat{R}(0.6)$ as our estimate of R.

Figure 3.4 shows the impact of regularization on the correlation matrix. Among penalized elements (top panel), we see significant shrinkage towards zero, although many penalized elements remain large in magnitude, even after regularization. The bottom panel shows the unpenalized elements of the correlation matrix before regularization (solid curve) and after

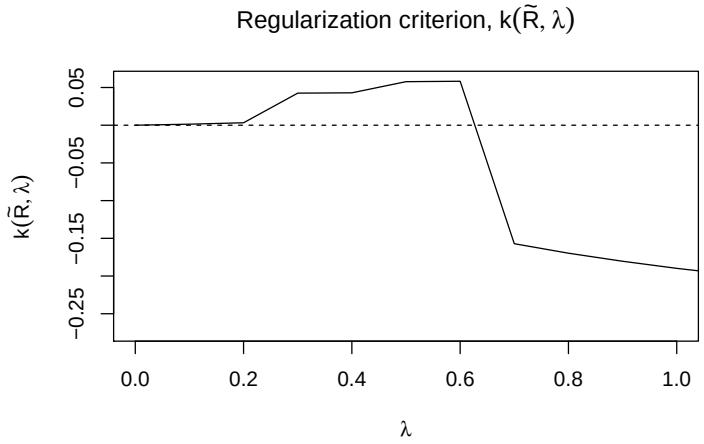

Figure 3.3: Regularization criterion, $k(\tilde{R}, \lambda)$ as a function of λ. The regularization criterion is the difference between the average shrinkage among shrunk elements of $\hat{R}(\lambda)$ and average inflation among inflated elements.

(dashed curve). On average we induce some shrinkage in the unpenalized elements, but the distribution is largely unchanged.

Projection and evaluation

We augment the BHM+IFE model with the LPoC estimate $\hat{R}(0.6)$ to produce probabilistic projections of migration for any collection of countries. Figure 3.5 contains medians and 80% prediction intervals of projected migration for all continents. In Africa, negative correlations narrow our projections. In Europe, positive correlations cause forecasts to widen. For the other continents, we see little change in projected migration.

For evaluation, we compare true migration rates for regional aggregates in 1995–2010 with projections of the same regional aggregates based only on migration data from 1950–1995. This procedure entails re-estimation of the BHM+IFE model using only the 1950–1995 data, followed by construction of an empirical correlation matrix, selection of λ, and extraction of $\hat{R}(\lambda)$. We compare the performance of the BHM+IFE model on regional aggregates to a model using the same sampled values of $\boldsymbol{\mu}$, $\boldsymbol{\phi}$, and $\boldsymbol{\sigma}$, but augmented with $\hat{R}(\lambda)$.

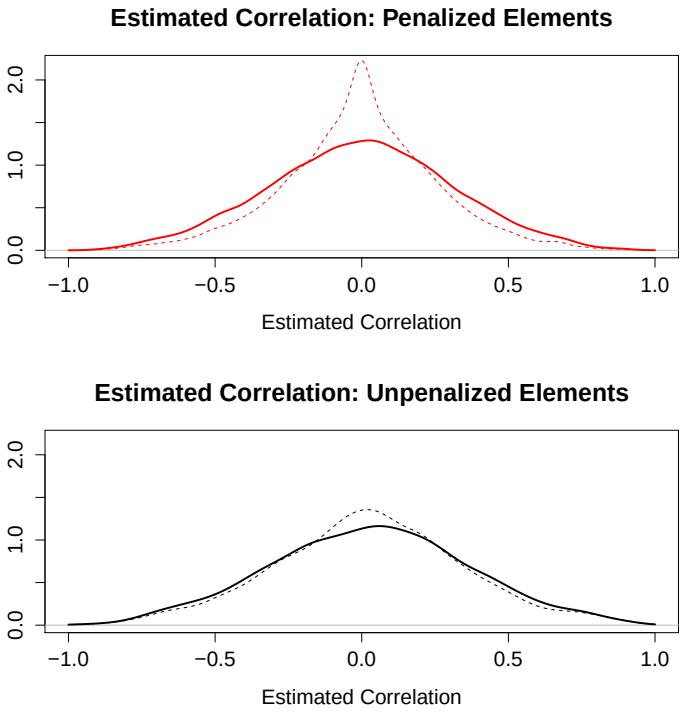

Figure 3.4: Comparison of elements of the correlation matrix before regularization (solid curves) and after (dashed curves). Top panel shows penalized elements; bottom panel shows unpenalized elements.

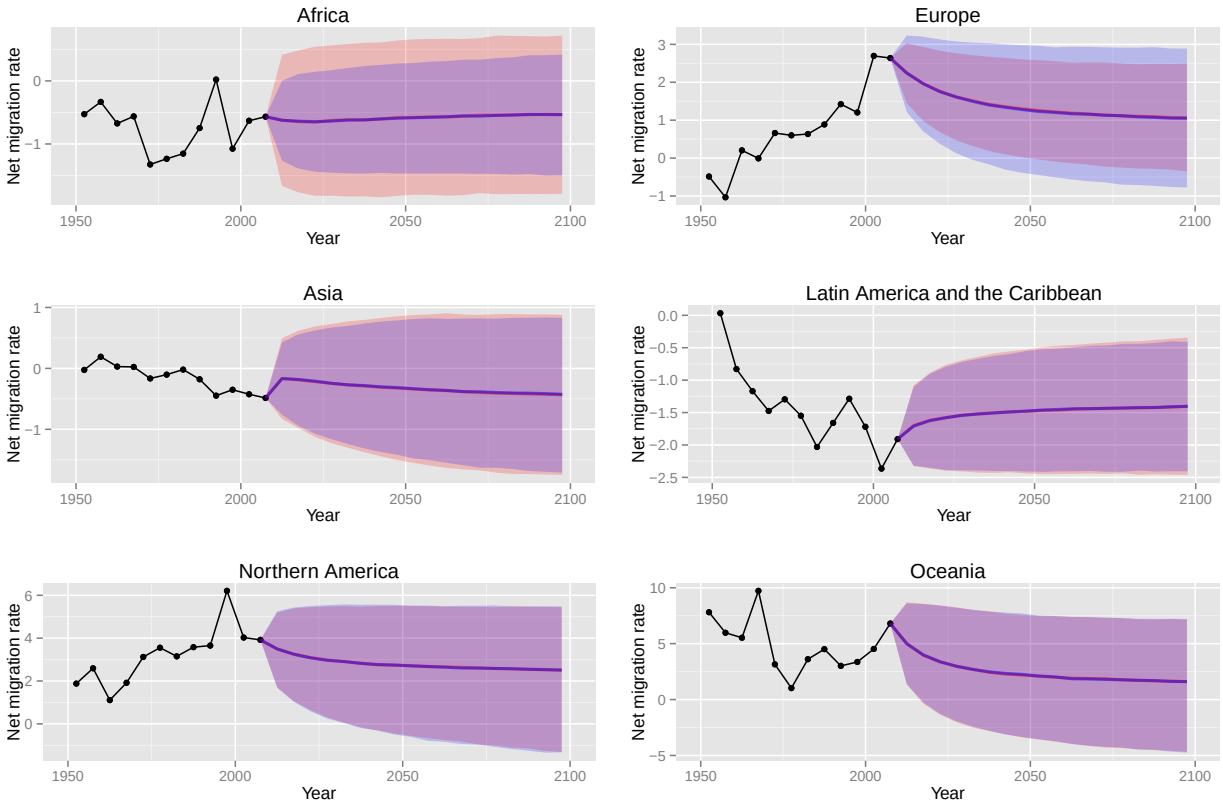

Figure 3.5: Medians and 80% prediction intervals for net migration among continents. Projections from the Bayesian hierarchical model with independent forecast errors (BHM+IFE) are given in red. Projections using our estimated correlation matrix (BHM+LPoC) are in blue. Overlap is in purple.

As an evaluation metric, we use the negatively oriented continuous ranked probability score (CRPS) (Hersbach, 2000; Gneiting & Raftery, 2007). The CRPS compares the cumulative distribution function, F, of a probabilistic forecast to an observation, x, and is defined by

$$\text{CRPS}(F, x) = \int_{-\infty}^{\infty} (F(y) - \mathbb{1}\{y \geq x\})^2 \, dy. \tag{3.20}$$

In our application the two probabilistic forecasts under consideration have the same mean as one other, by design. One approximate way of looking at CRPS in this setting is that when $g_{c,t}$ is close to the mean of the forecast, we reward F for having low variance; when $g_{c,t}$ is far from the mean, we reward F for having high variance.

Table 3.2 gives CRPS for projections of aggregate migration for the six continents. Our model improves the quality of projections in Africa and Europe, while projections for the other four continents are more or less unchanged. Figure 3.6 illustrates the change in projections of net migration in 1995–2010 for four subregions of Africa and Europe. Projections from the BHM+IFE model are in red; projections from BHM+LPoC are in blue. Our method narrows prediction intervals in Eastern and Western Africa, bringing the width of the 80% prediction intervals more into line with the range of observed variability. In both regions, true migration rates for the projected period stayed within our narrower intervals. In contrast, our method widens projections in Northern and Western Europe, where the 80% intervals from the BHM+IFE model either miss or nearly miss capturing some of the observed data points.

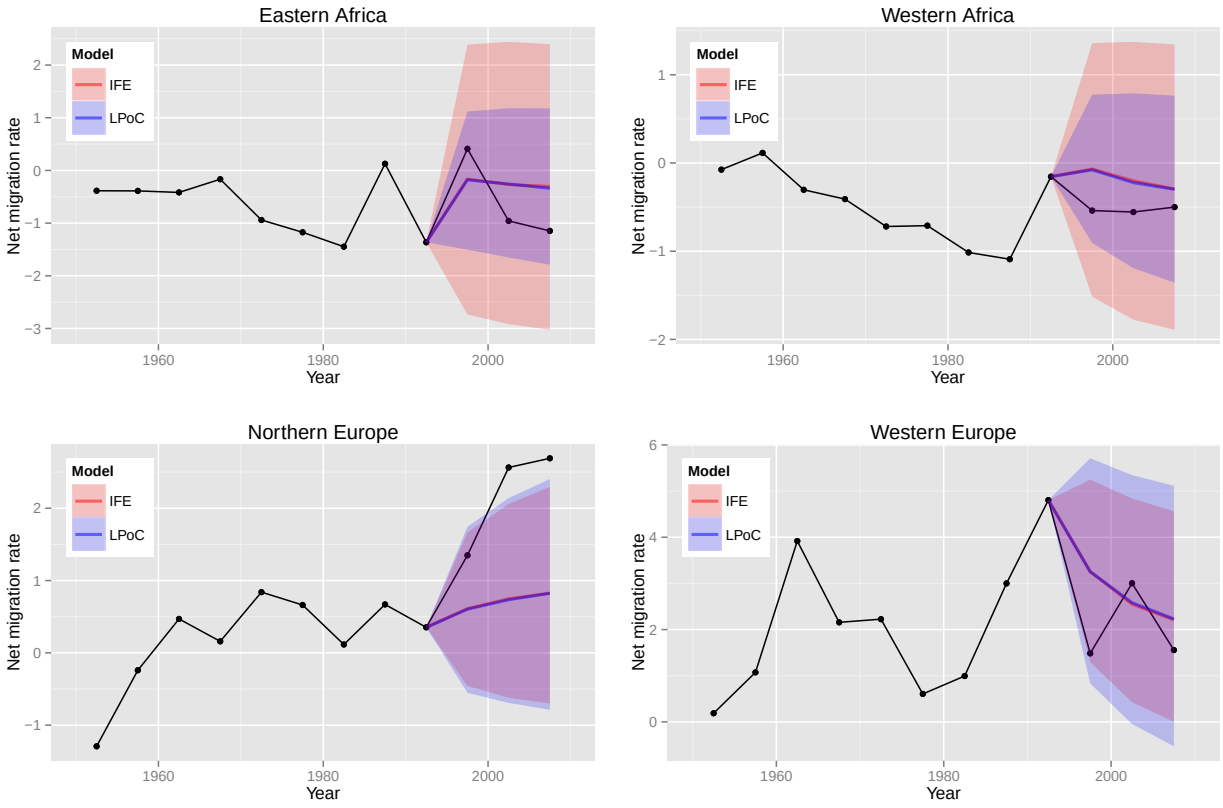

Figure 3.6: Medians and 80% prediction intervals for projections of net migration rates for regional aggregates. Projections from the Bayesian hierarchical model with independent forecast errors (BHM+IFE) are given in red. Projections using our estimated correlation matrix (BHM+LPoC) are in blue. Overlap is in purple.

Table 3.2: Continuous ranked probability score for all continents evaluated on projections of 1995-2010, where lower is better. Left column: Projections based on the Bayesian hierarchical model with independent correlation structure (BHM+IFE). Right column: Projections based on the Bayesian hierarchical model with our regularized correlation estimate (BHM+LPoC). Bolded entry in each row indicates the lower value.

	IFE	LPoC
Africa	1.66	**1.49**
Asia	**0.73**	0.74
Europe	3.92	**3.76**
Latin America and the Caribbean	**1.62**	**1.62**
Northern America	5.02	**4.99**
Oceania	8.53	**8.49**

3.3.2 Simulation study

In this section we show by simulation that our regularization procedure improves correlation estimates in a low-dimensional setting. To match the application of interest, we simulate 12 observed time points from an AR(1) process with correlated errors. For computational tractability, we decrease the number of simulated countries from 191 in the real data to 9 in the simulation. For each of 100 simulations, we perform the following procedure:

1. Generate a set of simulated migration rates $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_{12}$ from an AR(1) process with errors correlated as described below.

2. Produce point estimates of $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{11}$ via MCMC sampling of $\boldsymbol{\mu}$, $\boldsymbol{\phi}$, and $\boldsymbol{\sigma}$.

3. Convert $\boldsymbol{\varepsilon}_t$'s to a matrix $\tilde{R}$ using the procedure in Section 3.2.2.

4. Solve the minimization problem (3.12) to obtain a regularized estimate for the correlation matrix.

Since the procedure for selecting λ is computationally intensive, we perform this procedure only once and use the same value of λ for all subsequent simulations.

Simulation details

We simulate a collection of nine countries with true migration rates governed by the AR(1) process

$$\boldsymbol{g}_t - \boldsymbol{\mu} = \operatorname{diag}(\boldsymbol{\phi})(\boldsymbol{g}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t. \tag{3.21}$$

For simplicity we take $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\phi} = \frac{1}{2}\mathbf{1}$, and

$$\boldsymbol{\varepsilon}_t \overset{iid}{\sim} \mathcal{N}_9(\mathbf{0}, \Sigma). \tag{3.22}$$

We fix Σ to be block diagonal. Compound symmetric correlation structure within each 3×3 block is given by

$$\Sigma_{3\times 3} = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}, \tag{3.23}$$

and the full covariance matrix by

$$\Sigma = \begin{pmatrix} \Sigma_{3\times 3} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma_{3\times 3} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Sigma_{3\times 3} \end{pmatrix}. \tag{3.24}$$

We then simulate observations $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_{12}$ and attempt to make inference on the correlation structure of Σ.

Because we are basing inference on a small number of time points, Pearson estimates of correlation are highly variable. Solid curves in Figure 3.7 show the distributions of the off-diagonal elements of the unregularized Pearson correlation matrix in the ideal scenario where the values of $\boldsymbol{\varepsilon}_t$ can be perfectly estimated. The top panel shows the distribution of the elements for which the true correlation is zero. The bottom panel shows elements for which the true correlation is 0.5. In both cases, high variability makes inference difficult. Our method is designed to decrease variability among estimated correlations for those country pairs where prior knowledge suggests that correlation should be close to zero.

To illustrate a best case scenario, we choose a penalty matrix P which is well suited to the true correlation structure. The simplest such P is the one which penalizes the off-diagonal elements of the correlation matrix if and only if the true correlation is zero. That P is given by

$$P = \begin{pmatrix} \mathbf{0}_{3\times 3} & \mathbf{1}_{3\times 3} & \mathbf{1}_{3\times 3} \\ \mathbf{1}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{1}_{3\times 3} \\ \mathbf{1}_{3\times 3} & \mathbf{1}_{3\times 3} & \mathbf{0}_{3\times 3} \end{pmatrix}. \tag{3.25}$$

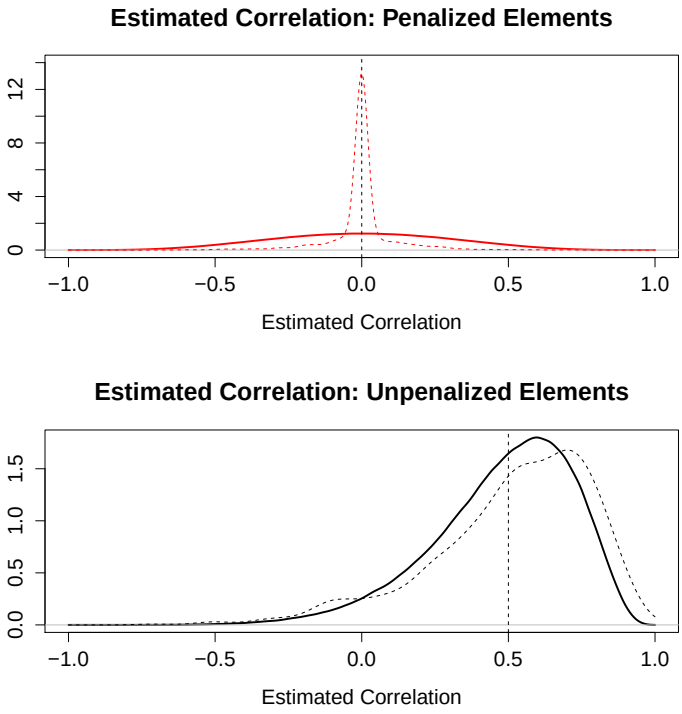

Figure 3.7: Simulation study results: Comparison of elements of the correlation matrix before regularization (solid curves) and after (dashed curves). Top panel shows penalized elements; bottom panel shows unpenalized elements. True correlations are indicated with dashed vertical lines.

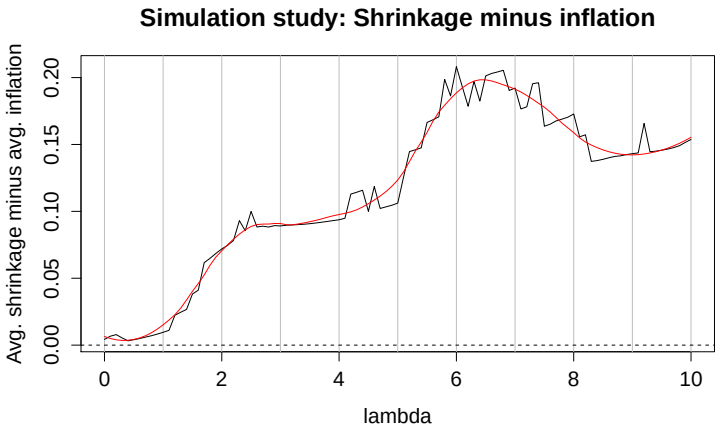

Figure 3.8: Average shrinkage minus average inflation of elements of $\hat{R}(\lambda)$ as λ varies from 0 to 10. Exact curve in black, Lowess-smoothed curve in red.

Initial run to select λ

Our procedure to select λ is computationally expensive, as it requires us to compute $\hat{R}(\lambda)$ repeatedly as λ varies. We therefore perform this procedure only once and use the same λ for estimation of R in all subsequent simulated data sets. Figure 3.8 plots our λ-selection criterion based on a single simulated data set over the range $\lambda = 0, 0.1, 0.2, \ldots, 10$. The exact curve, shown in black, exhibits some jumpiness in this low-dimensional setting, a problem which naturally becomes less severe in the high-dimensional setting of interest. Because of this jumpiness, we base our selection of λ on a Lowess-smoothed curve, selecting the maximizing value of $\lambda = 6.4$.

Evaluation of repeated estimation of R

We produced 100 estimates of $\hat{R}(\lambda = 6.4)$ from 100 different sets of simulated migration rates, all using the same block diagonal correlation structure. Dashed lines in Figure 3.7 show the distribution of off-diagonal elements of $\hat{R}$, split into those elements where the true correlation is 0 and elements where the true correlation is 0.5 (top and bottom panel, respectively).

Our method is successful in shrinking penalized elements towards zero. Among elements where the true correlation is zero, we correctly estimate an exact zero in 62% of cases in this simulation. Among unpenalized elements, our method produces estimates with slightly more variability (the standard deviation is 0.256 for Pearson correlations versus 0.272 for our estimates). Both methods produce estimates for unpenalized elements that are within two standard errors of the true mean value of 0.5. The mean estimated correlation is 0.489 for Pearson correlations (standard error 0.009) versus 0.514 for our estimates (standard error of 0.009). On the whole, the LPoC estimator greatly improves estimates of penalized elements at the expense of slightly increasing variability in unpenalized elements.

Table 3.3 compares mean absolute error and mean squared error from our method with two competing estimators. We compare our results against both Pearson correlation matrices and correlation matrices that have been regularized using the Ledoit-Wolf method, which shrinks Pearson estimates towards a spherical correlation structure (Ledoit & Wolf, 2003). In the top panel, we estimate $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{11}$ with a Bayesian hierarchical model, as is done in our real application to migration. In the bottom panel, we assume instead a scenario where we have direct access to $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{11}$, as would be suitable in other applications where the interest is in estimating correlations of directly observed quantities. In both cases, our method provides an overall reduction in mean squared error by at least two thirds when compared against the Pearson sample correlation matrix. A large reduction in error from shrinking penalized elements is offset by a mild increase in error among unpenalized elements. We also outperform the Ledoit-Wolf estimator in terms of overall error.

