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

Supplementary Materials

SM.1 Modelling of Region Effects

By using only the covariates for one year rather than using time-varying covariates, we assume that the relative regional distributions for the percent of population born overseas and the percent of population in full-time study remain stable over time. For region r and time t, let $\widetilde{\boldsymbol{X}}_{r,t}$ denote a vector consisting of the logarithm of percent of population born overseas and the logarithm of percent of population in full-time study. We have $\boldsymbol{X}_r = \widetilde{\boldsymbol{X}}_{r,2013}$. Assume $\boldsymbol{Z}_r = \widetilde{\boldsymbol{X}}_{r,t} - \widetilde{\boldsymbol{X}}_{1,t}$ remains the same over time, where $\boldsymbol{Z}_r$ denotes the vector consisting of the logarithm of the ratio of the percent of population born overseas for region r to that for region 1 and the logarithm of the ratio of the percent of population in full-time study for region r to that for region 1. We have

$$\boldsymbol{\gamma}^\top \widetilde{\boldsymbol{X}}_{r,t} = \boldsymbol{\gamma}^\top \boldsymbol{X}_r + (\boldsymbol{\gamma}^\top \widetilde{\boldsymbol{X}}_{1,t} - \boldsymbol{\gamma}^\top \boldsymbol{X}_1), \tag{1}$$

where the term $\boldsymbol{\gamma}^\top \widetilde{\boldsymbol{X}}_{1,t} - \boldsymbol{\gamma}^\top \boldsymbol{X}_1$ can be assimilated into the intercept β^0 and the time effect β_t^{time}. This means that we do not need to use time-varying covariates $\widetilde{\boldsymbol{X}}_{r,t}$ in modelling region effects.

SM.2 Details of the MCMC Algorithm

We use a Gibbs sampler to obtain a posterior sample for the λ_{asrt} and the hyperparameters. Because of the use of normal distributions, most of the hyperparameters can be updated easily, using standard methods (**?**). The exception is the state parameters in the random walk with noise models. These are updated using the forward filtering backward sampling (FFBS) algorithm, described in Chapter 4 of **?**.

The λ_{asrt} are updated one at a time using the Metropolis-Hastings algorithm. Let $\lambda_{asrt}^{(n)}$ denote the nth posterior draw of λ_{asrt}, and let $\eta_{asrt}^{(n)} = \log \lambda_{asrt}^{(n)}$. Draw a proposal of $\eta_{asrt}^* = \log \lambda_{asrt}^*$ from a normal distribution with mean $\eta_{asrt}^{(n)}$ and variance τ^2. Let $\lambda_{asrt}^{(n+1)} = \lambda_{asrt}^*$ with probability equal to p_{MH}, and let $\lambda_{asrt}^{(n+1)} = \lambda_{asrt}^{(n)}$ otherwise. When r is not in greater Auckland or t is in years other than 2011-2013 such that y_{asrt} is observed,

$$p_{\text{MH}} = \min\left(1, \frac{\text{Poisson}(y_{asrt}|\lambda_{asrt}^* x_{asrt})}{\text{Poisson}(y_{asrt}|\lambda_{asrt}^{(n)} x_{asrt})} \frac{N(\eta_{asrt}^*|\mu_{asrt}^{(n)}, \sigma_\epsilon^{2(n)})}{N(\eta_{asrt}^{(n)}|\mu_{asrt}^{(n)}, \sigma_\epsilon^{2(n)})}\right). \tag{2}$$

For $\lambda_{as,A_j,t}$ for a territorial authority within greater Auckland during 2011-2013,

$$p_{\text{MH}} = \min\left(1, \frac{\text{Poisson}(y_{as,Au,t}| \sum_{j' \neq j} \lambda_{as,A_{j'},t}^{(n)} x_{as,A_{j'},t} + \lambda_{as,A_j,t}^* x_{as,A_j,t})}{\text{Poisson}(y_{as,Au,t}| \sum_{j' \neq j} \lambda_{as,A_{j'},t}^{(n)} x_{as,A_{j'},t} + \lambda_{as,A_j,t}^{(n)} x_{as,A_j,t})} \frac{N(\eta_{as,A_j,t}^*|\mu_{as,A_j,t}^{(n)}, \sigma_\epsilon^{2(n)})}{N(\eta_{as,A_j,t}^{(n)}|\mu_{as,A_j,t}^{(n)}, \sigma_\epsilon^{2(n)})}\right). \tag{3}$$

