Statistica Sinica **26** (2016), 1337-1363
doi:http://dx.doi.org/10.5705/ss.2014.200t

BAYESIAN FORECASTING OF DEMOGRAPHIC RATES FOR SMALL AREAS: EMIGRATION RATES BY AGE, SEX, AND REGION IN NEW ZEALAND, 2014-2038

John Bryant and Junni L. Zhang

Statistics New Zealand and Peking University

Abstract: Population forecasts for small areas within a country are an important planning tool. Standard methods for forecasting demographic rates do not, however, perform well with the noisy data that are typical of small areas. We develop a Bayesian model that combines ideas from the demographic, time series, and small area estimation literatures. We apply the model to the problem of forecasting emigration rates, disaggregated by age and sex, for 73 regions within New Zealand for the period of 2014-2038. We also deal with missing regional information and a change of geographic boundary. We test the calibration of the model using held-out data, and present extensions to accommodate age profiles and regional shares that vary over time. A key advantage of our approach is to provide meaningful uncertainty measures about forecasting. The prediction intervals for long-term forecasting are necessarily wide, engaging users to confront the substantial uncertainty about long-term trends.

Key words and phrases: Bayesian hierarchical model, multiple imputation, small area estimation, time series.

1. Introduction

Population forecasts are an essential input to long-range planning, including decisions about schools, infrastructure, and housing (e.g., Bulatao et al. (2000); Siegel (2002)). To be useful, forecasts typically need to be disaggregated. A minimal requirement for many purposes is that the forecasts include age, sex, and region within the country. Once the data are disaggregated in this way, a single cell may contain only a few observations—not enough on their own to support reliable estimates or forecasts. Disaggregated population forecasting is thus an example of "small area estimation" (Rao (2003); Pfeffermann (2013)).

Because planning for infrastructure requires long time horizons, forecasts that extend many years into the future are needed. It is important for these forecasts to be as accurate as possible. It is also important that users of these forecasts be given reliable information on the extent of the forecasts' uncertainties.

The standard "cohort-component" approach to population forecasting is based on the accounting identity that population at the end of a period equals population at the beginning of the period plus the number of births minus the number of deaths plus the number of immigrants minus the number of emigrants. Forecasts are constructed for birth rates, death rates, and migration rates. The accounting identity is then repeatedly applied to give forecasted population in each period after the base year.

In this paper, we focus on a single component of population change: emigration, that is, migration from each small area to destinations outside the country. We also focus on a single country: New Zealand. Forecasting different components in different settings would raise different problems and require different model assumptions. Nevertheless, the technical challenges that we discuss in this paper are sufficiently common, and the solutions we propose are sufficiently general, that the methods could serve as a starting point for many problems in demographic forecasting.

We forecast emigration rates rather than counts. Forecasting counts would require either (i) ignoring the size of the population at risk of emigrating or (ii) obtaining the population at risk from a full cohort-component population forecast, which would require forecasts of all other migration flows, as well as births and deaths. Although approach (i) is sometimes taken, it has serious shortcomings, including ignoring the fact that population size is typically a good predictor of migration counts, and a tendency to produce demographically implausible outcomes. Approach (ii) is our long-term goal, but is outside the scope of this paper. Forecasting a single component is challenging enough. In this paper, we forecast emigration rates for 16 age groups, 2 sexes, and 73 regions over 25 years, which amounts to $16 \times 2 \times 73 \times 25 = 58{,}400$ rates.

Demographic forecasters often deal with many rates. They do, however, enjoy one crucial advantage over most other forecasters: demographic rates are often surprisingly regular. Mortality, fertility, and migration rates often have characteristic age-sex profiles that are stable over time or that change in consistent ways. The regularities in the age-sex profiles reflect regularities in the life course. For instance, migration rates typically peak in the late teenage years because these are the years when people reach adulthood and begin to leave home (Courgeau (1985)). Age-sex profiles are slow to change because the life course regularities underpinning them are slow to change. In addition, regions that have high or low rates in one period tend to have similarly high or low rates in later periods, reflecting long-lasting characteristics of regions, such as the mix of industries or amenities. As we discuss below, emigration rates for New Zealand are well behaved, though there is some evidence of age profiles or region effects changing over time.

Demographers have developed techniques that exploit regularities in demographic rates. The standard approach to forecasting age-sex-specific rates at the national level is to obtain an initial set of age-sex specific rates, and then scale them up or down based on a parameter representing overall level (Preston, Heuveline, and Guillot (2001)). The initial set of age-sex-specific rates may be smoothed or adjusted to deal with random variation or biases. In the case of migration, the most common approach to smoothing is the use of a "model migration schedule", a mathematical function super-imposing a set of exponential curves that represent components of a standard age profile, such as labor force peaks or student migration peaks (Rogers and Castro (1981); Wilson (2010)). Traditionally, forecasts of the parameter capturing overall level are based on some mixture of extrapolation and expert judgment.

The standard way of measuring and communicating uncertainty in population forecasts has been to produce "low", "medium", and "high" scenarios based on relatively simple mathematical extrapolations of current trends (Keilman (2008)). The existence of multiple variants reminds users that the forecasts are subject to uncertainty. However, it is seldom clear how much uncertainty these variants are supposed to encompass, though ex-post analyses of projection errors can provide some guidance (Stoto (1983)). Moreover, when population forecasts are assembled from low, median, and high variants for fertility, mortality and migration, the results are often counter-intuitive. For instance, combinations of variants that lead to large variation in population size may lead to small variation in the ratio of young people to old people (Lee (1998)).

The problem of modelling future uncertainty in demographic rates has received extensive attention from academic demographers and statisticians. The influential paper by Lee and Carter (1992) on mortality forecasting, for instance, uses a principal component analysis to obtain an age profile and a time effect, and then models the time effect using a random walk. A recent alternative is to use functional data models with time series coefficients to model age-specific demographic rates (e.g. Hyndman and Booth (2008)). Booth (2006) and Booth and Tickle (2008) review the large literature sparked by Lee and Carter (1992). In recent years, Bayesian approaches have started to appear. Raftery and colleagues, for instance, have developed a set of Bayesian methods for forecasting mortality, fertility, and international migration that have been adopted by the United Nations Population Division and used to produce population forecasts for all countries (Alkema et al. (2011); Raftery et al. (2012); Gerland et al. (2014)). Bijak and Wiśniowski (2010) use Bayesian methods, including informative priors elicited from migration experts, to estimate migration flows between European countries from highly imperfect data. However, with the exception of Statistics Netherland and Statistics New Zealand, national statistical agencies have not taken up the new methodologies, and have not published probabilistic population forecasts.

Moreover, almost all the research on population forecasts, Bayesian or otherwise, has focused on national-level forecasts. Despite its practical importance, small area population forecasting has received relatively little attention from academic researchers (Wilson (2014)). Methods developed for national-level forecasts typically do not work well when applied, unmodified, to small area forecasts. The main problem is the increasing prominence of random variation as the data become more disaggregated. Standard methods for obtaining age profiles can break down when the number of events is small and observed rates are subject to large fluctuations, as occurs with small area forecasts. Common solutions are to impose identical profiles across all small areas, or to manually intervene when profiles are implausible. However, neither solution is completely satisfactory. Imposing identical profiles reduces accuracy, and manual intervention reduces transparency and is time-consuming.

A further complication with small area forecasting is that virtually all geographically-disaggregated data contain gaps and breaks due to changes in administrative boundaries. Manipulating the data to achieve consistent historical time series is often the most labour-intensive part of small area estimation and forecasting (Rees (1985); Gregory, Marti-Henneberg, and Tapiador (2010)). In principle, population forecasts should incorporate the uncertainty generated by the gaps and breaks into the overall uncertainty measures. In practice, this is rarely done.

The approach to small area forecasting presented in this paper draws on ideas from the literature on small area estimation, in addition to those on demography and time series. We develop a fully Bayesian hierarchical model for demographic rates of small areas, apply the model to New Zealand emigration data disaggregated by age, sex, and region for the period of 1991-2013, and then construct forecasts for the period of 2014-2038. We use multiple imputation to deal with two types of missingness in the region variable, including that caused by changes in administrative boundaries (Heitjan and Little (1991)). We measure uncertainty coherently using credible intervals for quantities of interest. We validate the model using five years of hold-out data.

