Bayesian small area demography

Junni L. Zhang John Bryant Kirsten Nissen

Abstract

Demographers are facing increasing pressure to disaggregate their estimates and forecasts by characteristics such as region, ethnicity, and income. Traditional demographic methods were designed for large samples, and perform poorly with disaggregated data. Methods based on formal Bayesian statistical models offer better performance. We illustrate with examples from a long-term project to develop Bayesian approaches to demographic estimation and forecasting. In our first example, we estimate mortality rates disaggregated by age and sex for a small population. In our second example, we simultaneously estimate and forecast obesity prevalence disaggregated by age. We conclude by addressing two traditional objections to the use of Bayesian methods in statistical agencies.

1 Introduction

Demography has traditionally been a big-data and big-area discipline. Demographers have used censuses, registration data, and surveys to obtain national-level estimates and forecasts. Big sample sizes for national populations have meant that, in contrast to most of applied statistics, sampling variation is small. Demographers have instead concentrated on other problems, such as measurement errors, and developed their own techniques and terminology distinct from mainstream statistics. Classic demographic methods combine simple deterministic models with complex expert judgements. The models are simple enough to be implemented on computer spreadsheets, but require practitioners to intervene and correct for problems caused by violation of the underlying assumptions. These methods have had many successes. They have, for instance, been used to document the dramatic fall in mortality and fertility in developed countries, and have alerted policy makers to future population ageing.

Traditional demographic methods are, however, coming under strain. The reason is the rising demand for disaggregation. Policy makers, social scientists, market researchers, and other users of demographic estimates and forecasts require ever-more disaggregated numbers. The United Nations 2030 Agenda for Sustainable Development, for instance, calls for increasing significantly "the availability of high-quality, timely and reliable data disaggregated by

income, gender, age, race, ethnicity, migratory status, disability, geographic location and other characteristics relevant in national contexts" (United Nations General Assembly, 2015, Goal 17.18). Disaggregation is challenging to traditional demography because, even when the overall population is large, the number of people in each subpopulation can be small. With these small numbers, random variation in data collection, or in underlying demographic processes such as fertility, mortality, and migration, becomes prominent, and deterministic methods break down.

To deal with these problems, demographers have been turning to mainstream statistics for new ideas on ways to deal with random variation. Similarly, statisticians have been showing an increasing interest in demographic applications. The result has been a boom in statistical demography (Alho and Spencer, 2006).

Demographic phenomena are often highly regular. Mortality, fertility, and migration rates, for instance, have characteristic age-sex profiles that are stable over time or that change in consistent ways. These regularities reflect common events over individuals' life courses. Migration rates typically peak in the late teenage years, for instance, because these are the years when people reach adulthood and begin to leave home. The ability to model units that are similar but not identical is a particular strength of Bayesian methods. Bayesians build models with multiple layers that can capture multiple, overlapping types of variability. Bayesian models pool information from

across similar units, to improve accuracy and precision.

Bayesian methods have other advantages for demographic modelling. They can coherently combine uncertainty from many sources, including random variation, missing data, and uncertainty about future trends. Bayesian methods also make it easy to construct inferences about derived quantities. Life expectancy, for instance, is a complicated nonlinear deterministic function of age-specific mortality rates, but within a Bayesian framework, deriving inferences about life expectancy from inference about age-specific mortality rates is straightforward.

Because of advantages such as these, within the field of statistical demography, there has been particularly fast growth in *Bayesian* statistical demography (Bijak and Bryant, 2016). The most prominent example has been the adoption, by the United Nations, of Bayesian methods for population forecasting (Gerland et al., 2014).

In this paper, we illustrate how Bayesian methods, and particularly Bayesian hierarchical models, can be used to obtain disaggregated demographic estimates and forecasts. The examples are drawn from a long-term project to develop Bayesian demographic methods for use in official statistics, including the development of open source software implementing the methods. In the statistical literature, the problem of obtaining estimates for domains with small sample sizes have been referred to as small area estimation (Pfeffermann, 2013; Rao and Molina, 2015). The models that we consider are all 'area-level'

models, in that they use counts and rates for disaggregated cells, rather than individual-level data. With area-level models, we can use datasets in the form of confidentialized tables that individual-level models cannot use. Demands for disaggregated estimates and forecasts are also related to groups rather than individuals.

In Section 2, we present mortality estimates for Māori, the indigenous people of New Zealand. The main inferential challenge is to capture the complex relationship between mortality and age, despite small numbers and considerable random variation. In Section 3, we interpolate and forecast obesity rates in New Zealand by age, based on survey data. The main problem here is carrying out a time series analysis with data from only five years. We conclude, in Section 4, by addressing two traditional objections to the use of Bayesian methods in statistical agencies.

2 Mortality rates for Māori

2.1 The estimation problem

Mortality rates are a fundamental measure of human welfare, as well as a major performance indicator for the health sector. Mortality rates are also a key input for population forecasts, and for the life insurance industry.

Statistics New Zealand publishes estimates of mortality rates for Māori

by sex and by one-year age groups. These rates are 'super-population' estimates. Super-population mortality rates measure the underlying risk of dying. They can be contrasted with finite-population rates, which measure the actual number of deaths divided by the actual population at risk. Suppose, for instance, that no 6-year-old Māori die in a particular year. The finite-population mortality rate is exactly zero, but the underlying risk of dying, and hence the super-population mortality rate, is presumably non-zero.

