

# Bayesian Demographic Accounts: Subnational Population Estimation Using Multiple Data Sources

John R. Bryant<sup>\*</sup> and Patrick J. Graham<sup>†</sup>

**Abstract.** Demographic estimates for small areas within a country have many uses. Subnational population estimation is, however, difficult, requiring the synthesis of multiple inconsistent datasets. Current methods have important limitations, including a heavy reliance on ad hoc adjustment and limited allowance for uncertainty. In this paper we demonstrate how subnational population estimation can be carried out within a formal Bayesian framework. The core of the framework is a demographic account, providing a complete description of the demographic system. Regularities within the demographic account are described by a system model. The relationship between the demographic account and the observable data is described by an observation model. Posterior simulation is carried out using Markov chain Monte Carlo methods. We illustrate the methods using data for six regions within New Zealand.

**Keywords:** demography, official statistics, population estimation, hierarchical Bayesian model, MCMC

## 1 Introduction

Sensible decisions about the location of infrastructure, the distribution of health funding, the targeting of marketing, and much else besides require information about local demographic trends. This information can be obtained from subnational population estimates and projections—that is, demographic estimates disaggregated by age, sex, and geographical area. Except in a handful of countries with accurate population registers, estimates must be constructed from multiple data sources of varying quality and completeness. In many countries, the most important such source, the population census, faces an uncertain future because of funding pressures. Statistical agencies are seeking to make better use of ‘administrative’ data produced by non-statistical organizations such as tax authorities and schools. Due to improvements in information technology, such data are becoming increasingly plentiful. However, using administrative data for population estimation can be difficult because of non-standard target populations, the absence of important variables, and frequent definitional changes.

Demographers and statisticians have developed many methods for adjusting, interpolating, and reconciling data from diverse sources to construct subnational estimates (e.g. Rees et al. 2004; Smith and Morrison 2005; Boden and Rees 2010; Smith et al. 2010;

---

<sup>\*</sup>Statistics New Zealand [john.bryant@stats.govt.nz](mailto:john.bryant@stats.govt.nz)

<sup>†</sup>Bayesian Research [patrick.graham.br@gmail.com](mailto:patrick.graham.br@gmail.com)

Raymer et al. 2011; Wilson and Bell 2011). Although much has been achieved with these methods, they have important limitations. Most are not based on explicit models for random variation in the demographic processes or in the measurement of these processes. Instead, practitioners are required to use expert judgment to reconcile inconsistent data or to adjust values that are demographically implausible. The estimation process is accordingly labour-intensive and difficult to replicate.

In this paper we set out a Bayesian statistical framework for subnational population estimation that is more efficient and transparent than traditional methods, and that deals more satisfactorily with uncertainty. At the core of the framework is a ‘demographic account’ (Rees 1979; Stone 1984), a complete description of the demographic stocks and flows of interest, subject to accounting identities that relate stocks to flows. The stocks are population counts, disaggregated by age, sex, and region; the flows are births, deaths, and migrations. Regularities within the demographic account are captured by a ‘system model’. The relationship between the demographic system and the available data is captured by an ‘observation model’. Inference is carried out through Markov chain Monte Carlo methods. We illustrate the framework by estimating the populations of six subnational regions in New Zealand, using an eclectic mix of data sources.

Our approach is distinct from, and complementary to, capture-recapture and record linkage approaches to population estimation (e.g. Fienberg et al. 1999; Elliott and Little 2000; Liseo and Tancredi 2011). The smallest unit in our method is a cell count; the smallest unit in capture-recapture and record linkage methods is the individual animal, person, or event. Our methods can exploit a wider range of data, and raise fewer privacy concerns; capture-recapture and record linkage methods potentially yield a richer individual-level dataset.

Bayesian methods similar to ours are common in the study of animal demography (King et al. 2009; Link and Barker 2009). Some methods combine multiple data sources and include system and observation models (e.g. King et al. 2009, chapter 10). In human demography, use of Bayesian methods is less common, but is increasing rapidly. Examples include the estimation and forecasting of mortality (Dellaportas et al. 2001; Girosi and King 2008; Lynch and Brown 2010), fertility (Alkema et al. 2008; Rendall et al. 2009; Alkema et al. 2011, 2012), migration (Brierley et al. 2008; Congdon 2008; Bijak and Wiśniowski 2010), and population (Daponte et al. 1997; Abel et al. 2010; Raftery et al. 2012). Some important new features of our framework and application are the use of a demographic account to link multiple demographic series, and the scale of the application, as measured by the volume of data or the number of parameters and sub-models.

We begin by presenting a high-level overview of our framework. We then develop system and observation models specific to our New Zealand application, demonstrate how to simulate from the posterior distribution, and present results. We conclude by describing extensions, other possible applications, and potential objections to the use of Bayesian methods in the production of official population statistics.

## 2 The framework

Let  $\mathbf{Q}$  denote a demographic account.  $\mathbf{Q}$  consists of counts of population, births, deaths, and migration, all cross-classified by region, sex, age, and time. The series making up  $\mathbf{Q}$  are linked by an accounting identity, typically stating that population at the end of a period must equal population at the beginning of the period plus births minus deaths plus in-migrations minus out-migrations. We treat  $\mathbf{Q}$  as latent or unobserved.

Demographic systems exhibit strong regularities. For instance, death rates tend to be similar across regions within a country, and regional populations tend to have distinctive age-profiles that persist over time. Moreover, some of the regularities can be predicted from characteristics such as income levels or the presence of universities. The regularities, and their relationship to external characteristics, are represented by the system model  $p(\mathbf{Q}|\boldsymbol{\theta}_\mathbf{Q}, \mathbf{Z})$ . Quantity  $\boldsymbol{\theta}_\mathbf{Q}$  contains the parameters for the model, and  $\mathbf{Z}$  contains covariates that help predict the value the parameters. Conditional on  $\boldsymbol{\theta}_\mathbf{Q}$ ,  $\mathbf{Q}$  is independent of  $\mathbf{Z}$ , that is,  $p(\mathbf{Q}|\boldsymbol{\theta}_\mathbf{Q}, \mathbf{Z}) = p(\mathbf{Q}|\boldsymbol{\theta}_\mathbf{Q})$ .

Data  $\mathbf{X}$  consist of  $K$  datasets,  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K)$ . The relationship between  $\mathbf{X}$  and  $\mathbf{Q}$  is governed by the observation model  $p(\mathbf{X}|\mathbf{Q}, \boldsymbol{\theta}_\mathbf{X})$ . Each dataset  $\mathbf{X}_k$  is typically correlated with a different subset of  $\mathbf{Q}$ . It is possible for multiple datasets to be correlated with the same elements of  $\mathbf{Q}$ . In general, the strength of correlations varies from dataset to dataset. For some datasets, the strength may also vary across dimensions such as age or time.

The two observable quantities  $\mathbf{X}$  and  $\mathbf{Z}$  have different characteristics and roles. All cell values within  $\mathbf{X}$  must be non-negative counts, while no such restriction is placed on values within  $\mathbf{Z}$ .  $\mathbf{X}$  forms part of the observation model, while  $\mathbf{Z}$  forms part of the system model. Typically,  $\mathbf{X}$  contains far more values than  $\mathbf{Z}$ .

We assume that the system and observation models are distinct in the sense that  $\boldsymbol{\theta}_\mathbf{Q}$  and  $\boldsymbol{\theta}_\mathbf{X}$  do not share any parameters, that  $p(\boldsymbol{\theta}_\mathbf{Q}, \boldsymbol{\theta}_\mathbf{X}|\mathbf{Z}) = p(\boldsymbol{\theta}_\mathbf{Q}|\mathbf{Z})p(\boldsymbol{\theta}_\mathbf{X})$ , and that  $p(\mathbf{Q}, \mathbf{X}|\boldsymbol{\theta}_\mathbf{Q}, \boldsymbol{\theta}_\mathbf{X}, \mathbf{Z}) = p(\mathbf{Q}|\boldsymbol{\theta}_\mathbf{Q}, \mathbf{Z})p(\mathbf{X}|\mathbf{Q}, \boldsymbol{\theta}_\mathbf{X})$ . The relationship between  $\mathbf{Q}$ ,  $\mathbf{X}$ ,  $\mathbf{Z}$ ,  $\boldsymbol{\theta}_\mathbf{Q}$ , and  $\boldsymbol{\theta}_\mathbf{X}$  is summarised in Figure 1.

The goal of inference is to obtain the joint posterior  $p(\mathbf{Q}, \boldsymbol{\theta}_\mathbf{Q}, \boldsymbol{\theta}_\mathbf{X}|\mathbf{X}, \mathbf{Z})$ . Typically, a summary of this, the marginal distribution  $p(\mathbf{Q}|\mathbf{X}, \mathbf{Z})$ , is the main focus of attention. However, the marginal distribution  $p(\boldsymbol{\theta}_\mathbf{Q}|\mathbf{X}, \mathbf{Z})$  provides potentially useful information about the demographic system. In addition, as illustrated below,  $p(\boldsymbol{\theta}_\mathbf{X}|\mathbf{X}, \mathbf{Z})$  may be useful for diagnosing problems with the model or evaluating the datasets.

Given the assumptions described so far,

$$\begin{aligned} p(\mathbf{Q}, \boldsymbol{\theta}_\mathbf{Q}, \boldsymbol{\theta}_\mathbf{X}|\mathbf{X}, \mathbf{Z}) &\propto p(\mathbf{X}|\mathbf{Q}, \boldsymbol{\theta}_\mathbf{Q}, \boldsymbol{\theta}_\mathbf{X}, \mathbf{Z})p(\mathbf{Q}|\boldsymbol{\theta}_\mathbf{Q}, \boldsymbol{\theta}_\mathbf{X}, \mathbf{Z})p(\boldsymbol{\theta}_\mathbf{Q}, \boldsymbol{\theta}_\mathbf{X}|\mathbf{Z}) \\ &= p(\mathbf{X}|\mathbf{Q}, \boldsymbol{\theta}_\mathbf{X})p(\mathbf{Q}|\boldsymbol{\theta}_\mathbf{Q})p(\boldsymbol{\theta}_\mathbf{Q}|\mathbf{Z})p(\boldsymbol{\theta}_\mathbf{X}). \end{aligned} \tag{1}$$

The joint posterior  $p(\mathbf{Q}, \boldsymbol{\theta}_\mathbf{Q}, \boldsymbol{\theta}_\mathbf{X}|\mathbf{X}, \mathbf{Z})$  can be simulated using a Gibbs sampler alter-

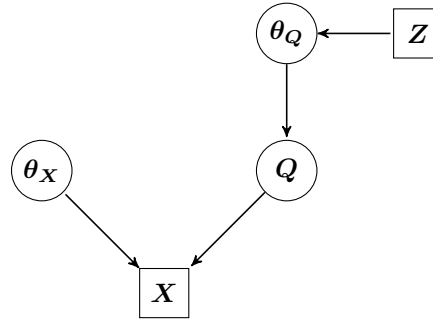


Figure 1: Our framework.  $Q$  denotes the demographic account,  $X$  the data,  $Z$  covariates for the system model,  $\theta_Q$  parameters for the system model, and  $\theta_X$  parameters for the observation model. Squares represent observed quantities and circles unobserved quantities.

nating between the following full conditional distributions:

$$\begin{aligned}
 p(Q|\theta_Q, \theta_X, Z, X) &\propto p(X|Q, \theta_X)p(Q|\theta_Q) \\
 p(\theta_Q|Q, \theta_X, Z, X) &\propto p(Q|\theta_Q)p(\theta_Q|Z) \\
 p(\theta_X|Q, \theta_Q, Z, X) &\propto p(X|Q, \theta_X)p(\theta_X).
 \end{aligned}$$

In the observation model  $p(X|Q, \theta_X)$ , the  $X$  plays the role of a response variable and the  $Q$  plays the role of a predictor. Within this setup, dataset  $X_k$  can be less detailed than the subset of  $Q$  to which it is linked. For instance,  $X_k$  need not have as many dimensions as  $Q$ , or have values for all age groups. In such cases, values from  $Q$  are subsetting or aggregated to obtain the predictor for  $X_k$ . This flexibility, combined with the ability to deal with systematic and random variation in the relationship between  $X_k$  and  $Q$ , means that messy or incomplete data can be accommodated much more easily within our framework than they can within traditional approaches.