Here

$$\mu_{asrt}^{(n)} = \beta^{0(n)} + \beta_a^{\text{age}(n)} + \beta_s^{\text{sex}(n)} + \beta_r^{\text{reg}(n)} + \beta_t^{\text{time}(n)} + \beta_{as}^{\text{age:sex}(n)} + \beta_{ar}^{\text{age:reg}(n)}, \tag{4}$$

in the basic model, with age-time or region-time interaction added in the two extensions; Poisson($\tilde{y}|\tilde{\lambda}$) denotes the probability function of a Poisson distribution with mean $\tilde{\lambda}$ evaluated at $\tilde{y}$; and N($\tilde{\eta}|\tilde{\mu}, \tilde{\sigma}^2$) denotes the probability density function of a normal distribution with mean $\tilde{\mu}$ and variance $\tilde{\sigma}^2$ evaluated at $\tilde{\eta}$.

When working with the basic model, we use a burnin of 30,000 draws and production of 30,000 draws with four independent chains, run using parallel processing. We run the model once for each of the M imputed datasets. Using a desktop computer and $M = 10D5$, the calculations take around 60 hours. When working with the extensions we increase the burnin and production to 50,000 each. The graphs and summary statistics are based on a random sample from the posterior sample. Using the whole posterior sample is not possible because of memory constraints. Most graphs, for instance, are constructed from 1,500 iterations.