To derive death rates we need death counts and measures of population at risk. New Zealand's death registration system is efficient and complete, and reporting of ethnicity on the death registrations is generally reliable (Bryant and Howard, 2017), so data on death counts can be treated as error-free. Finding appropriate measures of population at risk is more challenging. Population at risk is measured using person-years. For instance, if a person spends 9 months in New Zealand during the period of interest, then that person contributes 0.75 person-years to the population at risk. Ideally, population at risk would be obtained by summing up person-years contributed by each person in the population. However, such data can be difficult to obtain. Instead, demographers typically approximate population at risk using population count multiplied by length of period. Population counts for Māori in New Zealand are relatively accurate for census years (Bryant et al., 2016), but become less accurate away from census years, because it is not possible

to tell, from international migration data, how many Māori are entering and leaving the country. In addition, Statistics New Zealand does not treat ethnicity as a characteristic that is fixed at birth, but rather as an aspect of personal identity that individuals can change over their lifetimes.

In response to the difficulties in estimating Māori population counts outside census years, Statistics New Zealand focuses on periods centered on census years. Censuses are normally carried out every 5 years in New Zealand, though the 2011 census was postponed until 2013 because of an earthquake. The standard approach to mortality estimation is to use three-year periods, centered on a census year, such as 2012–2014. Using a three-year period gives larger numbers of death counts in each age-sex cell, and hence more stable estimates, than would be the case with single-year periods. To approximate the population at risk over a three-year period, Statistics New Zealand uses the population count at the middle of the period, that is on June 30 of the census year, multiplied by 3.

To give an idea of the modelling challenge, Figure 1 shows direct estimates of mortality rates on a log scale for Māori males in 2012–2014, for single-year age groups 0, 1, ..., 100+. Direct estimates of mortality rates are simply death counts for each age-sex cell divided by the population at risk for that cell. The diameter of each circle in Figure 1 is proportional to the square root of the number of deaths. Altogether, there were 9,170 deaths during the period, with the largest cell consisting of 130 deaths, two cells having 0

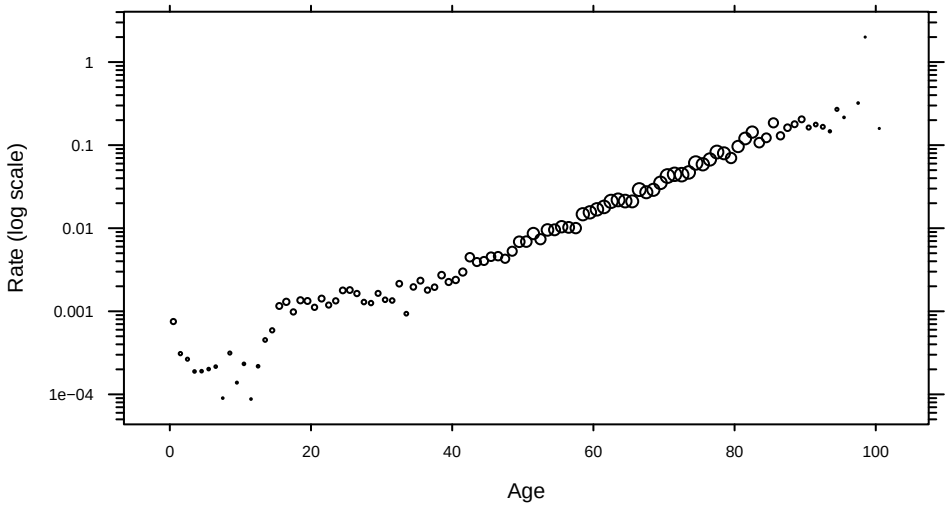

Figure 1: Direct estimates of mortality rates for Māori males in 2012–2014, by single year of age. The diameter of the circles are proportional to the square root of the number of deaths during the period.

deaths, and a median death count of 27. The Māori male population on June 30, 2013 was 328,000, giving a population at risk of $328,000 \times 3 = 984,000$ person-years.

Estimating underlying death rates between ages 40 and 90 is relatively easy. There are plenty of data, and, when shown on a log scale, the rates appear to fall on a straight line.

Somewhere between age 10 and age 20, death rates rise sharply, and then climb slowly up to about age 35. Many countries have a similar pattern of unusually high mortality rates in the late teenage years and 20s, particularly

among males. The phenomenon is referred to as the "accident hump" (meaning, mainly, car accidents), though in many places, including New Zealand, it would be more accurate to call it an accident and suicide hump.

Death rates are relatively high during the first year of life, before falling to very low levels. Exactly how low these rates go is difficult to tell, because death counts are small and the associated direct estimates are highly erratic. The same problem also exists above age 90, where trends in death rates are difficult to pin down.

The death rate for 99-year-olds is over 1. This implies that the number of deaths of 99-year-olds is greater than the (approximate) number of person-years lived during the period 2012–2014 by 99-year-olds. Rates, unlike probabilities, have no upper bound. Consider, for instance, a population consisting of one person, who dies 9 months into a one-year period. The implied death rate for that period is $1/0.75 \approx 1.33$.

2.2 The model

2.2.1 Model specification

Our input data are death counts y_{ast} and population at risk n_{ast}. Subscript a denotes age group; subscript s denotes sex; and subscript t denotes time, taking values 2005–2007 and 2012–2014. Using two periods allows us to borrow strength across time, and also to study change over time.