### 3 Application to subnational population estimation in New Zealand

#### 3.1 Scope and data

To illustrate our methods, we construct estimates and forecasts for six regions in New Zealand over the period 1996-2011. The regions are artificial creations of approximately equal population size, constructed by grouping contiguous ‘territorial authorities’. The estimates and forecasts are disaggregated by region, sex, single year of age, and time.

New Zealand holds population censuses every five years. Statistics New Zealand is, however, exploring alternatives to the traditional census, and has a long-term goal of increasing the use of administrative data. Data on births and deaths are obtained from

Table 1: Datasets used in the application

| Dataset  | Associated demographic series | Expectations about relationship between demographic series and dataset  | Sex, age, and time detail   |
|--|-------------------------------|---|---|
| Registered births  | Births                        | Very close  | Sex and single year of age for 1997-2011  |
| Registered deaths  | Deaths                        | Very close  | Sex and single year of age for 1997-2011  |
| Official estimates of resident population in census years  | Population                    | Very close  | Sex and single year of age, 1996, 2001, 2006  |
| Census data on 5-year transitions from elsewhere in NZ     | Internal in-migration         | Approximated by summing 5 years of annual migration counts, cohort by cohort. Relationship likely to vary by age.                   | Sex and single year of age, 1996-2001 and 2001-2006   |
| Census data on 5-year transitions to elsewhere in NZ       | Internal out-migration        | Approximated by summing 5 years of annual migration counts, cohort by cohort. Relationship likely to vary by age.                   | Sex and single year of age, 1996-2001 and 2001-2006   |
| Census data on 5-year transitions from overseas            | External in-migration         | Approximated by summing 5 years of annual migration counts, cohort by cohort. Relationship likely to vary by age.                   | Sex and single year of age, 1996-2001 and 2001-2006   |
| International ‘permanent and long-term’ arrivals           | External in-migration         | Close, though data on region of destination often unreliable, and some difficulty distinguishing long-term and short-term movements | Sex and single year of age, 1997-2011   |
| International ‘permanent and long-term’ departures         | External out-migration        | Close, though data on region of origin often unreliable, and some difficulty distinguishing long-term and short-term movements      | Sex and single year of age, 1997-2011   |
| Tax system data on changes of address from elsewhere in NZ | Internal in-migration         | Only includes people whose income is taxed at source. Relationship likely to vary by age and sex.                                   | For 2002-2008: sex and mainly 5-year age groups, ages 18+. For 2010-2011: sex and single year of age, ages 18+. |
| Tax system data on changes of address to elsewhere in NZ   | Internal out-migration        | Only includes people whose income is taxed at source. Relationship likely to vary by age and sex.                                   | For 2002-2008: sex and mainly 5-year age groups, ages 18+. For 2010-2011: sex and single year of age, ages 18+. |
| Electoral roll   | Population                    | Relationship known to vary by age   | No sex, mainly 5-year age groups, ages 18+, 2004-2011   |
| School roll  | Population                    | Relationship known to vary by age   | Sex and single year of age, ages 5-22, 2000-2010.   |

the vital registration system and are of excellent quality. Data on external migration are obtained from international arrival and departure cards and are of good quality. There are, however, no comparable data on internal migration. Moreover, New Zealand's internal migration rates are high by international standards (Greenwood 1997, Table 2). Estimation of internal migration is therefore the most difficult part of subnational population estimation in New Zealand.

The datasets  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K$  used in our application are summarised in Table 1. The table shows, for instance, that the relationship between the electoral roll data and the underlying population counts varies by age, and that the electoral roll data lack information on sex. All datasets are disaggregated by region. Information about the datasets listed in Table 1 is available in Dunstan and Ryan (2011).

Covariate vector  $\mathbf{Z}$  contains a single indicator variable taking a value of 1 if university students comprised more than one third of the population of that region and age group at the time of the 2006 census and 0 otherwise. In New Zealand, as elsewhere (Wilson 2010), regions with universities experience disproportionately high migration by young people.

### 3.2 The demographic account

In our application, demographic account  $\mathbf{Q}$  is composed of population, births, deaths, internal (i.e. domestic) in-migration, internal out-migration, external in-migration, and external out-migration,

$$\mathbf{Q} = (\mathbf{n}, \mathbf{b}, \mathbf{d}, \mathbf{m}_{\text{II}}, \mathbf{m}_{\text{IO}}, \mathbf{m}_{\text{EI}}, \mathbf{m}_{\text{EO}}).$$

Each component is a four-dimensional array with dimensions region, sex, age, and time. Cell  $n_{rsat}$ , for instance, records the population of region  $r$  belonging to sex  $s$  and age group  $a$  at time  $t$ . Definitions for all seven demographic series are given in Table 2. In Table 2, and throughout the paper, when referencing events, subscript  $t$  refers to the period between exact times  $t - 1$  and  $t$ . In addition, when referencing events, subscript  $a$  refers to the age group that the person experiencing the event will belong to at time  $t$ , *not* the age group that the person belongs to at the time of the event. Figure 6 in the Supplementary Materials illustrates the distinction. The former definition is useful for demographic accounting, since it unambiguously identifies the cohort that is affected by the event (Rees 1985). Age group  $\mathcal{A}$  is open, and includes everyone aged  $\mathcal{A}$  and older.

There are several standard ways of arranging data on internal migration, the most complete of which is to record the origin and destination of each move (Rees 1985). We use a simpler 'migrant pool' model (Wilson and Bell 2004; Alho and Spencer 2005), whereby in-migrations are classified only by region of destination and out-migrations only by region of origin. The number of internal in-migrations must by definition equal the number of internal out-migrations, so we impose the constraint that, for all  $s$ ,  $a$ , and  $t$ ,

$$\sum_{r=1}^{\mathcal{R}} m_{rsat}^{\text{II}} = \sum_{r=1}^{\mathcal{R}} m_{rsat}^{\text{IO}}. \quad (2)$$

The advantage of the migrant pool model is that the number of counts rises in proportion to the number of regions  $\mathcal{R}$ , in contrast to the full origin-destination model where it rises in proportion to  $\mathcal{R}^2$ . In addition, an exploratory analysis of the New Zealand migration data suggested that migration patterns have been relatively stable over time, implying that the extra flexibility of the origin-destination model would not be needed.

The seven demographic series are linked through an accounting identity stating that population at the end of a period must equal population at the beginning of the period, plus increments, minus decrements. For  $a = 1, 2, \dots, \mathcal{A} - 1$ , the identity takes the form

$$n_{rsat} = n_{r,s,a-1,t-1} - d_{rsat} + m_{rsat}^{II} - m_{rsat}^{IO} + m_{rsat}^{EI} - m_{rsat}^{EO}. \tag{3}$$

For the first age group, births during the period replace population at the beginning of the period,

$$n_{rs0t} = \sum_a b_{rsat} - d_{rs0t} + m_{rs0t}^{II} - m_{rs0t}^{IO} + m_{rs0t}^{EI} - m_{rs0t}^{EO}. \tag{4}$$

For the oldest age group, the population at the beginning of the period includes those who have already attained age  $\mathcal{A}$ ,

$$n_{rsAt} = n_{r,s,\mathcal{A}-1,t-1} + n_{r,s,\mathcal{A},t-1} - d_{rsAt} + m_{rsAt}^{II} - m_{rsAt}^{IO} + m_{rsAt}^{EI} - m_{rsAt}^{EO}. \tag{5}$$

### 3.3 The system model

#### Overall structure

The system model  $p(\mathbf{Q}|\boldsymbol{\theta}_Q, \mathbf{Z})$  consists of seven submodels, one for each of the demographic series in  $\mathbf{Q}$ . We assume *a priori* independence for the parameters of each submodel, so that

$$\begin{aligned} \boldsymbol{\theta}_Q &= (\boldsymbol{\theta}_N, \boldsymbol{\theta}_B, \boldsymbol{\theta}_D, \boldsymbol{\theta}_{II}, \boldsymbol{\theta}_{IO}, \boldsymbol{\theta}_{EI}, \boldsymbol{\theta}_{EO}) \\ p(\boldsymbol{\theta}_Q) &= p(\boldsymbol{\theta}_N)p(\boldsymbol{\theta}_B)p(\boldsymbol{\theta}_D)p(\boldsymbol{\theta}_{II})p(\boldsymbol{\theta}_{IO})p(\boldsymbol{\theta}_{EI})p(\boldsymbol{\theta}_{EO}). \end{aligned}$$

Our model for the demographic account has the form

$$\begin{aligned} p(\mathbf{Q}|\boldsymbol{\theta}_Q, \mathbf{Z}) &= C^{-1}p(\mathbf{n}|\boldsymbol{\theta}_N, \mathbf{Z})p(\mathbf{b}|\mathbf{n}, \boldsymbol{\theta}_B)p(\mathbf{d}|\mathbf{n}, \boldsymbol{\theta}_D)p(\mathbf{m}_{II}|\mathbf{n}, \boldsymbol{\theta}_{II}, \mathbf{Z}) \\ &\quad \times p(\mathbf{m}_{IO}|\mathbf{n}, \boldsymbol{\theta}_{IO}, \mathbf{Z})p(\mathbf{m}_{EI}|\mathbf{n}, \boldsymbol{\theta}_{EI})p(\mathbf{m}_{EO}|\mathbf{n}, \boldsymbol{\theta}_{EO}) \\ &\quad \times I(\mathbf{Q}). \end{aligned} \tag{6}$$

The constant  $C$  could be obtained by integrating over the remainder of the right hand side of (6), but is never needed in practice. The indicator  $I(\mathbf{Q})$  takes a value of 1 if all counts are non-negative and the demographic accounting and internal migration constraints (2)–(5) are satisfied, and 0 otherwise. Inconsistent or impossible values for the demographic account are thus given probability 0.

Our model captures dependencies between the component series in three ways. The first is through explicitly conditioning on population counts, as is done in the submodels for

Table 2: Demographic series recorded in the demographic account

| Variable               | Notation               | Definition   |
|------------------------|------------------------|--|
| Population             | $n_{rsat}$             | The number of people in region $r$ of sex $s$ and age group $a$ at time $t$ .  |
| Births                 | $b_{rsat}$             | The number of births in region $r$ of sex $s$ to women in age group $a$ during period $t$ .                                  |
| Deaths                 | $d_{rsat}$             | The number of deaths in region $r$ of people of sex $s$ and age group $a$ during period $t$ .                                |
| Internal in-migration  | $m_{rsat}^{\text{II}}$ | The number of moves into region $r$ from elsewhere in the country by people of sex $s$ and age group $a$ during period $t$ . |
| Internal out-migration | $m_{rsat}^{\text{IO}}$ | The number of moves out of region $r$ to elsewhere in the country by people of sex $s$ and age group $a$ during period $t$ . |
| External in-migration  | $m_{rsat}^{\text{EI}}$ | The number of moves into region $r$ from overseas by people of sex $s$ and age group $a$ during period $t$ .                 |
| External out-migration | $m_{rsat}^{\text{EO}}$ | The number of moves out of region $r$ to overseas by people of sex $s$ and age group $a$ during period $t$ .                 |



births, deaths, and migration. The second is through including the covariate vector  $\mathbf{Z}$  in several submodels. (In our application we only include  $\mathbf{Z}$  in the submodels for population and internal migration, but the technique could be carried further.) The third way of capturing dependencies is through the use of demographic accounting identities, which mean that a change in any component must be offset by changes in at least one other component.