SM.3 Additional Tables and Figures

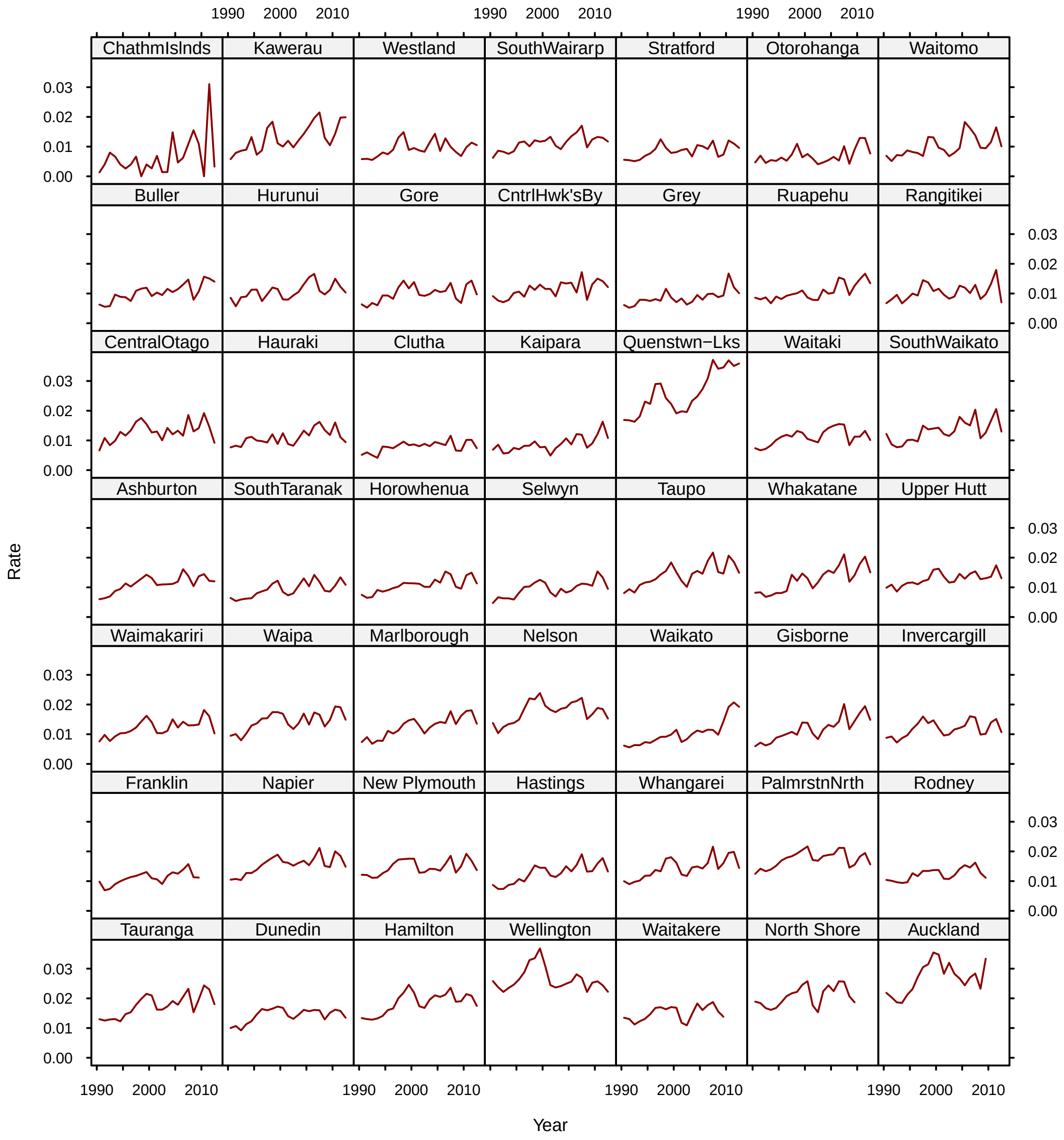

Figure A1: Direct estimates of emigration rates for 49 randomly-selected territorial authorities, 1991-2013. The territorial authorities are ordered by population size, with the smallest at the top left and the largest at the bottom right. Here "Auckland" refers to the pre-2010 territorial authority of Auckland City, not the post-2010 amalgamated region.

Table A1: Populations of territorial authorities in 2010

Territorial authority	Population	Territorial authority	Population
Chatham Islands	600	Whakatane	34,400
Kaikoura	3,800	Selwyn	39,500
Mackenzie	4,000	Upper Hutt	41,100
Kawerau	7,000	Wanganui	43,500
Carterton	7,500	Timaru	44,300
Waimate	7,600	Marlborough	45,300
Wairoa	8,400	Western Bay of Plenty	45,400
Westland	8,900	Nelson	45,500
Opotiki	9,000	Waipa	45,700
Stratford	9,200	Gisborne	46,500
Otorohanga	9,300	Tasman	47,300
South Wairarapa	9,300	Waimakariri	47,600
Waitomo	9,600	Waikato	48,300
Buller	10,000	Kapiti Coast	49,400
Hurunui	11,100	Papakura	49,800
Gore	12,300	Porirua	52,100
Central Hawke's Bay	13,500	Invercargill	52,400
Ruapehu	13,500	Napier	57,600
Grey	13,800	Far North	58,400
Rangitikei	14,900	Franklin	65,200
Clutha	17,500	Rotorua	68,600
Tararua	17,700	New Plymouth	73,200
Hauraki	17,900	Hastings	75,100
Central Otago	18,200	Whangarei	80,000
Kaipara	19,000	Palmerston North	81,300
Waitaki	20,800	Rodney	100,000
South Waikato	22,900	Lower Hutt	102,700
Masterton	23,400	Tauranga	114,300
South Taranaki	26,900	Dunedin	124,800
Thames-Coromandel	27,000	Hamilton	143,100
Queenstown-Lakes	27,800	Wellington	197,700
Ashburton	29,400	Waitakere	208,100
Southland	29,500	North Shore	229,000
Manawatu	29,700	Manukau	375,700
Horowhenua	30,600	Christchurch	376,700
Matamata-Piako	31,800	Auckland	450,200
Taupo	34,000		