We model death counts as draws from a Poisson distribution,

$$y_{ast} \sim \text{Poisson}(\gamma_{ast} n_{ast}), \tag{1}$$

where γ_{ast} is the super-population mortality rate. We calculate n_{ast} by multiplying the population at June 30 in the census years by 3, and treat it as error-free. The main goal of the modelling is to estimate γ_{ast}.

Traditionally, demographers have ignored the fact that, even after knowing the population at risk and the underlying death rate, the actual number of deaths is still random and therefore uncertain. With large cell counts, such as for national populations, this uncertainty is small, so ignoring it is sensible. With small cell counts, however, this uncertainty is substantial, and needs to be accounted for. We do this by treating y_{ast} as a random draw from a Poisson distribution.

We add to (1) assumptions about how γ_{ast} is likely to vary. In Bayesian terminology, we specify a prior model for the γ_{ast}. Because γ_{ast} is positive with no upper bound, we specify the model on a log scale. We assume that γ_{ast} varies systematically by age, sex, and time, with age patterns potentially differing between females and males,

$$\log \gamma_{ast} = \beta^0 + \beta_a^{\text{age}} + \beta_s^{\text{sex}} + \beta_t^{\text{time}} + \beta_{as}^{\text{age:sex}} + e_{ast}. \tag{2}$$

Here, β^0 is an intercept, capturing the overall level of log mortality rates, β_a^{age} is an age effect, capturing variation across age, β_s^{sex} is a sex effect,

capturing variation between sexes, β_t^{time} is a time effect, capturing common time trends, and $\beta_{as}^{\text{age:sex}}$ is an age-sex interaction, capturing variation between sexes in the age pattern. The presence of the error term e_{ast}, implies that we do not expect our prior model to predict $\log \gamma_{ast}$ with complete accuracy. Standard generalized linear models do not have an equivalent term, and thus are implicitly making stronger assumptions about the correctness of the model. We assume that the error term e_{ast} has a normal distribution with mean 0 and variance σ^2. The higher the value of σ^2, the less the implied accuracy of the prior model.

The most importance source of variation in mortality rates is age. As is apparent in Figure 1, mortality rates for people in the 90s are three or four orders of magnitude higher than mortality rates for young children. It is therefore crucial for accurate estimation that we capture the main features of the age pattern.

We model age effects using an approach originally developed for modelling change over time rather than age, a 'local trend' model (Prado and West, 2010, pp. 119-121),

$$\beta_a^{\text{age}} = \alpha_a^{\text{age}} + 1(a = 0)\psi + u_a^{\text{age}}, \tag{3}$$

$$\alpha_a^{\text{age}} = \alpha_{a-1}^{\text{age}} + \delta_{a-1}^{\text{age}} + v_a^{\text{age}}, \tag{4}$$

$$\delta_a^{\text{age}} = \delta_{a-1}^{\text{age}} + w_a^{\text{age}}. \tag{5}$$

Use of time series models to capture variation over age is relatively common

in statistical demography. The fundamental idea is that values for neighboring age groups, like values for neighboring time periods, are more highly correlated than values for age groups or time periods that are distant from one another.

Equation (3) says that age effects are a combination of underlying level, captured by α_a^{age}, and age-specific idiosyncratic effects, captured by error term u_a^{age}. Age group 0 typically has much higher mortality rates than those for other young age groups, reflecting the special risks faced by infants. This extra mortality is modelled by parameter ψ. Equation (4) says that the level effect at age a equals the level effect at age $a-1$, plus an increment $\delta_{a-1}^{\text{age}}$, plus an idiosyncratic error v_a^{age}. Equation (5) says that the increment at age a, δ_a^{age}, equals the increment at age $a-1$, $\delta_{a-1}^{\text{age}}$, plus an idiosyncratic error w_a^{age}. Under a local trend model, age effects are expected to rise or fall linearly, but the slope of the line can change, or even reverse direction, over the whole length of the age pattern. Our priors for the starting values of α^{age} and δ^{age} are $\alpha_0^{\text{age}} \sim \text{N}(0, 10^2)$ and $\delta_0^{\text{age}} \sim \text{N}(0, 1)$.

The age-sex interaction term $\beta_{as}^{\text{age:sex}}$ measures variation between sexes in the age pattern for mortality. We use a 'local level' model (Prado and West, 2010, pp. 119-121),

$$\beta_{as}^{\text{age:sex}} = \alpha_{as}^{\text{age:sex}} + u_{as}^{\text{age:sex}}, \tag{6}$$

$$\alpha_{as}^{\text{age:sex}} = \alpha_{a-1,s}^{\text{age:sex}} + v_{as}^{\text{age:sex}}, \tag{7}$$

$s \in \{\text{Female}, \text{Male}\}$. This model expresses the idea that, after accounting for age effects and sex effects, the residuals for mortality rates will be similar between neighbouring age groups, within each sex. The lack of a trend term (δ) implies that we do not expect these residuals to systematically trend upwards or downwards across the age range. We assume that any systematic trend will be shared by both sexes, and hence will be accounted for by the trend term in the age effect. Our prior for the starting value of $\alpha^{\text{age:sex}}$ is $\alpha_0^{\text{age:sex}} \sim \text{N}(0, 10^2)$.