Given values for initial population, and for births, deaths, and migration, values for population in subsequent years can be derived by repeated application of the demographic accounting identities. In fact, this is what demographic projections almost always do. However, in the absence of an explicit prior model for the population, values for births, deaths, and migrations are generated without regard to the plausibility of the implied population patterns—or are subjected to ad hoc adjustments when population values are obviously implausible. In contrast, by treating population symmetrically with the other demographic series, and setting out an explicit prior model, we are able to incorporate the plausibility of population patterns into the estimates and forecasts in a transparent and systematic way.

Each of the seven submodels making up the system model has a hierarchical Poisson-gamma structure. Section 3.3 describes the specification used for population, and Section 3.3 describes how the specification is varied to accommodate differences among the remaining six series.

The Supplementary Materials contain sensitivity tests of some details of the specifications. The tests suggest that our results are robust to alternative formulations. We emphasize, however, that our general framework can accommodate many specific models of demographic processes. We do not consider the models used here to be definitive or final, and we intend to explore alternative models in future research.

**Hierarchical Poisson-gamma model for population**

To model population counts, we assume, for  $r = 1, \dots, \mathcal{R}$ ,  $s = 1, 2$ ,  $a = 0, \dots, \mathcal{A}$ , and  $t = 0, \dots, \mathcal{T}$ , that

$$n_{rsat} | \lambda_{rsat}^N \overset{\text{indep}}{\sim} \text{Poisson}(\lambda_{rsat}^N) \tag{7}$$

$$\lambda_{rsat}^N | \xi^N, \mu_{rsat}^N \overset{\text{indep}}{\sim} \text{Gamma}(\xi^N, \xi^N / \mu_{rsat}^N) \tag{8}$$

$$\log \mu_{rsat}^N = \alpha_r^{N,R} + \alpha_s^{N,S} + \alpha_a^{N,A} + \alpha_t^{N,T} + \alpha_{rs}^{N,RS} + \alpha_{ra}^{N,RA}. \tag{9}$$

The  $\overset{\text{indep}}{\sim}$  symbol in equations (7) and (8) indicates that the variates on the left hand side are drawn independently from the probability models defined on the right hand side, conditionally on the entities after the conditioning symbol. The N superscripts indicate that the parameter values are specific to the model for population. The first four terms on the right hand side of (9) are region, age, sex, and time effects. The fifth term is an interaction between region and age, and the sixth term an interaction between region and sex. The interaction terms are included to account for the distinctive age-sex

profiles of regional populations. Additional interaction terms could be added to (9) to allow for more subtle dependencies. However, (9) only specifies the *a priori* form of the expected values. With sufficient data, the posterior for the  $\lambda_{rsat}^N$  should be drawn away from the prior model specified by (9), so that the extra interaction terms are not needed. Sensitivity analyses in the Supplementary Materials show that this is indeed the case. In other applications, more interaction terms might be warranted.

We model the region effects  $\alpha_r^{N,R}$  hierarchically, with the effects assumed *a priori* to be drawn from a common normal distribution,

$$\alpha_r^{N,R} | \eta^{N,R}, \tau^{N,R} \stackrel{\text{indep}}{\sim} N(\eta^{N,R}, \tau^{N,R}), \quad r = 1, \dots, \mathcal{R}. \quad (10)$$

A flat prior is used for the parameters of the common normal distribution,

$$p(\eta^{N,R}, \tau^{N,R}) \propto 1. \quad (11)$$

Gelman (2006) demonstrates that priors for variances of the form  $p(\tau) \propto 1$  perform better than the more standard  $p(\log \tau) \propto 1$  when sample sizes are small, as they often are for group-level effects. A sensitivity test described in the Supplementary Materials suggests that the choice of prior makes little difference in our particular application. Similar hierarchically-structured priors are assumed for the age effects and sex-region interactions,

$$\alpha_a^{N,A} | \eta^{N,A}, \tau^{N,A} \stackrel{\text{indep}}{\sim} N(\eta^{N,A}, \tau^{N,A}), \quad a = 0, \dots, \mathcal{A} \quad (12)$$

$$p(\eta^{N,A}, \tau^{N,A}) \propto 1 \quad (13)$$

$$\alpha_{rs}^{N,RS} | \eta^{N,RS}, \tau^{N,RS} \stackrel{\text{indep}}{\sim} N(\eta^{N,RS}, \tau^{N,RS}), \quad r = 1, \dots, \mathcal{R}, \quad s = 1, 2 \quad (14)$$

$$p(\eta^{N,RS}, \tau^{N,RS}) \propto 1. \quad (15)$$

For time effects, we use a random walk with drift. The specification can be defined using first differences,

$$(\alpha_t^{N,T} - \alpha_{t-1}^{N,T}) | \eta^{N,T}, \tau^{N,T} \stackrel{\text{indep}}{\sim} N(\eta^{N,T}, \tau^{N,T}), \quad t = 1, \dots, \mathcal{T} \quad (16)$$

together with

$$p(\alpha_0^{N,T}) \propto 1 \quad (17)$$

$$p(\eta^{N,T}, \tau^{N,T}) \propto 1. \quad (18)$$

We use flat priors for the sex main effect.

The prior structure for region-age interactions includes the covariate  $\mathbf{Z}$ , to allow regions with large student populations to have distinctive age profiles,

$$\alpha_{ra}^{N,RA} | \eta_0^{N,RA}, \eta_1^{N,RA}, \tau^{N,RA} \stackrel{\text{indep}}{\sim} N(\eta_0^{N,RA} + \eta_1^{N,RA} Z_{ra}, \tau^{N,RA}),$$

$$r = 1, \dots, \mathcal{R}, \quad a = 0, \dots, \mathcal{A} \tag{19}$$

with

$$p(\eta_0^{N,RA}, \eta_1^{N,RA}, \tau^{N,RA}) \propto 1. \tag{20}$$

Finally, we assume

$$p(\xi^N) = \frac{\mu_0^N}{(\mu_0^N + \xi^N)^2}, \tag{21}$$

which is a uniform shrinkage prior (Christiansen and Morris 1997; Daniels 1999). The parameter  $\mu_0^N$  is the prior median of  $\xi^N$ . Larger values of  $\mu_0^N$  encourage greater shrinkage towards the prior model (9). Strategies for choosing  $\mu_0^N$  are discussed in Daniels (1999) and Young et al. (2006). For all analyses reported in this paper we use  $\mu_0^N = 10$ , implying that we place equal weight on the prior model and on the direct estimate for a cell when that cell has size 10. In practice, our results are highly robust to alternative choices of  $\mu_0^N$ . Further detail on the uniform shrinkage prior and on the sensitivity test can be found in the Supplementary Materials.

### Hierarchical Poisson-gamma model for births, deaths, and migration

The principal difference between the models for births, deaths, and migration and the model for population is that (7) is replaced by a model that includes an offset  $o_{rsat}$ ,

$$b_{rsat} | \lambda_{rsat}^B \overset{\text{indep}}{\sim} \text{Poisson}(\lambda_{rsat}^B o_{rsat}) \tag{22}$$

$$d_{rsat} | \lambda_{rsat}^D \overset{\text{indep}}{\sim} \text{Poisson}(\lambda_{rsat}^D o_{rsat}) \tag{23}$$

$$m_{rsat}^{II} | \lambda_{rsat}^{II} \overset{\text{indep}}{\sim} \text{Poisson}(\lambda_{rsat}^{II} o_{rsat}) \tag{24}$$

$$m_{rsat}^{IO} | \lambda_{rsat}^{IO} \overset{\text{indep}}{\sim} \text{Poisson}(\lambda_{rsat}^{IO} o_{rsat}) \tag{25}$$

$$m_{rsat}^{EI} | \lambda_{rsat}^{EI} \overset{\text{indep}}{\sim} \text{Poisson}(\lambda_{rsat}^{EI} o_{rsat}) \tag{26}$$

$$m_{rsat}^{EO} | \lambda_{rsat}^{EO} \overset{\text{indep}}{\sim} \text{Poisson}(\lambda_{rsat}^{EO} o_{rsat}) \tag{27}$$

where

$$o_{rsat} = \frac{1}{2}(n_{r,s,a-1,t-1} + n_{rsat}) + \epsilon, \quad a = 1, \dots, \mathcal{A} - 1 \tag{28}$$

$$o_{rs0t} = \frac{1}{2}n_{rs0t} + \epsilon \tag{29}$$

$$o_{rsAt} = \frac{1}{2}(n_{r,s,\mathcal{A}-1,t-1} + n_{r,s,\mathcal{A},t-1} + n_{rsAt}) + \epsilon, \tag{30}$$

and  $\epsilon$  is a small positive constant that is included to avoid expected values being set to zero.

The justification for the inclusion of the  $o_{rsat}$  term is most straightforward in the case of the models for deaths (23), internal out-migration (25), and external out-migration (27). In the case of deaths,  $o_{rsat}$  measures person-years of exposure to the risk of dying, in the case of out-migration it measures exposure to the risk of out-migrating.

No such exposure-to-risk interpretation is possible for the offset terms in the models for internal in-migration (24) or external in-migration (26). The true population at risk of internal in-migration to a given region is the population in all other regions of New Zealand, and the population at risk of external in-migration is the population in the rest of the world. The former varies little from region to region, and the latter is constant, so neither has any predictive power. However, exploratory analyses showed that the inclusion of offsets identical to those used in the out-migration models increased the precision of the in-migration models almost as much as it increased the precision of the out-migration models. We therefore included the offsets in the in-migration models despite their less clear interpretation.

The model for births (22) states that the expected number of births of sex  $s$  in region  $r$  during period  $t$  is proportional to the number of person-years lived by females aged  $a$  in region  $r$  during period  $t$  (using  $s = 1$  to denote females). In accordance with standard ‘female-dominant’ model of fertility (Preston et al. 2001, p120), the number of person-years lived by males does not affect the expected number of births.

The hyper-prior structures for internal in-migration and internal out-migration take exactly the same form as the structure for population. The hyper-prior structures for births, deaths, and external migration take the same form, except that the models for region-age effects do not include covariate  $\mathbf{Z}$ . Instead, region-age effects are drawn from a common normal distribution, equivalent to that used for region-sex effects in the model for population (14).