Our model assumes a hierarchical prior structure for the rates disaggregated by age, sex, region, and time, which allows for borrowing of strength across age groups, sexes, regions, and times. For tractability, our basic model specifies that the age effect, the region effect, and the time effect are additive on the scale of log-rates. In extensions, we allow for an age-time interaction and a region-time interaction. Our models make no use of informative priors for hyperparameters. The main reason for not using informative priors is that national statistical agencies—an important potential user of our methods—have traditionally been averse to the use of such priors (Fienberg et al. (2011)). It is possible, however, that informative priors may bring sufficient improvements in model performance to outweigh such considerations; we return to the topic of informative priors in the Discussion section. A key advantage of our approach that it provides meaningful uncertainty measures for the forecasts, which is especially important for long-term forecasting. The prediction intervals are wide for the end of the forecast horizon. Users of population forecasts may sometimes dislike wide prediction intervals. But we argue that forecasters need to be prepared to defend wide intervals.

The remainder of the paper proceeds as follows. Section 2 discusses the data for the application. Section 3 presents the methods, including the basic model, forecasting, treatment of missing values, validation, and an extension to age-time or region-time interactions. Section 4 presents the results. Section 5 concludes with a discussion of limits of the current model, future developments, and potential advantages of the methods for statistical agencies.

2. Data

2.1. Data on emigration

Our emigration data are counts of "permanent and long-term" departures from New Zealand for 1991-2013. A permanent and long-term departure is a departure that entails a change in residence as opposed to a business trip or a holiday, which includes the departure of students studying at universities abroad. The counts are disaggregated by 5-year age-group, sex, region, calendar year, and citizenship. The information on citizenship is not of interest in itself, and the data cannot be publicly released with so many cross-classifying variables due to privacy concerns. However, as discussed in Section 3.3, citizenship is used when imputing for non-response in the region variable. Because New Zealand is an island with an efficient administrative system, its international migration data are unusually accurate.

The geographical unit used in the departures data is the "territorial authority". The territorial authority is the most important subnational administrative unit in New Zealand. In 2010 there were 73 territorial authorities in the country, giving an average population size of 60,000, though the smallest (Chatham Islands) had a population of less than 1,000. During 2010, seven territorial authorities within greater Auckland were amalgamated into a single unit, containing one third of the national population. From 2010, departures of people who had been living within the new amalgamated unit were simply coded to "Auckland", with no further geographical detail.

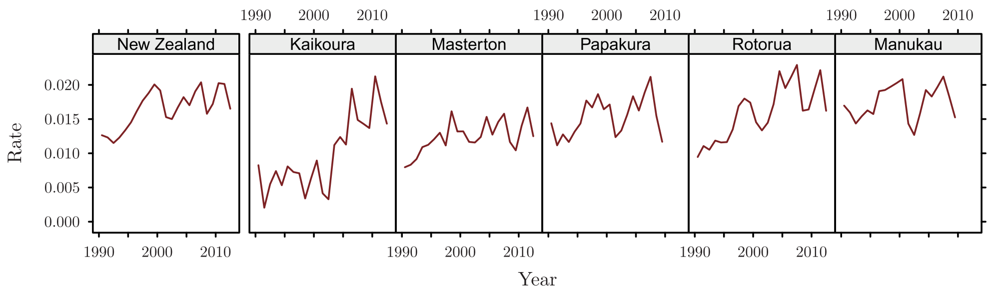

Figure 1. Direct estimates of emigration rates for New Zealand and five selected territorial authorities, 1991-2013. The territorial authorities are ordered by population size, with the smallest on the left. The series for Papakura and Manukau end in 2010, when these territorial authories were amalgamated into a new unified Auckland authority.

Altogether, 7.5% of individual emigration records have no regional information at all, either because the respondent did not provide it, or because the response could not be coded. The percent of missing values rose sharply around the year 2000, with an average of 2.2% in the 1990s, and 9.8% in 2000-2013.

Another challenge is sparsity. Once the emigration data are disaggregated by age, sex, territorial authority, and year, the median cell size is 7, and 14% of cells have a value of 0.

2.2. Other data

To capture the size of the population at risk of emigrating, we use population estimates disaggregated by age, sex, region, and year obtained from Statistics New Zealand as a customized tabulation. The population estimates are sufficiently disaggregated that population sizes for the seven pre-2010 territorial authorities within greater Auckland can be calculated directly for 2011-2013.

As discussed in Sections 3.1 and 3.3 below, we use data on regional-level characteristics to predict long-run regional migration levels. The characteristics are the percent of the territorial authority population born overseas and the percent of territorial authority population in full-time study. The data are publicly available at a low level of geographical detail (Statistics New Zealand (2014)).

2.3. Trends and patterns in emigration rates

Figure 1 shows direct estimates of emigration rates, that is, counts of emigrations divided by the corresponding population at risk. The figure shows rates for the country as a whole, and for five territorial authorities arranged by population size from smallest (Kaikoura, with a population in 2010 of 2,800) to largest (Manukau, with a population in 2010 of 375,700). Figure A1 in the Supplementary Materials shows equivalent rates for 49 randomly-selected territorial

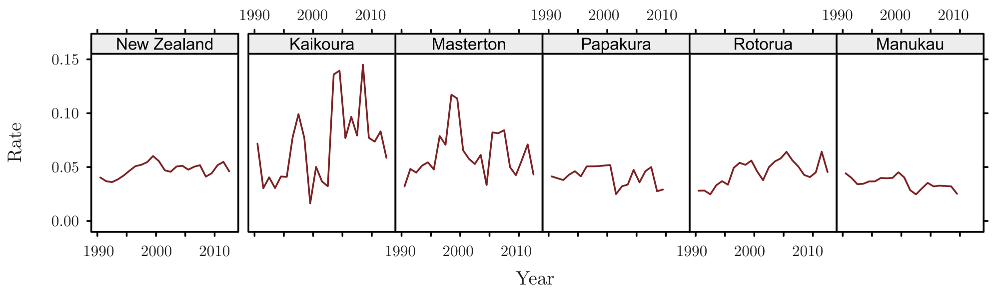

Figure 2. Direct estimates of emigration rates for females aged 20-24, for New Zealand and five selected territorial authorities. Note that the vertical scale is almost 6 times higher than that for Figure 1.

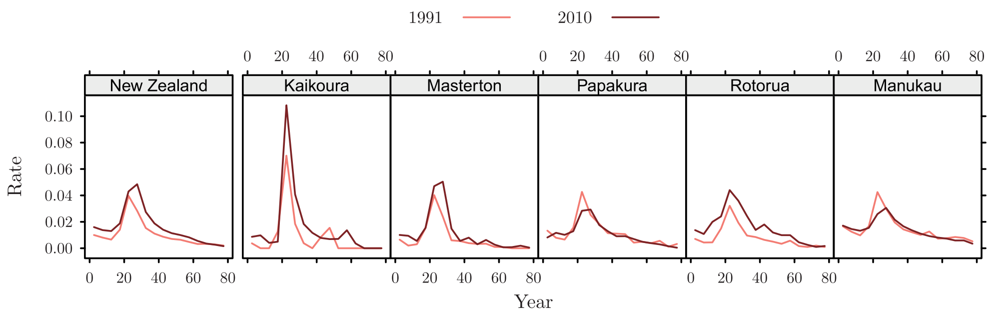

Figure 3. Direct estimates of emigration rates by age, for New Zealand and five selected territorial authorities, 1991 and 2010.

authorities. At the national level, the emigration rate appears to have increased until about the year 2000, then fluctuated around a constant level. None of the five selected territorial authorities exactly match this pattern, though all show evidence of fluctuating around a constant level since approximately 2,000. There is clear evidence of differences between territorial authorities in mean emigration rates. There is a hint of differences in slopes (rates of change) for emigration rates, but such differences are small relative to annual variation.

Rates for specific age-sex groups are subject to greater random fluctuations than rates for the population as a whole. To illustrate, Figure 2 shows rates for females aged 20-24. (Figure A2 in the Supplementary Materials presents equivalent rates for the 49 territorial authorities.) To appreciate the extent of the variability, it is important to note that the vertical scale is almost six times higher than that for Figure 1. The rates for Kaikoura are particularly noisy, which is to be expected, given its small size.

Figure 3 shows emigration rates by age in 1991 and 2010. There are hints of a change in age profile, though it is difficult to disentangle the effects of random variation. There is also some suggestion of variation in age profiles across territorial authorities.

3. Methods

3.1. The basic model

Small area methods typically use models to borrow strength from related areas or across time (Rao (2003); Pfeffermann (2013)). We use a hierarchical Bayesian model in which counts are Poisson and rates are log-normal. As will become apparent, assuming a log-normal distribution for rates greatly eases the task of modelling age, sex, region, and time effects, since it allows the use of linear models at this level. While a gamma distribution for rates would facilitate the estimation of the rates, since the Poisson and gamma are conjugate, it would complicate the modelling of age, sex, region, and time effects. An early example of a model similar to ours, though on a smaller scale, and dealing with education rather than migration, is Cargnoni, Müller, and West (1997).