We use a simple model for sex effects,

$$\beta_s^{\text{sex}} \sim \text{N}(0, 1), \tag{8}$$

$s \in \{\text{Female}, \text{Male}\}$. This implies that we expect that the mean difference between mortality rates for sex s and the average mortality rates for both sexes to lie within the range $(-2, 2)$ on a log scale. The variance of the female-male differences is $1 + 1 = 2$, so we expect this difference to lie within the range $(-2\sqrt{2}, 2\sqrt{2})$ on a log scale , or $(0.06, 16.9)$ on the original scale, which is a very large range compared to actual sex differences. This is an example of a 'weakly informative' prior, in that it understates the actual strength of existing scientific knowledge (Gelman et al., 2008). Weakly informative priors provide many of the benefits of strong priors, by ruling out implausible values, and speeding up computations. However, they are much more convenient, since they do not require the analyst to precisely summarize

external information about the parameter in question, which can be difficult.

As there are only two time periods, and hence insufficient information to warrant a complicated model for time effects, we simply assume that $\beta_t^{\text{time}} \sim \text{N}(0, 1)$. We assume $\beta^0 \sim \text{N}(0, 10^2)$.

All the error terms in our model (e_{ast}, u_a^{age}, v_a^{age}, w_a^{age}, $u_{as}^{\text{age:sex}}$, and $v_{as}^{\text{age:sex}}$) have normal distributions with mean 0. The standard deviation parameters for the error terms e_{ast}, u_a^{age}, v_a^{age} and w_a^{age} all have a half-t distribution, with 7 degrees of freedom and scale parameter 1. Figure 2 shows a half-t distribution with 7 degrees of freedom and scale parameter 1. The distribution puts a 65% probability on values below 1, and a 2% probability on values exceeding 3.

In practice, we expect the standard deviation of our error terms to be well under 1. The standard deviation governs the size of age-to-age, sex-to-sex, or time-to-time differences in rates. A standard deviation of 1 implies that we would often see differences of 100% or more, which we do not see in practice. Our prior for standard deviations is therefore weakly informative.

The standard deviation parameters for the error terms $u_{as}^{\text{age:sex}}$ and $v_{as}^{\text{age:sex}}$ in the age-sex interaction have a half-t distribution, with 7 degrees of freedom and scale parameter 0.5. We use a smaller scale for the interaction on the principle that interactions are typically smaller in size than main effects (Gelman et al., 2008).

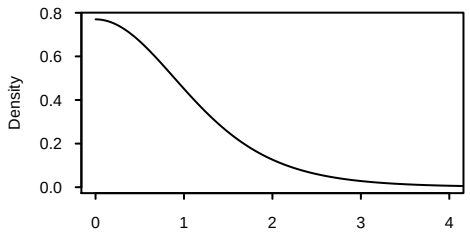

Figure 2: A half-t distribution with 7 degrees of freedom and scale 1.

2.2.2 Model output

The output from a Bayesian analysis is a sample from the posterior distribution of all of the parameters conditional on the data. The parameters include mortality rates, disaggregated by age, sex, and time, parameters in the model for age effects, parameters in the model for sex effects, and so on. Our main interest lies in the mortality rates. The mortality rates are well identified from the data. The main effects and interactions in the prior model, in contrast, are only weakly identified. We discuss the identification issue further in Appendix A.

We simulate draws from the posterior distribution using 4 independent chains, each with a burnin of 100,000 and production of 100,000. We keep every 250th iteration from each chain, yielding a combined sample of $S = 1,600$ draws. We monitor convergence using potential scale reduction factors (Gelman et al., 2014, Section 11.4). The calculations are done in our own open source R package **demest**. Sample code is shown in Appendix B.

For any given parameter, we use the median of its posterior draws as the

point estimate, and use the $100p\%$ $(0 < p < 1)$ credible interval formed by the $[50(1-p)]$ and $[50+50p]$ percentiles of its posterior draws to measure the uncertainty. For instance, a 95% credible interval with $p = 0.95$ is formed by the 2.5% and 97.5% percentiles of the posterior draws.

The posterior draws can easily be used to construct estimates of functions of the model parameters, together with measures of uncertainty. In the study of mortality, a particularly important example is life expectancy at birth. Life expectancy is a complicated nonlinear function of age-specific mortality rates (Preston et al., 2001). Let $f(\boldsymbol{\gamma})$ denote the nonlinear function that produces life expectancy from a set of age-specific mortality rates $\boldsymbol{\gamma}$. If $\boldsymbol{\gamma}^{(1)}, \cdots, \boldsymbol{\gamma}^{(S)}$ represent a sample from the posterior distribution of $\boldsymbol{\gamma}$, then $f(\boldsymbol{\gamma}^{(1)}), \cdots, f(\boldsymbol{\gamma}^{(S)})$ form a sample from the posterior distribution of life expectancy. We can summarize this sample to get point estimates and credible intervals of life expectancy.

Our approach is fully Bayesian in that, in addition to specifying a prior for γ_{ast}, we also specify priors for hyper-parameters, such as σ^2, that govern the prior for γ_{ast}. Inferences about the hyper-parameters are made together with inferences about γ_{ast}, using the joint posterior distribution. An alternative approach, known as Empirical Bayes, is to construct point estimates for the hyper-parameters and make inferences about γ_{ast} conditional on these point estimates (Rao and Molina, 2015, Chapter 9).