3.4 Discussion

Our method augments probabilistic projections of migration that are well-calibrated for individual countries, with a correlation structure that reflects prior knowledge of between-country correlations. By combining a high-dimensional empirical correlation matrix with an informative prior that shrinks spurious correlations, we produce an estimated correlation matrix that is in line with migration theory and improves projections of regional aggregates.

Table 3.3: Evaluation of correlation matrix estimates from simulation study. "LPoC" refers to our estimator, which uses Laplace priors on correlations. MAE is mean absolute error. MSE is mean squared error. Averages over "all elements" exclude diagonal elements, which are fixed at zero by all methods. The lowest (best) values are shown in bold.

Values of $\boldsymbol{\varepsilon}_t$ estimated with MCMC

	Estimator	MAE	MSE
All elements	Pearson	0.253	0.098
	Ledoit-Wolf	0.193	0.055
	LPoC	**0.090**	**0.028**
True correlation = 0	Pearson	0.270	0.109
	Ledoit-Wolf	0.190	0.053
	LPoC	**0.049**	**0.012**
True correlation = 0.5	Pearson	0.201	0.066
	Ledoit-Wolf	**0.200**	**0.060**
	LPoC	0.214	0.074

True values of $\boldsymbol{\varepsilon}_t$ used

	Estimator	MAE	MSE
All elements	Pearson	0.227	0.079
	Ledoit-Wolf	0.182	0.047
	LPoC	**0.078**	**0.022**
True correlation = 0	Pearson	0.244	0.089
	Ledoit-Wolf	0.162	0.039
	LPoC	**0.041**	**0.010**
True correlation = 0.5	Pearson	**0.176**	**0.051**
	Ledoit-Wolf	0.243	0.073
	LPoC	0.190	0.058

When compared with a simple model that assumes uncorrelated forecast errors, our method narrows projections of net migration for Africa and widens projections for Europe. Out-of-sample evaluation confirms that these changes produce better probabilistic forecasts as measured by continuous ranked probability score. Mechanically, the novelty of our method is our prior on correlations, which benefits from being interpretable and simple in form, and converts MAP estimation to an ℓ_1-penalized regularization problem which is computationally tractable.

Our analysis focuses on modeling net migration, but an attractive alternative would be to model a full matrix of bilateral migration flows. Such a model would naturally imply correlations in migration—if out-migrants from country i tend to go to country j, then net migration in countries i and j will be negatively correlated. However, modeling the global bilateral flow matrix is currently not feasible. Flows are hard to estimate, even in countries with good data (De Beer et al., 2010; Raymer et al., 2011). Abel (2013a) produces global estimates of migration flows based on migrant stock data, but for only a small number of time periods at which migrant stock data exist. His method involves minimizing the total number of migrants subject to the available data on migrant stocks. This induces many structural zeroes in his estimates, making modeling difficult. Because of the lack of good data on migration flows, we choose instead to work with net migration rates.

Although our method produces a MAP estimator in the presence of informative priors, we are not able to leverage any of the usual Bayesian machinery to produce a sample from the posterior distribution. While it would in theory be possible to use MCMC methods to produce a posterior sample by updating one element of the correlation matrix at a time, an updating procedure would need to iterate through some 18,000 elements of the correlation matrix, checking for positive definiteness after each proposed step. Such an algorithm is therefore likely to move around the parameter space too slowly to be of any use. In some settings a Laplace approximation centered at the posterior mode can provide a good approximation of marginal posterior distributions (Tierney & Kadane, 1986). However, the double-exponential priors in our setting render this procedure impracticable. Within each

orthant of the parameter space, a quadratic approximation to the log likelihood is reasonable, but because of the ℓ_1 penalty term, a different quadratic approximation is required for each of the roughly $2^{18,000}$ orthants, which is not feasible.

Given our interest in combining data with prior beliefs, an inverse Wishart prior on covariance is tempting because it allows easy sampling from the full posterior. However, the inverse Wishart distribution is restrictive in form (Barnard et al., 2000) and does not provide a straightforward way to describe prior beliefs about correlations.

Another tempting alternative is that of Liu et al. (2014), who give a simple thresholding method for producing a penalized correlation matrix that is guaranteed to be positive definite. Their estimator solves

$$\underset{\omega \succ \delta \cdot I}{\operatorname{argmin}} \frac{1}{2} \|\tilde{R} - \omega\|_F^2 + \lambda \|W * \omega\|_{1,\text{off}}, \tag{3.26}$$

to produce an estimator among the set of valid correlation matrices with minimum eigenvalue no smaller than δ. Although the weight matrix, W, is in principle arbitrary, they use W to induce greater shrinkage where empirical correlations are weakest, not as a means of conveying prior information. We would be hesitant to replace W with our penalty matrix P, as that off-license use of their method would not incorporate prior information in a principled way.

Our method can be generalized to shrink estimated correlations towards non-zero values by replacing the penalty term $\lambda \|P * R\|_1$ with $\lambda \|P * (R - S)\|_1$ for some target matrix S. This may be desirable in cases where heavily structured estimates of correlations are available, as is the case for modeling of fertility (Fosdick & Raftery, 2014).

Note that we have used the 2012 revision of the WPP here (United Nations, 2012). The more recent 2015 revision (United Nations, 2015b) contains one additional data point. It would be of interest to redo the analysis with the newer data, but we expect the results would be similar.

Chapter 4

A PSEUDO-BAYES ESTIMATOR FOR GLOBAL MIGRATION FLOW TABLES

In this chapter we propose a method for estimating migration flows between all pairs of countries, including breakdowns by place of birth. Our estimator is a Pseudo-Bayes estimator which smooths a set of state-of-the-art estimates of migration flows towards a simpler estimate which contains fewer structural zeroes. The smoothing process provides a natural way to bypass the state-of-the-art estimator's unrealistic assumption that the number of global migrants is as small as possible. We produce estimates of global migration flows at five-year intervals from 1990 to 2015, finding that the total number of individuals migrating has oscillated around 1% of the global population per five-year period. Our estimates contain substantial return migration flows, correcting a weakness inherent to the minimum-migrants assumption in previous estimates.

4.1 Introduction

In much of the developed world, volatility in population is now largely driven by migration, rather than fertility or mortality (Azose et al., 2016). However, migration remains difficult to estimate (Bilsborrow, 1997; Ratha & Shaw, 2007). This difficulty is especially pronounced in the context of return migration, whose estimation may be hindered by unauthorized migration and poor tracking of migration in the developing world (Massey & Capoferro, 2004; Zlotnik, 1987).

In this paper we propose a method for estimation of a complete matrix of global bilateral migration flow data using evidence from migrant stocks. We apply our method to United Nations data on migrant stocks (United Nations, 2015a) to produce estimates of migration

flows spanning the period from 1990 to 2015 in five-year intervals. Our method builds on Abel's (2013a) estimator of migration flows, which relies on an unrealistic assumption that the total number of global migrants is as small as possible in each time period. By weakening this assumption, we produce flow estimates with several desireable features. Firstly, the minimum-migration assumption introduces many structural zeroes into the estimated migration flow tables. The true migration data are unlikely to be so heavily structured, and the preponderance of exact zeroes may make the resulting flow estimates difficult to model. Our method is designed to smooth out these structural zeroes. Secondly, we produce flow estimates with substantially higher return migration flows, including flows from the United States to Latin America, and from the oil-producing countries of the Persian Gulf to their typical labor-supplying countries in South Asia. Examination of estimates of the flows between Mexico and the USA in Section 4.4.1 confirms that our higher estimates of return migration are realistic. Finally, our flow estimates may be aggregated into an estimate of the total number of global migrants, rather than simply a lower bound for this figure. We estimate that the total proportion of the world's population migrating has fluctuated at levels between 0.96% and 1.10% of the global population in each five-year period covering 1990–2015.

4.1.1 Background and motivation

There is a strong relationship between the level of aggregation in migration estimates and the ease of producing the estimates. On one end of the spectrum are net migration data—that is, estimates of the total inflow to country i minus the total outflow from country i. Net migration has the favorable quality of being the simplest form that migration can take in the demographic balancing equation. (i.e. Change in population = births – deaths + net migration.) Consequently, in countries where reliable estimates exist for all three of population, births, and deaths, it is possible to estimate net migration via residual methods. The estimates of net migration in the United Nations' World Population Prospects (WPP) are typically produced using a combination of residual methods and estimates of refugee

migration, where applicable (United Nations, 2015b). Since net migration can be viewed as the contribution of migration to population change, projections of net migration can be used with little disaggregation to produce population projections (Azose et al., 2016). Despite their relative ease of estimation and usefulness for projection, net migration data are less than ideal because they aggregate out all information about where in-migrants are coming from and where out-migrants are moving.

Migration flows are at the other end of the spectrum from net migration data; these data are highly disaggregated, but difficult to estimate. The basic quantity of interest is the number of individuals moving from origin country i to destination country j during time period t, which we will denote by m_{ijt}. These flows may be further disaggregated by place of birth, k, so that m_{ijkt} denotes the number of individuals born in country k who move from i to j during time period t. In contrast to net migration, flows cannot be estimated with a simple application of the demographic balancing equation. Typical sources of migration flow data are population registration systems, border crossing data, detailed population censuses, or alien registers (Nowok et al., 2006), all of which require a country to have a substantial data infrastructure. Consequently, many countries publish no estimates of their migration flows. In countries that do estimate flows, quality of those estimates is known to be mixed even in the developed world. Abel (2010) examined official estimates of flows between pairs of European countries and found that flow estimates often differed substantially depending on whether the sending country or the receiving country produced the estimates.

Nevertheless, some attempts have been made to produce and/or aggregate high quality migration flow estimates for limited groups of countries. The Organization for Economic Co-Operation and Development (OECD) compiles estimates of annual migration flows to and from 34 of the current 35 OECD member states[1] in their statistical database (Organization for Economic Co-Operation and Development, 2015). Another set of flow estimates was produced by the Migration Modeling for Statistical Analyses (MIMOSA) project (De Beer et al.,

[1]The missing country is Latvia, which joined in July 2016, after the data were published.

2010; Raymer et al., 2011). These consist of estimated total flows between 31 pairs of European countries generally based on information from national statistics offices (Kupiszewska & Wiśniowski, 2009). Notably, these two sets of estimates almost exclusively cover flows where either the origin or the destination country is economically developed. There is relatively little reliable data avialable on "South-South" flows for which both the origin and destination country are members of the global south.

In contrast to migration flows, high quality data on migrant stocks are relatively easy to come by. Although tracking migration flows often requires border monitoring or an up-to-date population register, stocks can be estimated by including a question about place of birth on census forms. The United Nations has compiled migrant stock data from various national sources for all countries at five-year intervals from 1990 to 2015 (United Nations, 2015a). A second set of migrant stock estimates are available from the World Bank at ten-year intervals from 1960 to 2000 (Özden et al., 2011), but we use the United Nations data for our analysis because of the greater recency and finer granularity in time.

Because of the global availability of migrant stock data, stocks are a good starting point for estimation of flows. The key observation is that changes in stock data provide information about the minimum number of people who move in a given time period. For instance, if we observe a large growth in the stock of individuals of French birth living in Germany then we know that an equivalent number of individuals of French birth must have moved to Germany. Furthermore, it seems likely that most of those individuals would have moved from places with a large existing stock of French nationals (e.g. France.)

Abel (2013a) uses the observation that stocks contain information about flows to produce global estimates of migration flows from place-of-birth data. He treats migration flows as coming from a log-linear Poisson model with an assumed quasi-independence structure. (We describe this model in more detail in Section 4.2.) Estimation of the model parameters, and consequently migration flows, requires knowledge of 1) migrant stocks at the start and end of a time period, 2) births and deaths during that time period, and 3) the number of individuals who *do not* migrate, broken down by place of birth and place of residence. Reliable estimates

for the first two are available from the United Nations migrant stock database (United Nations, 2015a) and WPP (United Nations, 2015b), respectively. However, the number of non-movers must be estimated separately. In practice, Abel handles this by making the simplifying assumption that the number of non-movers is maximized, or equivalently that the number of migrants is minimized. The resulting flow estimates can be seen as a lower bound on migration, in the sense that they are the estimates that have the lowest total number of migrants while retaining consistency with the observed changes in stocks. Through the remainder of this paper, we refer to estimates produced using Abel's (2013) methodology as "minimum migration" estimates, which we sometimes shorten to "MM" estimates.

Unfortunately, the requirement that the number of movers be minimized places strict limits on the structure of the estimated flow table. Most notably, in the MM estimates, either $\hat{m}_{ijkt}$ or its cross-flow, $\hat{m}_{jikt}$, must be zero. For instance, there can be a non-zero flow of Mexican-born individuals from Mexico to the USA or from the USA to Mexico, but not both. As a result, the MM method produces a matrix of flow estimates that is typically rather sparse. For instance, in 2010–2015, only 26% of flows m_{ijt} are estimated to be non-zero. Especially problematic are the flow estimates for country pairs where large flows may be present in both directions. For example, the MM method estimates the 2010–2015 flow from India to Australia to be 71,812 individuals, reflective of a large and growing population of Indian-born in Australia. In contrast, the reverse flow from Australia to India is estimated to be one individual. In fact, over the five-year time period, the MM method estimates a total of only 54 individuals migrating from Oceania to South Asia.

4.2 Methodology

We produce estimates of migration flows using a smoothed version of Abel's (2013) minimum-migration method to estimate migration flows from stock data. We begin with a brief description of the MM method before moving on to our extension.

4.2.1 Abel's (2013) Minimum-Migration Method

It is simplest to describe the MM method in an idealized context where there are no births or deaths and then extend the method to handle births and deaths. Assume for the moment that their are neither births nor deaths and that all change in population is attributable to migration. Section 4.2.2 describes the treatment of births and deaths.

At the finest level of granularity, the quantities of interest are the unknown flow tables M_{kt} defined by

$$M_{kt} := \begin{pmatrix} m_{11kt} & m_{12kt} & \cdots & m_{1Ckt} \\ m_{21kt} & m_{22kt} & \cdots & m_{2Ckt} \\ \vdots & \vdots & \ddots & \vdots \\ m_{C1kt} & m_{C2kt} & \cdots & m_{CCkt} \end{pmatrix}, \tag{4.1}$$

where the entry m_{ijkt} is a flow from country i to country j of individuals with place of birth k during time period t. Note that only the off-diagonal elements of M_{kt} represent genuine migration flows. The diagonal elements m_{iikt} represent individuals who resided in country i both at the beginning and end of the time period—that is, stayers rather than movers.

Strictly speaking, in the MM method, m_{ijkt} represents the number of individuals with place of birth k who resided in country i at time t and country j at time t'. That is, m_{ijkt} does not measure migration events, but changes in residence when comparing only the beginning and end of the time period. This distinction means that an individual who moves from country A to country B to country C during a single time period is only counted once under the MM method, as an A-to-C migrant. For short time periods, m_{ijkt} as we define it here should be a good approximation of the full migration flow.

A key observation is that in absence of births and deaths, the row and column sums of M_{kt} are known. Each row sum, $\sum_j m_{ijkt}$, gives the total population of individuals with place of birth k residing in country i at time t. That is, row sums are migrant stocks at time t, which we will denote by n_{ikt}. Likewise, column sums, $\sum_i m_{ijkt}$, give the population of individuals with place of birth k residing in country j at time t'—that is, the stocks $n_{jkt'}$.

The MM method amounts to fitting a separate Poisson log-linear model to each table, M_{kt}. For the moment we consider only a single table, M_{kt}, and suppress the k and t subscripts except where they are needed to draw a distinction between time points. The model assumes table elements m_{ij} are independently drawn from Poisson distributions,

$$m_{ij} \sim \text{Poisson}(y_{ij}). \tag{4.2}$$

Furthermore, the Poisson means are assumed to follow a quasi-independence structure,

$$y_{ij} = \alpha_i \beta_j \delta_{ij} o_{ij}, \tag{4.3}$$

where o_{ij} is a prespecified offset term and the δ_{ij} terms are constrained so that $\delta_{ij} = 1$ unless $i = j$. This condition on the δ_{ij} terms says that the Poisson means for the diagonal elements are unconstrained while the off-diagonal elements follow an independence structure. Conceptually, this constraint can be viewed a statement that moving is different from staying.

Sufficient statistics for inference on the Poisson model parameters are given by the known row margins (i.e. stocks at time t), the known column margins (i.e. stocks at time t'), and the unknown diagonal entries $\{m_{ii} \text{ for } i = 1, \ldots, C\}$. Once the diagonal values are fixed, the maximum likelihood estimates for the parameter vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}$, and $\boldsymbol{\delta}$ can be found via iterated proportional fitting (Bishop et al., 1975; Willekens, 1999). One implementation of the iterated proportional fitting procedure can be found in the `migest` package (Abel, 2013b).

The assumption that the number of stayers is maximized seems a reasonable first-order approximation to the truth. Moving between countries carries significant monetary and psychological costs (Sjaastad, 1962). In places where detailed estimates are available, it is clear that many more people stay put than move in any given time period (Organization for Economic Co-Operation and Development, 2015).

However, the requirement of *exact* rather than approximate maximization introduces

many structural zeroes into the estimates. The maximum possible value of m_{ii} conditional on the row and column margins is given by $\min(n_{ikt}, n_{ikt'})$ (Fréchet, 1951). Consequently, maximizing a diagonal entry forces all of the other entries in either the same row or column to be zero. This is a problematic side effect of the minimum-movers assumption, and one which our method addresses.

4.2.2 Treatment of births and deaths

The method from the previous section assumed no births or deaths. Abel (2013a) extends the MM method to incorporate known counts of births and deaths. Births in country i are assumed to increase the value of $n_{iit'}$, that is, the stock of individuals born in country i and residing in country i at the end of the time period of interest. Deaths in country i are assumed to occur proportionally to the place-of-birth composition in country i and decrease the relevant stocks. Both of these adjustments can be backed out to produce modified stock data which include only those individuals who lived through the complete time period. However, after adjusting stocks for births and deaths, stocks at the beginning and end of the time period may no longer be consistent with one another. A final adjustment brings these into concordance by adding additional migrant to or from an external region. (If not all countries are modeled, this external region is the rest of the world. If the model already includes all countries, this final step can be viewed as harmonization of stocks in light of imperfect data on births and deaths.)

In our results, we adopt Abel's (2013) treatment of births and deaths without modification.

4.2.3 Treatment of offsets

The Poisson model includes offset terms o_{ij}. In producing estimates on real data, Abel (2013a) sets these to

$$o_{ij} = d_{ij}^{-1}, \tag{4.4}$$

where d_{ij} is the distance between the capital cities of countries i and j. This offset term reflects the belief that migrant flows are more likely over shorter geographical distances. This belief is backed up by both migration theory (Lee, 1966; Sjaastad, 1962) and empirical evidence (Kim & Cohen, 2010).

In the subsequent analysis in this paper, we choose instead to omit the offset terms, or equivalently, to take $o_{ij} = 1$ for all country pairs. We see two main motivations for this decision. Firstly, distances between capital cities are readily available for most, but not all country pairs. Out of the 200 countries or regions in the United Nations data on migrant stocks, 190 have distance data available from CEPII's `geodist` database (Mayer & Zignago, 2011).[2] Omitting offsets allows us to produce estimates of migration flows for all regions rather than all but ten. Secondly, migration flow estimates produced using the MM method are extremely similar regardless of whether we choose $o_{ij} = 1$ or $o_{ij} = d_{ij}^{-1}$. In practice, we find that the correlation between estimated migration flows with the two different offsets is at least 0.999 for all time periods. This is in line with the finding by Abel (2016) that including inverse distance as an offset has little impact on estimates. Given that the more complex model returns nearly identical estimates to a simpler model without offsets, we have a preference for the simpler model.

Lanzieri (2014) examined Abel's method in relation to European migration statistics and suggested that Abel's estimates could be improved by introducing additional information in the form of offsets. Given that we find very little difference in estimates whether offset terms are set to inverse distance or omitted entirely, offsets may not be the most effective way to introduce additional information into this model.

4.2.4 A pseudo-Bayes estimator

We now construct a pseudo-Bayes estimator $\hat{M}_{kt}^{PB}$ for M_{kt} which has the same row and column totals as Abel's (2013) MM estimator $\hat{M}_{kt}^{A}$, but lacks the problematic structural

[2]The ten missing countries/regions are the Channel Islands, Curaçao, Guam, Mayotte, Montenegro, Serbia, South Sudan, the State of Palestine, Timor-Leste, and the United States Virgin Islands.

zeroes. The pseudo-Bayes estimator is produced using the following procedure:

1. Using data on stocks, births, and deaths, find the MM estimate, $\hat{M}^A_{kt}$ for each flow matrix M_{kt}. These estimates should include adjustments of stocks for births, deaths, and flows to external regions.

2. Compute the row and column totals of $\hat{M}^A_{kt}$ to extract adjusted stock estimates n^*_{ikt} and $n^*_{jkt'}$.

3. Construct a second set of estimates, $\hat{M}^I_{kt}$, with independence structure. Elements of $\hat{M}^I_{kt}$ are given by

$$\hat{m}^I_{ijkt} = \frac{n^*_{ikt} n^*_{jkt'}}{\sum_i n^*_{ikt}}. \tag{4.5}$$

   Note that this is the matrix of maximum likelihood estimates from a Poisson model on migration flows where $m_{ijkt} \overset{ind}{\sim} \text{Poisson}(\alpha_{ikt}\beta_{jkt})$ conditional on the given row and column totals with no additional assumptions made about the diagonal.

