A Bayesian Model for Estimating Population Size and Demographic Parameters With Uncertainty

Mark C. Wheldon[*], Adrian E. Raftery[†], Samuel J. Clark[‡] and Patrick Gerland[§]

March, 2010

1 Introduction

This paper presents a method of deriving probabilistic estimates of demographic parameters for a population of interest over a given period in the past. While methods have been developed to take account of uncertainty in demographic forecasts (see Booth, 2006, for a review), we focus on a different problem. That is, how should information on age-specific fertility, mortality and migration over some period of interest in the past, together with information about population size and age structure (e.g., from censuses), be combined to produce probabilistic interval estimates of these same demographic quantities? Moreover, how can this be done for countries and/or time periods with noisy and incomplete data?

In Section 2, the model is specified in detail. For simplicity, we restrict the application to a female-only population open to migration. The method uses the female dominant cohort-component method of population projection (CCMP projection; e.g., Preston, Heuveline & Guillot, 2001). In Section 3 we report the results of a simulation study designed to illustrate the output and estimate the achieved coverage of the reported probability intervals. Section 4 concludes by discussing the application to real-world datasets and other potential extensions.

2 Model Description

The estimation method proposed here can be viewed as a reconciliation of two different estimates of age-specific population counts, namely the estimate based on direct enumeration or survey (e.g., censuses) and the estimate derived from applying CCMP projection to estimates of fertility, mortality and migration. The reconciliation is through a Bayesian model which

[*]Graduate Student, Dept. of Statistics, Box 354320, University of Washington, Seattle, WA 98195-4320, ph: (206) 543 8797, fax: (206) 221 6873, email: mwheldon@uw.edu

[†]Blumstein-Jordan Professor Depts. of Statistics and Sociology, Box 354322, University of Washington, Seattle, WA 98195-4322, ph: (206) 543 4505, fax: (206) 685 7419, email: raftery@uw.edu

[‡]Assistant Professor, Dept. of Sociology, Box 353340, University of Washington, Seattle, WA 98195-3340, ph: (206) 616 3375, fax: (206) 543 2516, email: samclark@uw.edu

[§]Population Policy Section, United Nations Population Division, United Nations, New York, NY 10017, ph:(917) 367 7053, fax: (212) 963 2147, email: gerland@un.org.

This project is supported by Grant Number 1 R01 HD054511 01 A1 from the National Institute of Child Health and Human Development

provides posterior, probabilistic interval estimates of vital rates. In turn, these can be used to construct probabilistic estimates of summary measures such as total fertility rate (TFR), life expectancy at birth and population counts throughout the period of interest.

According to our model, baseline population counts and vital rates (i.e., age-specific fertility rates, survival proportions and migration counts) are unknown fixed quantities about which we specify prior beliefs. Prior beliefs are expressed in terms of probability distributions. The direct measurements of population counts (e.g., censuses) are treated as fixed quantities. A likelihood of the population counts is specified which is viewed as a function of the unknown vital rates. Posterior distributions for the vital rates are defined via Bayes' Theorem. Inference is via samples from the joint posterior which are drawn using an MCMC algorithm.

2.1 Notation and Indices

Index all demographic quantities by age group and time period in 5-year increments:

- Use x for age where $x = x_{(1)}, \ldots, x_{(K)}$. Let $x_{(2)}$ (i.e., the second x value) be the *beginning* of the second interval and $x_{(1)} = 0$, corresponding to zero years. The open ended interval is the $K+1^{\text{st}}$ interval. For example, for age groups 0–4 years, 5–9 years, ..., 80–84 years, 85+ years, $x_{(2)} = 5$ years, $x_{(K)} = 80$ years and $x_{(K+1)} = 85$ years.

- Use t for time where $t = t_{(1)}, \ldots, t_{(J)}$. Let $t_{(2)}$ be the *beginning* of the second interval and let $t_{(1)}$ denote the baseline (or "jump-off") year. This will probably the first year for which a population count (e.g., census) is available. For example, for time periods 1960–1964, 1965–1969, ..., let $t_{(1)} = 1960$ and $t_{(2)} = 1965$.