For age a, sex s, region r and time t, let y_{asrt} denote the emigration count, and x_{asrt} person-years of exposure to the risk of migrating. We assume that

$$y_{asrt} \overset{\text{ind}}{\sim} \text{Poisson}(\lambda_{asrt} x_{asrt}), \tag{3.1}$$

$$\log \lambda_{asrt} = \beta^0 + \beta_a^{\text{age}} + \beta_s^{\text{sex}} + \beta_r^{\text{reg}} + \beta_t^{\text{time}} + \beta_{as}^{\text{age:sex}} + \beta_{ar}^{\text{age:reg}} + \epsilon_{asrt}, \tag{3.2}$$

$$\epsilon_{asrt} \overset{\text{ind}}{\sim} \text{N}(0, \sigma_\epsilon^2). \tag{3.3}$$

Here λ_{asrt} is the underlying emigration rate, β^0 is an intercept, β_a^{age} is an age effect, β_s^{sex} is a sex effect, β_r^{reg} is a region effect, β_t^{time} is a time effect, and $\beta_{as}^{\text{age:sex}}$ and $\beta_{ar}^{\text{age:reg}}$ are respectively age-sex and age-region interactions. The $\overset{\text{ind}}{\sim}$ symbol indicates that the quantities on the left hand side are drawn independently, given the parameter values on the right.

The presence of the ϵ_{asrt} term means that posterior values for λ_{asrt} are a compromise between the predictions of the higher-level model (3.2) and the direct estimate y_{asrt}/x_{asrt}. The more observations there are in cell $asrt$, the closer the posterior distribution will be to the direct estimate. This is a standard feature of hierarchical Bayesian models (Gelman and Hill (2006)).

Including more interaction terms makes identification more difficult, and substantially increases the number of iterations required for the model to reach convergence. We chose interaction terms to include in the model by examining observed rates like those presented in Section 2.3, and by fitting models with many interactions and calculating the variance of the parameters for each interaction (with greater variance implying that the interaction was contributing more to the overall variance in $\log \lambda_{asrt}$) (Gelman et al. (2014, p.396)). In an extension described in Section 3.5, we add an age-time or a region-time interaction to the basic model.

We assume that the time effect β_t^{time} follows a random walk with noise, a special case of a non-stationary polynomial trend model (Prado and West (2010, pp.119-120)). We assume that

$$\beta_t^{\text{time}} = \theta_t + v_t, \tag{3.4}$$

$$\theta_t = \theta_{t-1} + w_t, \tag{3.5}$$

$$v_t \overset{\text{ind}}{\sim} \mathrm{N}(0, \sigma_v^2), \tag{3.6}$$

$$w_t \overset{\text{ind}}{\sim} \mathrm{N}(0, \sigma_w^2). \tag{3.7}$$

A random walk with noise reduces to a standard random walk when $v_t \equiv 0$. An attractive property of the random walk with noise is that it distinguishes between short-term idiosyncratic movements, captured by v_t, and permanent changes in the level of the distribution, captured by w_t. Following Petris (2010), our prior for the starting value of θ is $\theta_0 \sim \mathrm{N}(0,\ 1{,}000{,}000)$.

In initial work, we also used specifications like that of (3.4)-(3.7) to model age effects, on the grounds that values for consecutive age groups were likely to be correlated in the same way that values for consecutive periods were. However, further experimentation showed that assuming $\beta_a^{\text{age}} \sim \mathrm{N}(0, \sigma_{\text{age}}^2)$ gave almost identical results, presumably because there is enough information in the data on variation by age to overwhelm the effects of choice of prior. We therefore adopted the simpler exchangeable prior for age effects.

The region effects β_r^{reg} are modelled using linear regression on region-level covariates $\boldsymbol{X}_r$:

$$\beta_r^{\text{reg}} = \boldsymbol{\gamma}^\top \boldsymbol{X}_r + u_r, \tag{3.8}$$

$$u_r \overset{\text{ind}}{\sim} \mathrm{N}(0, \sigma_u^2), \tag{3.9}$$

where $\boldsymbol{X}_r$ consists of the logarithm of percent of population born overseas and the logarithm of percent of population in full-time study for region r in 2013. The elements of $\boldsymbol{\gamma}$ have improper uniform prior distributions. Section SM.1 in the Supplementary Materials discusses the set-up for region effects in more detail. We do not allow for time-varying covariates, as this would require forecasting covariates when forecasting migration rates. In future work, we intend to allow for correlations between neighbouring territorial authorities following, for instance, Congdon (2014). However, we do not attempt this here.

We assume an improper uniform prior distribution for the intercept term β^0 and for the sex effect β_s^{sex}. Interaction terms have normal priors:

$$\beta_{as}^{\text{age:sex}} \overset{\text{ind}}{\sim} \mathrm{N}(0, \sigma_{\text{age:sex}}^2), \tag{3.10}$$

$$\beta_{ar}^{\text{age:reg}} \overset{\text{ind}}{\sim} \mathrm{N}(0, \sigma_{\text{age:reg}}^2). \tag{3.11}$$

As with small area estimation models more generally, these priors allow estimates for related cells to “borrow strength”, alleviating problems of small sample sizes. The standard deviation terms for the normal priors all have improper uniform distributions over the positive real numbers. Section SM.2 in the Supplementary Materials provides details on the Gibbs sampler used to draw from the posterior distribution.

3.2. Forecasting

Having used data up to time T to estimate the model, we obtain forecasts for time $t = T+1, \ldots, T+K$. Many parameters, such as β^0 or $\beta_{as}^{\text{age:sex}}$, do not contain time indices and are time invariant. However, some parameters, including emigration rates λ_{arst}, do vary over time. We generate random values for these parameters by drawing on conditional distributions given values for the time-invariant parameters.

Estimation yields N posterior draws of the model parameters. We derive the nth $(n = 1, \ldots, N)$ set of forecasted parameters as follows.

1. For $k = 1, \ldots, K$, generate $v_{T+k}^{(n)}$ independently from a $\text{N}\left(0, \sigma_v^{2(n)}\right)$ distribution.
2. For $k = 1, \ldots, K$, generate $w_{T+k}^{(n)}$ independently from a $\text{N}\left(0, \sigma_w^{2(n)}\right)$ distribution.
3. Apply the observation equation (3.4) and the state equation (3.5) repeatedly to obtained time effects $\beta_{T+k}^{\text{time}(n)}$, $k = 1, \ldots, K$.
4. For $k = 1, \ldots, K$, generate $\epsilon_{a,s,r,T+k}$ independently from a $\text{N}\left(0, \sigma_\epsilon^{2(n)}\right)$ distribution for all a,s,r.
5. Use (3.2) to compute values for $\log \lambda_{a,s,r,T+k}^{(n)}$, for all a, s, r, and $k = 1, \ldots, K$.
6. Exponentiate to obtain $\lambda_{a,s,r,T+k}^{(n)}$, for all a, s, r, and $k = 1, \ldots, K$.

3.3. Missing values for region

The region variable is subject to two kinds of missingness: (i) some records are missing any information on region; and (ii) in 2011–2013, records with territorial authorities coded to “Auckland” are missing information on which pre-2010 territorial authority within greater Auckland the respondent departed from.

We address problem (i) by building a statistical model of departures and using the model to multiply impute region. Rubin (1988, p.1) argues that imputation of missing values in official data should ideally by carried out by national statistical office because there is often “information available to the data collector but not available to an external data analyst. ... This kind of information, even though inaccessible to the user of a public-use file, can often improve the imputed values." This description applies exactly to emigration data in New Zealand. Statistics New Zealand staff have access to detailed data on emigration by age, sex, region, time, and citizenship status that is not available to the public because of privacy concerns. One of the authors of this paper is a Statistics New Zealand employee, and therefore has access to the detailed data. Examination of the data shows that, even after stratifying on age, sex, and time, the proportion of missing values for region varies strongly by citizenship, as does the territorial authority of departure.

Let y^{obs}_{asrtc} denote counts of departures for age-group a, sex s, region r, time t and citizenship c where region is recorded. Let y^{mis}_{asrtc} denote counts where region is not recorded. Let $y^{\text{part}}_{astc} = \sum_r y^{\text{mis}}_{asrtc}$. Unlike y^{mis}_{asrtc}, y^{part}_{astc} is observed.

A hierarchical Poisson model similar to our main model is fitted to y^{obs}_{asrtc}. The model includes age, sex, region, time, and citizenship main effects, and a region-citizenship interaction, where the citizenship effects and region-citizenship interaction are assumed to be drawn from independent normal distributions. It does not include an exposure term, since population estimates disaggregated by citizenship are not available.