4. Construct the pseudo-Bayes estimator $\hat{M}^{PB}_{kt}$ as a convex combination of $\hat{M}^A_{kt}$ and $\hat{M}^I_{kt}$:

$$\hat{M}^{PB}_{kt}(w) = w \cdot \hat{M}^A_{kt} + (1 - w) \cdot \hat{M}^I_{kt} \tag{4.6}$$

   for some value $w \in [0, 1]$.

For any choice of $w \in [0, 1]$, the estimator $\hat{M}^{PB}_{kt}(w)$ is a pseudo-Bayes estimator of the unknown table entries. (See Appendix C.1 for mathematical details.) One way to think of the pseudo-Bayes estimator is as smoothing $\hat{M}^A_{kt}$ towards a matrix with smaller diagonal and fewer zero entries. This form of smoothing is a common approach to estimation of cell probabilities in contingency tables with many observed zeroes (Fienberg & Holland, 1973).

We note several advantageous properties of this estimator. First, row and column sums of each table will be identical to those of $\hat{M}^A_{kt}$, by construction. Second, this estimator lacks the structural zeroes of $\hat{M}^A_{kt}$. Before rounding to integer values, estimated flows will be non-zero

whenever countries i and j both have non-zero populations of individuals with place of birth k. The final estimate may round down to zero, but only when those populations are small, in which case migration flows are indeed likely to be small. Third, for values of w close to 1, the diagonal entries of $M_{kt}^{PB}(w)$ will be nearly maximized, but not exactly maximized. This allows us to retain the property that the number of non-movers is large. Finally, once $\hat{M}_{kt}^{A}$ is computed, finding the pseudo-Bayes estimator requires very little additional computation.

4.2.5 Estimating an optimal weight, w

Any choice of w between 0 and 1 will correspond to a valid pseudo-Bayes estimator. In practice, a sensible procedure is to select a value of w which minimizes the expected value of some loss function. Computation of a loss function requires us to have ground truth data to compare estimates against—for this application, we need a set of "true" migration flows which have been estimated via some other method. In this section we first discuss the data sources available for training, then motivate a choice of loss function, and finally present estimates of an optimal value of w.

Sources of flow data

To select w, we compare pseudo-Bayes flow estimates $M_{kt}^{PB}(w)$ against ground truth flow data from two sources: the OECD (Organization for Economic Co-Operation and Development, 2015) and MIMOSA (Raymer et al., 2011).

The OECD migration flow data are split into estimates of flows to and from OECD countries. Inflow data take the form of counts of the total number of foreign in-migrants to country j from all origins $\ell \neq j$ broken down by nationality k. (That is, the available quantities have the form $\sum_{\ell|\ell\neq j} m_{\ell jkt}$.) These data are on an annual time scale covering years from 2000 to 2013, although not all OECD countries have flow estimates for all years. Likewise, outflow data are estimates of the total number of foreign out-migrants from country i by nationality (i.e. $\sum_{\ell|\ell\neq i} m_{i\ell kt}$). Indices i and j are restricted to the 34 OECD countries, while nationality k covers all countries.

Note that the index k in the OECD data represents nationality rather than place of birth. For a large majority of individuals, nationality and place of birth will be in alignment. However, this will not always be the case, as individuals may change their nationality or may not receive citizenship in their country of birth. Rather than attempting to model nationality transitions on top of migratory transitions, we simply treat nationality and place of birth as interchangeable in this analysis. Flows by nationality and flows by place of birth should be reasonably similar so long as nationality transitions and birth outside the country of citizenship are relatively rare events.

In contrast to the OECD data, MIMOSA flow estimates are broken down by country of origin and country of destination, but aggregate out all data on nationality or place of birth. MIMOSA reports aggregated flow estimates $\sum_\ell m_{ij\ell t}$ for origin-destination country pairs (i, j) where both i and j come from a set of 31 European countries. Flow estimates are provided on an annual time scale covering 2002–2007.

Selecting a loss function

A loss function, $L(m, \hat{m})$, is a function that describes the cost associated with producing an estimate of $\hat{m}$ when the true migration flow is m. In principle, the choice of loss function should be motivated by the relative costs of making different errors in estimating migration flows. As such, we will only consider loss functions with the following two properties:

1. Correctly estimating a migration flow incurs no loss. That is,

$$L(m, m) = 0, \tag{4.7}$$

and

2. Any estimate other than the true flow incurs more loss than the truth. That is,

$$L(m, \hat{m}) > 0 \text{ whenever } \hat{m} \neq m. \tag{4.8}$$

A common, convenient choice of loss function is squared error loss in migration flows, defined as

$$L_{SE}(m, \hat{m}) = (m - \hat{m})^2. \tag{4.9}$$

Squared error loss in flows is likely to be useful only in limited circumstances. This loss function makes the statement that the cost associated with estimating a migration flow of 10,000 when the true flow is 20,000 is the same as the cost of estimating a migration flow of 990,000 when the true flow is 1,000,000. Squared error in flows is not reasonable as a choice for a global loss function, since it will favor results that are very accurate for a few of the largest flows at the expense of misestimating many smaller flows.

It is also tempting to consider squared error in log-flows, defined as

$$L_{SEL}(m, \hat{m}) = (\log(m + k) - \log(\hat{m} + k))^2, \tag{4.10}$$

where k is a small, positive constant. This loss function does not suffer the same weakness as squared error loss in migration flows; the cost of estimating a flow that is 50% too small is nearly equal whether the true flow is 1,000,000 or 10,000. However, losses can get unreasonably large as true flows move towards zero. For example, if we pick $k = 1$, then $L_{SEL}(10, 2)$ is more than three times as large as $L_{SEL}(2 \text{ million}, 1 \text{ million})$. For most practical purposes, the latter error is likely to be far more costly than the former. When dealing with very small flows, an ideal loss function would keep losses low for even large relative errors so long as absolute errors remain small.

Evaluating errors in flows on a rate scale generally allows us to avoid the weaknesses of squared error loss in flows or log-flows. We define a loss function of squared error in rates by

$$L_{SER}(m_{ij}, \hat{m}_{ij}) = \left(\frac{m_{ij}}{\text{population}_i} - \frac{\hat{m}_{ij}}{\text{population}_i} \right)^2, \tag{4.11}$$

where population_i denotes the total population of the origin country at the beginning of the time period. This loss function implies that there is a high cost to large errors in flow

estimates relative to the population in the country of origin. Errors in small flows are not penalized heavily, as they would be if we used squared error loss on log-flows. Errors which are large relative to the true flow are not necessarily costly; they are weighted heavily only when they are large relative to a country's population.

Because of the structure of the training data, it is not always feasible to put the population of the origin country in the denominator inside the loss function. For inflows to OECD countries, we use instead the population of the destination country as the denominator. For MIMOSA flows, where both origin and destination populations are available, we use the population of the country of origin.

Although we will use squared error in rates as our loss function for choosing a value of w, there may be contexts in which the other loss functions are more suitable. In Section 4.4.2, we evaluate our performance on simulated data with respect to the three loss functions proposed here and find that our method performs well under all three.

4.2.6 Estimating a five-year weight from annual data

The optimal value of w will depend on the length of the time period in consideration. In the limit as time period length decreases to zero, we would expect the MM estimates to be perfect. (If the time period is so short that only a single person migrates, that move will be reflected by a unitary change in migrant stocks, and the MM method will reproduce the single migration event.) In that limiting case, the optimal choice is to take $w = 1$ and use the correct, unsmoothed MM estimator. As the time period in consideration increases in length, we expect to see more bi-directional flows within country pairs, and therefore smaller optimal values of w.

We now derive an optimal five-year value of w based on annual data. Our derivation relies on three key assumptions. Firstly, we assume that births and deaths have been correctly accounted for so that all population change is due to migration. Secondly, we assume that migration flows occur uniformly across each five-year time period. Finally, we assume that neither country populations nor the composition of migrant stocks will change too dramat-

ically over the longer time period. In reality, we do not expect these assumptions to hold precisely, but they should be a good approximation of the truth for most migration flows over a five-year period.

The quantity to be estimated is an array of five-year migration flow rates, which we'll notate as M_{1-5} in order to distinguish it from annual migration flow rates $M_1, M_2, \ldots, M_5$. Finding an optimal five-year value of w is equivalent to finding a linear combination of the MM estimates $\hat{M}^A_{1-5}$ and independence estimates $\hat{M}^I_{1-5}$ which is optimal in that it minimizes mean squared error in rates. That is, our problem is expressible as

$$\text{minimize}_{w\in[0,1]} \; \|M_{1-5} - (w\hat{M}^A_{1-5} + (1-w)\hat{M}^I_{1-5})\|_2^2. \tag{4.12}$$

Under the assumption that flows occur uniformly across each five-year time period, we can express all of the five-year quantities in terms of one-year quantities. Since flows are assumed to occur uniformly across time, we have

$$M_1 \approx M_2 \approx \cdots \approx M_5 \approx \frac{1}{5} M_{1-5}. \tag{4.13}$$

The annual flow counts are identical because of the assumption of uniformity across time. Consequently, the annual flow *rates* are approximately equal because of the assumption that country population is nearly constant over time.

How do the estimates behave? The MM estimates are minimum flows to account for change in migrant stocks. Each annual flow will be estimated to be approximately one fifth of the quinquennial flow estimate.

$$\hat{M}^A_1 \approx \hat{M}^A_2 \approx \cdots \approx \hat{M}^A_5 \approx \frac{1}{5}\hat{M}^A_{1-5}. \tag{4.14}$$

In contrast, the independence estimates depend on the magnitude of migrant stocks, which we assume to be approximately constant over each five-year period. Rather than each annual estimate being roughly one fifth of the quinquennial estimate, each annual estimate

under independence structure will be nearly the same as the quinquennial estimate:

$$\hat{M}_1^I \approx \hat{M}_2^I \approx \cdots \approx \hat{M}_5^I \approx \hat{M}_{1-5}^I. \tag{4.15}$$

Substituting these expressions into Equation 4.12 allows us to rewrite our minimization in terms of annual quantities rather than quinquennial quantities. The minimization problem from Equation 4.12 is approximately expressible as:

$$\text{minimize}_{w\in[0,1]} \ \|5M_1 - (5w\hat{M}_1^A + (1-w)\hat{M}_1^I)\|_2^2. \tag{4.16}$$

Or, equivalently,

$$\text{minimize}_{w\in[0,1]} \ \left\| M_1 - \left(w\hat{M}_1^A + (1-w)\left(\frac{1}{5}\hat{M}_1^I\right)\right)\right\|_2^2. \tag{4.17}$$

This is nearly identical to the optimization problem we solve when selecting a value of w which minimizes errors in the annual data. To choose a w which works well for five-year data when we have annual data available, we substitute annual flow data along with annual MM flow estimates and independence-structured annual flow estimates into Equation 4.17. The minimization problem then reduces to a quadratic equation in w, which is easily solved.

In the next section we apply this optimization procedure to OECD and MIMOSA data and find annually and quinquennially optimal values of w.

4.3 Results

In this section, we first compare the MM estimates and independence-structured estimates to the OECD and MIMOSA data to illustrate the weak points of those two estimates in practice. We then present results of the procedure to select an optimal weighting factor, w. This is followed by summaries and visualizations of the resulting five-year flow estimates, and a comparison of our flow estimates to the MM estimates.

4.3.1 Comparison of minimum-movers estimates to independence-structured estimates

We used UN data on migrant stocks to compute MM flow estimates $\hat{M}^A_{kt}$ and independence-structured flow estimates $\hat{M}^I_{kt}$. Comparing these with annual flow estimates from the OECD (Organization for Economic Co-Operation and Development, 2015) and MIMOSA (Raymer et al., 2011) allows us to confirm the shortcomings of these simpler estimates and eventually to compute values of w which minimize mean squared error in migration rates. Since the OECD and MIMOSA flow estimates are annual estimates while the UN stock data are at five-year intervals, we linearly interpolated stocks to match the annual frequency of the flow estimates.

The MM estimates generally agree fairly well with the MIMOSA estimates of intra-European flows. The top left panel of Figure 4.1 plots the MM estimates versus the MIMOSA estimates on a log-log scale. The most noticeable disagreement between the two sets of estimates is that the MM estimates contain too many zeroes. (This is visible in the figure as a dark cloud along the horizontal axis.) MIMOSA reports 2.5% of the flows among these 31 European countries to be exactly zero. In contrast, the MM estimates are zeroes 12.3% of the time. The MM estimates are also slightly biased downwards. (The bulk of the cloud in the top left panel of Figure 4.1 lies below the $y = x$ line.) This is the anticipated result of the minimum-migration assumption in Abel's method, and it bears out when compared to the MIMOSA estimates.

In contrast, the independence-structured estimates $\hat{M}^I_{kt}$ are biased high, but largely lack the false zeroes of the MM estimates. The center left panel of Figure 4.1 plots the independence estimates versus the MIMOSA estimates. The independence estimates are typically too high by a factor of 10 to 20, but this overestimation is quite consistent.

A similar pattern emerges from the data on inflows to OECD countries, which are not depicted here because of their similarity to the MIMOSA plots. 32.0% of the MM estimates are zeroes, compared to 19.0% of the independence estimates and 13.3% of the reported true flows.

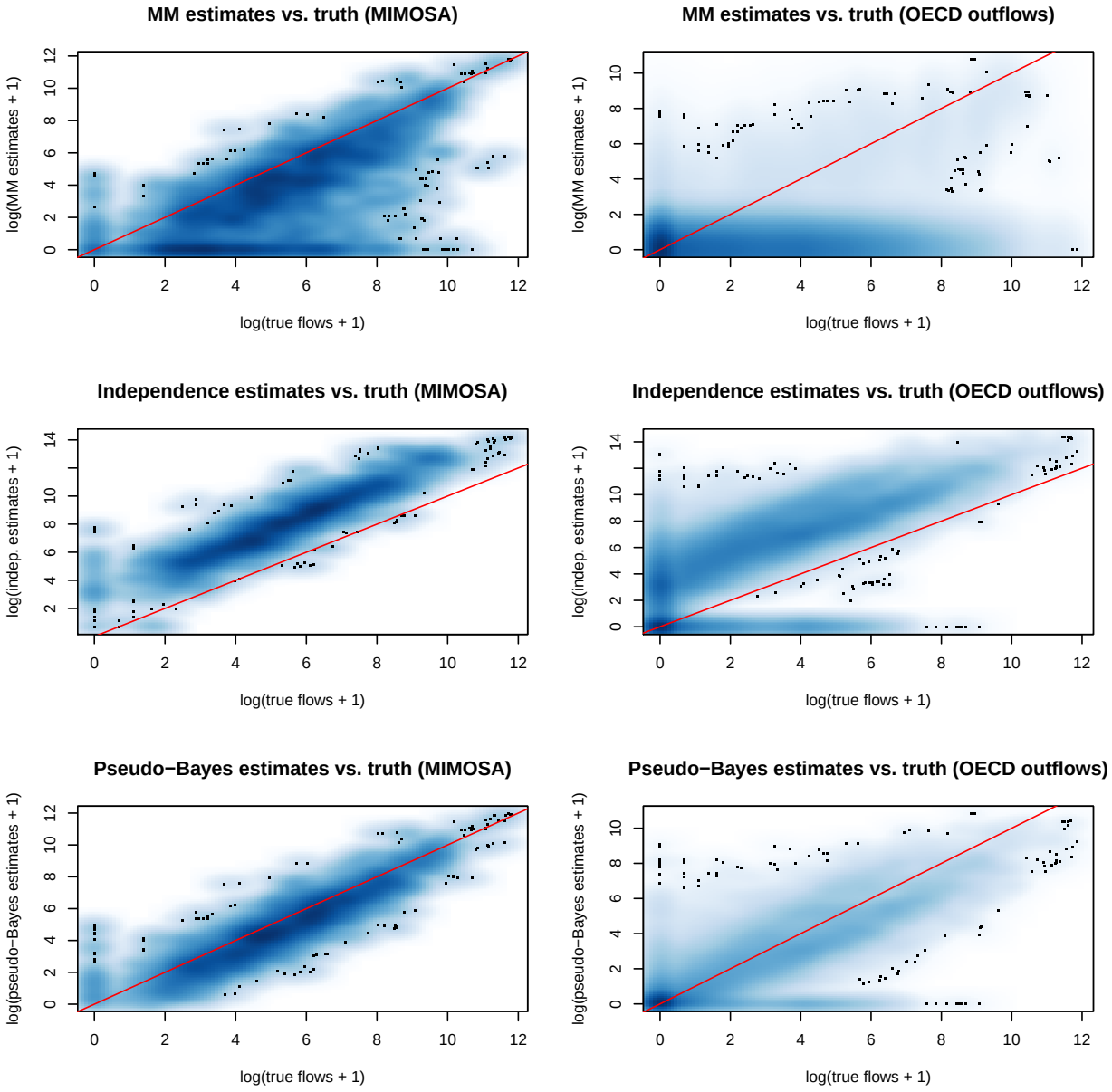

Figure 4.1: Comparison of gold standard bilateral flow estimates to estimates produced using three other methods: Abel's (2013) minimum-migration method (top row), independence-structured estimates (center row), and pseudo-Bayes estimates (bottom row). Gold standard data are MIMOSA data on European migration flows for 2002–2007 (left column) and OECD estimates of out-flows from OECD countries (right column). Red line is $y = x$.

The picture is much more striking for estimates of outflows from the OECD countries. Fully 95.7% of MM estimates of outflows from OECD countries are zeroes, compared to 18.3% of the independence-structured estimates and 23.0% of the reported flows. The top right and center right panels of Figure 4.1 plot, respectively, Abel's MM estimates and independence-structured estimates of outflows from OECD countries.

The top right panel of Figure 4.1 shows that the MM method is severely underestimating outflows from the OECD countries. This is a result of the minimum-mover constraint, which precludes cross-flows. If more people with place of birth k are moving into a country than out of it, as is the case for many OECD countries, then the MM outflow estimate will be zero. In contrast, the independence-structured matrix typically overestimates these flows.

In contrast, our pseudo-Bayes estimates in the bottom row of Figure 4.1 feature neither the overabundance of zeroes introduced by the MM method nor the systematic overestimation present in the independence estimates. (The pseudo-Bayes estimates in this figure were produced using a weighting value of $w = 0.9825$, the most conservative annual w value found in Table 4.1 in the following section.)

4.3.2 Results of w selection

Altogether, we estimated the optimal value of w separately using four different combinations of ground-truth data and optimization criterion. In all cases we choose w to minimize mean squared error in migration rates, but we test two definitions of migration rates which differ only in the population in the denominator. The three combinations of data and criterion are:

1. In-flows to OECD countries. Denominator on migration rates is the population of the destination country.

2. Out-flows from OECD countries. Denominator on migration rates is the population of the origin country.

3. MIMOSA flows. Denominator on migration rates is the population of the origin country.

Figure 4.2 plots optimal values of w, broken down by year as well as data and optimization criterion. All values are close to 1, meaning that the optimal pseudo-Bayes estimates will generally be close to estimates produced with the MM method. Furthermore, optimal w values are highly consistent across both data sources and time. This consistency in optimal values of w suggests that it is reasonable to use a single value of w to produce estimates over multiple time periods and multiple types of migration flows (i.e. South-North, North-North, and North-South.) The left column of Table 4.1 compares average optimal annual values of w across data sources. The right column extends these to five-year estimates using the method from Section 4.2.6.

4.3.3 Estimates of flows between regions

We computed our pseudo-Bayes estimates of global migration flows between all pairs of countries over five-year intervals covering the period from 1990 to 2015. Quinquennial migrant

Figure 4.2: One-year estimates of the optimal value of w. Optimal values show consistency across data source (OECD versus MIMOSA) and time. Values are also consistent across North-South flows (OECD out-flows), South-North flows (OECD in-flows), and North-North flows (MIMOSA).

stock data come from the United Nations international migrant stock dataset (United Nations, 2015a). Births and deaths in all countries are taken from WPP 2015 (United Nations, 2015b). For the weighting factor, w, we use a value of 0.8956, the most conservative of the five-year values in Table 4.1.

This method produces complete bilateral flow tables for each time period. Figures 4.3 through 4.7 illustrate these flows for periods from 1990–1995 through 2010–2015 using circular migration plots. This style of plot condenses large flow matrices into a form that makes it easy to distinguish the relative magnitudes of flows (Sander et al., 2014; Abel & Sander, 2014). International migrant flows are depicted with arrows pointing from the region of

Table 4.1: Estimated optimal values of w. In practice we use the most conservative w values (top row) when producing our flow estimates.

Data used for estimation	Annual w	Five-year w
OECD Inflows only	0.9825	0.8956
OECD Outflows only	0.9775	0.8848
MIMOSA flows (denominator: origin pop.)	0.9755	0.8613

origin to the region of destination. Colors distinguish the region of origin. Tick marks along the circumference give the size of the flow, in millions of migrants. Each figure contains two circular plots. The left plot shows flow estimates from the pseudo-Bayes estimator while the right plot show estimates produced with Abel's (2013) minimum-migration method. Within each figure, the two circular plots are scaled so that equal arc length corresponds to equal numbers of estimated migrants. Plots were produced using the `circlize` package (Gu et al., 2014) and demonstration code from the `migest` package (Abel, 2013b) in `R`.