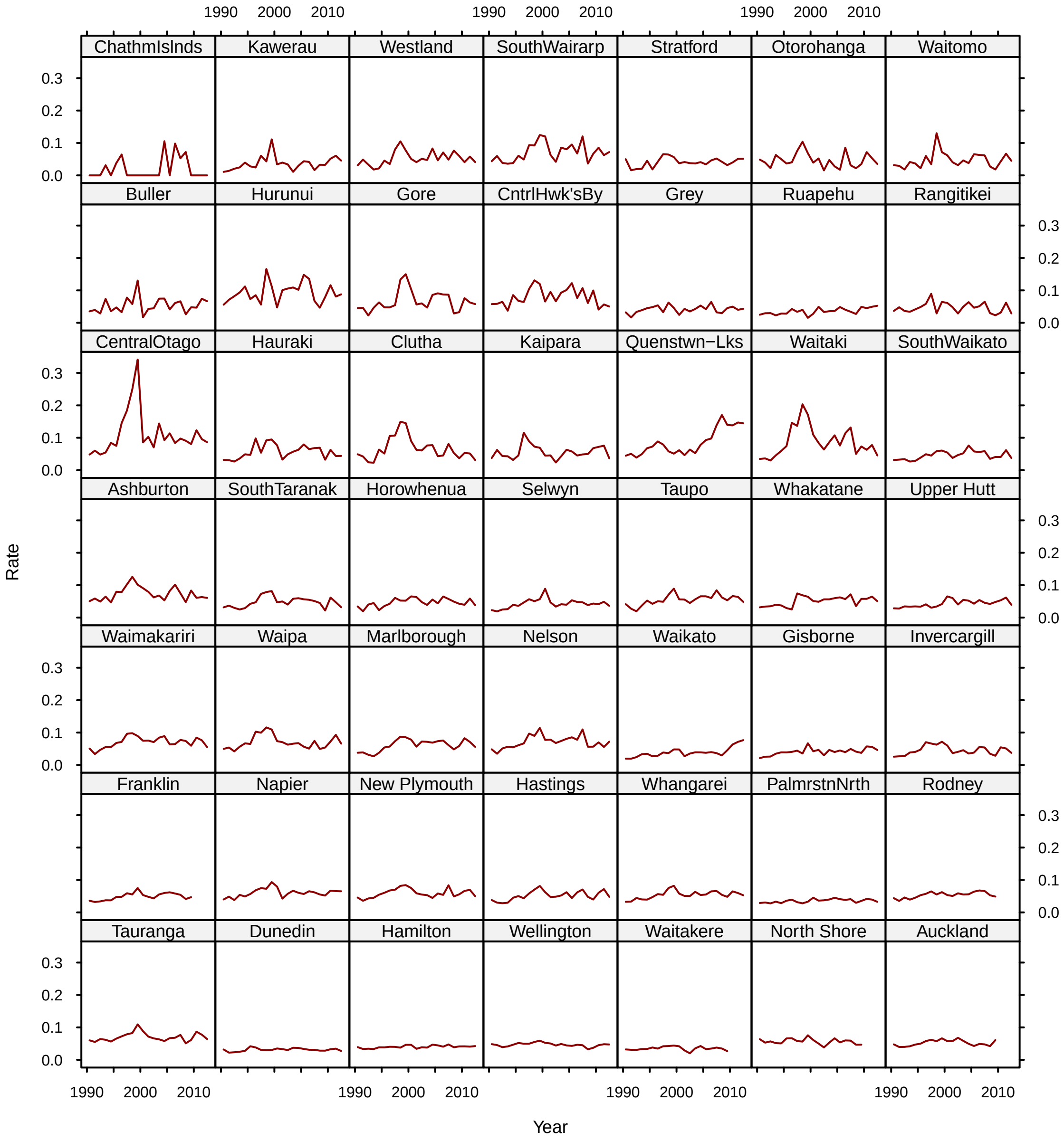

Figure A2: Direct estimates of emigration rates for females aged 20-24 in 49 randomly-selected territorial authorities, 1991-2013. The territorial authorities are ordered by population size, with the smallest at the top left and the largest at the bottom right. Here "Auckland" refers to the pre-2010 territorial authority of Auckland City, not the post-2010 amalgamated region.