Let ϕ_{arstc} denote the emigration rate generated by this model. The imputation procedure assumes that, conditioning on age, region, sex, time, and citizenship, the region variable is missing at random; that is, within y^{part}_{astc}, the probability of belonging to region r is proportional to ϕ_{arstc}. Based on this assumption, for each of M posterior draws of the ϕ_{asrtc}, values for y^{mis}_{asrtc} were generated using

$$(y^{mis}_{as1tc}, \dots, y^{mis}_{asRtc}) \sim \text{Multinomial}\left(y^{\text{part}}_{astc}, (\pi_{as1tc}, \dots, \pi_{asRtc})\right), \tag{3.12}$$

where $\pi_{asrtc} = \phi_{asrtc} / \sum_r \phi_{asrtc}$. Region r takes 73 values from 1991 to 2010, and 67 thereafter. Values

$$y_{asrt} = \sum_{c=1}^{C} (y^{obs}_{asrtc} + y^{mis}_{asrtc}) \tag{3.13}$$

were then calculated, providing M complete datasets. Because these datasets do not include the citizenship variable, they are sufficiently aggregated to fulfil Statistics New Zealand confidentiality requirements and can be released externally.

Rather than using Rubin's rules to combine estimates from the multiply imputed datasets (Rubin (1987)), we combined posterior samples. We applied our model to each of the M imputed data sets to obtain M samples from the M associated posterior distributions. We then combined the M samples into a single pooled sample, and derived all our posterior inferences from the pooled sample. The number of imputed datasets needed to adequately capture the uncertainty created by the imputation process depends on the fraction of data that are missing, and on the model being used, but often numbers as low as 3 or 5 are enough (Rubin (1987)). We used $M = 10$. We show in the Supplementary Material that using $M = 5$ would not substantially degrade performance (Figure A3 in the Supplementary Materials).

Problem (ii) is the allocation of departures from within greater Auckland to the seven pre-2010 territorial authorities. The simplest solution to this problem would be to combine data for the seven territorial authorities in every year, and work with 67 regions rather than 73. However, users of population projections require information on areas within greater Auckland. We therefore attempted to allocate emigrations from greater Auckland over the period 2011-2013 to the seven original territorial authorities.

Let $y_{as,\text{Au},t}$ denote the emigration count for age a and sex s for greater Auckland during 2011-2013. Let $A_1, \ldots, A_7$ denote the indices for the seven original territorial authorities within greater Auckland. Given $y_{as,\text{Au},t}$ and the rates, emigration counts for territorial authorities within greater Auckland can be imputed using

$$(y_{as,A_1,t}, \ldots, y_{as,A_7,t}) \sim \text{Multinomial}\left(y_{as,\text{Au},t}, (\psi_{as,A_1,t}, \ldots, \psi_{as,A_7,t})\right), \quad (3.14)$$

where $\psi_{as,A_j,t} = \lambda_{as,A_j,t} / \sum_{j'} \lambda_{as,A_{j'},t}$.

If the imputation was done within each step of the Gibbs sampler, strong correlation would be introduced between the imputed $y_{as,A_j,t}$'s and the corresponding $\lambda_{as,A_j,t}$'s, leading to slow convergence. We therefore exploited the fact that

$$y_{as,\text{Au},t} \overset{\text{ind}}{\sim} \text{Poisson}\Big(\sum_j \lambda_{as,A_j,t} x_{as,A_j,t}\Big), \quad (3.15)$$

and adopted instead the following strategy. First, model parameters were estimated based on the observed counts, including $y_{as,\text{Au},t}$ for years 2011-2013 and y_{asrt} for r not in greater Auckland or t in other years. Second, $y_{ar,A_j,t}$'s for years 2011-2013 were imputed given each posterior draw of the rates. Section SM.2 in the Supplementary Materials gives the details.

3.4. Validating the model

Following standard practice in time series modelling, we tested the performance of the model by constructing forecasts that were based on a subset of the available data, and assessing how well the forecasts match the data that were held out. Unfortunately, our time series are short, so it is only feasible to hold out a few years of data, and thus only test the short-run performance of the model.

Moreover, we wish to test the ability of the model to deal with the loss of information on region for the seven territorial authorities within greater Auckland. To do this, we need data with regional detail for these territorial authorities, which is only available up to 2010.

In an application like ours, where cell counts are small, the distinction between super-population and finite-population quantities introduces further complications. Our model produces forecasts for λ_{asrt}, which are unobservable super-population quantities. The data available for validation are y_{asrt}/x_{asrt}, which are observed, finite-population quantities. Because of sampling variation, finite-population quantities do not in general equal their super-population equivalents. When cell counts are large, so that sampling variation is minor relative to the rates, the differences between super-population and finite-population quantities are small enough to ignore. However, when cell counts are small, the distinction becomes important. We therefore need to convert our super-population quantities into their finite-population equivalents before making the comparison.

Our training dataset is data for 1991-2005, but with emigration counts for 2003-2005 for the seven territorial authorities within greater Auckland all coded to "Auckland", mimicking the effect of the actual amalgamation in 2010. The held-out data include emigration counts y_{arst} for all 73 regions for 2006-2010. Since there are records missing any information on region, we do not know the true values of y_{arst} in the held-out data, and we only have their imputed values. Hence our validation test were conditional on the 10 imputed datasets. The validation procedure was as follows.

1. Run the MCMC algorithm using the training data, including drawing the parameters and imputing the values for the seven territorial authorities within greater Auckland for 2003-2005, on 10 multiply imputed datasets, to obtain 10 sets of posterior samples.
2. For each of the 10 multiply imputed datasets and its associated set of posterior samples:
    - (2.1) Obtain a set of forecasts for the underlying emigration rates λ_{asrt} for the 73 territorial authorities for 2006-2010.
    - (2.2) Obtain a set of forecasts for the emigration counts y^{pre}_{asrt} for 2006-2010 by drawing from Poisson distributions with means equal to $\lambda_{asrt}x_{asrt}$. (This is equivalent to forecasting the finite population emigration rates y_{asrt}/x_{asrt} since the x_{asrt} are known constants.)
    - (2.3) Calculate the proportions of 50% and 90% prediction intervals for y^{pre}_{asrt} that do in fact contain y_{arst}.

(2.4) Use the median value for y^{pre}_{asrt} as a forecasted value for y_{arst}, and calculate an accuracy measure as the proportion of forecasted values falling within a given percentage of y_{arst}.

3. Average coverage rates and accuracy measures across the 10 datasets.

The accuracy measures obtained from step 2.4 are difficult to assess in isolation. We therefore compared them with equivalent measures using a naive model in which the forecasted value equals the last observed value, which is the value for 2002 for territorial authorities within greater Auckland and 2005 elsewhere.

3.5. Adding age-time and region-time interactions

Examination of observed emigration rates (e.g., Figure 3) provides some evidence for an interaction between age and time. More generally, the ability to accommodate changing age structures would greatly increase the number of situations to which our methods could be applied. We have therefore experimented with the use of an age-time interaction.

The age-time interaction consists of a random walk with noise within each age group,

$$\beta^{\text{age:time}}_{at} = \theta_{at} + v_{at}, \tag{3.16}$$

$$\theta_{at} = \theta_{a,t-1} + w_{at}, \tag{3.17}$$

$$v_{at} \overset{\text{ind}}{\sim} \text{N}(0, \tau_v^2), \tag{3.18}$$

$$w_{at} \overset{\text{ind}}{\sim} \text{N}(0, \tau_w^2). \tag{3.19}$$

To keep the calculations tractable, the specification does not allow for the correlations between neighbouring age groups.

We have also experimented with a region-time interaction. The specification mirrors that for the age-time interaction, but with region replacing age. A further extension is to combine age-time and region-time interactions in the same model. Doing so leads to extremely poor convergence, however, so we leave this for future research.

4. Results

4.1. The basic model

Figure 4 shows how estimated and forecasted emigration rates evolve over time, for females in selected age groups and territorial authorities. (Here, and throughout the paper, we use 'estimated' when referring to historical rates, and 'forecasted' when referring to future rates.) Results for females in the 20-24 age group in the 49 territorial authorities are shown in Figure A4 in the Supplementary Materials.