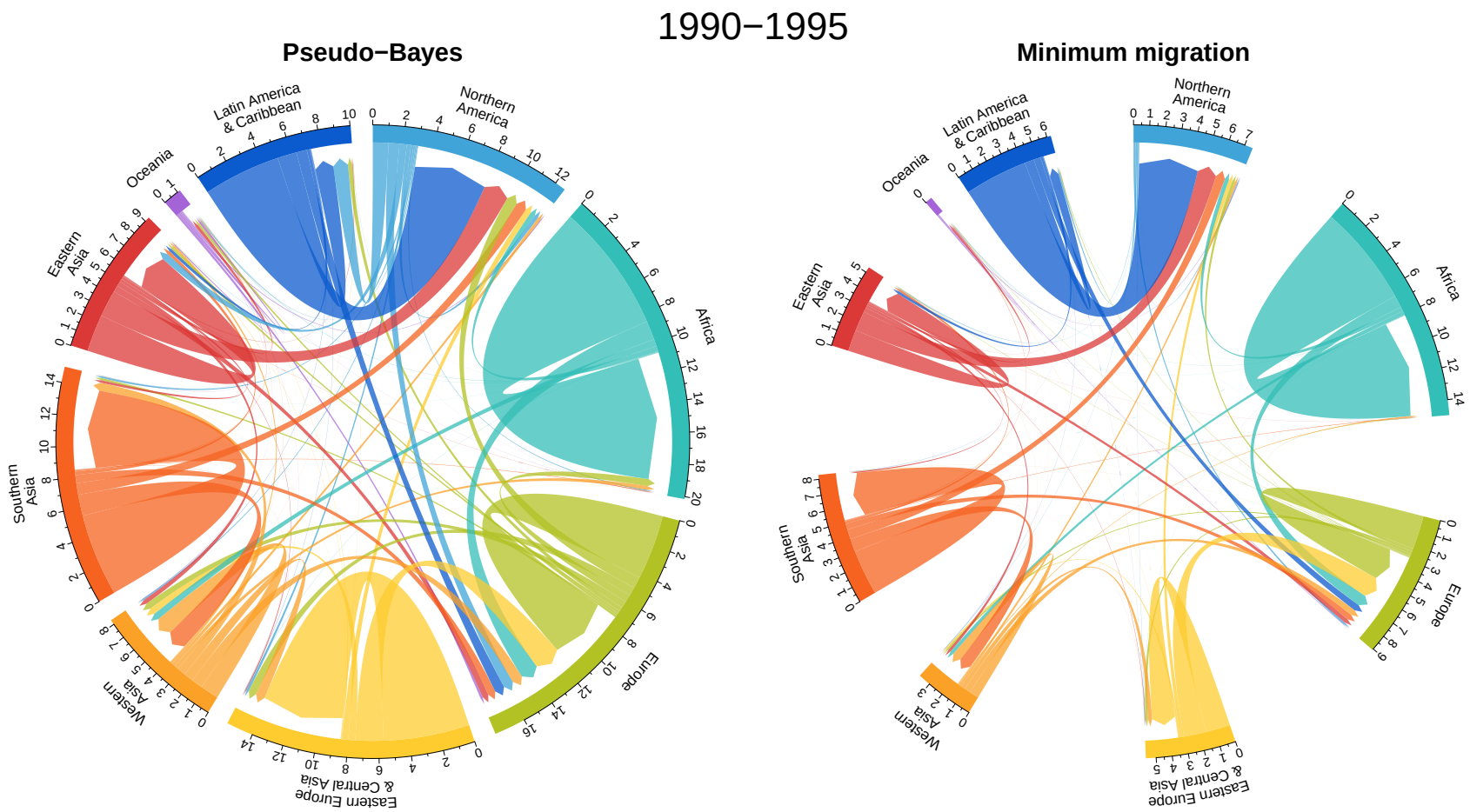

Figure 4.3: Estimated migration flows for 1990–1995. Left panel: Pseudo-Bayes estimates. Right panel: Estimates using the Abel (2013) method. Plots are scaled so that equal arc lengths represent equal numbers of migrants.

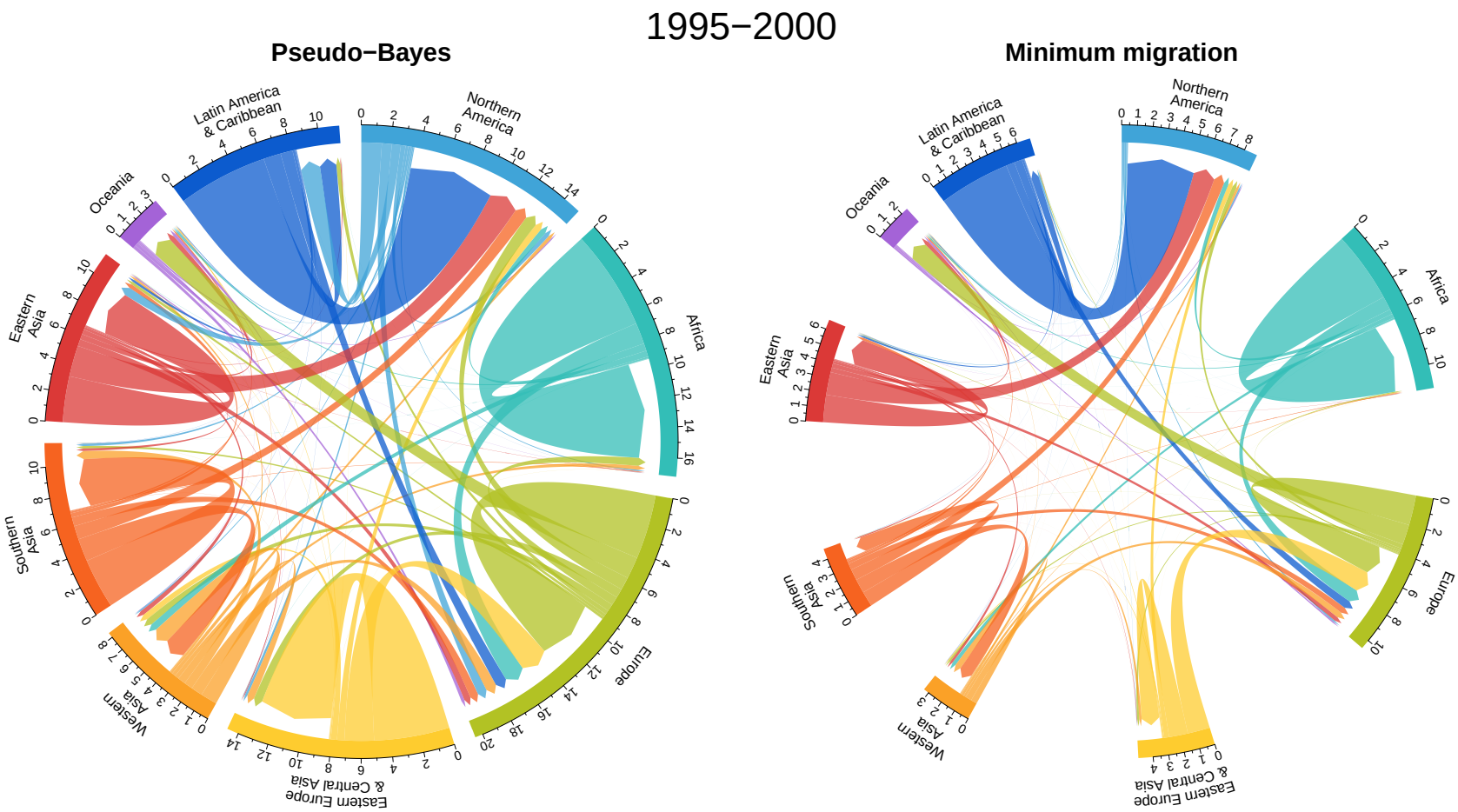

Figure 4.4: Estimated migration flows for 1995–2000. Left panel: Pseudo-Bayes estimates. Right panel: Estimates using the Abel (2013) method. Plots are scaled so that equal arc lengths represent equal numbers of migrants.

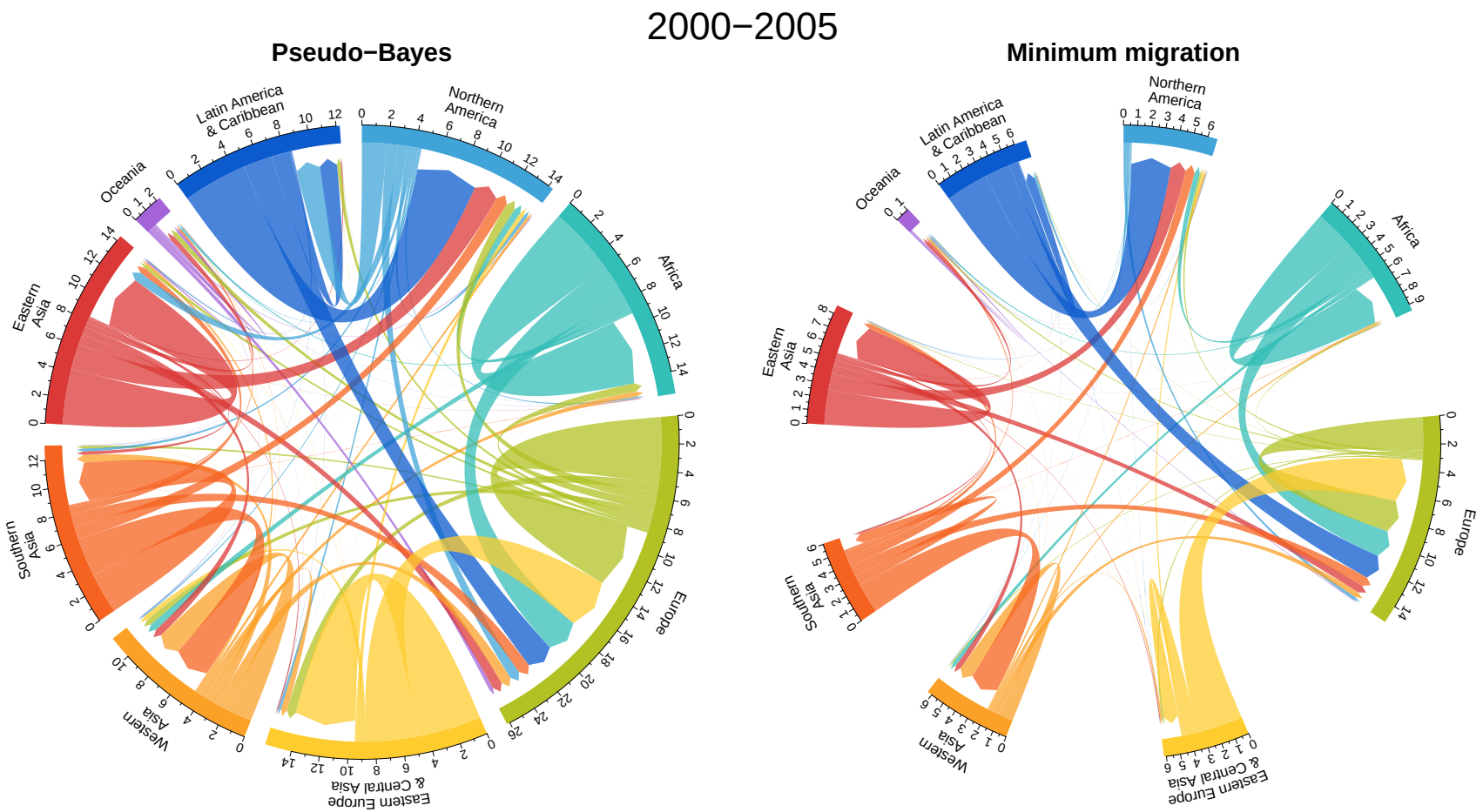

Figure 4.5: Estimated migration flows for 2000–2005. Left panel: Pseudo-Bayes estimates. Right panel: Estimates using the Abel (2013) method. Plots are scaled so that equal arc lengths represent equal numbers of migrants.

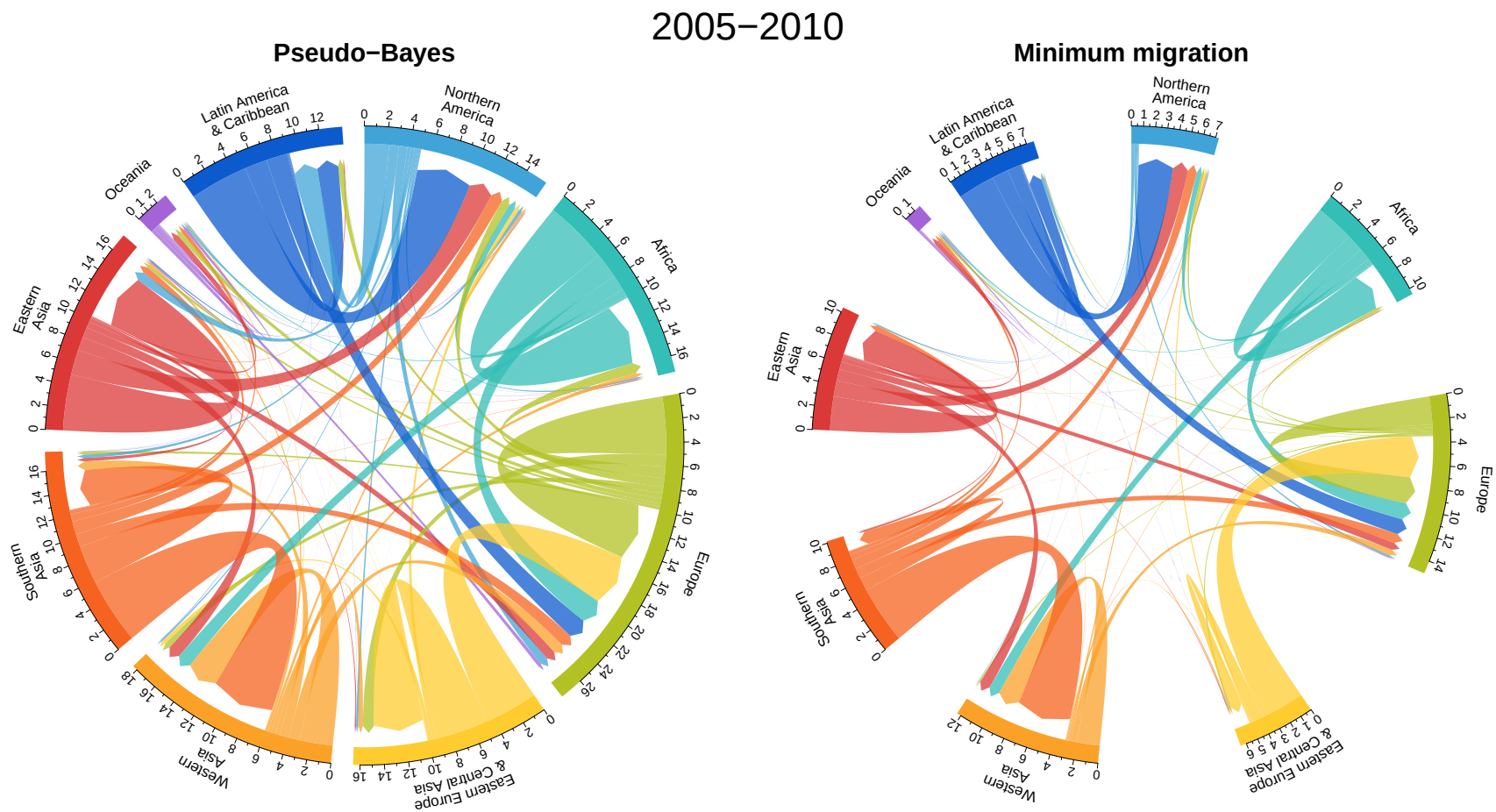

Figure 4.6: Estimated migration flows for 2005–2010. Left panel: Pseudo-Bayes estimates. Right panel: Estimates using the Abel (2013) method. Plots are scaled so that equal arc lengths represent equal numbers of migrants.

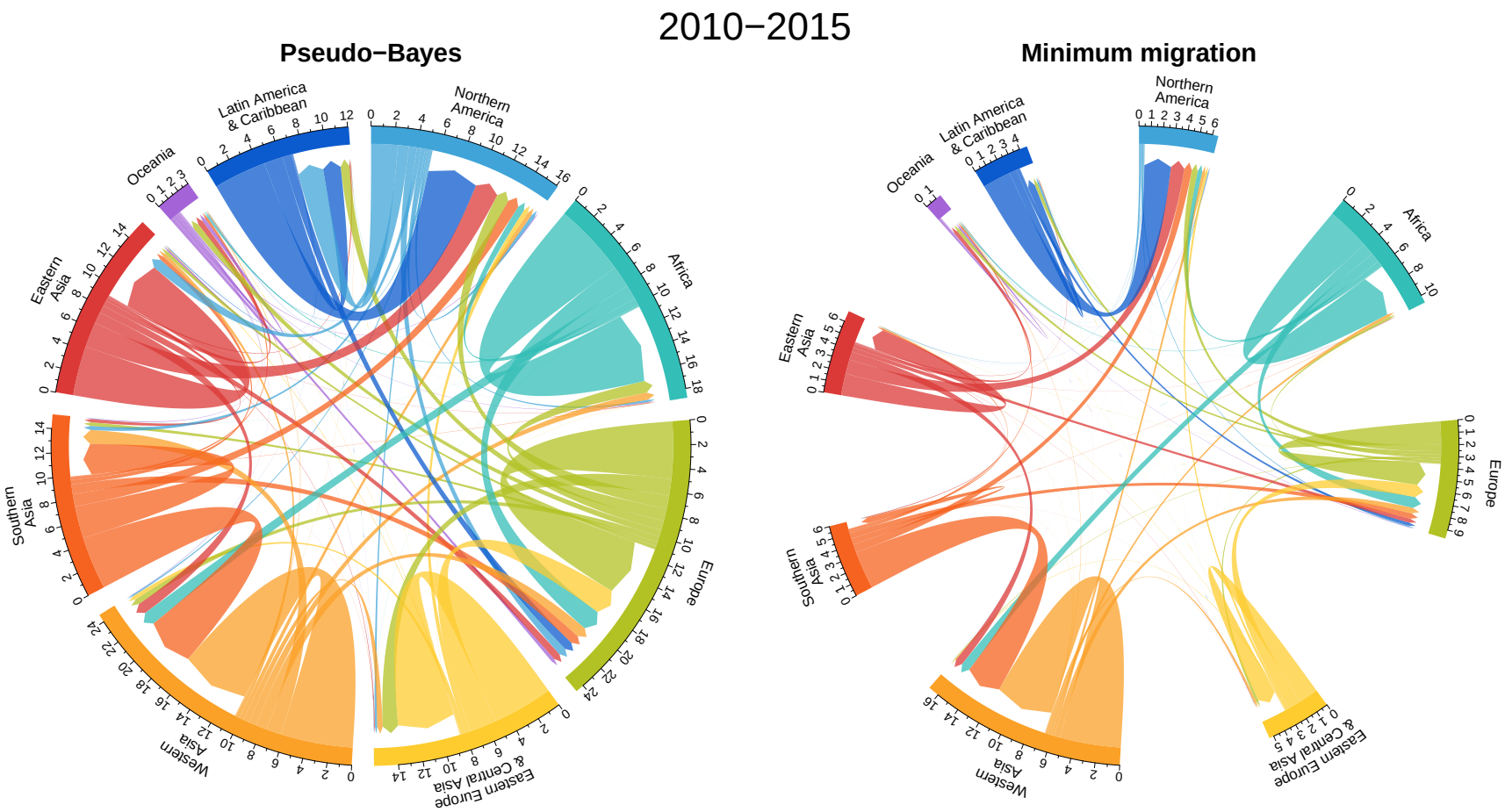

Figure 4.7: Estimated migration flows for 2010–2015. Left panel: Pseudo-Bayes estimates. Right panel: Estimates using the Abel (2013) method. Plots are scaled so that equal arc lengths represent equal numbers of migrants.

The most striking difference in the circular migration plots is that our estimates are substantially larger in magnitude than those produced with the minimum migration method. The top panel of Figure 4.8 plots the estimated total global migration flow over time for the two methods. The MM method estimates a total migration flow between 30 million and 41 million global migrants. In contrast, our method produces estimates of between 55 million and 74 million—at least 80% higher than the MM estimates. The bottom panel of Figure 4.8 converts the migration counts to a fraction of the world population. These estimates vary between 0.96% and 1.10% of the world's population migrating in each five-year time period. There is no evidence that the proportion of the world's population migrating has grown ($p = 0.32$ for a t-test of non-zero slope.)

Although the pseudo-Bayes estimates are higher than the MM estimates, the two sets of estimates are broadly similar in composition. For example, both sets of estimates contain large within-region flows in Africa and Western Asia. Also present in both estimates are large flows from Southern to Western Asia from 2005 onwards and from Latin America to Northern America, especially before 2000. The presence of a large flow in the MM estimates indicates a correspondingly large growth in migrant stocks. Consequently, the same flows will be large in our estimates as well, since the pseudo-Bayes estimates adhere to the same changes in migrant stocks.

However, the MM estimates of typical return migration flows are generally small. For example, in 2010–2015, the MM method estimates a flow of 2.4 million individuals from Latin America and the Caribbean to Northern America, but only 190,000 in the opposite direction. Our method increases both of these estimates. We estimate a flow of 4.6 million from Latin America and the Caribbean to Northern America and 2.3 million in the reverse direction. The net impact of Latin America/Northern America migration in the two sets of estimates is nearly identical; approximately 2.3 million individuals are gained by Northern America. The higher estimated return flows in our method are offset by higher inflows so that the net effect of migration is nearly unchanged. The same pattern is repeated for other typical return flows. For example, our method estimates substantial flows from Europe to

Africa and from Western Asia to Southern Asia.