- t^c indexes time periods for which independent population count estimates (e.g., census counts) are available. If $\mathscr{T} = \{t_{(1)}, t_{(2)}, \ldots, t_{(J)}\}$ and $\mathscr{T}^c$ is the set of all periods containing a census year, $\mathscr{T}^c \subseteq \mathscr{T}$.

Denote vital quantities as:

- $f_{xt} \equiv 5\times$ age-specific fertility rate for females aged $[x, x+5)$ for period $[t, t+5)$ (i.e., $f_{xt} = 5 \times {}_5F_{xt}([t+5]))$.

- κ_L, κ_U are the youngest and oldest age groups, respectively, with non-zero fertility.

- $s_{xt} \equiv$ proportion surviving from age $[x-5, x)$ to $[x, x+5)$, conditional on reaching age $x-5$, by time t. For $x = x_{(1)} = 0$, this is the proportion of births in period $[t-5, t)$ that survive to time $[t, t+5)$. For $x = x_{(K+1)}$ it is the proportion surviving for a further five years. This is estimated by (Preston et al., 2001)

$$s_{xt} = \begin{cases} {}_5L_0/l_0, & \text{for } x = x_{(1)} \\ {}_5L_{x+5}/{}_5L_x, & \text{for } x_{(2)} \le x \le x_{(K)} \\ {}_5T_{x+5}/{}_5T_x, & \text{for } x = x_{(K+1)} \end{cases}$$

- $g_{xt} \equiv$ net number of migrants aged $[x, x+5)$ in period $[t, t+5)$.

- $n_{xt} \equiv$ the population count for people aged $[x, x+5)$ at time t.

For a given x and t, each of the above can be arranged in a $(K+1) \times J$ matrix as shown in the next subsection.

2.1.1 CCMP Projection: Notation and Parameters

Female dominant CCMP projection requires the following inputs:

1. The baseline population or "jump-off" population:

$$\boldsymbol{n}_{\cdot t_{(1)}} = \begin{bmatrix} n_{x_{(1)}t_{(1)}} \\ \vdots \\ n_{x_{(K)}t_{(1)}} \end{bmatrix}$$

2. Fertility rates:

$$\mathsf{F} = \begin{bmatrix} f_{x_{(1)}t_{(1)}} & & f_{x_{(1)}t_{(J)}} \\ \vdots & \cdots & \vdots \\ f_{x_{(K)}t_{(1)}} & & f_{x_{(K)}t_{(J)}} \end{bmatrix}$$

3. The proportion of people surviving to the beginning of the next time period:

$$\mathsf{S} = \begin{bmatrix} s_{x_{(1)}t_{(1)}} & & s_{x_{(1)}t_{(J)}} \\ \vdots & \cdots & \vdots \\ s_{x_{(K+1)}t_{(1)}} & & s_{x_{(K+1)}t_{(J)}} \end{bmatrix}$$

4. The net number of migrants in the time period:

$$\mathsf{G} = \begin{bmatrix} g_{x_{(1)}t_{(1)}} & & g_{x_{(1)}t_{(J)}} \\ \vdots & \cdots & \vdots \\ g_{x_{(K)}t_{(1)}} & & g_{x_{(K)}t_{(J)}} \end{bmatrix}$$

5. Sex ratio at birth at each time period (i.e., the number of male births for every female birth): SRB.

SRB is currently fixed at 1.05 but could be allowed to vary across years.