### 3.4 The observation model

The observation model  $p(\mathbf{X}|\mathbf{Q}, \boldsymbol{\theta}_{\mathbf{X}})$  consists of  $K$  submodels, each relating a dataset  $\mathbf{X}_k$  to an associated demographic series. The demographic series associated with dataset  $k$  is denoted  $\mathbf{Q}_{[k]}$ . Multiple datasets can be associated with the same demographic series. For instance, of the 12 datasets used in our application, three are associated with population (Table 1). We assume that the submodels are distinct in the sense that

$$\begin{aligned}\boldsymbol{\theta}_{\mathbf{X}} &= (\boldsymbol{\theta}_{\mathbf{X}_1}, \boldsymbol{\theta}_{\mathbf{X}_2}, \dots, \boldsymbol{\theta}_{\mathbf{X}_K}) \\ p(\boldsymbol{\theta}_{\mathbf{X}}) &= \prod_{k=1}^K p(\boldsymbol{\theta}_{\mathbf{X}_k}) \\ p(\mathbf{X}|\mathbf{Q}, \boldsymbol{\theta}_{\mathbf{X}}) &= \prod_{k=1}^K p(\mathbf{X}_k|\mathbf{Q}_{[k]}, \boldsymbol{\theta}_{\mathbf{X}_k}).\end{aligned}$$

Dataset  $\mathbf{X}_k$  is composed of counts cross-classified by one or more of the dimensions used by  $\mathbf{Q}$ , that is, region, sex, age, and time. The school roll data, for instance, consist of counts of people enrolled in school, classified by region of school, sex, and time, for ages 5 to 22. For each dataset  $\mathbf{X}_k$ , we construct a predictor  $\mathbf{q}_k$  that has the same dimensions as  $\mathbf{X}_k$ . The predictor is constructed by subsetting and aggregating values from the associated demographic series  $\mathbf{Q}_{[k]}$ . Let  $x_{jk}$  denote the  $j$ th cell of  $\mathbf{X}_k$ , and  $q_{jk}$  the  $j$ th

cell of  $\mathbf{q}_k$ . Then

$$q_{jk} = \sum_{r,s,a,t} Q_{rsat}^{[k]} I_{rsat}^{jk} + \epsilon \tag{31}$$

where  $I_{rsat}^{jk}$  equals 1 if cell  $x_{jk}$  falls within region  $r$ , sex  $s$ , age group  $a$ , or time  $t$ , and 0 otherwise. As will become apparent, the  $\epsilon > 0$  is necessary to allow for the possibility that the dataset takes a positive value when the corresponding cells from  $\mathbf{Q}_{[k]}$  are all 0.

We assume that the  $x_{jk}$  are independent conditional on  $\mathbf{q}_k$  and  $\boldsymbol{\theta}_{\mathbf{X}_k}$ , so that submodel  $k$  can be written as

$$p(\mathbf{X}_k | \mathbf{Q}_{[k]}, \boldsymbol{\theta}_{\mathbf{X}_k}) = p(\mathbf{X}_k | \mathbf{q}_k, \boldsymbol{\theta}_{\mathbf{X}_k}) = \prod_j p(x_{jk} | q_{jk}, \theta_{jk}),$$

where  $\theta_{jk}$  is the  $j$ th cell of  $\boldsymbol{\theta}_{\mathbf{X}_k}$ . We vary the specification for  $p(x_{jk} | q_{jk}, \theta_{jk})$ , and for any priors for  $\theta_{jk}$ , to reflect what is known about the relationship between the dataset and the associated demographic series. Registered births should be close to actual births, for instance, since birth registration is rigorously enforced in New Zealand. In contrast, the number of people on the school roll for a region should be close to number of people in that region at ages 5 to 15, but lower at subsequent ages, reflecting the fact that New Zealand children begin school at age 5 and start to leave from age 15.

Our baseline specification is

$$x_{jk} | q_{jk}, \theta_{jk} \overset{\text{indep}}{\sim} \text{Poisson}(\theta_{jk} q_{jk}). \tag{32}$$

The  $\theta_{jk}$  term can usually be interpreted as a measure of coverage, with values greater than 1 indicating overcount and values less than 1 indicating undercount. In the case of the model for the school roll, for instance, we would expect  $\theta_{jk}$  to be less than 1 for ages 16 to 22.

The hyperprior structures for census migration counts, the school roll, the electoral roll, and tax data are similar to those for demographic series. The main difference is that the models for cell means are simpler than (9). In the case of census migration counts and the school roll, the model consists of a single age effect,

$$\log \mu_{rsat} = \alpha_a^A.$$

For each of these datasets there is no reason to expect marked differences in coverage by region, sex, or time. The model for the electoral roll is identical, except that  $\mu_{rat}$  replaces  $\mu_{rsat}$  since the electoral roll data do not include information on sex. The model for tax data is

$$\log \mu_{rsat} = \alpha_a^A + \alpha_s^S,$$

reflecting the possibility that sex differences in employment rates lead to sex differences in coverage.

In the case of the international arrivals and departures, we assume that  $\theta_{jk} \sim \text{Gamma}(100, 100)$ , which implies a mean of 1 and variance of 0.01. This reflects

our belief that the arrivals and departures data are likely to be unbiased and accurate. The specification does have some flexibility, however, and permits coverage rates that depart from 1 given sufficiently strong evidence.

Using a Poisson distribution for the first stage of a model for  $x_{ij}$  means that the variance of  $x_{jk}$  cannot be lower than  $\theta_{jk}q_{jk}$ . This implies, for instance, that in a model for registered births where  $\theta_{jk} = 1$ , and  $q_{jk} = 5$ , the probability that the registration data are correct to within one birth is only about 0.5. This understates the accuracy of a dataset such as registered births. To model such datasets, a distribution is needed that is under-dispersed relative to the Poisson.

We model highly accurate datasets using a Poisson-binomial mixture,

$$x_{jk} = g_{jk} + h_{jk}$$

where  $g_{jk} \sim \text{Poisson}(q_{jk}(1 - p_k))$  and  $h_{jk} \sim \text{Binomial}(q_{jk}, p_k)$ . Parameter  $p_k$  can be interpreted as the probability that a person or event is detected and appropriately enumerated by the dataset. The  $h_{jk}$  term is then the number of people or events that are correctly included in cell  $j$ . The  $g_{jk}$  term measures ‘overcount’, the number of people or events that are double-counted or that are incorrectly included in cell  $j$ , which is assumed to be proportional to true cell count. The expected value of the Poisson-Binomial mixture is  $q_{jk}$ , and the variance is  $q_{jk}(1 - p_k^2)$ . When  $p_k$  approaches 1, as it does with an accurate dataset, the variance is substantially less than the mean. We use the Poisson-binomial mixture to model registered births, registered deaths, and official census-year population counts. For all three datasets we set  $p_k = 0.98$ , based on discussions with Statistics New Zealand staff about likely levels of accuracy.

### 3.5 Posterior calculation

As stated in Section 2, our approach to simulating from the joint posterior distribution  $p(\mathbf{Q}, \boldsymbol{\theta}_{\mathbf{Q}}, \boldsymbol{\theta}_{\mathbf{X}} | \mathbf{X}, \mathbf{Z})$  is to use a Gibbs sampler that alternates between updates of  $\mathbf{Q}$ ,  $\boldsymbol{\theta}_{\mathbf{Q}}$ , and  $\boldsymbol{\theta}_{\mathbf{X}}$ . Updating  $\boldsymbol{\theta}_{\mathbf{Q}}$  and  $\boldsymbol{\theta}_{\mathbf{X}}$  can be done using standard methods for hierarchical models, with Metropolis Hastings updates applied in cases where it is not possible to draw directly from a full conditional distribution (Gelman et al. 2004). Updating  $\mathbf{Q}$ , however, is more challenging. Not only is  $\mathbf{Q}$  very large, but, because of constraints (2)–(5), entries within it do not follow standard distributions. Our approach is to use Metropolis-Hastings updates that are highly customized for our application.

The first step is to generate a proposal  $\mathbf{Q}^*$ . The proposal must satisfy the accounting identities, but also be sufficiently close to the current value  $\mathbf{Q}^i$  that it has a reasonable chance of being accepted. Early versions of our algorithm attempted to change hundreds of cells per update and had acceptance rates well below 1%. Our current version attempts to change only about 10–20 cells per update but has acceptance rates of 40–50%.

To describe the procedure for generating proposals, it is helpful to extend the subscript notation to provide a compact way of representing changes within cohorts. If  $u_{rsat}$  is a

generic count or parameter such as  $b_{rsat}$  or  $\lambda_{rsat}^D$ , then

$$u_{rsatv} = u_{r,s,\min(a+v,A),t+v}. \tag{33}$$

We also need the notion of an ‘expected offset’, which is defined in the same way as an actual offset ((29)-(30)), but with expected population at the end of the period  $\lambda_{rsat}^N$  taking the place of actual population at the end of the period  $n_{rsat}$ .

It is simplest to begin with proposals that do not include changes to internal migration. The process for generating such proposals is as follows. At each iteration  $i$ ,

- i. Randomly select a series to update, where the choices are initial population (i.e. population at time 0), births, deaths, external in-migration, and external out-migration.
- ii. Randomly select a cell from within that series—that is, select a region  $r$ , sex  $s$ , age group  $a$ , and (except in the case of initial population) period  $t$ .
- iii. Draw a proposed value for the cell using the Poisson models (7), (22), (23), (26), or (27), but with expected offset  $\hat{o}_{rsat}^i$  taking the place of actual offset  $o_{rsat}^i$  when the series being updated is deaths, external in-migration, or external out-migration.
- iv. Revise subsequent population counts to be consistent with the proposed value. If the series chosen for updating is initial population, the subsequent population counts are  $n_{rsa0u}^*$ ,  $u = 1, 2, \dots, \mathcal{T}$ . If births, the population counts are  $n_{rs1tu}^*$ ,  $u = 0, \dots, \mathcal{T} - t$ . If deaths or external migration, the population counts are  $n_{rsatu}^*$ ,  $u = 0, 1, \dots, \mathcal{T} - t$ .
- v. Check that the subsequent population contains no negative values. If they do, return to step 3.

A simplified example of the application of this algorithm is given in the Supplementary Materials.

Using expected offsets to generate deaths and migration in step 3 is simpler than using actual offsets because actual offsets depend on the number of deaths or external migrations during the period. Using expected offsets is not necessary in the case of births, since a change to the number of births affects the size of the cohort being born, not the size of the cohort having the births.

Generating proposals that involve changes to internal migration is more complicated, because internal migration is subject to constraint (2), stating that total in-migrations equal total out-migrations. An obvious way to satisfy this constraint would be to match any change to in-migration into a region with an equivalent change to out-migration out of that region. However, this would mean that net migration to that region would remain forever at its initial value. To allow net migration to vary freely, while ensuring that (2) is satisfied, we generate proposals that change in-migration and out-migration in two regions simultaneously, under the constraint that

$$\sum_{r=r_1,r_2} m_{rsat}^{II*} - \sum_{r=r_1,r_2} m_{rsat}^{II,i} = \sum_{r=r_1,r_2} m_{rsat}^{IO*} - \sum_{r=r_1,r_2} m_{rsat}^{IO,i}. \tag{34}$$

Once values for three of the migration flows have been drawn from Poisson models (24) and (25), a value for the fourth, residual, flow can be obtained from constraint (34). The algorithm uses the flow with the highest expected value as the residual flow, with expected value calculated using expected offsets. If this flow takes a negative value, new draws are made for the other three flows. The remainder of the algorithm for internal migration is identical to that for the other series, except that subsequent population counts are changed in two regions rather than one.