Section 4.3.4 provides numerical details on the largest estimated flows according to our method as well as the largest changes from the MM estimates.

4.3.4 Largest flows and largest changes in estimates

Tables in this section provide answers to two questions: First, between which countries or regions do we estimate the largest flows? Second, in which countries or regions do we observe the largest changes between the MM estimates and our estimates? For this analysis, we focus only on the most recent time period, covering 2010–2015.

Tables 4.2 and 4.3 answer these two questions in the context of international flows rather than regionally aggregated flows. Table 4.2 lists the ten largest estimated flows under our method. Estimates of migration out of Syria are very similar under the two methods. Flows out of Syria in 2010–2015 would largely be comprised of refugees, with little return migration during this time period. The other entries on this list are flows between the USA and Mexico, flows within the former Soviet Union, and flows of labor migrants from India to the United Arab Emirates. We estimate these non-refugee flows to be substantially larger than the MM estimates—more than twice as large for all but the flow from India to the United Arab Emirates. Table 4.3 lists the ten international flows with the largest disparity between our estimates and Abel's (2013) estimates. The largest disparities are found in estimates of USA/Mexico flows, flows within the former Soviet Union, and flows within the South Asian members of the Commonwealth of Nations.

In tables 4.4, 4.5, and 4.6, we repeat the analysis of largest flows and largest differences in flows with regional aggregates of flows rather than flows between pairs of countries. In our estimates, as well as the MM estimates, many of the largest flows after regional aggregation are within-region rather than between-region flows (Table 4.4.) We estimate that six of the ten largest flows are within-region flows. Eight out of the ten largest flows appear again in the list of the largest differences between our estimates and the MM estimates (Table 4.5.) This reflects the general pattern that the pseudo-Bayes estimates look like a scaled-up

Table 4.2: Comparison of Abel's minimum migration (MM) estimates to pseudo-Bayes (PB) estimates (thousands of migrants). Largest estimated international flows according to the pseudo-Bayes estimator in 2010–2015.

Origin	Destination	MM	PB
Mexico	United States of America	946	2,065
Syrian Arab Republic	Turkey	1,512	1,505
United States of America	Mexico	145	1,265
Syrian Arab Republic	Lebanon	1,190	1,188
Ukraine	Russian Federation	432	984
India	United Arab Emirates	607	908
Syrian Arab Republic	Iraq	821	855
Russian Federation	Ukraine	214	767
Russian Federation	Kazakhstan	204	597
Kazakhstan	Russian Federation	196	587

version of the MM estimates in many cases.

Table 4.3: Comparison of Abel's minimum migration (MM) estimates to pseudo-Bayes (PB) estimates (thousands of migrants). Largest differences between Abel's (2013) estimates and pseudo-Bayes estimates of international flows in 2010–2015.

Origin	Destination	MM	PB	Difference	Ratio
United States of America	Mexico	145	1,265	1,120	8.7
Mexico	United States of America	946	2,065	1,119	2.2
Russian Federation	Ukraine	214	767	553	3.6
Ukraine	Russian Federation	432	984	552	2.3
Russian Federation	Kazakhstan	204	597	393	2.9
Kazakhstan	Russian Federation	196	587	391	3.0
Bangladesh	India	41	357	316	8.7
India	Bangladesh	3	317	314	113.3
Pakistan	India	2	314	312	175.2
India	Pakistan	58	369	312	6.4

Table 4.4: Comparison of Abel's minimum migration (MM) estimates to pseudo-Bayes (PB) estimates (millions of migrants). Largest estimated regional flows according to the pseudo-Bayes estimator in 2010–2015.

Origin	Destination	MM	PB
Africa	Africa	3.85	6.35
Western Asia	Western Asia	5.28	6.26
Eastern Europe & Central Asia	Eastern Europe & Central Asia	1.68	5.21
Europe	Europe	2.03	4.97
Latin America & Caribbean	Northern America	2.45	4.58
Southern Asia	Western Asia	3.30	4.57
Eastern Asia	Eastern Asia	1.88	4.31
Southern Asia	Southern Asia	0.52	2.79
Eastern Europe & Central Asia	Europe	1.15	2.35
Northern America	Latin America & Caribbean	0.19	2.31

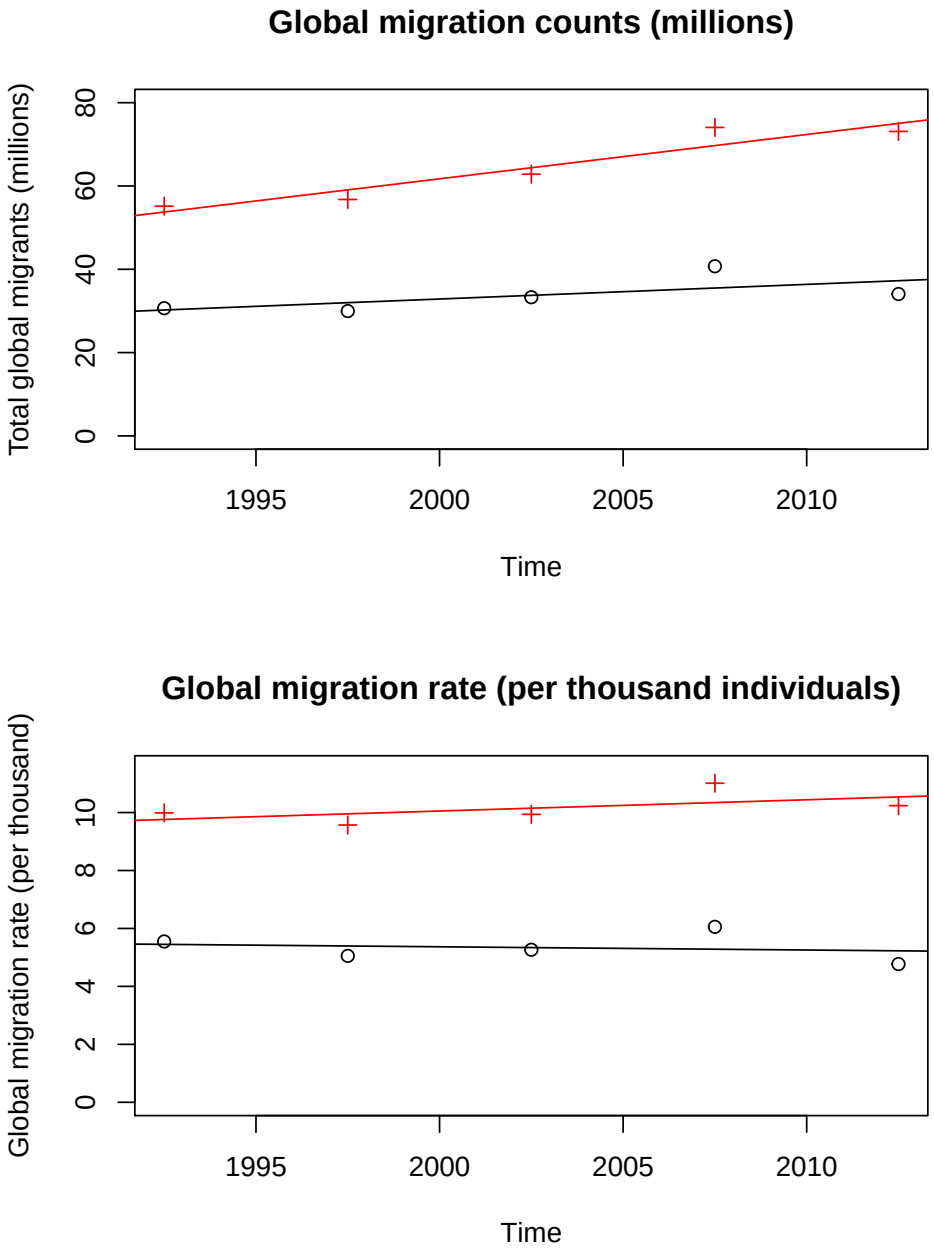

Figure 4.8: Top panel: Estimated global migration counts in millions of migrants. Bottom panel: Estimated global migration rate per thousand individuals. Plots compare pseudo-Bayes estimates ('+') to estimates produced with Abel's (2013) method (circles). Solid lines are OLS regression lines.

Table 4.5: Comparison of Abel's minimum migration (MM) estimates to pseudo-Bayes (PB) estimates (millions of migrants). Largest differences between Abel's (2013) estimates and pseudo-Bayes estimates of regional flows in 2010–2015.

Origin	Destination	MM	PB	Difference	Ratio
Eastern Europe & Central Asia	Eastern Europe & Central Asia	1.68	5.21	3.53	3.1
Europe	Europe	2.03	4.97	2.94	2.4
Africa	Africa	3.85	6.35	2.50	1.7
Eastern Asia	Eastern Asia	1.88	4.31	2.43	2.3
Southern Asia	Southern Asia	0.52	2.79	2.27	5.4
Latin America & Caribbean	Northern America	2.45	4.58	2.13	1.9
Northern America	Latin America & Caribbean	0.19	2.31	2.12	12.4
Western Asia	Southern Asia	0.02	1.30	1.27	57.1
Southern Asia	Western Asia	3.30	4.57	1.27	1.4
Europe	Eastern Europe & Central Asia	0.23	1.42	1.19	6.3

The magnitude of the increase in estimates is more dramatic in some regions than in others. The largest proportional increases are to typical return migration flows, which are estimated to be of very low magnitude under Abel's method due to the minimum-migrants assumption. For instance, the flow from Northern America to Latin America makes our list of the ten largest flows, with an estimated 2.31 million migrants under our method, compared to 190,000 under the MM method. Even more dramatic is the flow from Western Asia to Southern Asia, which is largely comprised of a return flow of labor migrants to the Gulf States. The MM estimate is 23,000 individuals in 2010–2015, while ours is 1.30 million. Table 4.6 gives the ten regional flows where the ratio of our estimates to the MM estimates is largest. All are instances where the MM method estimates small return flows despite the presence of a large migrant population.

Table 4.6: Comparison of Abel's minimum migration (MM) estimates to pseudo-Bayes (PB) estimates (thousands of migrants). Largest ratios of pseudo-Bayes estimates to Abel's estimates of regional flows in 2010–2015.

Origin	Destination	MM	PB	Difference	Ratio
Oceania	Southern Asia	0.05	66	65.49	1213.7
Africa	Southern Asia	0.10	12	12.10	123.2
Northern America	Southern Asia	4	370	366.89	102.9
Europe	Southern Asia	4	335	330.64	77.8
Western Asia	Southern Asia	23	1297	1274.13	57.1
Oceania	Western Asia	0.70	28	27.57	40.5
Northern America	Eastern Europe & Central Asia	7	216	208.90	32.2
Oceania	Eastern Europe & Central Asia	0.60	18	17.42	30.0
Oceania	Eastern Asia	6	171	165.37	29.8
Northern America	Eastern Asia	33	932	898.10	27.9

4.4 Validation

In this section we validate our model in two ways. We first conduct a case study on one of the largest and highest profile flows, between the USA and Mexico. We find that the pseudo-Bayes estimates of flows from the USA to Mexico align well with existing estimates,

while the MM estimates are probably an order of magnitude too low. We then conduct a simulation study to determine some approximate conditions under which the pseudo-Bayes estimator outperforms the minimum migration estimator.

4.4.1 Case study: Flows between USA and Mexico

The two flows for which our estimates differ the most from Abel's (2013) estimates in 2010–2015 are the two directions of flows between Mexico and the USA. Migration between the USA and Mexico is well studied, although difficult to estimate (Durand et al., 1999; Hanson, 2006; Massey & Espinosa, 1997). In this section we compare the pseudo-Bayes and minimum migration estimates, which are based only on stock data, to estimates from the Pew Research Center which incorporate information from demographic surveys in both the USA and Mexico (Gonzalez-Barrera, 2015; Passel et al., 2012). Figure **??** compares our estimates of the flows from the USA to Mexico (blue) with Abel's (2013) minimum migration estimates (red) and the Pew estimates (green). The three sets of estimates don't quite cover identical time periods; we compare the Pew estimate for 2009–2014 with pseudo-Bayes and minimum migration estimates for 2010–2015. Our estimates approximately agree with the Pew estimates, while the MM estimates are too low by a factor of somewhere between 6 and 10.

In the Mexico-to-USA direction, Pew provides estimates of flows of only Mexican-born individuals. Figure 4.10 compares the Pew estimates to the pseudo-Bayes and MM estimates. In this direction, the MM estimates agree with the Pew estimates while ours are too high by a factor that's between 1.2 and 2.4.

Note that flow estimates are constrained to match observed changes in stocks. In the MM estimates of USA/Mexico flows, a large flow from Mexico to the USA is offset by a very small flow in the opposite direction, resulting in a large increase in the stock of Mexican-born in the USA. Our estimates produce the same large change in stocks, but do so by offsetting a very large Mexico-to-USA flow with a moderate flow in the opposite direction.

Some of the disparity with the Pew flow estimates is due to a difference in estimates of

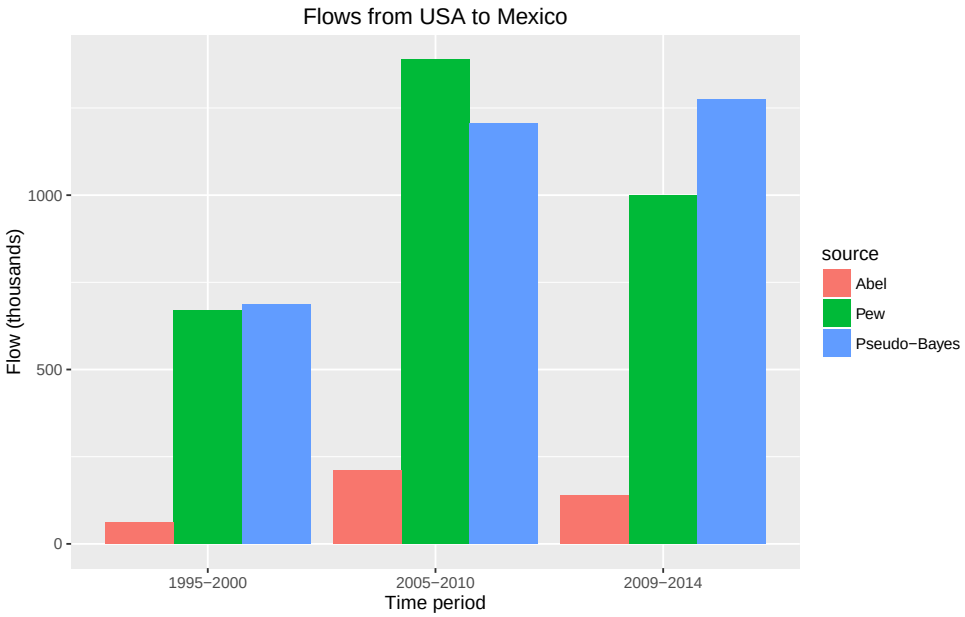

Figure 4.9: Estimates of flows from USA to Mexico, aggregated across all places of birth. Estimates from Pew for 2009–2014 are genuinely for 2009–2014. Pseudo-Bayes and Abel's minimum migration estimates for that time period use 2010–2015 instead.

migrant stocks. The minimum migration and pseudo-Bayes estimates replicate the changes in the UN stock data, which disagree substantially with the Pew estimates of migrant stocks (Passel et al., 2012). The UN stock data estimate that the population of Mexican-born in the US increased from 10.31 million in 2005 to 11.57 million in 2010—an increase of 1.26 million, which is about 90% larger than Pew's estimated increase of 670,000. Because of this disparity, estimates based on UN stock data are not able to align with the Pew estimates.

This comparison to external estimates provide evidence that the MM estimates for USA-to-Mexico flows are probably too low. Our Pseudo-Bayes estimates of the Mexico-to-USA flow don't agree with the Pew estimates, but seem to be at least within the realm of plausibility. This is especially true in light of the fact that the stock of Mexican-born in the US grew by nearly 600,000 more in the UN estimates than in the Pew estimates over 2005–2010.

It would be interesting to perform case studies on more of the large flows estimated by our method, but reliable flow estimates are unfortunately sparse. Studies on flows within the

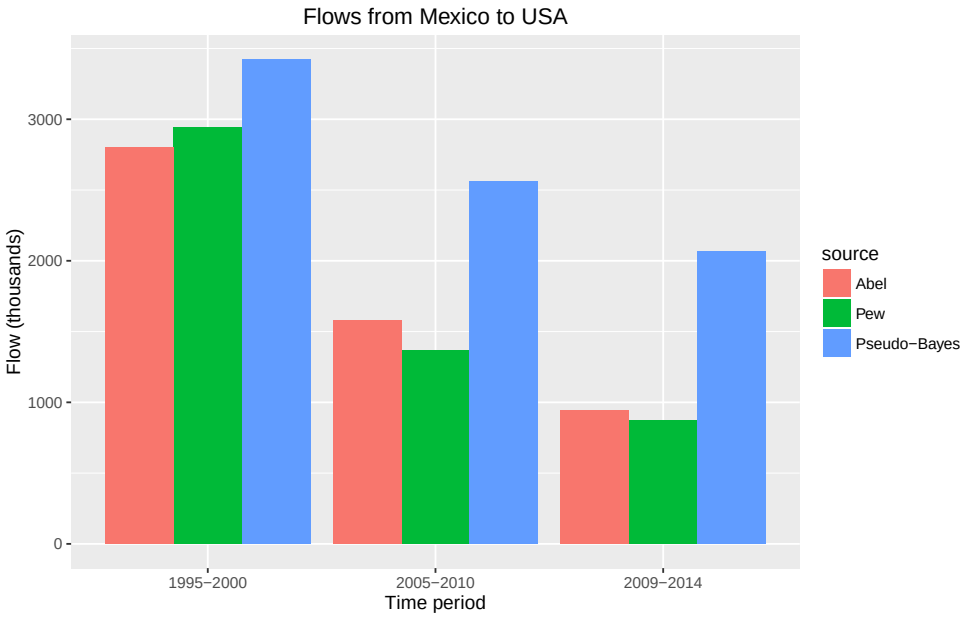

Figure 4.10: Estimates of flows from Mexico to the USA. Only includes Mexican-born individuals. Estimates from Pew for 2009–2014 are genuinely for 2009–2014. Pseudo-Bayes and Abel's minimum migration estimates for that time period use 2010–2015 instead.

countries of the former Soviet Union suggest that return migration is common (Buckley et al., 2008), but also that official statistics may be misstating flows by a factor of two or more (Karachurina, 2013). Within the South Asian countries, although data on migrant stocks exist, it is not clear whether any flow estimates are available.

4.4.2 Simulation study

Abel's (2013) flow estimates rely on the assumption that the diagonals of the flow table are maximized—that is, that the minimum possible number of individuals migrate in each time period. This assumption is likely to be a good approximation of the truth, but will not be strictly true in real data. In this section we examine simulated flow tables when the minimum migration assumption is weakened. We find that if the number of non-movers is smaller than 99.6% of its maximum value, then our flow estimates outperform the MM estimates.

Method for simulating flow tables

We aim to simulate migration flow tables which approximately replicated observed global stock data. To do so, we start with the UN estimates of migrant stocks for the 2010–2015 time period (United Nations, 2015a) broken down by place of birth and adjusted for births and deaths. These adjusted stock data will be the target row and column totals for our simulated flow tables. We then simulate flow tables under weakened assumptions about the proportion of stayers. (We will use the terminology "mover" and "stayer" to denote individuals who either make or do not make an international move during the time period of interest. The group of stayers will include everyone whose country of residence is the same at the start and end of the time period, regardless of whether they previously migrated to that country of residence.) We denote by r the ratio of the actual number of stayers in a flow table to the maximum possible number of stayers across all flow tables with the same marginal totals. A value of r near 1 indicates that the number of stayers is nearly maximized, and the number of movers nearly minimized.

We use the following procedure to simulate a flow table with marginal total similar to the marginal totals from the 2010–2015 data and with a ratio of true stayers to maximum stayers of approximately r:

1. Based on the adjusted stock data for each place of birth k in 2010–2015, compute the minimum and maximum possible diagonal entries of the table $M_{k,2010}$. Call these diagonal vectors $\boldsymbol{d}_k^{(min)}$ and $\boldsymbol{d}_k^{(max)}$.

2. For each place of birth, construct a diagonal

$$\boldsymbol{d}_k := \omega_k \boldsymbol{d}_k^{(min)} + (1 - \omega_k)\boldsymbol{d}_k^{(max)} \tag{4.18}$$

by choosing a weight ω_k such that all entries of $\boldsymbol{d}_k$ are non-negative and

$$\|\boldsymbol{d}_k\|_1 = r \cdot \|\boldsymbol{d}_k^{(max)}\|_1. \tag{4.19}$$

This diagonal is constructed to satisfy the requirement about the number of stayers relative to the maximum possible number of stayers. Furthermore, this choice of diagonal tends to concentrate movers more heavily in rows and columns where stocks are high, which is also a feature we expect from the true tables.

Note that if the entries of $\boldsymbol{d}_k^{(min)}$ are too large or p is too small, there may be no possible value of $\boldsymbol{d}_k$ that satisfies these conditions. In that case, take $\boldsymbol{d}_k = \boldsymbol{d}_k^{(min)}$, which is the choice of diagonal which comes as close as possible to satisfying the desired proportion of stayers.

3. For each place of birth, fit a quasi-independent log-linear model to the table with diagonal $\boldsymbol{d}_k$ and margins given by the adjusted stock data. Denote the entries of this table by $\Lambda_{k,2010}$.