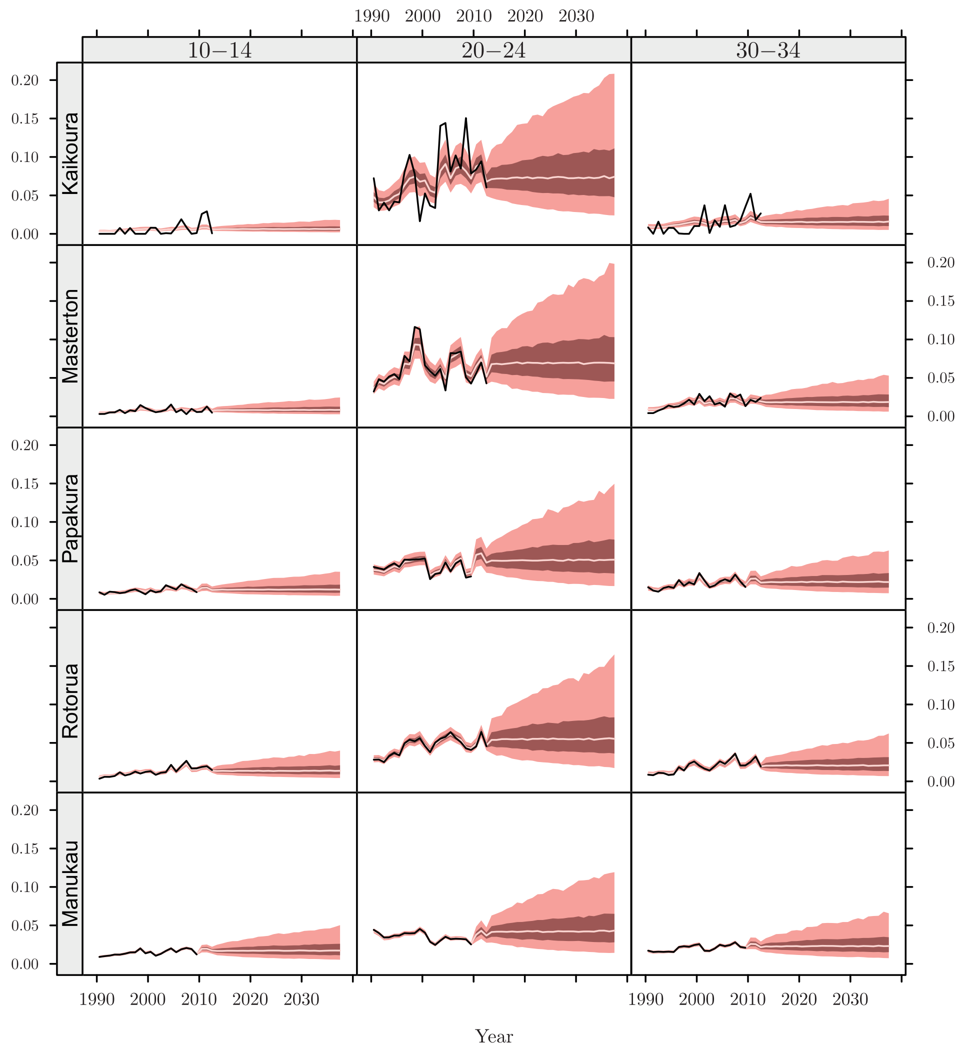

Figure 4. Estimated and forecasted emigration rates. Each panel shows emigration rates by time, for females, for selected territorial authorities and selected age groups. The territorial authorities are ordered by population size, with the smallest at the top. The light shading represents 90% credible intervals, the dark shading represents 50% credible intervals, and the light lines in the center show posterior medians. The black lines are direct estimates.

The black lines in the figures depict observed finite-population emigration rates, while the shaded regions depict credible intervals for the underlying super-population λ_{asrt}. It is to be expected that less than 90% of the finite-population rates lie inside the 90% credible intervals for λ_{asrt}, since the finite-population

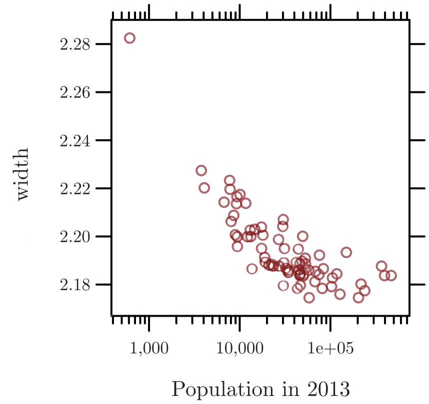

Figure 5. The average length of 95% credible intervals for $\log \lambda_{asrt}$ in 2038 versus population size in 2013 on a logarithmic scale for all 73 regions.

rates are subject to Poisson variability.

During the estimation period, when y_{asrt} is observed, the direct estimates for small territorial authorities are subject to greater smoothing, and the credible intervals are wider for small territorial authorities than for large territorial authorities. This is sensible behavior, since cell counts are lower in smaller territorial authorities. This behavior is typical of small area estimation models.

In Papakura and Manukau, both of which are within greater Auckland, there is an increase in uncertainty between 2010 and 2011. The increase in uncertainty reflects the fact that information on areas within greater Auckland is no longer available after 2010. It is again sensible behavior, but would not occur with traditional deterministic methods for imputation.

Between 2013 and 2014, when moving from estimates to forecasts, in all territorial authorities there is a sharp increase in uncertainty. The credible intervals then widen more gradually over time. This pattern is typical of time series models for demographic rates.

The relationship between regional population size and uncertainty is less clear in the forecast period than in the estimation period. But a relationship does exist. Figure 5 shows widths of 90% prediction intervals for $\log \lambda_{asrt}$ in 2038, averaged across age and sex, versus log population size in 2013. There is a negative relationship between width and population size, hence larger regions on average have less uncertainty about $\log \lambda_{asrt}$. As can be seen from the vertical scale, however, the differences are small. When prediction intervals that have nearly constant average width on a log scale are translated back into natural units for λ_{asrt}, averaged across age and sex, the prediction intervals for territorial authorities with high average λ_{asrt} are wider than those for territorial authorities with low average λ_{asrt}. But for a particular age and sex group, the pattern of regional difference in uncertainties about λ_{asrt} is not clear-cut.

Table 1. Percents of prediction intervals containing true values: The basic model.

	50% intervals	90% intervals
In Greater Auckland	51.5	89.6
Outside Greater Auckland	58.7	90.7
Overall	58.0	90.6

In most territorial authorities shown in Figure 4 and Figure A4, the means for the forecasted rates are plausible. An exception is Queenstown-Lakes (abbreviated to "Quenstwn-Lks") in Figure A4. The mean for Queenstown Lakes for the whole period 1991-2013 is a poor guide to future rates, as migration rose sharply in the years leading up to 2013. We return to the Queenstown-Lakes case in our extension to region-time interactions.

Figure 6 shows estimated and forecasted age profiles in four selected years, for males in the five selected territorial authorities. Figure A5 in the Supplementary Material shows results for 49 territorial authorities. The forecasted profiles have the same shape as the smoothed historical ones, but with much greater uncertainty. When assessing this uncertainty, it is important to bear in mind that the graphs show annual rates, which are more variable than long-term averages.

4.2. Validation of the basic model

Table 1 shows the results of the calibration exercise described in Section 3.4. The prediction intervals referred to in the table are for the period 2006-2010, and the true values are values held out from the model. For the country as a whole the actual coverage of 50% prediction intervals is 58.0%, and for 90% intervals it is 90.6%. The 90% intervals are well calibrated, but the 50% intervals are too wide.

Figure 8 shows coverage disaggregated by region and time. Auckland city has the worst performance in two of the five years. Figure 7 helps explain why. The figure shows emigration rates for the seven territorial authorities within greater Auckland. The "Auckland" panel in the figure shows results for the territorial authority "Auckland City" rather than the wider region. The solid grey line marks the start of the imputation period, and the dashed grey line marks the start of the forecast period. Over the years for which the model has data, emigration rates for the seven territorial authorities in greater Auckland move up and down together. During the imputation and forecast periods, in contrast, the rates for the Auckland territorial authority move upwards while rates for everywhere else move downwards.