Having generated a proposal, the next step is to calculate the Metropolis-Hastings ratio  $r_{\text{MH}}$ . Using  $J$  to denote the proposal density,  $r_{\text{MH}}$  can be calculated as follows,

$$\begin{aligned} r_{\text{MH}} &= \frac{p(\mathbf{Q}^* | \boldsymbol{\theta}_{\mathbf{Q}}^i, \boldsymbol{\theta}_{\mathbf{X}}^i, \mathbf{X}, \mathbf{Z}) J(\mathbf{Q}^i | \mathbf{Q}^*, \boldsymbol{\theta}_{\mathbf{Q}}^i)}{p(\mathbf{Q}^i | \boldsymbol{\theta}_{\mathbf{Q}}^i, \boldsymbol{\theta}_{\mathbf{X}}^i, \mathbf{X}, \mathbf{Z}) J(\mathbf{Q}^* | \mathbf{Q}^i, \boldsymbol{\theta}_{\mathbf{Q}}^i)} \\ &= \frac{p(\mathbf{X} | \mathbf{Q}^*, \boldsymbol{\theta}_{\mathbf{X}}^i) p(\mathbf{Q}^* | \boldsymbol{\theta}_{\mathbf{Q}}^i) J(\mathbf{Q}^i | \mathbf{Q}^*, \boldsymbol{\theta}_{\mathbf{Q}}^i)}{p(\mathbf{X} | \mathbf{Q}^i, \boldsymbol{\theta}_{\mathbf{X}}^i) p(\mathbf{Q}^i | \boldsymbol{\theta}_{\mathbf{Q}}^i) J(\mathbf{Q}^* | \mathbf{Q}^i, \boldsymbol{\theta}_{\mathbf{Q}}^i)}. \end{aligned} \quad (35)$$

The third term on the right of (35) is the probability of proposing  $\mathbf{Q}^i$  starting from  $\mathbf{Q}^*$ , divided by the probability of proposing  $\mathbf{Q}^*$  starting from  $\mathbf{Q}^i$ . A complete proposal density would include the probability that a cell is chosen for updating. However, this probability does not depend on the starting point, so it cancels out when the ratio is taken. The range of valid values for the cell is constrained by the requirement that subsequent population counts be non-negative. A complete proposal density would include a normalizing constant to reflect this constraint, calculated by integrating over the valid range. However, the range of valid values for the cell is identical for  $\mathbf{Q}^i$  and  $\mathbf{Q}^*$ , so a normalizing constant would again cancel out. Subsequent population counts do not appear in  $J$  since they depend deterministically on the value of the cell chosen for updating.

When the initial population is updated, the third term on the right of (35) equals  $\frac{p(n_{rsa0}^i | \lambda_{rsa0}^{N,i})}{p(n_{rsa0}^* | \lambda_{rsa0}^{N,i})}$ , with  $p$  denoting the Poisson distribution. When births are updated, the term equals  $\frac{p(b_{rsat}^i | o_{r1at}^i \lambda_{rsat}^{B,i})}{p(b_{rsat}^* | o_{r1at}^i \lambda_{rsat}^{B,i})}$ , and when deaths are updated, it equals  $\frac{p(d_{rsat}^i | \hat{o}_{rsat}^i \lambda_{rsat}^{D,i})}{p(d_{rsat}^* | \hat{o}_{rsat}^i \lambda_{rsat}^{D,i})}$ . The expressions for external migration have the same form as the one for deaths. The expression for internal migration contains terms for the three flows in (34) that are randomly generated. Each of these terms has the same form as the one for deaths. No term is needed for the fourth flow in (34) since it is obtained deterministically from the other three.

The second term on the right of (35) measures the effect on  $p(\mathbf{Q} | \mathbf{X})$  of moving from  $\mathbf{Q}^i$  to  $\mathbf{Q}^*$ . Using  $\mathbf{n}_0$  to denote population at  $t = 0$  and  $\mathbf{n}_{1+}$  to denote population at  $t = 1, \dots, \mathcal{T}$ , the term can be decomposed as follows,



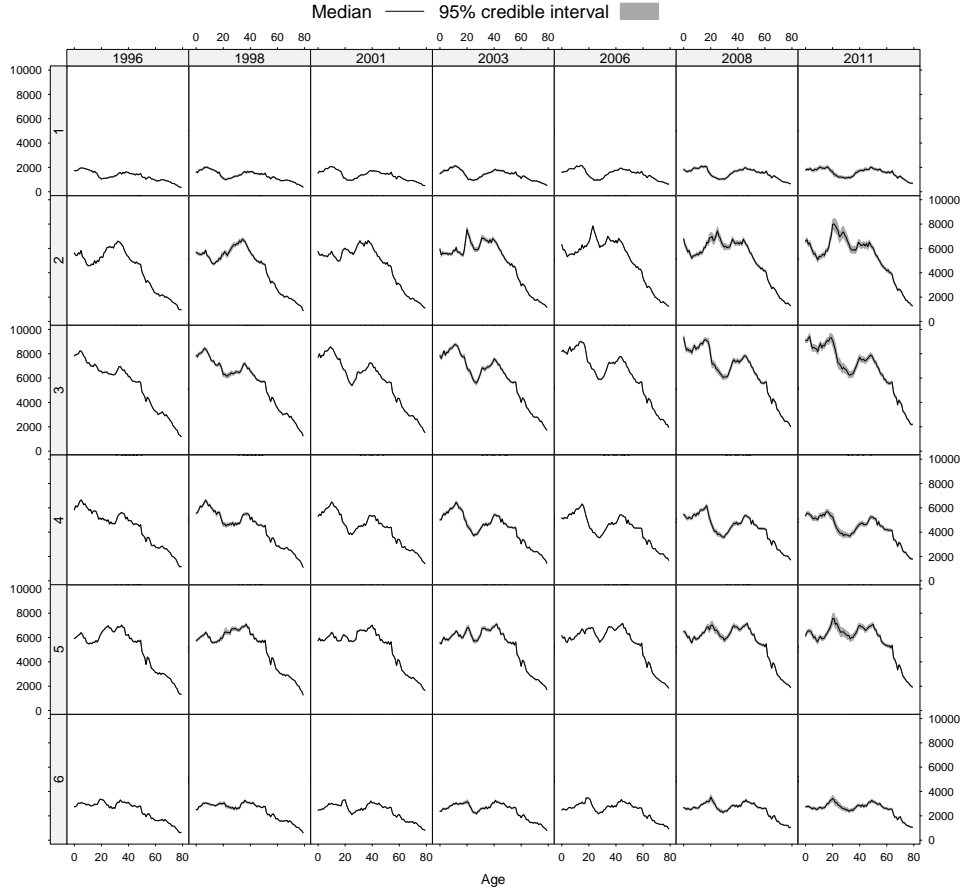


Figure 2: Estimated population counts, by region, age, and year, for males. The top left panel, for instance, shows the number of males, by age, in region 1 in 2001.

$$\begin{aligned}
 \frac{p(Q^*|\theta_Q^i)}{p(Q^i|\theta_Q^i)} &= \frac{p(n^*|\theta_Q^i)p(b^*, d^*, m^*|n^*, \theta_Q^i)}{p(n^i|\theta_Q^i)p(b^i, d^i, m^i|n^i, \theta_Q^i)} \\
 &= \frac{p(n_0^*, n_{1+}^*|\theta_Q^i)p(b^*, d^*, m^*|n^*, \theta_Q^i)p(b^i, d^i, m^i|n^*, \theta_Q^i)}{p(n_0^i, n_{1+}^i|\theta_Q^i)p(b^i, d^i, m^i|n^i, \theta_Q^i)p(b^i, d^i, m^i|n^*, \theta_Q^i)} \\
 &= \frac{p(n_0^*|\theta_Q^i)p(b^*, d^*, m^*|n^*, \theta_Q^i)p(n_{1+}^*|\theta_Q^i)}{p(n_0^i|\theta_Q^i)p(b^i, d^i, m^i|n^*, \theta_Q^i)p(n_{1+}^i|\theta_Q^i)} \\
 &\quad \times \frac{p(b^i, d^i, m^i|n^*, \theta_Q^i)}{p(b^i, d^i, m^i|n^i, \theta_Q^i)}. \tag{36}
 \end{aligned}$$

The first two terms on the right of (36) together represent the effect on  $p(\mathbf{Q}|\mathbf{X})$  of the change in initial population, births, deaths, or migration, holding all else constant, including subsequent population. Changes in subsequent population affect  $p(\mathbf{Q}|\mathbf{X})$  in two ways. The first effect occurs via changes to  $\mathbf{Q}$  itself. This effect is measured by the third term in (36). The second effect occurs via changes to offsets, and hence expected numbers of births, deaths, and migrations. This effect is measured by the fourth term in (36). For instance, this term would be greater than 1 if, at iteration  $i$ , a cohort was experiencing more deaths and migrations than would be expected for a cohort of that size, and the proposal implied an increase in the cohort's size. When the terms in (36) are calculated, almost all the elements cancel out, since few cells differ between  $\mathbf{Q}^*$  and  $\mathbf{Q}^i$ . Formulas are provided in the Supplementary Materials.

Finally, the first term on the right of (35) measures the effect on  $p(\mathbf{X}|\mathbf{Q}, \boldsymbol{\theta}_{\mathbf{X}})$  of moving from  $\mathbf{Q}^i$  to  $\mathbf{Q}^*$ . We have

$$\begin{aligned} \frac{p(\mathbf{X}|\mathbf{Q}^*, \boldsymbol{\theta}_{\mathbf{X}}^i)}{p(\mathbf{X}|\mathbf{Q}^i, \boldsymbol{\theta}_{\mathbf{X}}^i)} &= \prod_k \frac{p(\mathbf{X}_k|\mathbf{Q}_{[k]}^*, \boldsymbol{\theta}_{\mathbf{X}_k}^i)}{p(\mathbf{X}_k|\mathbf{Q}_{[k]}^i, \boldsymbol{\theta}_{\mathbf{X}_k}^i)} \\ &= \prod_{j,k} \frac{p(x_{jk}|q_{jk}^*, \theta_{jk}^i)}{p(x_{jk}|q_{jk}^i, \theta_{jk}^i)} \\ &= \prod_{q_{jk}^* \neq q_{jk}^i} \frac{p(x_{jk}|q_{jk}^*, \theta_{jk}^i)}{p(x_{jk}|q_{jk}^i, \theta_{jk}^i)}. \end{aligned} \quad (37)$$

Because there are few differences between  $\mathbf{Q}^*$  and  $\mathbf{Q}^i$ , there are few cases where  $q_{jk}^* \neq q_{jk}^i$ , and few terms appearing in (37). Moreover, the procedure for obtaining  $q_{jk}$  from  $\mathbf{Q}$ , set out in (31), is designed to make it easy to identify cases where  $q_{jk}^* \neq q_{jk}^i$ . Calculating (37) is therefore simple.

Once the calculation of  $r_{\text{MH}}$  is complete, proposal  $\mathbf{Q}^*$  is adopted with probability  $\min(r_{\text{MH}}, 1)$ .

The current version of our software, written in a mixture of R and C, carries out about 100,000 updates of  $\mathbf{Q}$  per minute on a modest desktop computer. The results for six regions shown in this paper come from simulation consisting of two chains, each of which involves 200 million updates of  $\mathbf{Q}$  and 20,000 updates of  $\boldsymbol{\theta}_{\mathbf{Q}}$  and  $\boldsymbol{\theta}_{\mathbf{X}}$ , preceded by a burn-in of the same length. Convergence was assessed using potential scale reduction factors for randomly-chosen cells within  $\mathbf{Q}$  (Gelman and Rubin 1992).

### 3.6 Results

We present a small selection from the voluminous output produced by the model. The first set of results, shown in Figure 2, are population estimates and forecasts, for males, in all six regions, in selected years. Each regional population has a distinctive age profile, which evolves over time. The grey bands in the figure denote 95% credible intervals. The bands are widest for ages 15 to 39, reflecting the uncertainty created by these ages'

high migration rates. The bands are too narrow to be visible during the census years of 2001 and 2006, but widen from 2006 onwards. (A census was due to have been held in 2011 but was cancelled because of an earthquake.)

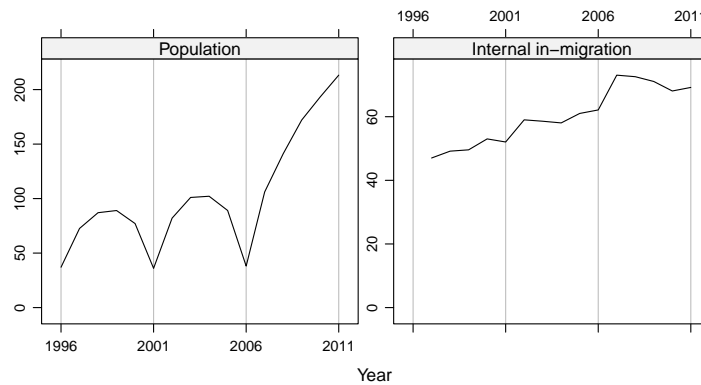


Figure 3: Changes over time in the widths of credible intervals for population and internal in-migration. The panels show the median width of region-sex-age-specific credible intervals in each year. The vertical scales differ between the two panels.