Empirical Bayes approaches can be less computationally intensive than

fully Bayesian ones, which means they sometimes scale better to large datasets. They can, however, be difficult to implement with complicated models containing many levels, such as ours. Using probability distributions, rather than point estimates, for hyper-parameters also leads to a more complete representation of uncertainty.

2.3 Results

Figure 3 shows results from the model. The light blue band represents 95% credible intervals. If the assumptions of the model are met, then each vertical slice of the band has a 95% probability of containing the true value for γ_{ast}. The pale line in the middle is the median of the posterior distribution, which can be used for point estimates. The black circles are the direct estimates from Figure 1.

The age pattern obtained from the model is approximately linear over ages 40–80. The model successfully smooths through the random variation in the direct estimates. Around age 18, however, the slope changes abruptly, marking the beginning of the accident hump. The smoothness at ages 40 and over does not come at the expense of an ability to detect local changes in the teenage years. The model also makes no attempt to smooth away the spike in mortality at age 0. This is a result of the inclusion of a covariate for age 0: models that do not have this term do partly smooth away the spike

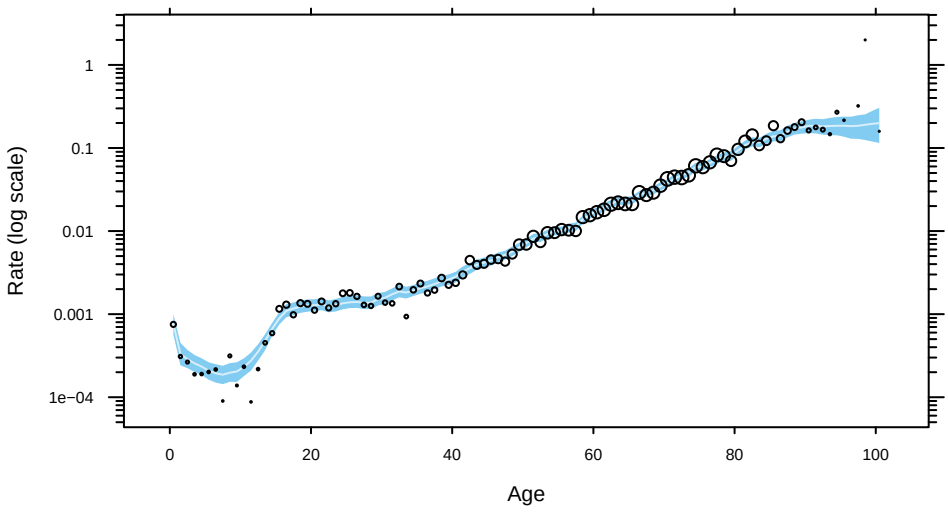

Figure 3: Modelled estimates of mortality rates for Māori males in 2012–2014, by single year of age. The light blue band represents 95% credible intervals, and the white line represents posterior medians. The black circles represent direct estimates.

(results not shown.)

Estimates around age 10 are, on a log scale, less precise than for most other age groups, reflecting the small cell counts for children. In other words, the model produces uncertainty measures that reflect local availability of data.

Uncertainty also increases steadily beyond age 90, as death counts become smaller and smaller. The posterior median suggests little increase in death rates beyond age 90. The apparent plateauing in mortality rates may be genuine: Māori males who survive to age 90 may systematically differ from ones who do not, so that the flat mortality for people at high ages reflects a kind of selection effect (Vaupel et al., 1979). However, it is also possible that the plateauing reflects problems with the input data, such as inaccurate recording of ages of very old peopleg.

Figure 4 shows life expectancies derived from the model. Unlike Figures 1 and 3, it shows results for both sexes and both periods. Female life expectancy at birth increased from 75.1 years, with a 95% credible interval of (74.8, 75.5), in 2005–2007 to 76.7 years, with a credible interval of (76.4, 77.0), in 2012–2014. The corresponding estimates for males are 70.8 (70.5, 71.1) and 72.5 (72.2, 72.9). It is still rare in demography for life expectancies to be accompanied by uncertainty measures. Using Bayesian methods, however, uncertainty measures can be calculated routinely.

Figure 4: Modelled estimates of life expectancy at birth, for Māori, 2005–2007 and 2012–2014. The light blue bands represent 95% credible intervals, and the white lines represent posterior medians.

3 Interpolating and forecasting obesity prevalence

3.1 The estimation problem

In New Zealand, as in most countries, obesity rates are rising. Public health researchers and policy makers monitor and forecast obesity prevalence, to assess the success, or otherwise, of obesity-reduction measures, and to gauge future demand for services.

The main source of data on obesity prevalence in New Zealand is the New Zealand Health Survey, a nationally-representative survey of around 19,000 people (Ministry of Health, 2013). Like most household surveys, it has a complex design, with stratification and clustering. Obesity is measured using body mass index (BMI). A person is defined as being obese if he or she has a BMI of 30 or higher.

Surveys were carried out in 1997, 2003, 2007, 2012, and 2013. We use data for all these years. Our objective is to obtain prevalence estimates for the period 1997–2013, including non-survey years, and then forecast for the period 2014–2023. Our estimates and forecasts are disaggregated into age groups 15–24, 25–34, 35–44, 45–54, 55–64, 65–74, and 75+.