For brevity, let $\mathsf{V} \equiv [\mathsf{F}, \mathsf{S}, \mathsf{G}, \boldsymbol{n}_{\cdot(1)}, SRB]$ (where "v" is for "vitals") and $\mathscr{M} : (\mathbb{R}^K \times \mathbb{R}^{KJ} \times \mathbb{R}^{(K+1)J} \times \mathbb{R}^{KJ} \times \mathbb{R}) \mapsto \mathbb{R}^{KJ}$ stand for the CCMP model. As indicated, $\mathscr{M}$ is a deterministic mapping from the space of CCMP inputs to the space of age-time specific population counts:

$$\mathscr{M}(\mathsf{V}) = n_{x_{(1)}t_{(2)}} \ldots, n_{x_{(K)}t_{(J)}} \equiv \mathsf{N}_{t_{(2)}+}$$

2.2 Graphical representation

The model is represented graphically in Figure 1. Migration has been omitted to save space, but contributes an additional cluster of nodes at the top level similar to those of the other vital rate parameters.

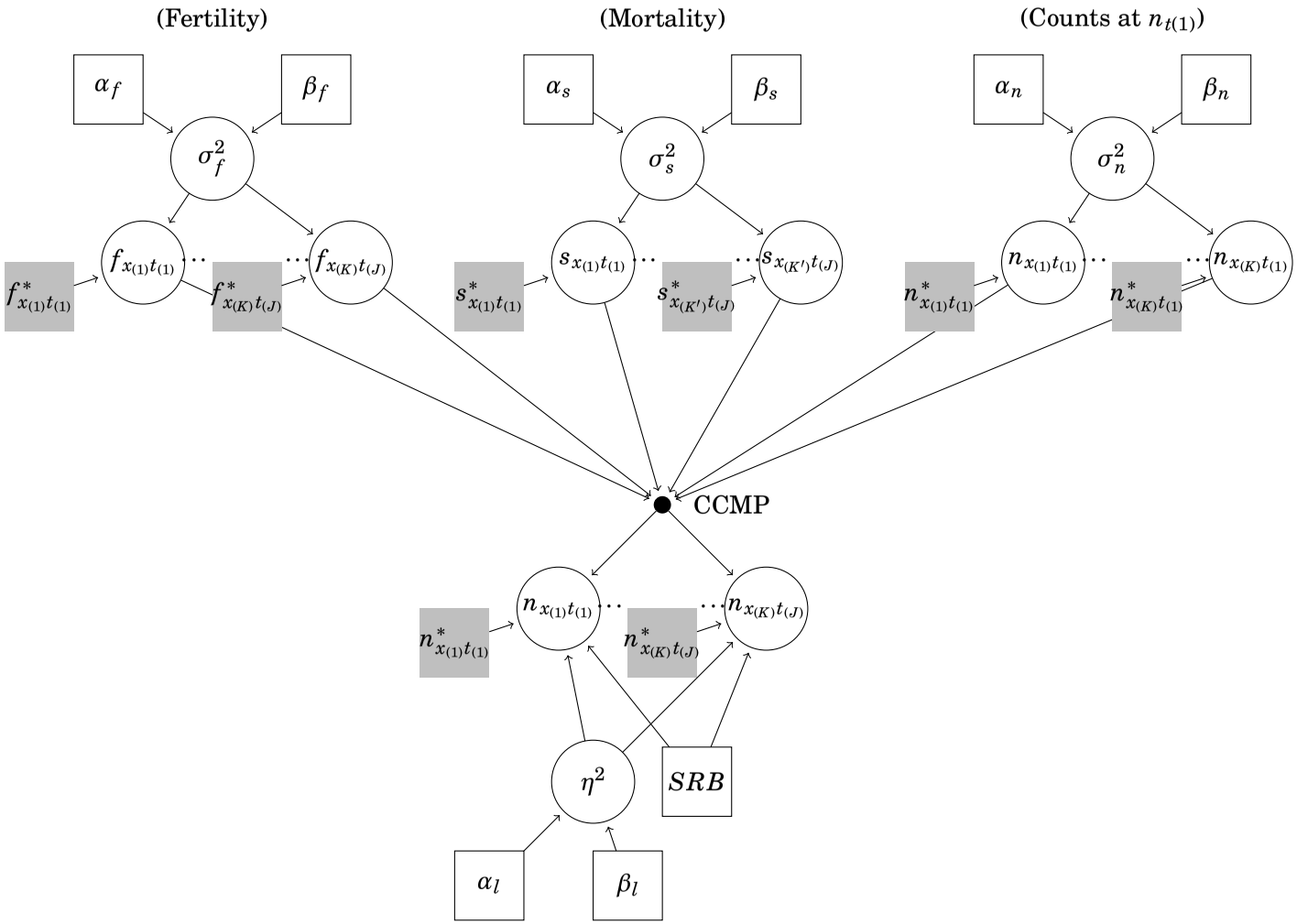

Figure 1. Graphical representation of model: circles indicate stochastic quantities, empty squares fixed parameters, filled squares indicate data. $K' = K + 1$. To save space, migration is not included but contributes a cluster of nodes very similar to those for the other three vital parameters.

2.3 Distributions

2.3.1 Likelihood

Let n^*_{xt} be the measured population count (e.g., a census count) for age group $[x, x+5)$ in last year t. Asterisks (*) are used to distinguish measured quantities from their unknown random variable counterparts. Then let the likelihood of these counts depend on $\mathbf{V}$ only through $\mathbf{N}_{t_{(2)}+}$:

$$\log n^*_{xt} | \mathbf{V}, \eta^2 \overset{d}{=} \log n^*_{xt} | n_{xt}, \eta^2 \sim \text{Normal}\left(\log n_{xt}, \eta^2\right) \tag{1}$$