Figure 3 provides more detail on how uncertainty changes over time. As can be seen in the left panel, uncertainty about population counts is low throughout the 1996-2001 and 2001-2006 periods, when estimates are anchored by a census at each end. Uncertainty rises steadily after 2006.

Figure 4 shows estimates and forecasts for internal in-migration, that is, annual migrations into each region from elsewhere in New Zealand. The credible intervals for internal migration are much wider, in relative terms, than the credible intervals for population. Internal migration estimates are less certain because the data available for estimating migration are less complete and less reliable: in particular, none provides accurate annual counts for all age groups.

As can be seen in Figure 3, uncertainty about internal in-migration follows a different trajectory from uncertainty about population. Uncertainty about internal in-migration does not fall to low levels in census years because the census question on residence 5 years earlier does not provide definitive information on annual migration counts. Instead, uncertainty trends gradually upwards over time. We suspect that the slight fall between 2007 and 2010 reflects the fact that, for the years 2010 and 2011, highly detailed data on changes of addresses of taxpayers were available (see Table 1).

Finally, Figure 5 shows output from the submodel relating the number of people on the school roll to the number of people in that age group. As expected, coverage rates are generally about 1 from ages 5 to 15, and fall in subsequent ages. Regions 1 and 2 are an interesting exception: coverage rates are lower than expected in region 1 and higher than expected in region 2. Region 1 is next to region 2, which contains Auckland, the

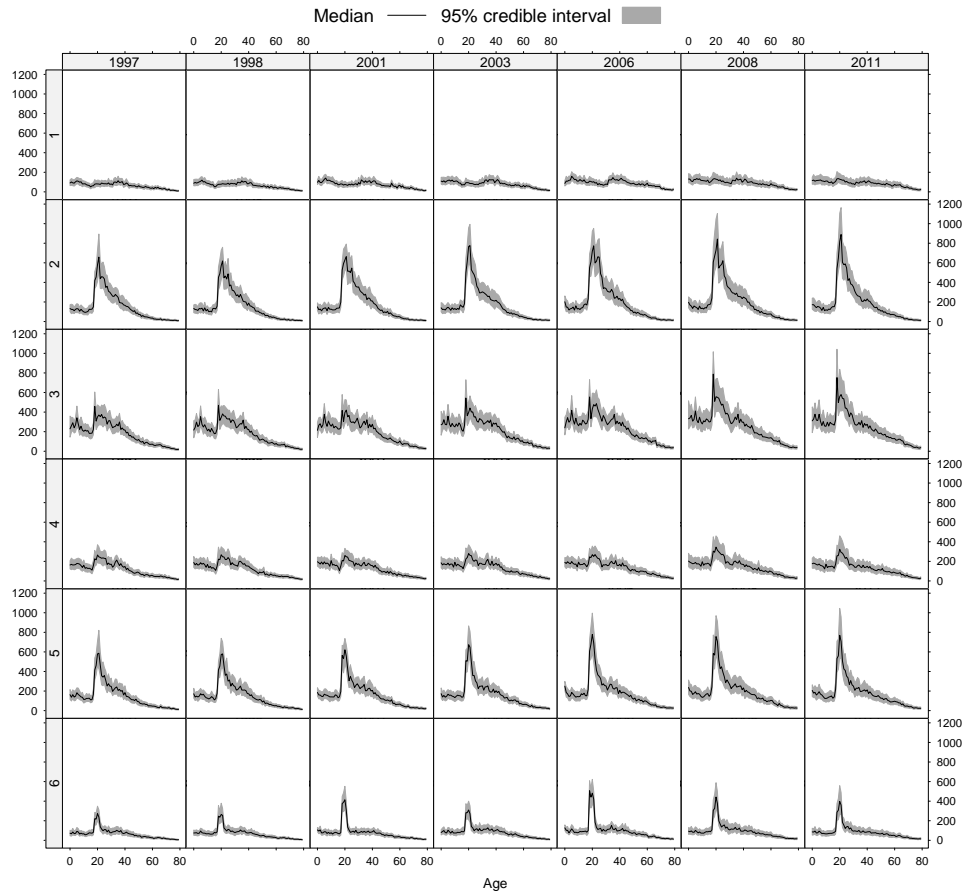


Figure 4: Estimated internal in-migration by age, males, 2002-2011.

largest city in New Zealand. The most likely explanation for the anomalous coverage rates is that children travel into region 2 to study, though the possibility of problems with the data or the model also needs to be considered.

## 4 Discussion

We have shown how subnational population estimation, including the evaluation and synthesis of multiple data sources, can be placed in a Bayesian statistical framework. Expressing the problem in statistical terms permits greater transparency and efficiency, as well as more satisfactory treatment of uncertainty. It also leads naturally to ideas about possible extensions.

One such extension is to apply the framework to forecasting future values of demographic

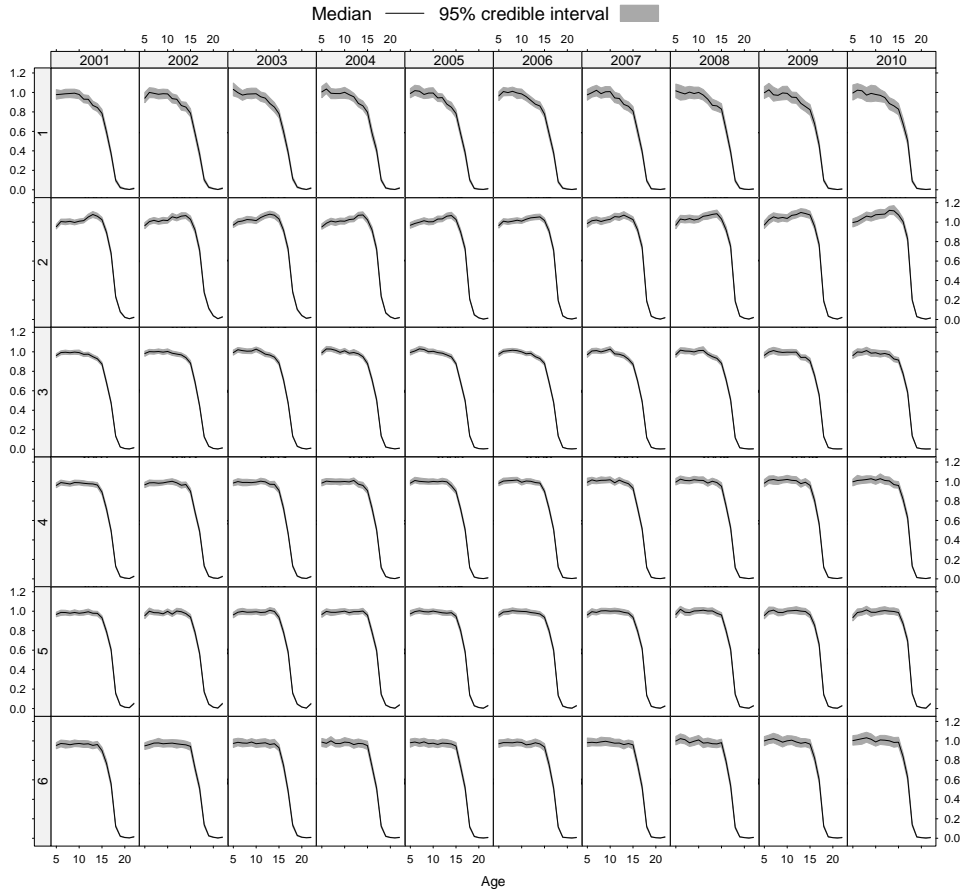


Figure 5: Numbers of people on the school roll for each person estimated to live in the region, by age and year, females.

series. Forecasts can be obtained by defining  $Q$  to include future periods and then letting the Gibbs sampler generate values for both historical and future periods. Data series do not extend into the future (unless lagged values are used), but from the point of view of the model, this is just another type of missingness. Because the system model does not distinguish between past and future demographic values, both can be governed by the same set of demographic assumptions. Moreover, uncertainty about historical values can be carried through into forecast values.

Another extension is to develop special-purpose priors for the system and observation models. Demographers have, for instance, developed parsimonious models for representing change over time in age profiles for fertility, mortality, and migration, which could be used to construct prior distributions (Preston et al. 2001; Raymer and Rogers

2006; Booth and Tickle 2008). If data from a sample survey such as a labor force survey were used, then the priors for the data model could include information about the design of the survey. In models with many small geographical units, precision could be increased by using priors that recognized similarity across space, and exploited spatial correlations in demographic rates (Congdon 2010, chapter 4).

Covariates could be added to the observation model, in the same way that they are currently added to the system model. Using age-specific employment rates, for instance, might predict how coverage by tax data varied across age group. In some cases, covariates in the system or observation models might be included to investigate substantive questions, and not just for their ability to improve the precision of the population estimates.

A final way of extending the framework would be to expand the choice of dimensions beyond that of region, sex, age, and time. Disaggregating by educational status, for instance, might be an effective way to account for regional differences in migration patterns, as well as having substantive interest in its own right. More generally, the framework could be applied to moves that were entirely non-geographical, such as changes in labor force status or health status (Schoen 1988; Rogers 1995). This would allow the framework to be applied to a large class of estimation problems beyond that of subnational population estimation, such as labor force projections, or measurement of trends in population health. The common element would be production of internally-consistent demographic estimates from multiple imperfect datasets.

Before our proposed methods could be used for our original application, the production of official subnational population statistics, they would need to be tested on larger datasets, benchmarked against traditional methods, and perhaps refined further. Fienberg (2011) and Little (2012) observe that, even when Bayesian methods perform well, official statistical agencies are often reluctant to adopt them, because of concerns that the use of priors compromises objectivity. Whether or not these concerns are justified in general, we think that they are less relevant to the case of subnational population estimation. As discussed in the Introduction, current subnational population estimation methods rely heavily on expert judgment and ad hoc adjustments. The methods proposed in this paper offer the possibility of reducing this reliance by making it easier to exploit multiple data sources and hence bring more data to bear. Moreover, when expert judgment is required, such as when specifying priors or selecting models for data sources, analysts using our methods would specify their judgements explicitly, in the form of computer code. This would provide a far clearer audit trail than is provided by current population estimation methods.

## Supplementary material

### Allocating events to age groups

As discussed in Section 3.2, we allocate events to age groups based on age at the end of the period, not age at the time of the event. Figure 6, a Lexis diagram (Preston et al. 2001, pp. 31-34), illustrates the distinction.

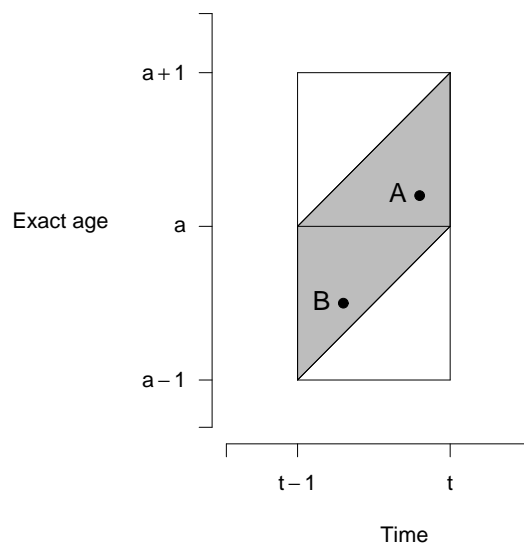


Figure 6: Allocating events to age groups. The shaded parallelogram shows our definition of age group  $a$ . Our definition and the standard age-at-event definition both imply that event A should be allocated to age group  $a$ . However, our definition implies that event B should also be allocated to age group  $a$  while the age-at-event definition implies that it should be allocated to age group  $a - 1$ .