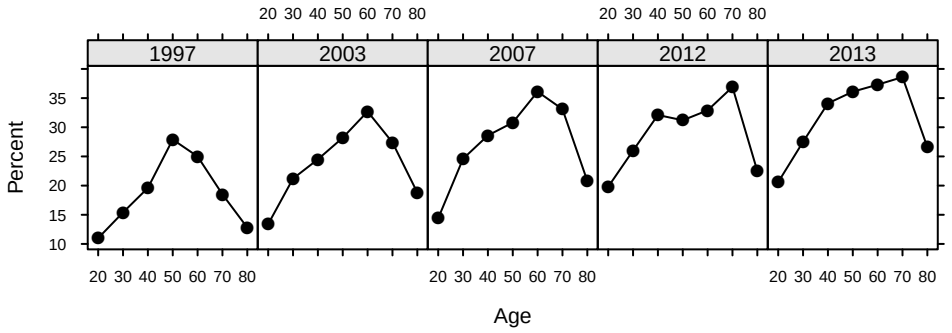

Figure 5: Proportion of obesity in New Zealand, by age and year, as estimated in the New Zealand Health Survey.

3.2 The model

Our main input data are published estimates for the proportion of New Zealanders aged a at time t who are obese, which we denote p_{at}, and the published standard errors for the p_{at}, denoted s_{at}. The p_{at} are graphed in Figure 5.

When individual-level data are available, the standard Bayesian approach towards accounting for complex survey design is to include as many features of the design as possible in the estimation model (Gelman et al., 2014, Chapter 8). Chen et al. (2014) show, however, how the full individual-level approach can be approximated by an aggregate-level approach that starts from designed-based estimates such as p_{at} and s_{at}. Chen et al. (2014) assume that the design-based estimates are constructed so as to reflect all the important

features of the survey design, and show how these estimates can be converted into a form suitable for inclusion in an aggregate-level model.

Applying the approach of Chen et al. (2014), we approximate the individual-level approach using a Binomial likelihood. We obtain counts of individuals with obesity y_{at} and total counts of individuals n_{at} by finding y_{at} and n_{at} such that $\frac{y_{at}}{n_{at}} \approx p_{at}$ and $\frac{y_{at}}{n_{at}}\left(1 - \frac{y_{at}}{n_{at}}\right) \approx s_{at}^2$. The likelihood is

$$y_{at} \sim \text{Binomial}(n_{at}, \gamma_{at}). \tag{9}$$

Here γ_{at} is the super-population probability of obesity: the probability that a person aged a at time t is obese. Our objective is to estimate γ_{at} for past years, including years without survey data, and to forecast γ_{at} for future years.

Our prior model for γ_{at} is

$$\text{logit}(\gamma_{at}) = \beta^0 + \beta_a^{\text{age}} + \beta_t^{\text{time}}, \tag{10}$$

which includes age and time effects, but not an age-time interaction. We experimented with an age-time interaction, but found that its size was small enough to omit (results not shown).

As with the mortality model of Section 2, we use a local trend model for the age effect, though in the obesity case we do not have an infant covariate. The rationale for using a local trend model is, once again, to capture the correlations between neighbouring age groups. We also use the same prior

for the intercept as we do in Section 2, a normal distribution with mean 0 and standard deviation 10.

We use a local trend model for time,

$$\beta_t^{\text{time}} = \alpha_t^{\text{time}} + u_t^{\text{time}} \tag{11}$$

$$\alpha_t^{\text{time}} = \alpha_{t-1}^{\text{time}} + \delta_{t-1}^{\text{time}} + v_t^{\text{time}} \tag{12}$$

$$\delta_t^{\text{time}} = \delta_{t-1}^{\text{time}} + w_t^{\text{time}}, \tag{13}$$

but with two different sets of assumptions about innovation terms v_t^{time} and w_t^{time}.

In our first version, we assume that v_t^{time} and w_t^{time} are always very close to 0, which we implement by using extremely tight priors on the standard deviations for these terms. The standard deviations for both terms have half-t priors with scales of 0.001. This version of the local trend model essentially fits a straight line through the data. Aside from assuming no change, this is perhaps the most common approach to forecasting future rates in epidemiology and demography. We refer to this model as the 'straight line' model.

Our second version is a generalization of the first. Rather than assuming that v_t^{time} and w_t^{time} are always close to 0, we allow them to take values that imply year-on-year changes in obesity rates of a few percentage points. We do this by setting the scale of the prior for the standard deviation of v_t^{time} to 0.05 and setting the scale of the prior for standard deviation of w_t^{time} to

0.025. We use a larger scale for v_t^{time} than for w_t^{time} on the basis that levels change more rapidly than systematic trends. We refer to the model based on this version of the time effect as the 'flexible' model.

We carry out the estimation using our package **demest**, with the same settings for burnin, production, chains, and thinning as for the mortality application.

3.3 Results

Figure 6 shows results based on the 'straight line' model. Estimates for survey years are shown in red, and estimates and forecasts for the remaining years are shown in blue. As is conventional with forecasting, we use 80% credible intervals, rather than 95%.

Estimates for years with survey data are more precise than those for years without survey data, as we would expect. Estimates for years between surveys are more precise than those for forecasts. The differences in precision between estimates and forecasts are, nevertheless, small. Strong assumptions about linearity lead to precise forecasts.