4. For each place of birth, generate a simulated table $\tilde{M}_{k,2010}$ whose entries $\tilde{M}_{ijk,2010}$ are sampled independently from Poisson distributions with intensity parameters given by $\Lambda_{ijk,2010}$.

Because of the Poisson variability introduced in Step 4, neither the marginal sums nor the proportion of stayers will precisely match their pre-specified values. In practice, however, both are very close to the intended values. By construction, each marginal sum of $\tilde{M}_{k,2010}$ follows a Poisson distribution with mean given by the corresponding marginal sum of the true data, $M_{k,2010}$. The Poisson distribution typically produces values close to its mean; when the Poisson mean is large, the coefficient of variation is small. For example, if a true row sum is 1,000,000 individuals, our simulated tables will have a row sum that is within 1 million $\pm$ 2,000 individuals roughly 95% of the time. Consequently, marginal sums of our simulated tables match quite well with the true marginal sums.

With respect to r, we verified after simulation that our simulated tables produce ratios of true stayers to maximum possible stayers that are close to the desired value. Our simulations

look at values of r ranging from 0.99 to 0.9999. In all cases, we found that the value of r in the simulated table was within 0.00002 of its intended value.

Simulation Results

We simulated 100 complete flow tables broken down by place of birth using the above procedure. All simulations approximately replicate the adjusted stock data by place of birth for the 2010–2015 time period. Simulated tables have values of r ranging from 0.9999 (almost no violation of the minimum migration assumption) to 0.99 (mild violation of the minimum migration assumption). These values of r cover our estimated values of r on the real data, which lie between 99.41% and 99.51% in each five-year time period. For each set of tables, we then used the simulated row and column margins to produce two sets of flow estimates: one based on Abel's (2013) minimum migration method with the maximum-stayers assumption, and a second based on the pseudo-Bayes estimator using the same weight we use to produce our flow estimates on the true quinquennial stock data, $w = 0.8956$.

Figure 4.11 summarizes performance of the MM method and the pseudo-Bayes method according to the loss functions discussed in Section 4.2.5. Evaluation occurs on the estimates of total bilateral flows, rather than flows broken down by place of birth. Points that lie below the $y = x$ line represent simulations where the pseudo-Bayes estimator outperforms the MM estimator. When evaluating on the basis of root mean squared error in flow counts (bottom left) or in flow rates (top row, left and right), the MM method has lower error only when $r \geq 0.9963$. That is, unless the true number of stayers is at least 99.63% of its maximum value, the pseudo-Bayes method is the better choice. Over the five quinquennial periods studied, the pseudo-Bayes estimator places the true number of stayers between 99.41% and 99.51% of its maximum value; in all of this range, the pseudo-Bayes estimator outperforms the MM estimator. Expressed in terms of numbers of movers, the pseudo-Bayes method outperforms the MM method when the number of migrants over the 2010–2015 time period is at least 58.4 million, as compared to a minimum possible value of 34.1 million.

For some application, we may instead be interested in estimating flows on a log scale.

The lower right panel of Figure 4.11 shows root mean squared error in log(migration flow + 1). On the log scale, the pseudo-Bayes estimator outperforms the MM estimator whenever $r < 0.9995$. Equivalently, the Pseudo-Bayes estimator is better so long as the number of movers over this time period is at least 37.4 million.

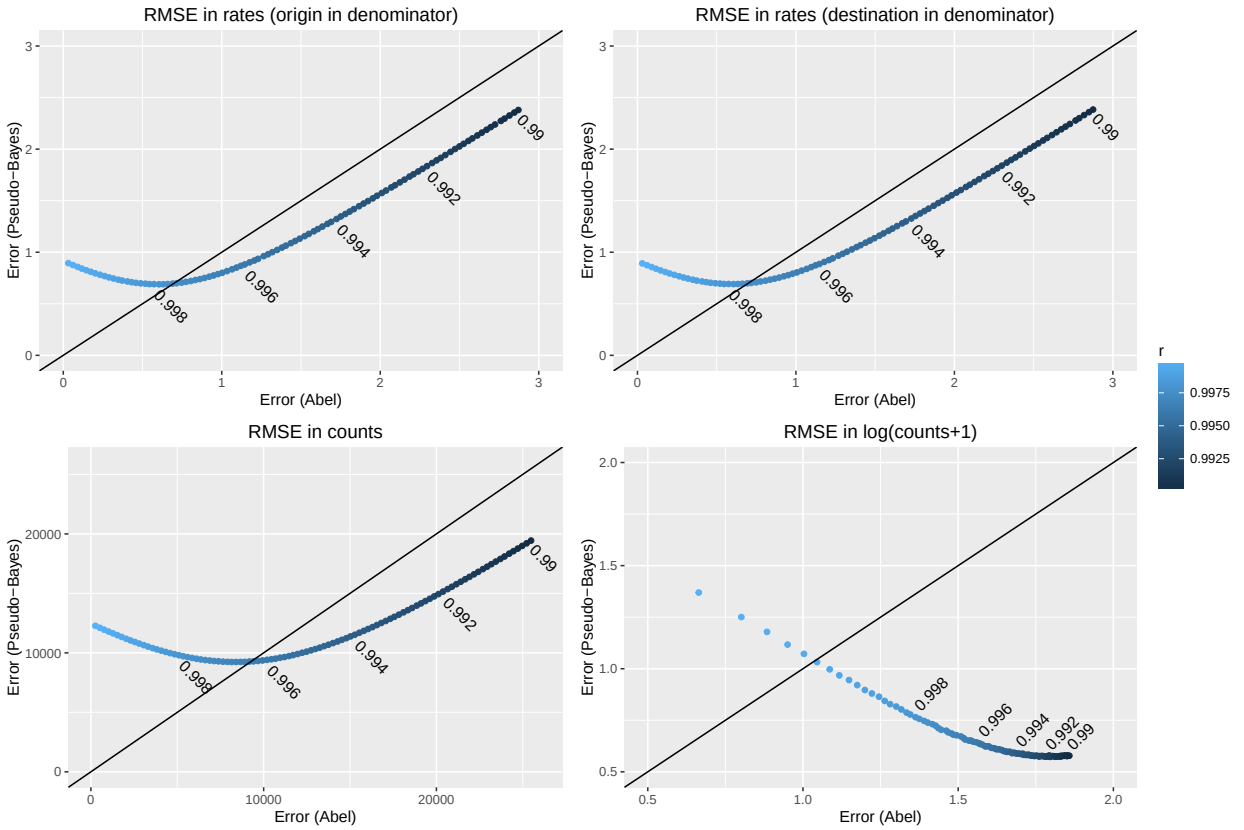

Figure 4.11: Performance of Abel's (2013) method and the Pseudo-Bayes estimator with $w = 0.8956$ when estimating simulated migration flows. Each dot represents a single simulated global migration flow table. Dot color indicates the ratio of the number of stayers in the true data to the maximum possible number of stayers, with lighter dots representing more stayers. Evaluation metrics are root mean squared error in migration flows (top left), in log(migration flow+1) (top right), in migration rates (bottom row). To calculate rates, we divide flows by either the population of the origin country (bottom left) or the destination country (bottom right)

There is some Monte Carlo variability in these error curves due to the Poisson-distributed flows, but that variability is very slight relative to the variability introduced by changing

the number of stayers. Errors would be more variable if we either weakened the quasi-independence assumption or selected the diagonal values of $\Lambda_{k,2010}$ via some non-deterministic process. As such, these results cannot speak to the two methods' robustness to violation of quasi-independence or of the assumption that movers tend to be concentrated in rows or columns with high migrant stocks. However, both assumptions should be reasonably good approximations of reality.

4.5 Discussion

Our method produces estimates of global migration flows from migrant stocks. We incorporate empirical evidence in the form of OECD and MIMOSA estimates of migration flows which allows us to weaken the minimum-migration assumption present in Abel's (2013) method. As a result, we estimate anywhere from 80% to 115% more total migrants than the MM method. Especially large increases are present in estimates of return migration flows such as those from Northern America to Latin America or from Western Asia to Southern Asia. These larger estimated return flows mitigate a weakness introduced by the necessary structural zeroes in estimates which satisfy the minimum migration assumption. A simulation study suggests that estimates produced under the minimum migration assumption are not very robust to violation of that assumption. If the true number of non-movers is lower than 99.6% of its maximum value, then the pseudo-Bayes estimates will be preferable to the MM estimates.

Our estimates of migration flows are broken down by place of birth. In principle, the same procedure could be applied to flows which are broken down by other characteristics in addition to or instead of birthplace. At a global level, more detailed breakdowns of stock data by other characteristics are not generally available for all countries. One exception is breakdowns by sex, which are addressed in (Abel, 2016). However, the same technique developed here may be useful in producing estimates of either disaggregated international migration flows for a subset of the world's countries or disaggregated subnational migration flows.

The issue of changing national borders provides an interesting edge case for our method. The UN stock data upon which we base our estimates classify individuals' "country of birth" based on the modern country in which their birth location resides. For instance, an individual born in the former Estonian Soviet Socialist Republic would be counted in the stock data as having been born in Estonia rather the now-defunct USSR. Futhermore, if that individual moved from Estonia to Russia before the breakup of the Soviet Union, she would appear in the stock data as an international migrant, even though the move was an internal migration at the time.[3] These stocks of "international migrants" who may not have actually migrated internationally are potentially problematic. This misclassification of individuals may cause spuriously high flow estimates in regions where countries have split, especially in the former USSR and the former British India. Some evidence from the Baltic States suggests that our model is overestimating out-migration there. There were at one time large flows of ethnic Russians from the Baltics to Russia, but these flows have decreased greatly since the 1990s (OECD, 2013).

One weakness of this method is that our weighting factor, w, is empirically determined based on existing flow estimates, and those flow estimates have limited coverage. In particular, we have good coverage of flows to and from developed countries, but not of so-called South-South flows between developing countries. While we did find that optimal values of w differed very little between the flow types we had available, South-South flows may be systematically different from flows to or from more developed countries. Unfortunately, it is not clear whether we should expect that the minimum migration assumption will be more suitable or less suitable for South-South flows. Absent additional evidence on South-South flows, and given that we did not see systematic differences between the optimal value of w for flows *to* OECD countries, flows *from* OECD countries, and flows within Europe, we operate on the assumption that South-South flows do not need special treatment.

---

[3]The subject of mismatch between place of birth, nationality, and ethnic identity is explored in depth in T.S. Eliot's *The Waste Land* (Eliot, 1922). e.g. "*Bin gar keine Russin, stamm' aus Litauen, echt deutsch.*" (I'm no Russian at all; I'm from Lithuania, a real German.)

As with our work on projecting net migration in Chapter 2, our outputs should be interpreted as being conditional on the correctness of the inputs to our model. In this case, our flow estimates operate on the assumption that all migrant stocks are correctly estimated. The documentation accompanying the United Nations estimates of migrant stocks (United Nations, 2015a) indicates that in reality these stocks are not known precisely. As was the case with estimates of net migration, many stock estimates incorporate data on refugee flows, and some stocks are imputed rather than measured. Although this document doesn't report any measure of uncertainty for individual flow estimates, it is important to note that such uncertainty does exist and that both model uncertainty and the unknown variability in the stock estimates may be important contributors to overall uncertainty in flow estimates.

Although the pseudo-Bayes estimator performs well as a point estimator for migration flows, it lacks quantified uncertainty in those estimates. A natural way to obtain confidence intervals for flows would be to take a fully Bayesian approach, estimating posterior distributions on the entries in flow tables conditional on the known table margins and a prior distribution on table entries. In principle, Bayesian analysis on contingency tables with known margins is possible (Dobra et al., 2006). This approach would entail sampling from the posterior distribution via Markov Chain Monte Carlo updating of table entries. However, in each time period we are trying to estimate a total of 200 tables, each of dimension 200×200. Given the relatively high dimension of the problem, it is not clear whether the sampling procedure would be prohibitively costly in terms of computational power.

Chapter 5

CONCLUSION

This chapter contains a summary of the main contributions of this work as well as some possible directions for future research. Additional details can be found in the main chapters' Discussion sections (Sections 2.4, 3.4, and 4.5.)

5.1 Contributions to Research

I have developed a method for probabilistic projection of net international migration rates for all countries. This method can be used to project migration as far into the future as is desired. Validation of migration projections on available data shows that the resulting projections remain well calibrated over at least a 30-year period. The Bayesian hierarchical modeling approach gives a natural quantification of uncertainty in projections. It also allows for principled incorporation of expert beliefs in the form of prior distributions. For example, priors may be used to put empirical thresholds on cumulative migration rates over several time periods so that countries never experience prolonged periods of in- or out-migration on a scale that is historically unprecedented.

A key outcome of my work is the collection of projections of net international migration for all countries, available as online supplements to Azose & Raftery (2015). The migration projections produced by my model are of a form that can be used to incorporate probabilistic migration into the United Nations' probabilistic population framework. Other work by our research group has shown that probabilistic migration greatly improves the calibration of probabilistic population projections (Azose et al., 2016).

Another main contribution of my work is new statistical methodology for estimation of a correlation structure in the presence of prior beliefs about strength of correlations. One

weakness of my initial hierarchical model for net migration is that disturbances in migration rates are assumed to be independent. This assumption is convenient, but not fully realistic. Countries which are close to one another may have migration rates that tend to move in tandem (as they both attract or repel migrants at the same time), or rates that move in opposite directions (as migrants move from one country to the other). While standard estimates of correlations show evidence of these phenomena, Pearson estimates are unsuitable in this application because the estimated correlation matrix is very large relative to the amount of available data. I developed a technique for estimating a maximum *a posteriori* correlation matrix in the presence of informative prior information on correlations. My method greatly reduces the presence of spurious correlations while retaining the potential for strong correlations between nearby countries. The resulting estimated correlation matrix can be used to augment the initial Bayesian hierarchical model for migration rates. This allows us to retain the calibration of the base model on individual countries, while improving calibration of projections for some regional aggregates.

The third major component of my research is a new method for estimating migration flows from data on migrant stocks. I have produced estimates of bilateral migration flows between all pairs of countries. Flow estimates produced with Abel's (2013a) method may be difficult to model, as they contain a large number of zeroes, especially in common corridors of return migration. My flow estimates are broadly similar to Abel's, but lack the problematic structural zeroes inherent to his method. Estimated flows under my method have lower error than Abel's in places where reliable flow data are available, especially with respect to return migration. They may prove to be more conducive to efforts at projection.

5.2 Future Work

5.2.1 Topics of statistical interest

In the course of my research, I encountered several unexpected statistical results. This section briefly discusses these topics, which range from unusual (but probably known) properties of

common probability distributions to open statistical questions.

One surprising outcome of my work on net migration projections was that a model with normal forecast errors outperformed a model with t errors despite the presence of large outliers in the migration data. According to the usual metrics of maximized data likelihood or the Bayesian Information Criterion, I found that an AR(1) model with t errors is a better description of the data than the same model with normal errors. However, for projection purposes, we were interested in a model with good prediction intervals (i.e. intervals which are as tight as possible subject to calibration) rather than a model under which future points will have high likelihood. It turned out in this case that normal errors improved interval scores at the expense of lower maximized likelihood. In some sense this discrepancy is not surprising. From a decision theoretic perspective, we could have instead framed our model selection problem as one of choosing a model and its parameters in such a way to maximize expected interval scores for future observations rather than maximizing log-likelihood. In general, the decision theoretic approach could be used to optimize for any score function. However, this approach seems to be incompatible with the Bayesian hierarchical modeling approach, which implicitly favors models with higher data likelihood. As far as I know, it is an open question whether the two approaches can be integrated.

Another unexpected outcome was the discovery of limitations in commonly used probability distributions. This happened twice in the course of this research. The first was in the context of correlation modeling, where it was tempting to try a Bayesian approach with an inverse Wishart prior on covariance. As I mentioned in Section 3.4, the inverse Wishart is known to be quite restrictive in form (Barnard et al., 2000). The precise weakness that was relevant to our case is that the Wishart distribution has a parameter for degrees of freedom which can be no lower in our model than (# of countries-1). This is problematic because a high value for the degrees of freedom parameter forces tight marginal prior distributions on the variances σ_c^2, which is incompatible with our original Bayesian hierarchical model in which we placed minimally informative priors on the marginal variances. This is a known weakness in the inverse Wishart distribution, and some suggestions have been made for al-

ternative, minimally informative matrix priors (A. Huang & Wand, 2013), but it is not clear whether any of the available alternatives would be compatible with our Bayesian hierarchical model for net migration.

The second distributional issue I encountered was with the use of the negative binomial distribution as a sort of variance-inflated Poisson. When modeling count data for which the Poisson distribution understates variability, it is fairly common practice to treat the data as coming from a negative binomial distribution. The negative binomial distribution is effectively the simplest discrete probability distribution on the natural numbers which decouples mean and variance. Because of this property, I looked into using the negative binomial distribution to model overdispersion in the pseudo-Bayes estimates of migration flows. Despite the overdispersion present in the data, this distribution turned out not to be suitable. The key issue was that the negative binomial mode moves towards zero as the variance of the distribution increases. Because of this, a negative binomial model that accurately described the amount of variance in the migration flow data ended up vastly overstating the frequency with which our model underestimates the true flows. This issue could potentially be rectified with a more flexible distribution on count data. However, in practice, an approach that models logged counts as if they were continuous data seems to work well.

Finally, my work on migration flow estimation revealed an issue with estimation of sums of random variables. Errors in my estimates of individual migration flows, m_{ij+t}, appear to be approximately log-normal. The core issue is that when errors are asymmetric, the sum of individual estimates of random variables can be a very poor estimate of the sum of the random variables. For example, suppose that $X_1, \ldots, X_{100}$ are independent and identically distributed log-normal random variables. Each $\log(X_i)$ has a $N(\log(1{,}000), 1)$ distribution. For each X_i individually, a good estimate of X_i is 1,000. (It's the median of the distribution, and X_i is as likely to be too low by a factor of 2 as it is to be too high by a factor of 2.) However, 100,000 turns out to be a terrible estimate for $\sum_{i=1}^{100} X_i$. An 80% confidence interval for the sum is $(139{,}000, 193{,}000)$. The intuition for why the sum of the estimates is a poor

estimate of the sum is that a value of X_i which is higher than its median by a factor of 2 is much more influential on the sum than a value which is lower than its median by a factor of 2.

In the context of my migration flow estimates, the implication for misestimation of aggregates is somewhat problematic. On the one hand, we would like estimates of regionally aggregated flows to be consistent with estimates of individual flows. On the other hand, the observed asymmetry in errors suggests a potential underestimation of aggregated flows. There is not a clear path to reconciling the conflicting goals of retaining additivity in estimates and unbiased estimation of aggregates. A partial solution may be to define a loss function by assigning weights to the two conflicting goals and use loss function optimization to look for an intermediate estimate which is both nearly additive and nearly unbiased in aggregate.

5.2.2 Migration modeling

My work on both projection and estimation has focused only on international migration, but the same techniques could be used to model migration at a sub-national level. Sub-national migration projections may be of interest to national statistics offices in countries where reliable inter-region migration data are available. For instance, the US Census Bureau no longer publishes population projections at the state level, but they have in the past produced such projections, most recently in 2005 (US Census Bureau, 2005). Their methodology for projection of internal migration fit an ARIMA model to migration flows for the first five years of projections, followed by a deterministic return to historical mean migration flows in the medium- to long-term. Applying a Bayesian hierarchical modeling approach to the US between-region migration data could produce quite similar mean projections, but with the added benefit of quantified uncertainty. It remains an open question whether my hierarchical model for migration flows is well suited to sub-national data, as internal migration may have unforeseen characteristics which are not present in international migration.

Sub-national data may also prove useful as a means of validating and extending my

method for estimation of flow tables. While international migration flows are difficult to estimate, there are relatively good quality, detailed estimates of both state-to-state and county-to-country migration within the United States available from the Internal Revenue Service (United States Internal Revenue Service, 2016). Furthermore, since these flow estimates are based on tax records, they contain at least partial information about the ages of migrants. These data could potentially allow for a breakdown of migration flow tables by not only place of birth, but also by age. In addition to being of methodological interest, this may provide insight into the drivers of sub-national migration as well as serving as a stepping stone towards projections of sub-national flows.

One obvious extension of my research would be to project bilateral international migration flows using my improved flow estimates. There are advantages and disadvantages associated with producing projections of net migration, as I have done, rather than projections of bilateral flows or of summed in- and out-migration totals. If the ultimate goal is to study population dynamics, then the simplicity of net migration is an attractive feature. On the other hand, net migration condenses all information about a country's migration down to a single number, and necessarily obscures information about the origins and destinations of migrants—information that can help answer questions about where and why people migrate. My work on estimation of bilateral migration flows partially addresses this issue, but projections of flows would also be of great interest. While I anticipate that a Bayesian hierarchical modeling approach may prove fruitful when applied to flow data, the model I used for net migration rates would not be suitable without substantial modification. A more likely starting point would be to fit a Poisson model to migration flows, allowing the Poisson means to vary over time.

Modeling the distribution of migrants by age and sex has been a persistent difficulty. Substantial work has been done to identify typical age patterns of migration flows (Rogers & Castro, 1981). However, these patterns are not theoretically well suited to *net* migration (Rogers, 1990). Examination of age distributions of migrants in the WPP data bears this out, with age patterns of migration being particularly volatile when net migration is near

zero. Our current practice is to apply deterministic age and sex schedules to projected net migration counts, using schedules that we derive from the WPP data. While this leads to age and sex schedules of projected migration which are largely plausible, it also causes understatement of uncertainty in the impact that migration will have on a population's age distribution. This could be improved by applying probabilistic rather than deterministic migration schedules, perhaps derived from plausible combinations of in- and out-migration schedules.