Our main accuracy measure is the proportion of forecasted values falling within 25% of the true values. This measure is 42.6% for our model and 37.3%

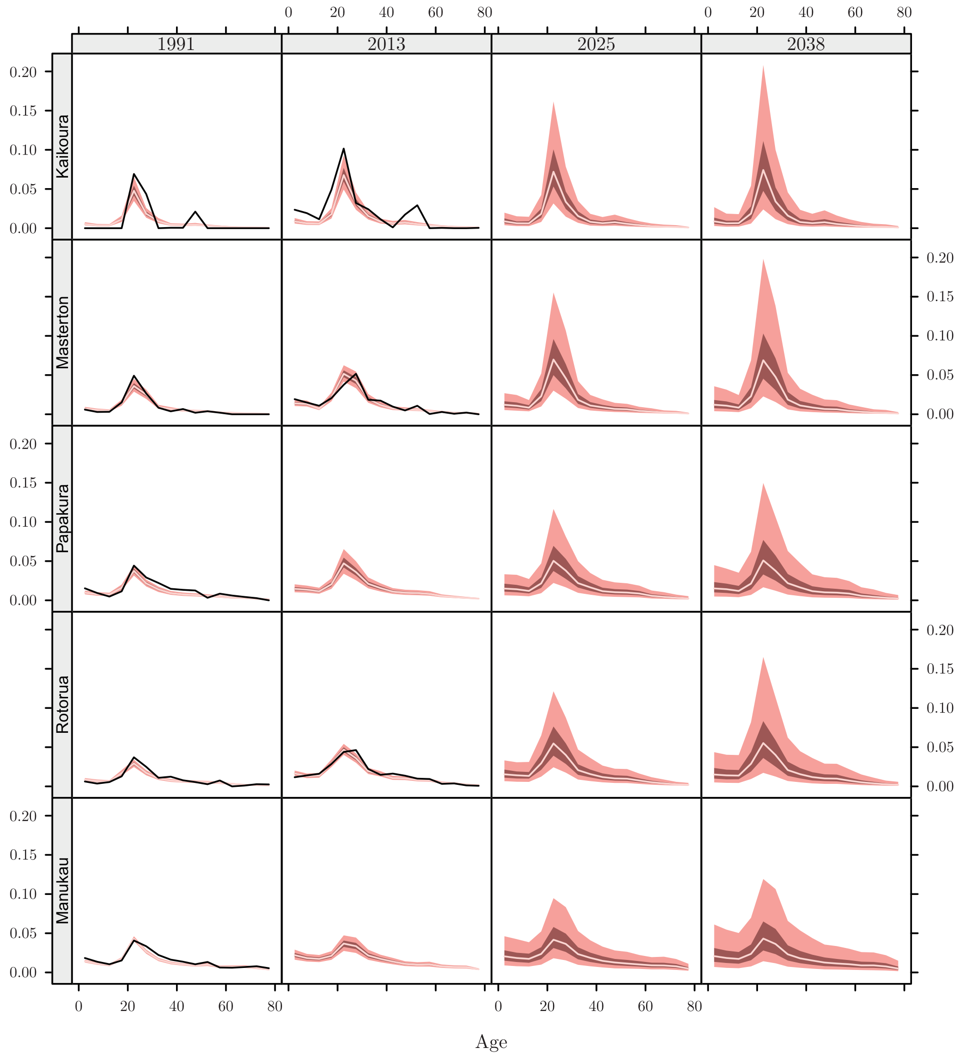

Figure 6. Estimated and forecasted emigration rates. Each panel shows emigration rates by age, for males, for a selected region and selected years. The territorial authorities are ordered by population size, with the smallest at the top. The light shading represents 90% credible intervals, the dark shading represents 50% credible intervals, and the light lines in the center show posterior medians. The black lines are direct estimates.

when using the last observed value as a forecast. Further detail on forecast accuracy is given in Figure A6 in the Supplementary Materials.

4.3. Adding age-time interaction or region-time interaction

Figure 9 presents the estimated and predicted emigration rates from the

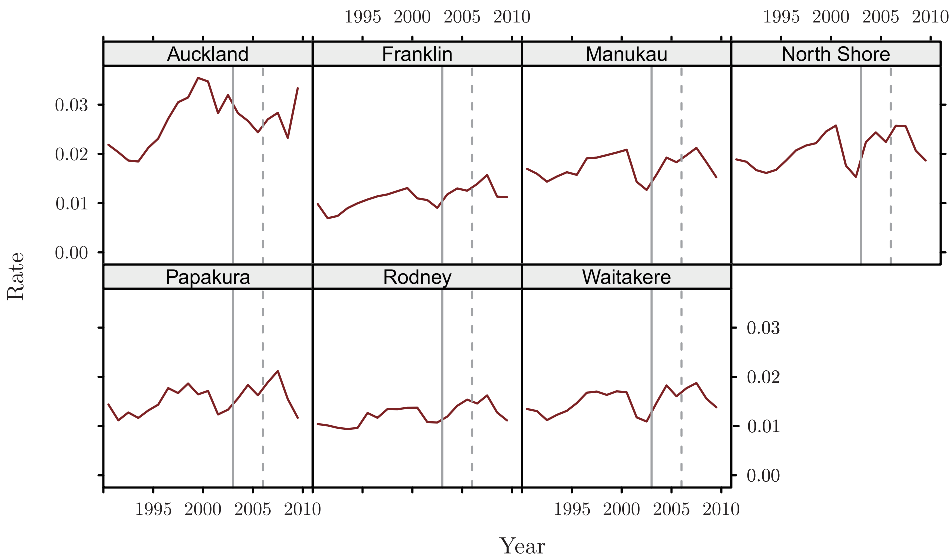

Figure 7. Direct estimates of emigration rates for seven regions within the amalgamated Auckland. Here "Auckland" in the figure refers to the pre-2010 territorial authority of Auckland City. The solid grey line indicates year 2003, the start of the imputation period in the validation exercise; and the dotted grey line indicates year 2006, the start of the prediction period in the validation exercise.

model including age-time interactions, for females in selected age groups and regions. For comparison, it also shows 90% credible intervals from the basic model. Adding age-time interactions appears to have little effect. This is presumably because age profiles are relatively stable in our dataset; different datasets could be expected to yield different results.

Figure 10 presents the estimated and predicted emigration rates from the model including region-time interaction. In contrast to the age-time case, adding a region-time interaction has a substantial effect on posterior distributions. In particular, it increases mean forecasted values for Kaikoura and Queenstown-Lakes to levels that are much more plausible. Adding the region-time interaction increases the widths of prediction intervals. This makes sense: if it cannot be assumed that regional share of emigration will remain constant over time, then there is more uncertainty.

The model with region-time interactions exhibits slightly greater over-coverage than does the basic model. As can be seen in Table 2, the 50% prediction intervals contain the true value 60.1% of the time, and the 90% prediction intervals contain the true value 92.3% of the time. Coverage remains lower in greater Auckland than in other parts of the country.

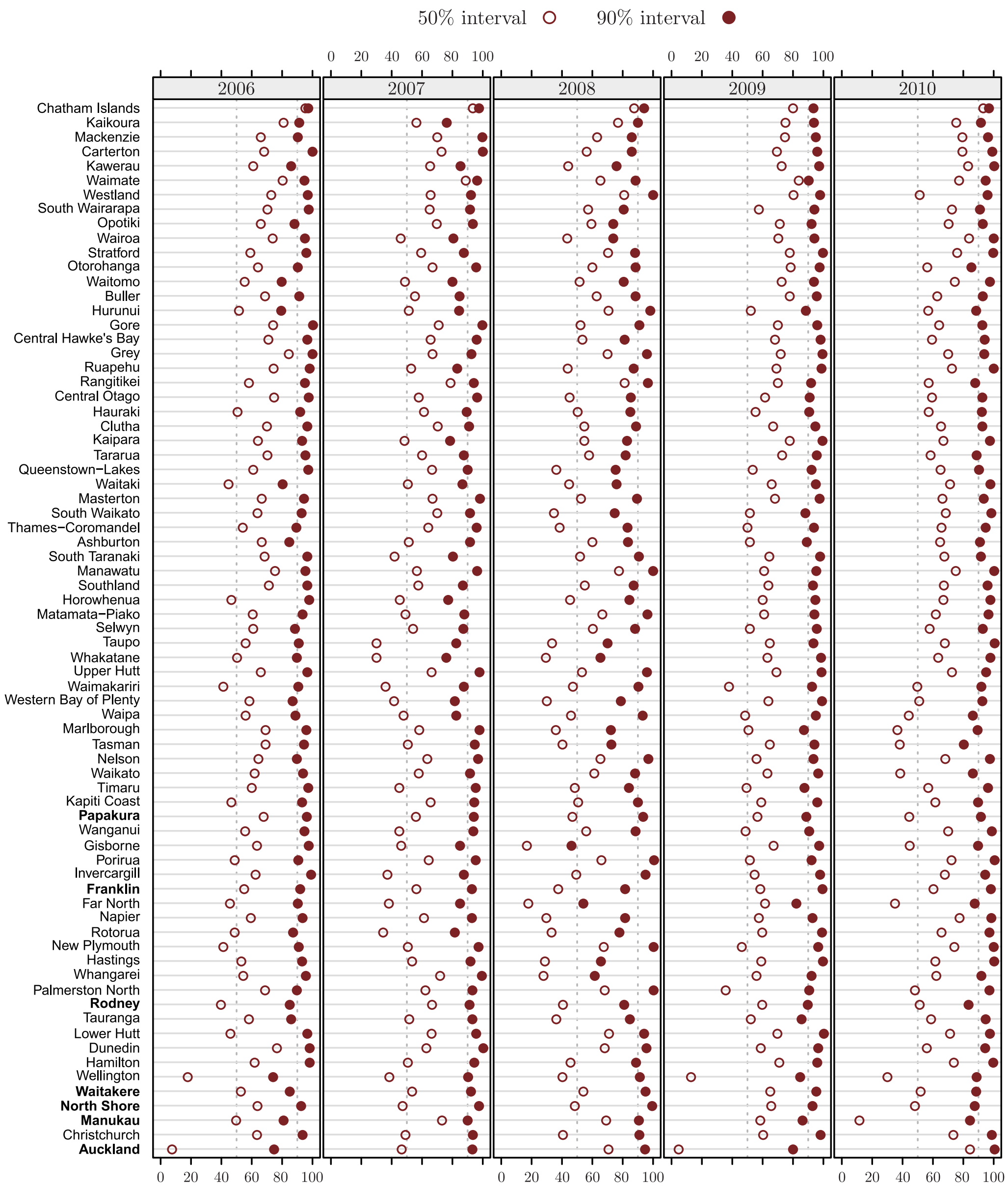

Figure 8. Percent of held-back y_{asrt} values falling within 50% and 90% prediction intervals, by territorial authority and time. The territorial authorities are order from smallest to largest, with the smallest at the top. The seven regions in bold are the ones within the amalgamated Auckland.