### Further details on the uniform shrinkage prior

Under the hierarchical Poisson Gamma model given by (7)–(9), and treating  $n_{rsat}$  as directly observed, the cell-specific Poisson parameters follow a Gamma distribution with expectation

$$E[\lambda_{rsat} | n_{rsat}, \alpha, \xi] = W_{rsat} \mu_{rsat} + (1 - W_{rsat}) n_{rsat}$$

where

$$W_{rsat} = \frac{\xi}{\xi + \mu_{rsat}}$$

(See, for example, Christiansen and Morris (1997, Equation 9)).

We refer to the weight on the prior mean  $W_{rsat}$  as the shrinkage parameter for the *rsat* cell. For the purposes of deriving a prior for  $\xi$ , the cell-specific prior mean can be replaced by a constant  $\mu_0$  to give

$$W_0 = \frac{\xi}{\xi + \mu_0},$$

which can be viewed as the shrinkage parameter for a hypothetical cell with prior mean  $\mu_0$ . A uniform prior on  $[0, 1]$  for  $W_0$  implies

$$p(\xi) = \frac{\mu_0}{(\mu_0 + \xi)^2},$$

which has median  $\mu_0$  (Christiansen and Morris 1997).

In our framework, the  $n_{rsat}$  are not directly observed. Instead, the observed data  $\mathbf{X}$  contain coarsened versions of demographic variables. Consequently, the conditional posterior for the cell-specific Poisson parameters is

$$\begin{aligned} E[\lambda_{rsat} | \mathbf{X}, \boldsymbol{\alpha}, \xi] &= E[E[\lambda_{rsat} | n_{rsat}, \mathbf{X}\boldsymbol{\alpha}, \xi]] \\ &= E[W_{rsat}\mu_{rsat} + (1 - W_{rsat})n_{rsat} | \mathbf{X}, \boldsymbol{\alpha}, \xi] \\ &= W_{rsat}\mu_{rsat} + (1 - W_{rsat})E[n_{rsat} | \mathbf{X}, \boldsymbol{\alpha}, \xi]. \end{aligned}$$

### Sensitivity tests

We carried out the following sensitivity tests:

- i. *All interactions.* In all models for demographic series we used the complete set of second-order and third-order interactions between region, age, and sex, rather than just interactions between region and age, and region and sex.
- ii. *Alternative variance prior.* In all hierarchical Poisson-gamma models, including those for demographic series and for data sources, we used a prior for the top-level variance (e.g. (11)) of  $p(\log \tau) \propto 1$ , rather than  $p(\tau) \propto 1$ .
- iii. *Uniform shrinkage median 5 and 20.* In all hierarchical Poisson-gamma models, those for demographic series and for data sources, we used medians of 5 and 20, rather than 10, in the uniform shrinkage prior for  $\xi$ .

The results of the sensitivity tests are summarized in Figure 7. The tests use a burnin of 10,000 iterations and production run of 10,000 iterations. This is only half the lengths used for the results in the main body of the paper, and the tests are subject to some Monte Carlo error. Nevertheless, the clear message of the figure is that our results are robust to the alternative specifications that were investigated.



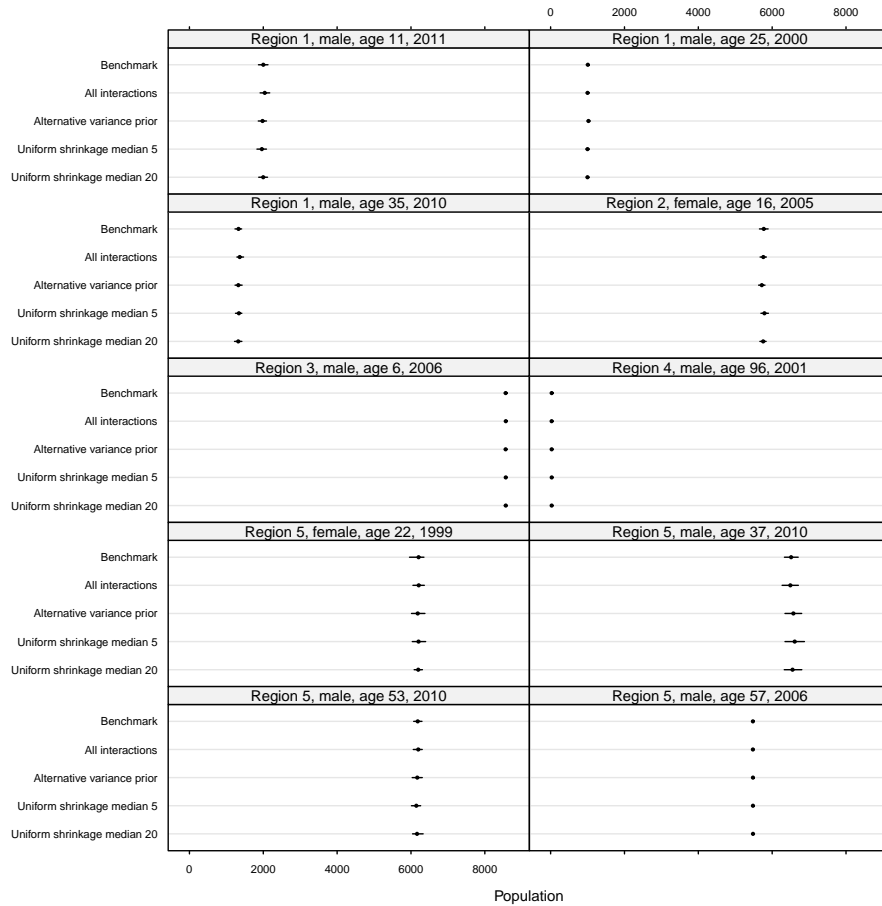


Figure 7: Sensitivity tests. The figure shows posterior distributions for population counts, for 10 randomly-chosen region-sex-age-time cells. The dots represent the medians of the distributions, and the bars represent 95% credible intervals. In many cases, the credible intervals are too narrow to be distinguished from the medians. The ‘benchmark’ results come from the specification described in the main body of the paper. The ‘all interactions’ results come from a specification in which all demographic models contain a full set of interactions between region, sex, and age. The ‘alternative variance prior’ results come from a specification in which the prior for group-level effects is  $p(\log \tau) \propto 1$ . The ‘uniform shrinkage median 5’ and ‘uniform shrinkage median 20’ results come from specifications in which  $\mu_0$  values for the prior for  $\xi$  are 5 and 20.

**Simplified example of generating a proposal for  $Q$**

Figure 8 presents a simplified example of the generation of a proposal for  $Q$ .

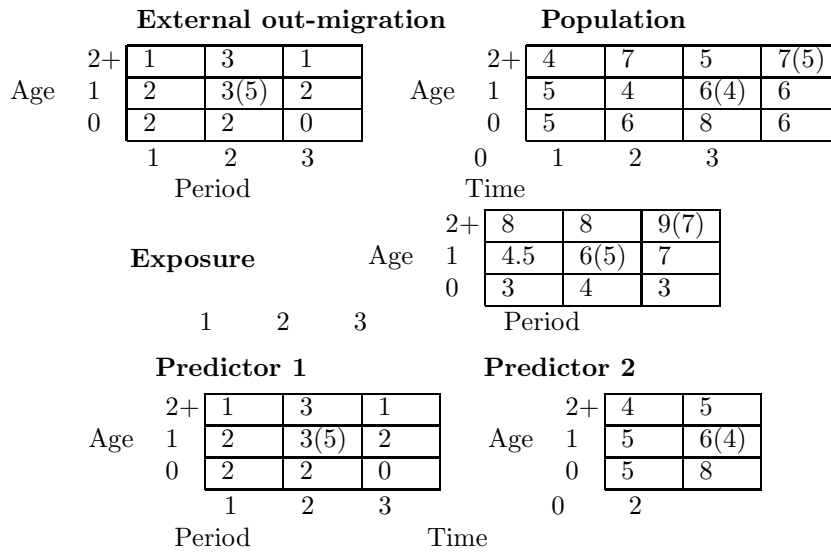


Figure 8: A simplified example of the updating of  $Q$ . The cell to be updated is external out-migration for age-group 2 during period 2, for a single region and sex. Only the region and sex that were selected for updating are shown. Births, deaths, internal migration, and external in-migration, which all remain unchanged, are not shown. The current value for the cell is 3 and the proposed value is 5. Exposure is calculated using  $\epsilon = 0$ . Predictor 1 is calculated from external out-migration, and Predictor 2 is calculated from population.

**Additional formulas used in posterior calculation**

Section 3.5 discusses how  $\frac{p(Q^*|\theta_Q^*)}{p(Q^*|\theta_Q)}$  can be decomposed into three terms. Formulas for the three terms are given in Table 3. The  $p$  in the second column denotes Poisson density. The  $C_N$  in the third column and  $C_{BDM}$  in the fourth column are introduced

to save space. They are defined as follows, using the cohort notation set out in (33),

$$\begin{aligned}
 C_N(r, s, a, t) &= \prod_{v=0}^{\mathcal{T}-t} \frac{p\left(n_{rsatv}^* | \lambda_{rsatv}^{N,i}\right)}{p\left(n_{rsatv}^i | \lambda_{rsatv}^{N,i}\right)} \\
 C_B(r, s, a, t) &= \prod_{v=0}^{\mathcal{T}-t} \prod_{w=1}^{\mathcal{S}} \frac{p\left(b_{rwatv}^i | o_{rsatv}^* \lambda_{rwatv}^{B,i}\right)}{p\left(b_{rwatv}^i | o_{rsatv}^i \lambda_{rwatv}^{B,i}\right)} \text{ if } s = 1 \text{ and } 0 \text{ otherwise,} \\
 C_D(r, s, a, t) &= \prod_{v=0}^{\mathcal{T}-t} \frac{p\left(d_{rsatv}^i | o_{rsatv}^* \lambda_{rsatv}^{D,i}\right)}{p\left(d_{rsatv}^i | o_{rsatv}^i \lambda_{rsatv}^{D,i}\right)} \\
 C_M(r, s, a, t) &= \prod_{v=0}^{\mathcal{T}-t} \frac{p\left(m_{rsatv}^{II,i} | o_{rsatv}^* \lambda_{rsatv}^{II,i}\right)}{p\left(m_{rsatv}^{II,i} | o_{rsatv}^i \lambda_{rsatv}^{II,i}\right)} \frac{p\left(m_{rsatv}^{IO,i} | o_{rsatv}^* \lambda_{rsatv}^{IO,i}\right)}{p\left(m_{rsatv}^{IO,i} | o_{rsatv}^i \lambda_{rsatv}^{IO,i}\right)} \\
 &\quad \times \frac{p\left(m_{rsatv}^{EI,i} | o_{rsatv}^* \lambda_{rsatv}^{EI,i}\right)}{p\left(m_{rsatv}^{EI,i} | o_{rsatv}^i \lambda_{rsatv}^{EI,i}\right)} \frac{p\left(m_{rsatv}^{EO,i} | o_{rsatv}^* \lambda_{rsatv}^{EO,i}\right)}{p\left(m_{rsatv}^{EO,i} | o_{rsatv}^i \lambda_{rsatv}^{EO,i}\right)} \\
 C_{BDM} &= C_B C_D C_{II} C_{IO} C_{EI} C_{EO}.
 \end{aligned}$$