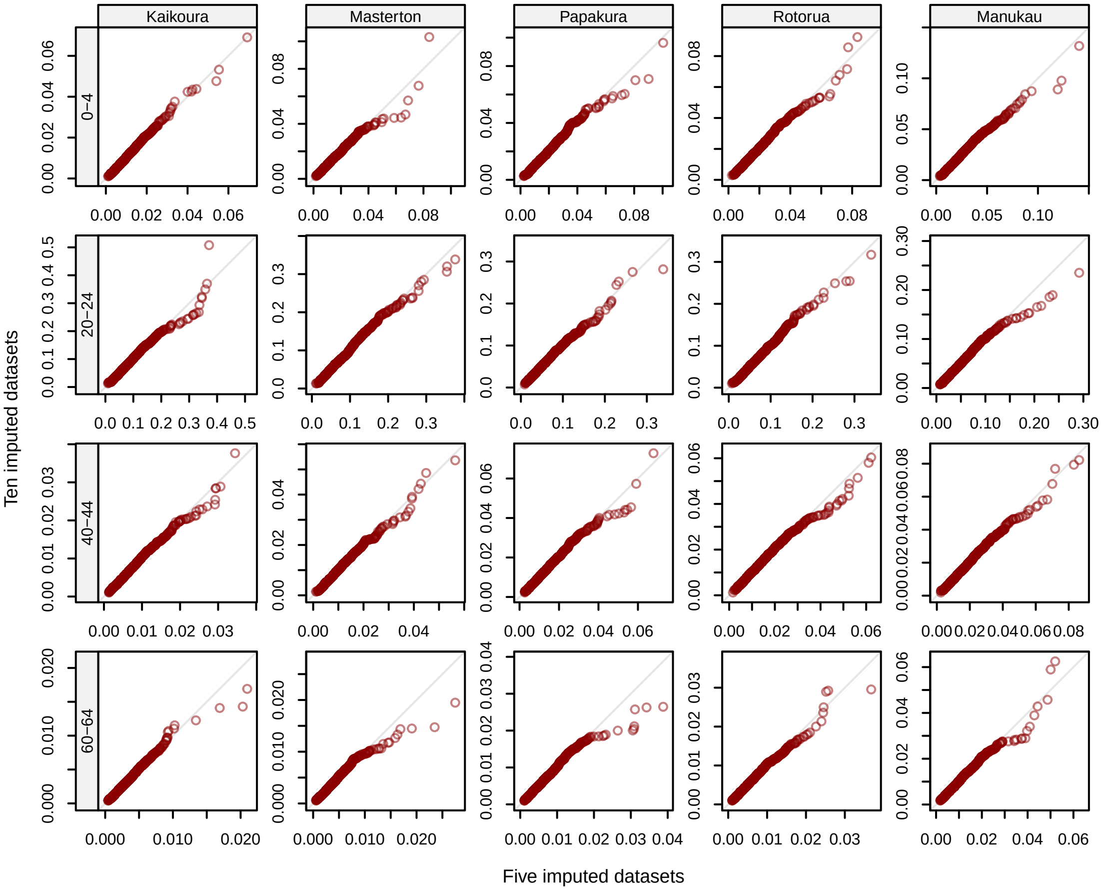

Figure A3: Quantile-quantile (qq) plots of posterior distributions of λ_{asrt} based on $M = 5$ and $M = 10$ imputed datasets. The plots are for females in 2026 (the midpoint of the forecast period), for four selected age groups and five selected territorial authorities; results for other subsets of the posterior sample were similar. A qq plot shows the quantiles of one distribution against the quantiles of a second distribution. When all points lie on the 45-degree line, the two distributions are identical. In the panels above, the points lying away from the line are all in the tails of the distributions, and are a small fraction of the 1,500 points shown per panel. The plots imply that the posterior distributions derived from $M = 5$ and $M = 10$ datasets are very similar.