5. Discussion

In this paper, we have presented a new approach to subnational demographic

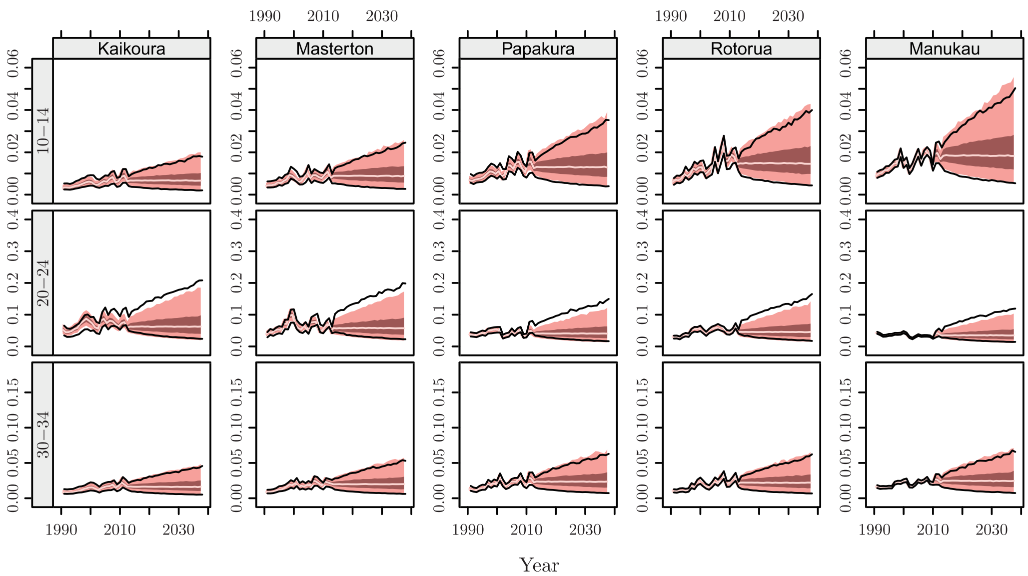

Figure 9. Estimated and predicted emigration rates from the model including age-time interaction, for females in selected age groups and territorial authorities. The light shading represents 90% credible intervals, the dark shading represents 50% credible intervals, and the light lines in the center show posterior medians. The black lines show 90% credible intervals from the basic model.

Table 2. Percents of prediction intervals containing true values: Model including region-time interaction.

	50% intervals	90% intervals
In Greater Auckland	53.7	92.5
Outside Greater Auckland	60.8	92.3
Overall	60.2	92.3

forecasting that combines ideas from demography, small area estimation, and time series statistics. We have constructed forecasts that behave sensibly in the face of very small cells counts, that avoid much of the manual adjustment required by traditional methods, and that acknowledge and combine multiple source of uncertainty. The new approach is, however, computationally demanding, which imposes difficult tradeoffs between realism and tractability. For instance, the results from the validation suggest that the model needs to be further refined and extended. However, computation times are already long enough to place practical constraints on model-building.

In our model, when t is a time in the past, the posterior distribution for rate λ_{asrt} is a compromise between the higher-level model for $\log \lambda_{rsat}$, and the direct estimate y_{rsat}/x_{rsat}. However, when t is a time in the future, the posterior

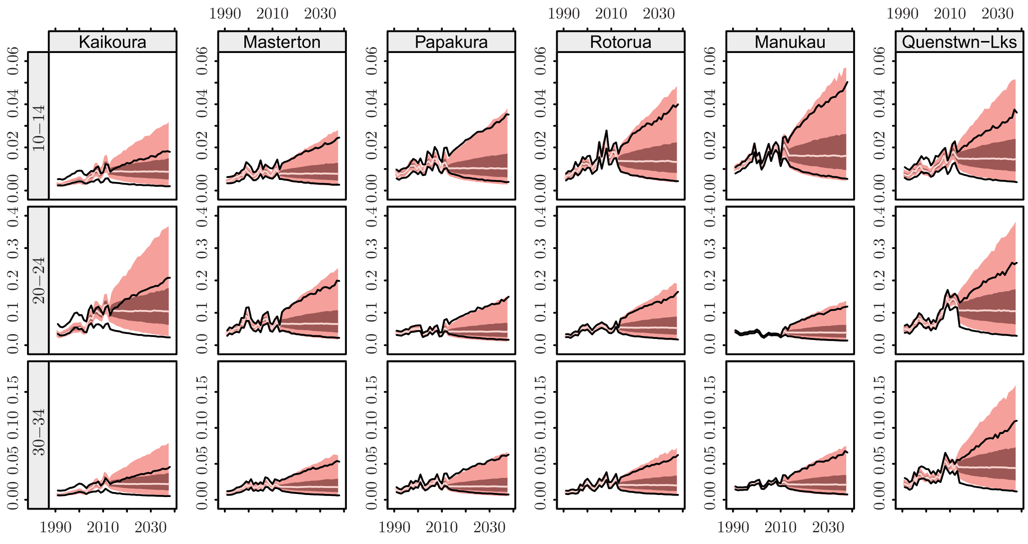

Figure 10. Estimated and predicted emigration rates from the model including region-time interaction, for females in selected age groups and territorial authorities. The territorial authority "Queenstown-Lakes" is included because it appears to have a particularly strong region-time interaction. The light shading represents 90% credible intervals, the dark shading represents 50% credible intervals, and the light lines in the center show posterior medians. The black lines show 90% credible intervals intervals from the basic model.

distribution reflects only the higher-level model. When forecasting for small areas, even more than when estimating, it is important that the higher-level model adequately represents the major data characteristics. Even the highest-level parameters must have realistic values. Consider, for example, the error terms v_t and w_t in the prior model for the time effect (3.4)–(3.5). If the variances for these terms are poorly estimated, then the historical estimates of the time effect will borrow too much or too little strength. Whether this matters depends on the data and application; if there are abundant data, it may have little effect on the posterior distribution of the λ_{asrt}. In contrast, obtaining good estimates of the variances of v_t and w_t is essential to forecasting, since these variances are key determinants of the widths of the prediction intervals. However, higher levels of hierarchical models are typically more difficult to estimate than lower levels, because the data provide less information on which to base the estimates (Gelman et al. (2006, p.516)). In our dataset, for instance, there are only 23 years from which to estimate time effects, even though there are tens of thousands of data points.

Future progress on small area demographic forecasting will require further development of priors representing age, sex, region, and time effects, and interactions between these effects. Examples include region priors that account for spatial correlations, and time priors that account for persistent upward or downward trends. Adapting the model to other settings or other demographic series will require new priors. When modelling internal migration, for instance, the most conceptually attractive approach is to use origin-destination matrices, where every possible combination of origin and destination receives its own effect. Under such specifications, however, the number of origin-destination effects increases with the square of the number of regions. Given that it can be difficult to fit models to large disaggregated datasets even with existing priors, future work on more complex priors will need to be mindful of computational issues. A related priority for future research is finding ways to reduce the computing time. We are, for instance, investigating the use of Hamiltonian Monte Carlo methods to speed up convergence of the Gibbs sampler (Neal (2011)).