Another unresolved difficulty in the modeling of net migration is handling shocks. Historically, many large-scale refugee movements have followed a similar pattern of migration flows. In many cases, including the Rwandan genocide of the early 1990s, a large outflow of refugees to neighboring countries is followed within the next five to fifteen years by repatriation of most refugees. This pattern is not present in the net migration trajectories projected by our Bayesian hierarchical model. Modeling this trend would produce more realistic trajectories for net migration. One option for incorporating this into our model would be to break down the historical net migration data into refugee and non-refugee migration and model total net migration as a sum of the two components. I believe shocks are less problematic in the context of migration flow estimation. In fact, the scenario in which there is a large outflow to a country with a very small migrant stock is ideal for flows-from-stocks methodology, since the absence of a large pre-existing stock heavily constrains the set of possible flows.

Under my net migration model, projected net migration rates have a tendency to decay on average towards a historical mean migration rate. This feature of the model is well suited to projections with a long time horizon, such as those produced by the UNPD. However, some evidence suggests that in the short term it is appropriate to assume that current growth or decrease in migration will continue (Bijak & Wiśniowski, 2010; Kim & Cohen, 2010). A model which allows migration to follow current trends in the short term while returning to historic levels in the long term could improve projections. One possibility for how to operationalize this would be an approach akin to Bayesian melding, where projected migration rates are allowed to come from two or more different models, with probability

corresponding to a time-varying posterior model probability.

Prior work by other members of our research group has led to the publication of code packages for Bayesian probabilistic projection of fertility (`bayesTFR`), mortality (`bayesLife`) and population (`bayesPop`) in the CRAN repository (Ševčíková et al., 2011; Ševčíková & Raftery, 2015; Ševcıková et al., 2016; Ševcıková, 2014). A `bayesMig` package for probabilistic projection of net migration would be a natural addition to this family of packages.

Finally, by eschewing covariates in our model for estimating migration flows, we have omitted a lot of potentially useful information. Although my work agrees with Abel's findings that including a distance offset leaves his estimates essentially unchanged (Abel, 2016), there are many other untested covariates that may be of use to us. For example, economic development and population age structure are both theoretical drivers of migration (Billari & Dalla-Zuanna, 2011; Harris & Todaro, 1970). Our model for migration flows would allows us to first test whether these covariates are useful as predictors of migration conditional on known stocks and then use any predictive variables to improve estimates. Regularities in the age structure of migration could also serve as a means of partially validating my flow estimates, especially for the South-South flows where good data are sparse.

REFERENCES

Abel, G. J. (2010). Estimation of international migration flow tables in Europe. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*, 797–825.

Abel, G. J. (2013a). Estimating global migration flow tables using place of birth data. *Demographic Research*, *28*(article 18), 505–546. doi: 10.4054/DemRes.2013.28.18

Abel, G. J. (2013b). migest: Useful R code for the estimation of migration. *The CRAN Project. Retrieved December*, *12*, 2014.

Abel, G. J. (2016). *Estimates of global bilateral migration flows by gender between 1960 and 2015* (Tech. Rep.). Vienna Institute of Demography.

Abel, G. J., Bijak, J., Forster, J. J., Raymer, J., Smith, P. W., & Wong, J. S. (2013). Integrating uncertainty in time series population forecasts: An illustration using a simple projection model. *Demographic Research*, *29*(article 43), 1187–1226. doi: 10.4054/DemRes.2013.29.43

Abel, G. J., & Sander, N. (2014). Quantifying global international migration flows. *Science*, *343*, 1520–1522.

Al Jazeera America. (2016, January). *Germany registers record 1.1 million asylum seekers in 2015.* Retrieved from http://america.aljazeera.com/articles/2016/1/6/refugees-germany-more-than-1million.html ([Online; posted 2016/01/06])

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

Antoniadis, A., & Fan, J. (2001). Regularization of wavelet approximations. *Journal of the American Statistical Association*, *96*(455), 939–955.

Azose, J. J., & Raftery, A. E. (2015). Bayesian probabilistic projection of international migration. *Demography*, *52*(5), 1627–1650.

Azose, J. J., Ševčíková, H., & Raftery, A. E. (2016). Probabilistic population projections with migration uncertainty. *Proceedings of the National Academy of Sciences*, *113*(23), 6460–6465.

Barnard, J., McCulloch, R., & Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, *10*(4), 1281–1312.

Beck, A., & Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal of Imaging Sciences*, *2*(1), 183–202.

Bickel, P. J., & Levina, E. (2008a). Covariance regularization by thresholding. *The Annals of Statistics*, *36*(6), 2577–2604.

Bickel, P. J., & Levina, E. (2008b). Regularized estimation of large covariance matrices. *The Annals of Statistics*, *36*(1), 199–227.

Bien, J., & Tibshirani, R. J. (2011). Sparse estimation of a covariance matrix. *Biometrika*, *98*(4), 807.

Bijak, J. (2006). *Forecasting international migration: Selected theories, models, and methods.* Dordrecht, Netherlands: Springer.

Bijak, J. (2010). *Forecasting international migration in Europe: A Bayesian view.* Springer Science & Business Media.

Bijak, J., Kupiszewska, D., Kupiszewski, M., Saczuk, K., & Kicinger, A. (2007). Population and labour force projections for 27 European countries, 2002–2052: Impact of international migration on population ageing. *European Journal of Population*, *23*(1), 1–31.

Bijak, J., & Wiśniowski, A. (2010). Bayesian forecasting of immigration to selected European countries by using expert knowledge. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*, 775–796.

Billari, F. C., & Dalla-Zuanna, G. (2011). Is replacement migration actually taking place in low fertility countries? *Genus*, *67*, 105–123.

Bilsborrow, R. E. (1997). *International migration statistics: Guidelines for improving data collection systems*. International Labour Organization.

Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: theory and practice*. The MIT Press.

Brown, S. K., & Bean, F. D. (2012). Population growth. In J. Gans, E. M. Replogle, & D. J. Tichenor (Eds.), *Debates on U.S. Immigration.* Thousand Oaks, California: SAGE.

Brücker, H., & Siliverstovs, B. (2006). On the estimation and forecasting of international migration: How relevant is heterogeneity across countries? *Empirical Economics*, *31*, 735–754.

Buckley, C. J., Ruble, B. A., & Hofmann, E. T. (2008). *Migration, homeland, and belonging in Eurasia*. Woodrow Wilson Center Press.

Castles, S., & Miller, M. J. (2003). *The age of migration: International population movements in the modern world*. London: Macmillan.

Chaudhuri, S., Drton, M., & Richardson, T. S. (2007). Estimation of a covariance matrix with zeros. *Biometrika*, *94*(1), 199–216.

Chen, X., Xu, M., Wu, W. B., et al. (2013). Covariance and precision matrix estimation for high-dimensional time series. *The Annals of Statistics*, *41*(6), 2994–3021.

Chi, E. C., & Lange, K. (2014). Stable estimation of a covariance matrix guided by nuclear norm penalties. *Computational Statistics & Data Analysis*, *80*, 117–128.

Cohen, J. E. (2012). Projection of net migration using a gravity model. In *Proceedings of the XXVII IUSSP International Population Conference.* Paris, France: International Union for the Scientific Study of Population. (Retreived from http://www.iussp.org/sites/default/files/event_call_for_papers/IUSSPsession020CohenProjectionNetMigrationGravityModelUNPopDiv2012corrected.pdf)

Cohen, J. E., Roig, M., Reuman, D. C., & GoGwilt, C. (2008). International migration beyond gravity: A statistical model for use in population projections. *Proceedings of the National Academy of Sciences*, *105*, 15269–15274.

Committee for Development Policy and United Nations Department of Economic and Social Affairs. (2008). *Handbook on the least developed country category: Inclusion, graduation, and special support measures.* New York, NY: United Nations.

Congdon, P. (2010). Random-effects models for migration attractivity and retentivity: A Bayesian methodology. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*, 755–774.

Crush, J. (1999). Fortress South Africa and the deconstruction of Apartheid's migration regime. *Geoforum*, *30*(1), 1–11.

Cui, Y., Leng, C., & Sun, D. (2016). Sparse estimation of high-dimensional correlation matrices. *Computational Statistics & Data Analysis*, *93*, 390–403.

De Beer, J., Raymer, J., Van der Erf, R., & Van Wissen, L. (2010). Overcoming the problems of inconsistent international migration data: A new method applied to flows in Europe. *European Journal of Population*, *26*(4), 459–481.

Deng, X., & Tsui, K.-W. (2013). Penalized covariance matrix estimation using a matrix-logarithm transformation. *Journal of Computational and Graphical Statistics*, *22*(2), 494–512.

Dobra, A., Tebaldi, C., & West, M. (2006). Data augmentation in multi-way contingency tables with fixed marginal totals. *Journal of Statistical Planning and Inference*, *136*(2), 355–372.

Durand, J., Massey, D. S., & Parrado, E. A. (1999). The new era of Mexican migration to the United States. *The Journal of American History*, *86*(2), 518–536.

El Karoui, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *The Annals of Statistics*, *36*(6), 2717–2756.

Eliot, T. S. (1922). *The Waste Land*. Horace Liveright.

Esipova, N., Ray, J., & Publiese, A. (2011). *Gallup world poll. The many faces of global migration* (No. 43). Geneva, Switzerland: International Organization for Migration.

Faiola, A. (2016, March). *A right-wing party in Germany hopes to capitalize on anti-migrant anger.* Retrieved from https://www.washingtonpost.com/world/europe/in-germany-a-rising-voice-on-the-right/2016/03/10/c8582e08-e54f-11e5-a9ce-681055c7a05f_story.html ([Online; posted 2016/03/10])

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review*, *1*(2), 293–314.

Fan, J., Huang, T., & Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association*, *102*(478), 632–641.

Fan, J., Liao, Y., & Liu, H. (2015). An overview on the estimation of large covariance and precision matrices. *arXiv preprint arXiv:1504.02995*.

Fan, J., Liao, Y., & Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *75*(4), 603–680.

Fassmann, H., & Munz, R. (1994). European East-West migration, 1945-1992. *International Migration Review*, *28*(3), 520–538.

Fertig, M., & Schmidt, C. M. (2000). *Aggregate-level migration studies as a tool for forecasting future migration streams* (Tech. Rep. No. 183). Bonn, Germany: Institute for the Study of Labor.

Fienberg, S. E., & Holland, P. W. (1973). Simultaneous estimation of multinomial cell probabilities. *Journal of the American Statistical Association*, *68*(343), 683–691.

Fosdick, B. K., & Raftery, A. E. (2014). Regional probabilistic fertility forecasting by modeling between-country correlations. *Demographic Research*, *30*(35), 1011.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Annals de l'Université de Lyon*, *3*(14), 53–77.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441.

Furrer, R., & Bengtsson, T. (2007). Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *Journal of Multivariate Analysis*, *98*(2), 227–255.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*, 359–378.

Gonzalez-Barrera, A. (2015, November). *More Mexicans leaving than coming to the US* (Tech. Rep.). Pew Research Center.

Grobecker, C., & Brückner, G. (2016, March). *Nettozuwanderung von Ausländerinnen und Ausländern im Jahr 2015 bei 1,1 Millionen.* Retrieved from https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2016/03/PD16_105_12421.html ([Online; posted 2016/03/21])

Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). circlize implements and enhances circular visualization in R. *Bioinformatics*, btu393.

Hanson, G. H. (2006). Illegal migration from Mexico to the United States. *Journal of Economic Literature*, *44*(4), 869–924.

Harris, J. R., & Todaro, M. P. (1970). Migration, unemployment and development: A two-sector analysis. *American Economic Review*, *60*(1), 126–142.

Hatton, T. J., & Williamson, J. G. (1998). *The age of mass migration: Causes and economic impact*. Oxford University Press.

Hatton, T. J., & Williamson, J. G. (2002). *What fundamentals drive world migration?* (Tech. Rep.). Cambridge, MA: National Bureau of Economic Research.

Hatton, T. J., & Williamson, J. G. (2005). *Global migration and the world economy: Two centuries of policy and performance*. Cambridge, UK: Cambridge University Press.

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, *15*(5), 559–570.

Huang, A., & Wand, M. P. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, *8*(2), 439–452.

Huang, J. Z., Liu, N., Pourahmadi, M., & Liu, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika*, *93*(1), 85–98.

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

James, W., & Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 361–379).

Karachurina, L. (2013). *Migration in post-Soviet countries* (Tech. Rep.). Russian International Affairs Council. Retrieved from http://russiancouncil.ru/en/inner/?id_4=2369#top-content

Kim, K., & Cohen, J. E. (2010). Determinants of international migration flows to and from industrialized countries: A panel data approach beyond gravity. *International Migration Review*, *44*, 899–932.

Kupiszewska, D., & Wiśniowski, A. (2009, May). *Availability of statistical data on migration and migrant population and potential supplementary sources for data estimation* (Tech. Rep.). Central European Forum for Migration and Population Research.

Lanzieri, G. (2014). *Test of an estimation method for annual migration flows between EU-EFTA countries* (Working Paper). Eurostat.

Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, *10*(5), 603–621.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411.

Ledoit, O., & Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, *40*(2), 1024–1060.

Lee, E. S. (1966). A theory of migration. *Demography*, *3*(1), 47–57.

Leonard, T., & Hsu, J. S. (1992). Bayesian inference for a covariance matrix. *The Annals of Statistics*, *20*(4), 1669–1696.

Leslie, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika*, *33*, 183–212.

Levina, E., Rothman, A., Zhu, J., et al. (2008). Sparse estimation of large covariance matrices via a nested lasso penalty. *The Annals of Applied Statistics*, *2*(1), 245–263.

Liechty, J. C., Liechty, M. W., & Müller, P. (2004). Bayesian correlation estimation. *Biometrika*, *91*(1), 1–14.

Liu, H., Wang, L., & Zhao, T. (2014). Sparse covariance matrix estimation with eigenvalue constraints. *Journal of Computational and Graphical Statistics*, *23*(2), 439–459.

Lutz, W., & Goldstein, J. R. (2004). Introduction: How to deal with uncertainty in population forecasting? *International Statistical Review*, *72*, 1–4.

Massey, D. S. (1990). Social structure, household strategies, and the cumulative causation of migration. *Population Index*, *56*, 3–26.

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (1993). Theories of international migration: A review and appraisal. *Population and Development Review*, *19*(3), 431–466.

Massey, D. S., & Capoferro, C. (2004). Measuring undocumented migration. *International Migration Review*, 1075–1102.

Massey, D. S., & Espinosa, K. E. (1997). What's driving Mexico-US migration? A theoretical, empirical, and policy analysis. *American Journal of Sociology*, *102*, 939–999.

Mayer, T., & Zignago, S. (2011). Notes on CEPII's distances measures: The GeoDist database.

Myers, N. (2002). Environmental refugees: a growing phenomenon of the 21st century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *357*, 609–613.

Nocedal, J., & Wright, S. (2006). *Numerical Optimization*. Springer Science & Business Media.

Nowok, B., Kupiszewska, D., & Poulain, M. (2006). Statistics on international migration flows. In M. Poulain, N. Perrin, & A. Singleton (Eds.), *Thesim: Towards harmonised european statistics on international migration* (pp. 203–231). Louvain University Press.

Öberg, S. (1996). Spatial and economic factors in future South-North migration. In W. Lutz (Ed.), *The future population of the world: What can we assume today?* (pp. 336–357). Earthscan Publications.

OECD. (2013). *Coping with emigration in Baltic and East European countries* (Tech. Rep.). OECD Publishing. Retrieved from http://dx.doi.org/10.1787/9789264204928-en

Okólski, M. (1998). Regional dimension of international migration in Central and Eastern Europe. *Genus*, *54*(1), 11–36.

Organization for Economic Co-Operation and Development. (2015). *OECD international migration database*. Organization for Economic Co-operation and Development, Paris. Retrieved from http://stats.oecd.org/Index.aspx?DataSetCode=MIG

Özden, Ç., Parsons, C. R., Schiff, M., & Walmsley, T. L. (2011). Where on earth is everybody? The evolution of global bilateral migration 1960–2000. *The World Bank Economic Review*, *25*(1), 12–56.

Parker, G., Chazan, G., Chassany, A.-S., & Brunsden, J. (2016, June). *Cameron pins Brexit on EU failure to grant UK brake on migration.* Retrieved from https://www.ft.com/content/3901dd48-3cee-11e6-9f2c-36b487ebd80a ([Online; posted 2016/06/28])

Passel, J., Cohn, D., & Gonzalez-Barrera, A. (2012, April). *Net migration from Mexico falls to zero—and perhaps less* (Tech. Rep.). Pew Research Center.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd international workshop on distributed statistical computing* (pp. 20–22). Vienna, Austria: Austrian Association for Statistical Computing and the R Foundation for Statistical Computing.

Portes, A., & Walton, J. (1981). *Labor, class, and the international system.* New York: Academic Press.

Pourahmadi, M. (2011). Covariance estimation: The glm and regularization perspectives. *Statistical Science*, *26*(3), 369–387.

Raftery, A. E., Chunn, J. L., Gerland, P., & Ševčíková, H. (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography*, *50*, 777–801.

Raftery, A. E., Li, N., Ševčíková, H., Gerland, P., & Heilig, G. K. (2012). Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences*, *109*, 13915–13921.

Ratha, D., & Shaw, W. (2007). *South-South migration and remittances* (No. 102). World Bank Publications.

Raymer, J., de Beer, J., & van der Erf, R. (2011). Putting the pieces of the puzzle together: Age and sex-specific estimates of migration amongst countries in the EU/EFTA, 2002–2007. *European Journal of Population*, *27*(2), 185–215.

Raymer, J., & Rogers, A. (2007). Using age and spatial flow structures in the indirect estimation of migration streams. *Demography*, *44*, 199–223.

Reuveny, R., & Moore, W. H. (2009). Does environmental degradation influence migration? Emigration to developed countries in the late 1980s and 1990s. *Social Science Quarterly*, *90*, 461–479.

Richmond, A. H. (1988). Sociological theories of international migration: The case of refugees. *Current Sociology*, *36*, 7–25.

Rogers, A. (1990). Requiem for the net migrant. *Geographical analysis*, *22*, 283–300.

Rogers, A., & Castro, L. J. (1981). *Model migration schedules*. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Sander, N., Abel, G. J., Bauer, R., & Schmidt, J. (2014). *Visualising migration flow data with circular plots* (Working Paper). Vienna Institute of Demography.

Ševcıková, H. (2014). bayesDem: Graphical user interface for bayesTFR, bayesLife, and bayesPop. *R package version 2.4-1*.

Ševcıková, H., Raftery, A., & Buettner, T. (2016). bayesPop: Probabilistic population projection. *R package version 6.0-0*. Retrieved from http://CRAN.R-project.org/package=bayesPop

Ševčíková, H., Alkema, L., & Raftery, A. E. (2011). bayesTFR: An R package for probabilistic projections of the total fertility rate. *Journal of Statistical Software*, *43*, 1–29.

Ševčíková, H., & Raftery, A. (2015). bayesLife: Bayesian projection of life expectancy. *R package version 3.0-0*.

Sjaastad, L. A. (1962). The costs and returns of human migration. *Journal of Political Economy*, *70*(5), 80–93.

Stark, O. (1991). *The migration of labor*. Blackwell Oxford.

Stark, O., & Bloom, D. E. (1985). The new economics of labor migration. *American Economic Review*, *75*(2), 173–178.

ter Heide, H. (1963). Migration models and their significance for population forecasts. *The Milbank Memorial Fund Quarterly*, *41*, 56–76.

Tierney, L., & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, *81*(393), 82–86.

United Nations. (2011). *World Population Prospects: The 2010 Revision*. United Nations, New York.

United Nations. (2012). *World Population Prospects: The 2012 Revision*. United Nations, New York.

United Nations. (2015a). *International migrant stock: The 2015 Revision*. United Nations, New York.

United Nations. (2015b). *World Population Prospects: The 2015 Revision*. United Nations, New York.

United Nations Population Division. (2014). *Probabilistic population projections based on the World Population Prospects: The 2012 revision*. New York, NY: United Nations.

United States Internal Revenue Service. (2016). *U.S. population migration data*. IRS, Washington D.C. Retrieved from https://www.irs.gov/uac/soi-tax-stats-migration-data

US Census Bureau. (2005). *2005 interim state population projections*. United States Census Bureau.

U.S. Social Security Administration. (2013). *The 2013 annual report of the board of trustees of the federal old-age and survivors insurance and federal disability insurance trust funds.* Board of Trustees, Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds.

Valverde, M. (2016, July). *How Trump plans to build, and pay for, a wall along U.S.-Mexico border.* Retrieved from http://www.politifact.com/truth-o-meter/article/2016/jul/26/how-trump-plans-build-wall-along-us-mexico-border/ ([Online; posted 2016/07/26])

Wallerstein, I. (1974). *The modern world system, capitalist agriculture and the origins of the European world economy in the sixteenth century.* New York: Academic Press.

Wei, G. C., & Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, *85*(411), 699–704.

Wheldon, M. C., Raftery, A. E., Clark, S. J., & Gerland, P. (2013). Estimating demographic parameters with uncertainty from fragmentary data. *Journal of the American Statistical Association*, *108*, 96–110.

Willekens, F. (1999). Modeling approaches to the indirect estimation of migration flows: From entropy to EM. *Mathematical Population Studies*, *7*(3), 239–278.

Wiśniowski, A., Smith, P. W., Bijak, J., Raymer, J., & Forster, J. J. (2015). Bayesian population forecasting: Extending the Lee-Carter method. *Demography*, *52*(3), 1035–1059.