Figure 7 shows results based on the 'flexible' model. Point estimates and forecasts from the flexible model are indistinguishable from those of the straight line model. The level of uncertainty, however, is clearly different. Compared with the straight line model, there is a modest increase in uncer-

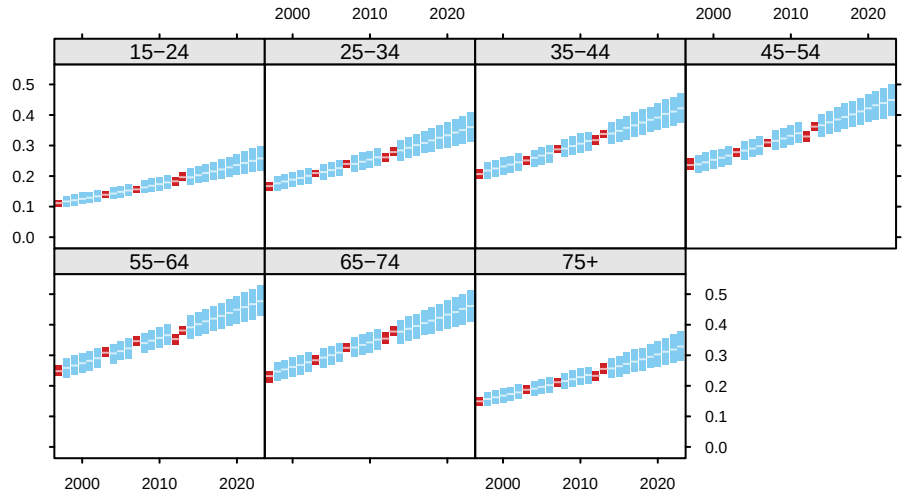

Figure 6: Estimates and forecasts of obesity prevalence in New Zealand—'straight line' model. The bands represent 80% credible intervals, and the pale lines represent posterior medians. The red bands are for years with survey data and the light blue bands are for years without data.

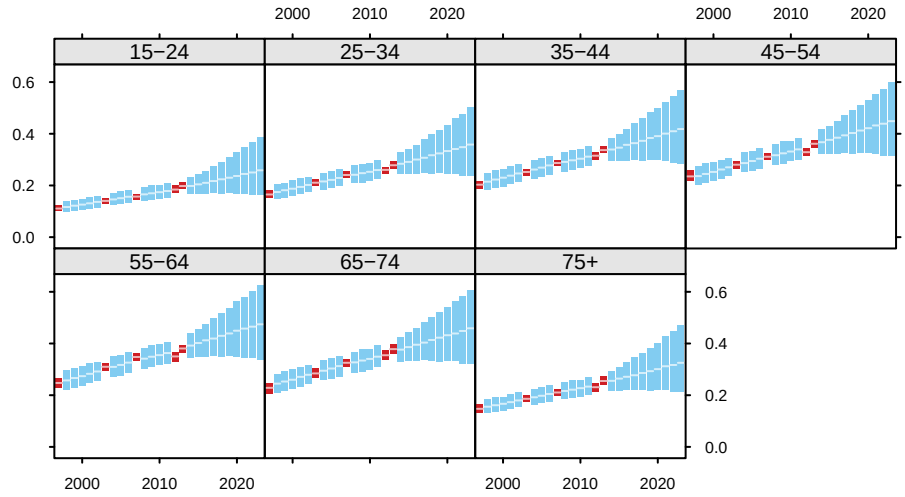

Figure 7: Estimates and forecasts of obesity prevalence in New Zealand—'flexible' model. The bands represent 80% credible intervals, and the pale lines represent posterior medians. The red bands are for years with survey data and the light blue bands are for years without data.

tainty for years between surveys and a large increase in uncertainty for the forecast period, particularly in later years.

The flexible model, arguably, gives a better representation of knowledge about obesity trends in New Zealand than the linear model. The linear assumption is conventional, but does not have any strong theoretical basis. Over-reliance on the linear assumption can produce over-confidence. The flexible model illustrates the implications of weaker assumptions.

4 Discussion

Despite the increasing popularity of Bayesian methods in the research community, national statistical agencies and policy analysts have been wary of these methods. National statistical agencies are particularly concerned about two aspects of Bayesian methods: their use of prior distributions, and their complexity.

The use of prior distributions to represent external information is indeed a distinctive feature of Bayesian analyses. Little (2012) has argued that national statistical agencies should use 'noninformative' priors, which avoid the impression of subjectivity, and which form a bridge to classical methods, in that they often lead to similar results. Among Bayesian statisticians, however, weakly informative priors have been gradually displacing noninformative priors as the default for most analyses. Compared with noninformative priors, weakly informative priors can stabilize estimates, and speed up calculations. But because they rule out only the most implausible values, they are generally no more controversial, and require little more work or justification, than noninformative priors.

However, in cases where the data to hand do not permit sufficiently strong answers to the questions of interest, there may be advantages to using priors that are strongly, rather than weakly, informative. In our obesity example, for instance, it may be possible to improve on both of our forecasting models

by specifying priors for the standard deviation parameters that accurately reflected likely year-on-year variation in obesity rates.

If statistical agencies were to use strongly informative priors, they would need to spell these priors out clearly, justify their choices, and test sensitivity to alternative choices. But, in most cases, this would be an improvement on current practice. Current practice with analyses such as population forecasts is often to apply informal adjustments, or to retrospectively adjust assumptions, until a plausible result is obtained. Bayesian methods provide analysts with a more transparent and systematic way of bringing in external information and expert judgement.