We suspect that as the models are developed and applied, increasing use will be made of informative priors. One way that informative priors are likely to prove valuable is in estimating higher-level parameters that are difficult to estimate using only the information contained in the data to hand. For instance, informative priors may be helpful in estimating the variance of innovation terms in time series priors. A logical place to start is "weakly informative" priors (Gelman et al. (2006)). Another use for informative priors would be forecasting in situations where the past was not a reliable guide to the future. Our basic model is essentially a sophisticated form of extrapolation. This is arguably the best approach when forecasting emigration from New Zealand, where historical patterns have been relatively stable and there are no strong grounds for expecting a dramatic break in the future. It would not be the best approach when forecasting fertility in a poor country experiencing rapid socioeconomic development, where the past experience of the country is likely to be a poor guide to the future and there is a large body of international research on which to base predictions. A third possible use for informative priors is testing hypothetical scenarios, such as exploring the implications of policies to reduce emigration. Scenarios could be constructed by placing informative priors on some combination of time, age, and region effects, depending on how the the policy was to be implemented. Such scenario building would have much in common with conventional population projections.

Once the basic computational problems are addressed, questions of model selection and model checking are likely to become increasingly prominent. Standard model selection strategies, such as fitting multiple models and choosing the best model on some criterion, are not feasible at present because of long computation times. The large number of parameters complicates the task of model checking, though we have found graphs like those presented in this paper to be useful.

Another important task is studying the calibration and accuracy of longer-term forecasts from the model. The main obstacle to doing so is the absence of long time series of disaggregated data. Without such data, longer-term forecasts need to be treated with caution. In our view the best approach is to produce forecasts based on transparent statistical methods, and to be open about their limitations. There is strong demand for long-term forecasts. If statisticians refrain from producing long-term forecasts, methods that are less transparent and less statistical will be used instead.

Assuming that the methods described in this paper continue to improve, they offer important practical advantages to national statistical offices. One advantage is efficiency. Moving from traditional forecasting methods to methods like those described here would require resources for model-building, testing, and retraining of staff. However, once these costs are paid, the marginal cost of producing a forecast would be much lower than at present, because more of the process would be automated.

Another potential advantage for national statistical offices would be improved transparency. The methods proposed here do not escape the need for expert judgment. However, the judgments concern high-level modelling questions such as which interactions to include, rather than specific substantive matters, such as whether to allow a particular region to continue growing indefinitely. Decisions on high-level modelling are generally easier to explain and justify than decisions on specific substantive questions.

Another advantage of the methods proposed here is that they provide measures of uncertainty. Estimating uncertainty is difficult: for instance, small changes to specifications can induce large changes in the widths of prediction intervals. Furthermore, wider intervals will be obtained if the uncongeniality between the imputation model class and the analysis model class in multiple-imputation inference is taken into account, an issue formally addressed by Xie and Meng (2014). Much more research and experience with small area forecasting models is required. However, methods for carrying out this work are available, including validation exercises like the one presented in this paper. It is reasonable to expect that progress in measuring uncertainty can be made.

Prediction intervals from probabilistic population forecasts are often criticised as being "too wide". Such criticisms can be interpreted in several ways. One interpretation is that the model is miscalibrated, in that the true coverage of the prediction intervals is greater than the nominal coverage. A second interpretation is that the model does not use available information efficiently, and could have achieved the same level of coverage with narrower intervals. Prediction intervals that are too wide in either of these senses can potentially be improved.

A third interpretation is that the prediction intervals are wider than users of the forecasts expect or feel comfortable with. Demographers and statisticians need to be careful when responding to criticisms of this kind. Users' intuitions about reasonable sizes for prediction intervals are not necessarily well informed. Few people have experience with making small area population forecasts and then comparing them with actual trajectories, especially when the forecast horizons are 20 years are more. Moreover, decision-makers who insist on forecasts with narrow prediction intervals may regret doing so in the long run. Expanding a bridge or a hospital in response to higher-than-expected population growth is expensive.

Supplementary Materials

Section SM.1 discusses the modelling of region effects (given by (3.8) and (3.9)) in more detail. Section SM.2 provides details on the MCMC algorithm used to draw from the posterior distribution. Section SM.3 presents additional tables and figures.

Acknowledgement

The authors would like to thank the editor and the referees for their insightful and constructive comments, and also thank Jenny Harlow for C programming and discussions about software design.

References

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**, 815-839.

Bijak, J. and Wiśniowski, A. (2010). Bayesian forecasting of immigration to selected European countries by using expert knowledge. *J. Roy. Statist. Soc. Ser. A* **173**, 775-796.

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *Internat. J. Forecasting* **22**, 547-581.

Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Ann. Actuarial Sci.* **3**, 3-43.

Bulatao, R. A., Bongaarts, J., et al. (2000). *Beyond Six Billion: Forecasting the World's Population*. National Academies Press.

Cargnoni, C., Müller, P. and West, M. (1997). Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models. *J. Amer. Statist. Assoc.* **92**, 640-647.

Congdon, P. (2014). Estimating life expectancies for US small areas: a regression framework. *J. Geograph. Systems* **16**, 1-18.

Courgeau, D. (1985). Interaction between spatial mobility, family and career life-cycle: a French survey. *European Sociolog. Rev.* **1**, 139-162.

Fienberg, S. E. et al. (2011). Bayesian models and methods in public policy and government settings. *Statist. Sci.* **26**, 212-226.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2014). *Bayesian Data Analysis. Third Edition.* CRC press.

Gelman, A. et al. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **1**, 515-534.

Gelman, A. and Hill, J. (2006). *Data Analysis using Regression and Multilevel/Hierarchical Models.* Cambridge University Press.

Gerland, P., Raftery, A. E., Ševčíková, H., Li, N., Gu, D., Spoorenberg, T., Alkema, L., Fosdick, B. K., Chunn, J., Lalic, N., et al. (2014). World population stabilization unlikely this century. *Science* **346**, 234-237.

Gregory, I. N., Marti-Henneberg, J. and Tapiador, F. J. (2010). Modelling long-term pan-European population change from 1870 to 2000 by using geographical information systems. *J. Roy. Statist. Soc. Ser. A* **173**, 31-50.

Heitjan, D. F. and Little, R. J. (1991). Multiple imputation for the fatal accident reporting system. *Appl. Statist.* **40**, 13-29.

Hyndman, R. J. and Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *Internat. J. Forecasting* **24**, 323-342.

Keilman, N. (2008). Using deterministic and probabilistic population forecasts. In *Complexity: Interdisciplinary Communications* (Edited by W. Østreng), 22-28. Centre for Advanced Study, Oslo.

Lee, R. D. (1998). Probabilistic approaches to population forecasting. *Population and Development Review*, 156-190.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting us mortality. *J. Amer. Statist. Assoc.* **87**, 659-671.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2.

Petris, G. (2010). An R package for dynamic linear models. *J. Statist. Software* **36**, 1-16.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statist. Sci.* **28**, 40-68.

Prado, R. and West, M. (2010). *Time Series: Modeling, Computation, and Inference.* CRC Press.

Preston, S., Heuveline, P. and Guillot, M. (2001). *Demography: Modelling and Measuring Population Processes.* Blackwell, Oxford.

Raftery, A., Li, N., Ševčíková, H., Gerland, P. and Heilig, G. (2012). Bayesian probabilistic population projections for all countries. *Proc. Nat. Acad. Sci.* **109**, 13915-13921.

Rao, J. N. (2003). *Small Area Estimation.* Wiley, New York.

Rees, P. (1985). Population structure and models. In *Choices in the Construction of Regional Population Projections* (Edited by R. Woods and P. Rees). Allen and Unwin, London.

Rogers, A. and Castro, L. J. (1981). *Model Migration Schedules.* International Institute for Applied Systems Analysis [IIASA] Laxenburg Austria.

Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys.* Wiley.

Rubin, D. B. (1988). An overview of multiple imputation. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 79-84.

Siegel, J. S. (2002). *Applied Demography: Applications to Business, Government, Law, and Public Policy.* Academic Press.

Statistics New Zealand (2014). 2013 census meshblock dataset. [Online; accessed 25 April 2014].

Stoto, M. A. (1983). The accuracy of population projections. *J. Amer. Statist. Assoc.* **78**, 13-20.

Wilson, T. (2010). Model migration schedules incorporating student migration peaks. *Demographic Research* **23**, 191-222.

Wilson, T. (2014). New evaluations of simple models for small area population forecasts. *Population, Space, and Place* **21**, 335-353.

Xie, X. and Meng, X.-L. (2014). Dissecting multiple imputation from a multi-phase inference perspective: What happens when god's, imputer's and analyst's models are uncongenial? *Statist. Sinica.*

Statistics New Zealand, 120 Hereford Street, Private Bag 4741, Christchurch 8011, New Zealand.

E-mail: john.r.bryant@gmail.com

Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, P.R.China 100871.

E-mail: zjn@gsm.pku.edu.cn

(Received May 2014; accepted January 2016)