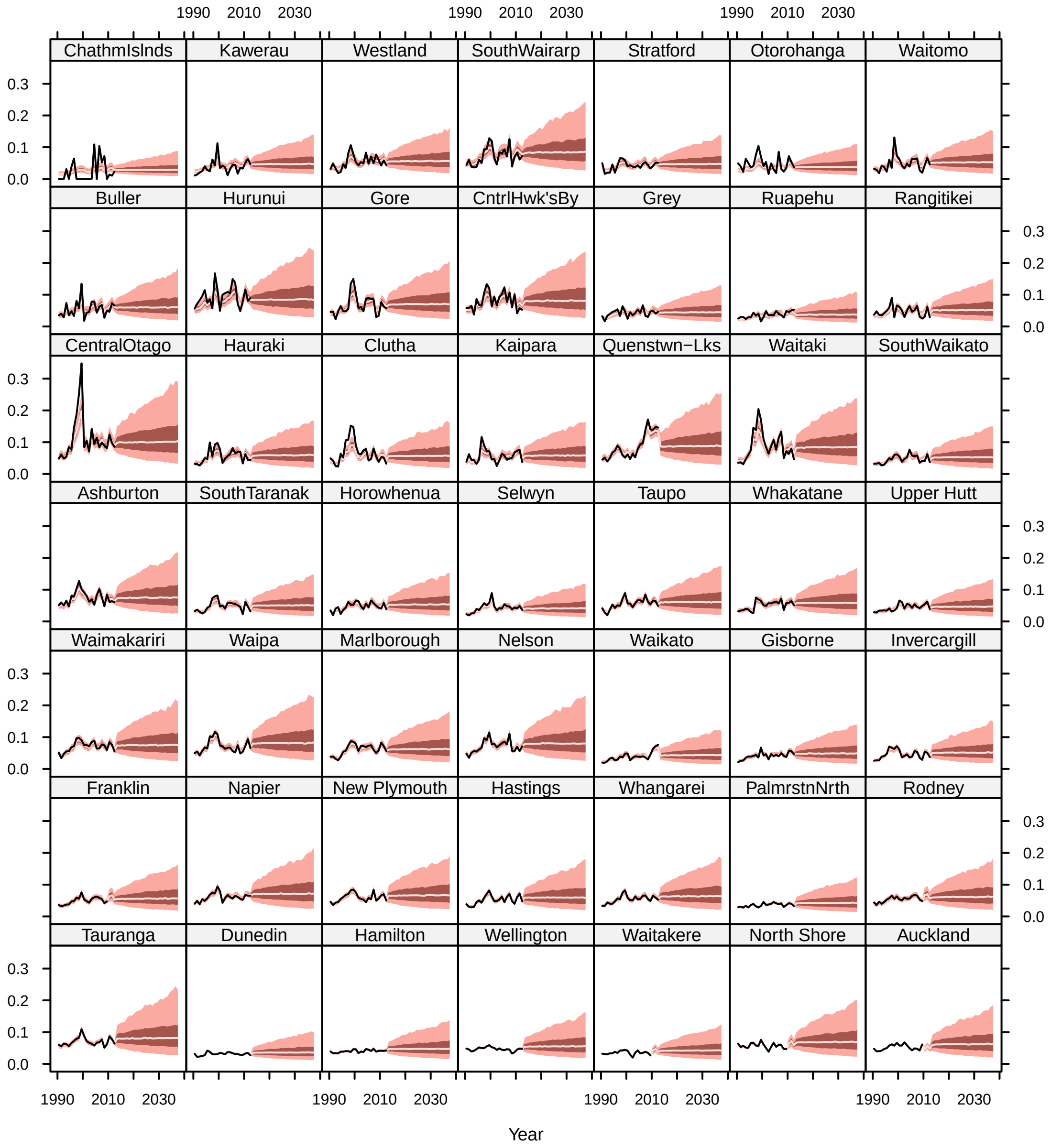

Figure A4: Estimated and forecasted emigration rates over time, for females aged 20-24 in 49 randomly-selected territorial authorities. The territorial authorities are ordered by population size, with the smallest at the top left and the largest at the bottom right. Here “Auckland” refers to the pre-2010 territorial authority of Auckland City, not the post-2010 amalgamated region. 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.