Table 3: Formulas for components of  $\frac{p(\mathbf{Q}^*|\theta_Q^i)}{p(\mathbf{Q}^i|\theta_Q^i)}$ 

| Updated         | $\frac{p(\mathbf{n}_0^* \theta_Q^i)}{p(\mathbf{n}_0^i \theta_Q^i)}$  | $\frac{p(\mathbf{b}^*, \mathbf{d}^*, \mathbf{m}^* \mathbf{n}^*, \theta_Q^i)}{p(\mathbf{b}^i, \mathbf{d}^i, \mathbf{m}^i \mathbf{n}^*, \theta_Q^i)}$ | $\frac{p(\mathbf{n}_{1+}^* \theta_Q^i)}{p(\mathbf{n}_{1+}^i \theta_Q^i)}$ | $\frac{p(\mathbf{b}^i, \mathbf{d}^i, \mathbf{m}^i \mathbf{n}^*, \theta_Q^i)}{p(\mathbf{b}^i, \mathbf{d}^i, \mathbf{m}^i \mathbf{n}^i, \theta_Q^i)}$ |
|-----------------|--|---|---|---|
| $n_{rsa0}$      | $\frac{p(n_{rsa0}^* \lambda_{rsa0}^{N,i})}{p(n_{rsa0}^i \lambda_{rsa0}^{N,i})}$                                |   | $C_N(r, s, \min(a+1, \mathcal{A}), 1)$                                    | $C_{BDM}(r, s, \min(a+1, \mathcal{A}), 1)$  |
| $b_{rsat}$      | $\frac{p(b_{rsat}^* o_{rsat}^*\lambda_{rsat}^{B,i})}{p(b_{rsat}^i o_{rsat}^*\lambda_{rsat}^{B,i})}$            |   | $C_N(r, s, 1, t)$   | $C_{BDM}(r, s, 1, t)$   |
| $d_{rsat}$      | $\frac{p(d_{rsat}^* o_{rsat}^*\lambda_{rsat}^{D,i})}{p(d_{rsat}^i o_{rsat}^*\lambda_{rsat}^{D,i})}$            |   | $C_N(r, s, a, t)$   | $C_{BDM}(r, s, a, t)$   |
| $m_{rsat}^{II}$ | $\frac{p(m_{rsat}^{II*} o_{rsat}^*\lambda_{rsat}^{II,i})}{p(m_{rsat}^{II,i} o_{rsat}^*\lambda_{rsat}^{II,i})}$ |   | $C_N(r, s, a, t)$   | $C_{BDM}(r, s, a, t)$   |
| $m_{rsat}^{IO}$ | $\frac{p(m_{rsat}^{IO*} o_{rsat}^*\lambda_{rsat}^{IO,i})}{p(m_{rsat}^{IO,i} o_{rsat}^*\lambda_{rsat}^{IO,i})}$ |   | $C_N(r, s, a, t)$   | $C_{BDM}(r, s, a, t)$   |
| $m_{rsat}^{EI}$ | $\frac{p(m_{rsat}^{EI*} o_{rsat}^*\lambda_{rsat}^{EI,i})}{p(m_{rsat}^{EI,i} o_{rsat}^*\lambda_{rsat}^{EI,i})}$ |   | $C_N(r, s, a, t)$   | $C_{BDM}(r, s, a, t)$   |
| $m_{rsat}^{EO}$ | $\frac{p(m_{rsat}^{EO*} o_{rsat}^*\lambda_{rsat}^{EO,i})}{p(m_{rsat}^{EO,i} o_{rsat}^*\lambda_{rsat}^{EO,i})}$ |   | $C_N(r, s, a, t)$   | $C_{BDM}(r, s, a, t)$   |

## References

- Abel, G., Bijak, J., Forster, J., Raymer, J., and Smith, P. (2010). “What do Bayesian methods offer population forecasters?” Technical report, ESRC Centre for Population Change. 592
- Alho, J. and Spencer, B. (2005). *Statistical demography and forecasting*. Springer Verlag. 596
- Alkema, L., Raftery, A., Gerland, P., Clark, S., and Pelletier, F. (2008). “Estimating the Total Fertility Rate from Multiple Imperfect Data Sources and Assessing Its Uncertainty.” Technical report, Centre for Statistics and the Social Sciences, University of Washington. Working paper 89. 592
- (2012). “Estimating trends in the total fertility rate with uncertainty using imperfect data: Examples from West Africa.” *Demographic Research*, 26: 331–362. 592
- Alkema, L., Raftery, A., Gerland, P., Clark, S., Pelletier, F., Buettner, T., and Heilig, G. (2011). “Probabilistic projections of the total fertility rate for all countries.” *Demography*, 43(3): 815–839. 592
- Bijak, J. and 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(4): 775–796. 592
- Boden, P. and Rees, P. (2010). “Using administrative data to improve the estimation of immigration to local areas in England.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(4): 707–731. 591
- Booth, H. and Tickle, L. (2008). “Mortality Modelling and Forecasting: a Review of Methods.” *Annals of Actuarial Science*, 3(1-2): 3–43. 612
- Brierley, M., Forster, J., McDonald, J., and Smith, P. (2008). “Bayesian estimation of migration flows.” In Raymer, J. and Willekens, F. (eds.), *International Migration in Europe: Data, Models and Estimates*. Wiley and Sons. 592
- Christiansen, C. and Morris, C. (1997). “Hierarchical Poisson regression modelling.” *Journal of the American Statistical Association*, 92: 618–632. 601, 613, 614
- Congdon, P. (2008). “Models for migration age schedules: a Bayesian perspective with an application to flows between Scotland and England.” In Raymer, J. and Willekens, F. (eds.), *International Migration in Europe: Data, Models and Estimates*. Wiley and Sons. 592
- (2010). *Applied Bayesian Hierarchical Models*. Boca Raton: CRC Press. 612
- Daniels, M. (1999). “A prior for the variance in hierarchical models.” *Canadian Journal of Statistics*, 27(3): 567–578. 601

- Daponte, B., Kadane, J., and Wolfson, L. (1997). “Bayesian demography: projecting the Iraqi Kurdish population, 1977-1990.” *Journal of the American Statistical Association*, 92(440): 1256–1267. [592](#)
- Dellaportas, P., Smith, A. F. M., and Stavropoulos, P. (2001). “Bayesian analysis of mortality data.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164(2): 275–291. [592](#)
- Dunstan, K. and Ryan, M. (2011). *Evaluation of Alternative Data Sources for Population Estimates*. Wellington: Statistics New Zealand. [596](#)
- Elliott, M. and Little, R. (2000). “A Bayesian approach to combining information from a census, a coverage measurement survey, and demographic analysis.” *Journal of the American Statistical Association*, 95(450): 351–362. [592](#)
- Fienberg, S. (2011). “Bayesian models and methods in public policy and government settings.” *Statistical Science*, 26(2): 212–226. [612](#)
- Fienberg, S., Johnson, M., and Junker, B. (1999). “Classical, multilevel and Bayesian approaches to population size estimation using multiple lists.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(3): 383–405. [592](#)
- Gelman, A. (2006). “Prior distributions for variance parameters in hierarchical models.” *Bayesian Analysis*, 1(3): 515–533. [600](#)
- Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis*. New York: Chapman and Hall. [604](#)
- Gelman, A. and Rubin, D. (1992). “Inference from iterative simulation using multiple sequences.” *Statistical Science*, 7(4): 457–472. [608](#)
- Girosi, F. and King, G. (2008). *Demographic Forecasting*. Princeton: Princeton University Press. [592](#)
- Greenwood, M. (1997). “Internal migration in developed countries.” In *Handbook of Population and Family Economics*, 647–720. Amsterdam: Elsevier. [596](#)
- King, R., Morgan, B., Gimenez, O., and Brooks, S. (2009). *Bayesian Analysis for Population Ecology*. Boca Raton: Chapman & Hall. [592](#)
- Link, W. and Barker, R. (2009). *Bayesian Inference with Ecological Applications*. Academic Press. [592](#)
- Liseo, B. and Tancredi, A. (2011). “Bayesian estimation of population size via linkage of multivariate Normal data sets.” *Journal of Official Statistics*, 27(3): 491–505. [592](#)
- Little, R. J. (2012). “Calibrated Bayes, an Alternative Inferential Paradigm for Official Statistics.” *Journal of Official Statistics*, 28(3): 309–334. [612](#)

- Lynch, S. and Brown, J. (2010). “Obtaining multistate life table distributions for highly refined subpopulations from cross-sectional data: A Bayesian extension of Sullivan’s method.” *Demography*, 47(4): 1053–1077. [592](#)
- Preston, S., Heuveline, P., and Guillot, M. (2001). *Demography: Modelling and Measuring Population Processes*. Oxford: Blackwell. [602](#), [611](#), [613](#)
- Raftery, A., Li, N., Ševčíková, H., Gerland, P., and Heilig, G. (2012). “Bayesian probabilistic population projections for all countries.” *Proceedings of the National Academy of Sciences*, 109(35): 13915–13921. [592](#)
- Raymer, J., de Beer, J., and 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/Revue européenne de Démographie*, 22(2): 1–31. [592](#)
- Raymer, J. and Rogers, A. (2006). “Applying Model Migration Schedules to Represent Age-Specific Migration Flows.” Technical report, Research Program on Population Processes, Institute of Behavioral Science, University of Colorado at Boulder. [611](#)
- Rees, P. (1979). “Regional Population Project Models and Accounting Methods.” *Journal of the Royal Statistical Society. Series A (General)*, 142(2): 223–255. [592](#)
- (1985). “Population Structure and Models.” In Woods, R. and Rees, P. (eds.), *Choices in the construction of regional population projections*. London: Allen and Unwin. [596](#)
- Rees, P., Norman, P., and Brown, D. (2004). “A framework for progressively improving small area population estimates.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 167(1): 5–36. [591](#)
- Rendall, M., Handcock, M., and Jonsson, S. (2009). “Bayesian estimation of Hispanic fertility hazards from survey and population data.” *Demography*, 46(1): 65–83. [592](#)
- Rogers, A. (1995). *Multiregional demography: Principles, methods, extensions*. New York: Wiley. [612](#)
- Schoen, R. (1988). *Modelling multigroup populations*. New York: Plenum Press. [612](#)
- Smith, P. W. F., Raymer, J., and Guilietti, C. (2010). “Combining available migration data in England to study economic activity flows over time.” *Journal of the Royal Statistical Society. Series A (General)*, 173(4): 733–754. [591](#)
- Smith, S. and Morrison, P. (2005). “Small-area and business demography.” *Handbook of Population*, 761–785. [591](#)
- Stone, R. (1984). “The Accounts of Society.” *Nobel Prize in Economics documents*. [592](#)
- Wilson, T. (2010). “Model migration schedules incorporating student migration peaks.” *Demographic Research*, 23(8): 191–222. [596](#)

- Wilson, T. and Bell, M. (2004). “Comparative empirical evaluations of internal migration models in subnational population projections.” *Journal of Population Research*, 21(2): 127–160. [596](#)
- (2011). “Advances in local and small-area demographic modelling.” *Journal of Population Research*, 28: 103–107. [592](#)
- Young, J., Graham, P., and Blakely, T. (2006). “Modelling the relationship between socioeconomic status and mortality in a mixture of majority and minority ethnic groups.” *American Journal of Epidemiology*, 164(3): 282–291. [601](#)

**Acknowledgments**

The authors wish to thank Jakub Bijak, the associate editor, and two anonymous referees for helpful comments on the paper.

The views expressed in the paper are those of the authors, and do not necessarily reflect the views of Statistics New Zealand.