Wright, E. (2010). 2008-based national population projections for the United Kingdom and constituent countries. *Population Trends*, *139*, 91–114.

Yoshida, R. (2015, November). *Japans immigration policy rift widens as population decline forces need for foreign workers.* Retrieved from http://www.japantimes.co.jp/

news/2015/11/25/national/politics-diplomacy/japans-immigration-policy-rift-widens-population-decline-forces-need-foreign-workers/#.V-6nTclWbrl ([Online; posted 2015/11/25])

Zhang, T., & Zou, H. (2012). Sparse precision matrix estimation via lasso penalized D-trace loss. *Biometrika*, *99*(1), 1–18.

Zipf, G. K. (1946). The $p_1 p_2/d$ hypothesis: on the intercity movement of persons. *American sociological review*, *11*(6), 677–686.

Zlotnik, H. (1987). Measuring international migration: theory and practice. *International Migration Review*, *21*(4), v–xii.

Appendix A

APPENDICES TO CHAPTER 2

A.1 Data Sources

One source of uncertainty in our migration projections is uncertainty in the estimates of net migration themselves. However, the practice of the World Population Prospects is to publish only point estimates of net migration with no measure of uncertainty. As a partial stand-in for uncertainty, we might instead look at the method used to produce estimates of migration. The United Nations Population Division publishes a WPP methodology document that outlines data sources for migration estimates for all countries. These data sources are summarized by Table A.1.

Table A.1: Summary of sources of net migration data as reported in the documentation for the 2015 revision of the World Population Prospects. "Residual" indicates the number of countries that used an estimate of net migration as the difference between population growth and natural increase. "Refugee" indicates the number of countries in a region for which estimates of refugee migration were used in estimating total net migration. "Destination" indicates the number of countries for which data from typical receiving countries was used to estimate out-migration. "Modeled" indicates countries for which some reported net migration values were based on models or assumptions rather than measurements.

Region	Total # of countries	Residual	Refugee	Destination	Modeled
Africa	57	30	38	19	10
Asia	51	40	17	15	11
Europe	40	40	4	1	1
Lat. Am. & Carib.	38	37	0	23	4
Northern America	2	2	0	0	0
Oceania	13	13	0	0	3

In much of the developed world, net migration is estimated as the difference between population growth and natural increase. When this is the case, the quality of migration estimates will be controlled by the quality of estimates of population, births, and deaths. As such, in places with good censuses and means of accurately tracking births and deaths, net migration estimates should be reasonably good. However, in 39 out of 201 countries in the 2015 revision of WPP (United Nations, 2015b), migration is estimated entirely using other methods. The most common alternative techniques for estimating net migration in the WPP data are using estimates of refugee migration (a common technique in Africa), and using information about either changes in migrant populations or work permits along typical migration corridors (a common technique in Latin America). In 29 out of 201 countries, some or all of the net migration data was based on models or assumptions rather than measurements of migration. This is especially prevalent among the earlier time points, where WPP sometimes reports zero net migration as an indicator of missing data. A particularly prominent example is North Korea, for which net migration is estimated to be zero at all time points since 1955.

The issue of variability in estimates of net migration is particularly insidious, since not only our model fitting but also our model evaluation rely on these estimates. Consequently, if the WPP estimates of net migration are systematically different from the true migration counts, our projections could appear to be well calibrated while consistently failing to capture the truth. Although there is no evidence that this is the case, we should be vigilant of the fact that in many cases we are modeling the sum of migration and the error in the residual estimation method. Theoretically, this missing source of variability should cause undercoverage by our projections. However, in absence of trustworthy true migration counts, it is difficult to assess the extent to which this missing variability is an issue.

A.2 Gravity Model Implementation

We implemented a version of Cohen's 2012 gravity model, which projects net migration counts for five-year intervals starting at 2010 and ending at 2100. Projections are made for

each country independently, with no redistribution step to ensure zero global net migration. For each country, projections are produced as follows. Let $L(t)$ be the population of country c at time t (in millions) and $M(t)$ be the population of the rest of the world at time t (in millions). Then expected in-migration to country c is given by $a \times L(t)^{\alpha} M(t)^{\beta}$, where a is a country-specific proportionality constant, and the exponents α and β are constant across countries, with values estimated by Kim & Cohen (2010). Similary, expected out-migration from country c has the form $b \times L(t)^{\gamma} M(t)^{\delta}$, where b is to be estimated, and γ and δ come from Kim & Cohen (2010).

The constants of proportionality a and b for each country are chosen to minimize the sum of squared deviations between estimates of net migration from the gravity model and WPP estimates of net migration (United Nations, 2011) given in units of millions of net annual migrants. We use the values $\alpha = 0.728$, $\beta = 0.602$, $\gamma = 0.373$, and $\delta = 0.948$, reported by Cohen (2012). For each country, having estimated a and b, we calculate net migration projections by $a \times L(t)^{\alpha} M(t)^{\beta} - b \times L(t)^{\gamma} M(t)^{\delta}$, where $L(t)$ and $M(t)$ are now projected populations, also taken from WPP's 2010 revision (United Nations, 2011).

Our implementation appears to reproduce Cohen's 2012 results. Cohen reported the values of the proportionality constants, a and b, obtained for the United States, and provided a plot of the projections from his implementation of the gravity model. Using these, we are able to confirm that our results agree with those from Cohen's implementation. Cohen reported $a = 3.43 \times 10^{-4}$ and $b = -8.28 \times 10^{-4}$. We find very similar values of $a = 3.42 \times 10^{-4}$ and $b = -8.33 \times 10^{-4}$. The slight discrepancies may come from having used only three decimal places for the values of α, β, γ, and δ in our implementation. Furthermore, Fig. A.1 shows the projected net migration counts for the United States using our implementation of the gravity model. Our projections appear to be essentially the same as the gravity model projections plotted in Cohen (2012:figure 1(b)).

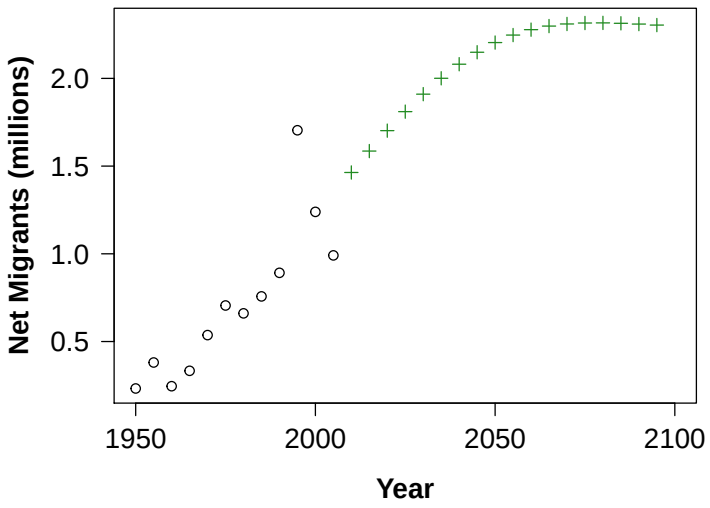

Figure A.1: Gravity model based projections of net international migration counts for the United States

A.3 Regional Performance Tables

Here, we present tables of evaluation results split up by region. Because some countries in the Middle East have had much higher migration rates than other regions during the past several decades, we split out Western Asia[1] from the rest of Asia.

When making predictions over a five-year time period, our model outperforms the gravity model in all regions and the persistence models in three out of six regions. Over longer horizons, our model outperforms both the gravity and persistence models for all regions except Oceania.

[1]In the WPP 2010 data set, Western Asia comprises Armenia, Azerbaijan, Bahrain, Cyprus, Georgia, Iraq, Israel, Jordan, Kuwait, Lebanon, Occupied Palestinian Territory, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Turkey, United Arab Emirates, and Yemen (United Nations, 2011).

Table A.2: Five-year predictive performance of different methods: Mean absolute errors (MAE) and prediction interval coverage for our Bayesian hierarchical model, the gravity model, and the persistence models across different regions

Region	Model	MAE	80% Cov. (%)	95% Cov. (%)
Africa	Bayesian	1.61	94.5	100
	Gravity	3.22	—	—
	Persistence (of rates)	2.56	—	—
	Persistence (of counts)	2.16	—	—
Europe	Bayesian	1.73	85.0	90.0
	Gravity	3.39	—	—
	Persistence (of rates)	2.01	—	—
	Persistence (of counts)	2.02	—	—
Americas	Bayesian	1.38	94.9	100
	Gravity	2.58	—	—
	Persistence (of rates)	1.39	—	—
	Persistence (of counts)	1.23	—	—
Oceania	Bayesian	2.23	91.7	100
	Gravity	2.86	—	—
	Persistence (of rates)	1.82	—	—
	Persistence (of counts)	1.75	—	—
Western Asia	Bayesian	17.54	77.8	88.9
	Gravity	21.28	—	—
	Persistence (of rates)	17.27	—	—
	Persistence (of counts)	19.16	—	—
Rest of Asia	Bayesian	2.37	97.0	97.0
	Gravity	2.91	—	—
	Persistence (of rates)	2.87	—	—
	Persistence (of counts)	2.80	—	—
World	Bayesian	3.24	91.4	96.4
	Gravity	4.70	—	—
	Persistence (of rates)	3.57	—	—
	Persistence (of counts)	3.58	—	—

Table A.3: Fifteen-year predictive performance of different methods: Mean absolute errors (MAE) and prediction interval coverage for our Bayesian hierarchical model, the gravity model, and the persistence models across different regions

Region	Model	MAE	80% Cov. (%)	95% Cov. (%)
Africa	Bayesian	4.60	84.8	95.2
	Gravity	7.14	—	—
	Persistence (of rates)	7.45	—	—
	Persistence (of counts)	6.38	—	—
Europe	Bayesian	3.44	78.3	87.5
	Gravity	5.87	—	—
	Persistence (of rates)	5.49	—	—
	Persistence (of counts)	5.62	—	—
Americas	Bayesian	2.44	89.7	93.2
	Gravity	3.78	—	—
	Persistence (of rates)	2.79	—	—
	Persistence (of counts)	2.57	—	—
Oceania	Bayesian	4.25	83.3	94.4
	Gravity	4.68	—	—
	Persistence (of rates)	3.53	—	—
	Persistence (of counts)	3.58	—	—
Western Asia	Bayesian	15.23	85.2	92.6
	Gravity	18.59	—	—
	Persistence (of rates)	19.46	—	—
	Persistence (of counts)	19.21	—	—
Rest of Asia	Bayesian	3.86	87.9	97.0
	Gravity	3.92	—	—
	Persistence (of rates)	5.99	—	—
	Persistence (of counts)	5.34	—	—
World	Bayesian	4.76	84.9	93.4
	Gravity	6.57	—	—
	Persistence (of rates)	6.74	—	—
	Persistence (of counts)	6.30	—	—

Table A.4: Thirty-year predictive performance of different methods: Mean absolute errors (MAE) and prediction interval coverage for our Bayesian hierarchical model, the gravity model, and the persistence models across different regions

Region	Model	MAE	80% Cov. (%)	95% Cov. (%)
Africa	Bayesian	5.63	77.3	87.3
	Gravity	17.06	—	—
	Persistence (of rates)	10.40	—	—
	Persistence (of counts)	7.33	—	—
Europe	Bayesian	3.25	73.3	86.7
	Gravity	5.50	—	—
	Persistence (of rates)	3.27	—	—
	Persistence (of counts)	3.25	—	—
Americas	Bayesian	4.17	81.2	94.9
	Gravity	8.44	—	—
	Persistence (of rates)	5.29	—	—
	Persistence (of counts)	4.72	—	—
Oceania	Bayesian	5.16	79.2	94.4
	Gravity	7.53	—	—
	Persistence (of rates)	4.36	—	—
	Persistence (of counts)	4.23	—	—
Western Asia	Bayesian	11.08	76.9	88.0
	Gravity	27.17	—	—
	Persistence (of rates)	14.21	—	—
	Persistence (of counts)	11.58	—	—
Rest of Asia	Bayesian	4.42	76.3	87.9
	Gravity	10.94	—	—
	Persistence (of rates)	5.91	—	—
	Persistence (of counts)	5.17	—	—
World	Bayesian	5.12	77.2	89.3
	Gravity	12.32	—	—
	Persistence (of rates)	7.17	—	—
	Persistence (of counts)	5.82	—	—

Appendix B

APPENDICES TO CHAPTER 3

B.1 Determining step size

Step size selection is necessary in high dimensions for the general gradient descent algorithm to converge quickly enough to be useful. Complex methods for step size selection are available, but we obtained reasonable results with the backtracking line search algorithm, which starts with a large step size and decreases step size whenever a proposed step results in too little improvement in the objective function.

Say we have an objective function $f(x)$ which we are trying to minimize. The core of the backtracking line search algorithm is as follows (Nocedal & Wright, 2006).

1. Fix a backtracking coefficient $\beta \in (0, 1)$, a starting step size α_0, and a starting location x_0.

2. Propose a step of length α_k in direction p_k. (The backtracking line search algorithm is a generic algorithm that will work regardless of how the direction p_k is determined.)

3. If the improvement in the objective function is enough to meet the Armijo condition given in (B.1) below, then take the proposed step. That is, take $x_{k+1} = x_k + \alpha_k p_k$. Keep the step size constant (i.e., $\alpha_{k+1} = \alpha_k$).

4. Otherwise, if there is any improvement in the objective function, take the proposed step, but also decrease the step size for the next iteration (specifically, set $\alpha_{k+1} = \beta\alpha_k$).

5. Otherwise, there must have been no improvement in the objective function. Don't take a step, but do decrease step size. ($x_{k+1} = x_k$ and $\alpha_{k+1} = \beta\alpha_k$.)

6. Repeat steps 2-5 until convergence.

The Armijo condition, which is used to determine whether to decrease step size, is as follows. The Armijo condition is met if the following inequality is satisfied:

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_i \nabla f_k^T p_k. \tag{B.1}$$

(c_1 is a constant chosen from $(0, 1)$ that controls how strictly the change in f must match the gradient at x_k.)

In our application, there's a missing component—we can't actually compute the gradient of our objective function. The relevant objective function is given by

$$f(R) = \text{tr}(R_i^{-1} R) + \text{tr}(R^{-1} \tilde{R}) + \lambda \|P * R\|_1. \tag{B.2}$$

The first two terms in the sum are differentiable, but the third is not.

We rewrite the Armijo condition as

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_i p_k^T p_k - c_1 \alpha_i (p_k - \nabla f_k)^T p_k \tag{B.3}$$

and then approximate $(p_k - \nabla f_k)$ with

$$-2 \cdot \nabla(\text{tr}(R_i^{-1} R) + \text{tr}(R^{-1} \tilde{R})). \tag{B.4}$$

B.2 Inflation of correlation estimates

We provide here an example of our correlation estimation procedure which produces inflation in some unpenalized elements of the correlation matrix. We solved the minimization problem in (3.12) with three different methods, finding identical answers each time, up to small numerical tolerances. Those methods are:

1. Estimate R using our code, which appeals to the generalized gradient descent algorithm.

2. Estimate R using a black-box numerical optimization algorithm, which has access to the function we're minimizing, but not its derivative.

3. Estimate R by finding an analytic expression for the gradient of the function we're minimizing, and solve for a point where the gradient is zero.

One case in which inflation manifests if we take our evidence from the data to be given by

$$\tilde{R} = \begin{pmatrix} 1 & 0.8 & 0.5 \\ 0.8 & 1 & 0.1 \\ 0.5 & 0.1 & 1 \end{pmatrix} \tag{B.5}$$

and the penalty matrix by

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \tag{B.6}$$

We denote the unknown true correlation matrix by

$$R = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{pmatrix}. \tag{B.7}$$

We fix the regularization parameter at $\lambda = 0.5$. The problem is then to estimate the three parameters ρ_1, ρ_2, and ρ_3.

With all three methods we find an estimate of

$$\hat{\boldsymbol{\rho}} = \begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \hat{\rho}_3 \end{pmatrix} = \begin{pmatrix} 0.8211 \\ 0.1542 \\ -0.1813 \end{pmatrix}. \tag{B.8}$$

Note that the second element, which is penalized, experiences shrinkage towards zero, as expected. The first element is inflated, while the third is both inflated and changes sign.

Appendix C

APPENDICES TO CHAPTER 4

C.1 Bayes and pseudo-Bayes estimators for entries in contingency tables

This appendix outlines the details of standard Bayes and pseudo-Bayes estimators for multinomial probabilities, as well as how our estimator of migration counts differs from the standard estimators. The results in this section are primarily drawn from Chapter 12 of Bishop, Fienberg, and Holland's *Discrete Multivariate Analysis: Theory and Practice* (Bishop et al., 1975).

C.1.1 Standard Bayes and pseudo-Bayes estimators for multinomial cell probabilities

A typical estimation scenario for multinomial tables is that the observations consist of a complete matrix of cell counts, X, with entries $\{x_{ij}\}$, which sum to a grand total, n. Cell counts are assumed to be drawn from a multinomial distribution. The quantity to be estimated is the true underlying cell probability matrix, P, with entries $\{p_{ij}\}$.

In the presence of an assumed prior distribution for the cell probabilities, P, we can derive a posterior distribution for P given the observed cell counts, X. A standard choice is to assume a Dirichlet distribution on P with parameter matrix B, with entries $\{b_{ij}\}$. It is also convenient to decompose the Dirichlet parameters by defining

$$K := \sum_{ij} b_{ij} \tag{C.1}$$

and

$$a_{ij} := \frac{b_{ij}}{K}. \tag{C.2}$$

Under this decomposition, the prior mean for p_{ij} is given by a_{ij}, while K is an overall

concentration parameter that controls how tight the prior distribution is around its mean.

Then we can express the posterior mean for P in the following form:

$$\mathrm{E}[P|B,X] = \frac{n}{n+K}(X/n) + \frac{K}{n+K}A. \tag{C.3}$$

That is, the Bayes estimator for P is a convex combination of the sample cell probabilities and the prior probabilities.

The *bona fide* Bayes estimator can then be extended to a pseudo-Bayes estimator in a number of ways. Note that a true Bayes estimator requires a prior belief about cell probabilities to be specified in advance, while it may be preferable in practice to allow these prior probabilities to be influenced by observed data.

One possible extension is to apply some additional structure to the prior mean matrix, A. A common empirical choice of prior recommended by Bishop et al. (1975) is to choose $A_{ij} \propto X_{i+}X_{+j}$. This is the independence-structured prior which produces the expected cell counts $\hat{M}^I$ which appear in our estimator.

Furthermore, in a genuine Bayes estimator, the concentration parameter K is a property of the prior distribution on P. However, one may prefer to specify only the prior means, A, while choosing K in a way which minimizes the expectation of some loss function. The value of K which minimizes Bayes risk under squared error loss can be shown to be

$$K(P,A) = \frac{1 - \|P\|_2^2}{\|P - A\|_2^2}. \tag{C.4}$$

Of course, this optimal value for K depends on the unknown value of P. Instead, one might estimate the optimal value of K with

$$\hat{K}(\hat{P},A) = \frac{1 - \|\hat{P}\|_2^2}{\|\hat{P} - A\|_2^2}, \tag{C.5}$$

where $\hat{P}$ is an initial estimate of P. This gives a pseudo-Bayes estimator which often lowers

risk in practice.

C.1.2 Our pseudo-Bayes estimator

Again, the true Bayes estimator for cell probabilities in a contingency table with known entries X is given by

$$\frac{n}{n+K}(X/N) + \frac{K}{n+K}A, \tag{C.6}$$

or equivalently,

$$w(X/n) + (1-w)A. \tag{C.7}$$

The typical extension to a pseudo-Bayes estimator is to allow K (or equivalently w) and A to depend on the data.

However, in our migration application we introduce two further modifications. Firstly, we have an alternate route to estimating w. Instead of leveraging an initial set of estimates into a guess at a better value for w, we choose a value for w by validating against external data sources (i.e. OECD and MIMOSA flow estimates). The motivation is still minimization of expected loss, but we accomplish this via validation against gold standard data rather than analytical results. Secondly, we don't have a set of observed cell counts X. What we do have are estimated cell counts $\hat{M}^A$ produced using Abel's (2013) minimum migration method.

Our pseudo-Bayes estimator for cell probabilities takes the form

$$w(\hat{M}^A/n) + (1-w)(\hat{M}^I/n). \tag{C.8}$$

Taking the place of observed cell counts, X, are the MM estimates of cell counts, $\hat{M}^A$. In place of prior cell probabilities, A, we use an empirical prior, $\hat{M}^I/n$, that places an independence structure on the table. Finally, w is a weight estimated by minimizing some loss function on OECD and MIMOSA data.

There is some cost associated to introducing these modifications to the Bayes estimator for cell probabities. Most Bayes estimators have the property of converging to some sample

statistic asymptotically, which is a desirable feature because the sample statistic is usually a consistent estimator of the quantity of interest. That is not the case here. As n increases, our weight w doesn't change because we estimated w from an external data source. However, the MM estimator $\hat{M}^A$ is itself asymptotically biased, so asymptotic convergence to $\hat{M}^A$ would not guarantee consistency anyhow.

While our estimator fits into the morphological family of pseudo-Bayes estimators, it's worth keeping in mind that the motivation for using it is somewhat different from the typical Bayes estimator. The scenario is not one of trying to adjust a consistent estimator in a way that lowers Bayes risk. Nor are we improving a low-sample-size estimate by incorporating prior information. Instead, we're making up for a structural weakness in the MM estimator (namely, that no cross-flows exist) by shrinking towards a prior mean which lacks that weakness.