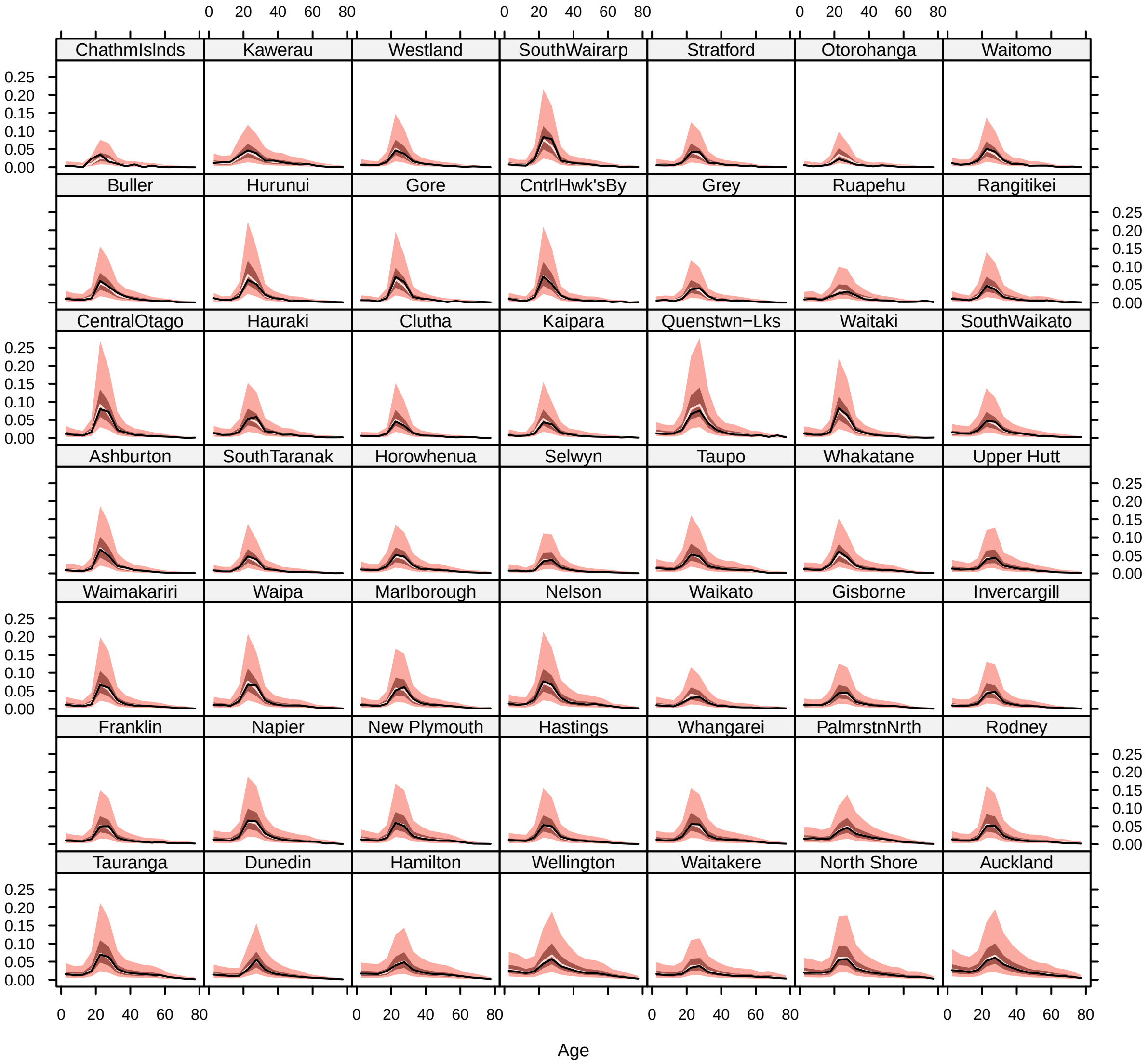

Figure A5: Forecasted age profiles for 2038, for males in 49 randomly-selected territorial authorities. The territorial authorities are ordered by population size, with the smallest at the top left and the largest at the bottom right. Here “Auckland” refers to the pre-2010 territorial authority of Auckland City, not the post-2010 amalgamated region. The light shading represents 90 % credible intervals, the dark grey shading represents 50% credible intervals, and the light lines in the center show posterior medians. For comparison, the black lines are mean direct estimates of emigration rates for the years 2000-2009.

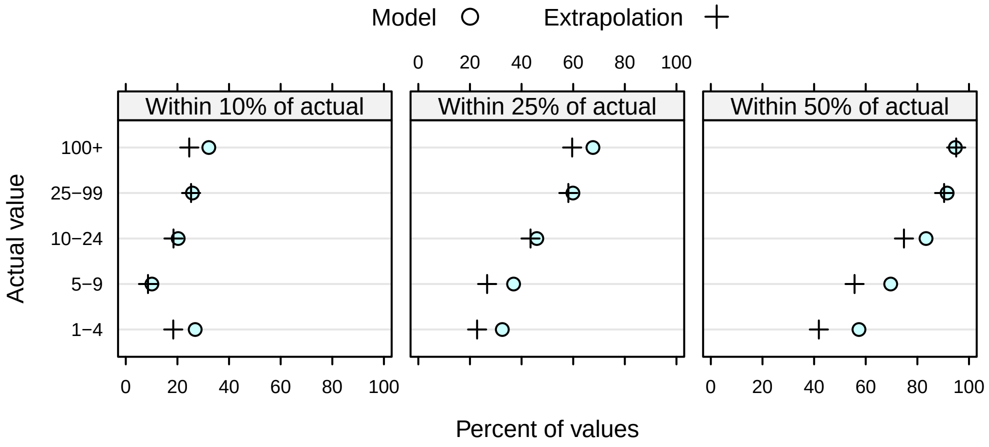

Figure A6: Percentage of forecasted values falling within the stated percentage of the actual value. The 'model' forecasts are obtained from the basic model, and the 'extrapolation' forecasts equal the most recent observed value before the forecast period. The comparisons are stratified by the size of the actual value. The figure shows, for instance, that when the actual value is between 1 and 4, 27% of model forecasts fall within 10% of this value, compared to 18% of extrapolation forecasts.