Objections about Bayesian models being complicated are partly true. Many Bayesian models *are* complicated, in that, like the models presented in this paper, they use many layers and many parameters. At the same time, however, the individual components of these models are often simple and intuitive. To make sense of our model for mortality rates, for instance, we can start with the likelihood, move on to the prior model, and then consider the priors for main effects and interactions one by one. With this divide-and-conquer approach, even complicated models are accessible. Moreover, the main assumptions behind the models can often be described in nontechnical language, even if the mathematical techniques cannot.

Similarly, the traditional objection that Bayesian modelling require advanced computing skills is gradually losing force. Packages such as ours

allow analysts to fit specific classes of demographic estimation models relatively easily. General-purpose Bayesian programming languages such as Stan (Carpenter et al., 2016) offer greater flexibility in exchange for slightly more programming effort. These tools allow practitioners to easily fit complicated Bayesian models.

Appendix A: identification of our model

In the prior model, each main effect or interaction includes all possible categories of the classifying dimensions. For instance, the sex effect includes separate female and male effects, and the age-sex interaction includes effects for every possible combination of age and sex. Because all of our priors are proper (i.e., are genuine probability distributions that integrate to 1), the posterior distribution is proper. All parameters are therefore identified in the broadest sense, and a Bayesian analysis can be carried out.

The main effects and interactions are, however, only weakly identified. For instance, adding a value λ to the female and male effects $\beta^{\text{sex}}_{\text{Female}}$ and $\beta^{\text{sex}}_{\text{Male}}$, and subtracting λ from the intercept β^0, will produce exactly the same expected value for the γ_{ast} as the original parameter settings. The data do not allow us to distinguish between the two possibilities. Identification is achieved entirely through the differences in prior densities for the original and shifted parameters.

The γ_{ast}, in contrast, cannot be arbitrarily shifted without affecting the likelihood Poisson$(y_{ast}|\gamma_{ast}n_{ast})$. In others words, the γ_{art} are well identified from the data. Shifting the values of the main effects and interactions does not affect inferences about standard deviation terms, as inferences about standard deviations depend on variation across effects, rather than absolute levels. The standard deviation terms are therefore also well identified.

In this paper we only report the γ_{ast}. In some applications, however, the main effects and interactions are also of interest. In such cases, one approach is to systematically shift the parameter estimates to achieve identification (Gelman, 2005).

Appendix B: R code

We have developed a set of *R* packages for implementing Bayesian small area demographic estimation and forecasting. The packages are available at `github.com/statisticsnz/R`. Package **dembase** contains data structures for demographic data and functions for manipulating these data structures. The basic data structure is a 'demographic array', which, in addition to the counts or rates themselves, also holds metadata such as age groups or time periods, and units of measurement. Bayesian estimation and forecasting is carried out by functions in package **demest**. The estimation functions use metadata from the demographic arrays to assign sensible default values. As a result, complex models can be specified and run relatively simply. For instance, the key parts of the code for our model in the mortality example are set out in Figure 8. Package **demlife** contains tools for creating life tables and extracting life table functions.

```
model <- Model(y ~ Poisson(mean ~ age * sex + period),
                   age ~ DLM(covariates = Covariates(infant = TRUE),
                             damp = NULL),
                   age:sex ~ DLM(trend = NULL,
                                 damp = NULL),
                   jump = 0.05)
filename <- ''out/mortality_model.est''
estimateModel(model = model,
              y = deaths,
              exposure = 3 * population,
              filename = filename,
              nBurnin = 100000,
              nSim = 100000,
              nChain = 4,
              nThin = 250)
```

Figure 8: *R* code to specify and run the mortality model, using package **demest**.

References

Alho, J. and Spencer, B. (2006). *Statistical demography and forecasting*. Springer Science & Business Media.

Bijak, J. and Bryant, J. (2016). Bayesian demography 250 years after bayes. *Population studies*, 70(1):1–19.

Bryant, J., Dunstan, K., Graham, P., Matheson-Dunning, N., Shrosbree, E.,

and Speirs, R. (2016). Measuring uncertainty in the 2013-base estimated resident population.

Bryant, J. and Howard, A. (2017). Estimating infant mortality by ethnicity: New methods for dealing with inconsistent ethnic reporting and small numbers. Working Paper 17-01, Statistics New Zealand.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, 20:1–37.

Chen, C., Wakefield, J., and Lumely, T. (2014). The use of sampling weights in bayesian hierarchical models for small area estimation. *Spatial and spatio-temporal epidemiology*, 11:33–43.

Gelman, A. (2005). Analysis of variancewhy it is more important than ever (with discussion). *The annals of statistics*, 33(1):1–53.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2014). *Bayesian Data Analysis. Third Edition*. Chapman and Hall, New York.

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, pages 1360–1383.

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(6206):234–237.

Little, R. J. (2012). Calibrated Bayes, an alternative inferential paradigm for official statistics. *Journal of Official Statistics*, 28(3):309.

Ministry of Health (2013). New zealand health survey: Annual update of key findings 2012/13. Technical report, Ministry of Health.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, 28(1):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.

Rao, J. N. K. and Molina, I. (2015). *Small area estimation, Second edition*. John Wiley & Sons.

United Nations General Assembly (2015). Transforming our world: the 2030 agenda for sustainable development.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of het-

erogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454.