$$n_{xt} = \mathscr{M}(\mathbf{V})_{xt} \tag{2}$$

$$x = x_{(1)}, \ldots, x_{(K)},\ t = t_{(2)}, \ldots, t_{(J)}$$

Thus the population counts (censuses) are modeled as "truth plus noise" where the "truth" is the CCMP projection over the projection interval.

2.3.2 Priors

Prior distributions of the parameters $f_{xt}, s_{xt}, g_{xt}, n_{xt_{(1)}}, n_{xt_{(2)}+}$ are defined on different scales:

- $f_{xt}, n_{xt_{(1)}}, n_{xt_{(2)}+}$ are modeled on the log scale.
- s_{xt} is modeled on the logit scale.
- g_{xt} is modeled as a proportion of the previous year's population count. Define

  $$\gamma_{xt_{(j)}} = g_{xt_{(j)}} / n_{xt_{(j-1)}},\ j = 2, \ldots, J.$$

  Estimation will be done in terms of γ_{xt}.

Let $n^*_{xt_{(1)}}, f^*_{xt}, s^*_{xt}, \gamma^*_{xt}$ be measured vital rates and migration proportions. The following are assumed to be conditionally independent.

$$\log(f_{xt}) | f^*_{xt}, \sigma^2_f \sim \text{Normal}\left(\log(f^*_{xt}), \sigma^2_f\right),\ x = x_{(1)}, \ldots, x_{(K)},\ t = t_{(1)} \ldots, t_{(J)} \tag{3}$$

$$\text{logit}(s_{xt}) | s^*_{xt}, \sigma^2_s \sim \text{Normal}\left(\text{logit}(s^*_{xt}), \sigma^2_s\right),\ x = x_{(1)}, \ldots, x_{(K+1)},\ t = t_{(1)} \ldots, t_{(J)} \tag{4}$$

$$\gamma_{xt} | \gamma^*_{xt}, \sigma^2_\gamma \sim \text{Normal}\left(\gamma^*_{xt}, \sigma^2_\gamma\right),\ x = x_{(1)}, \ldots, x_{(K)},\ t = t_{(1)} \ldots, t_{(J)} \tag{5}$$

$$\log(n_{xt_{(1)}} | n^*_{xt_{(1)}}, \sigma^2_n) \sim \text{Normal}\left(\log(n^*_{xt_{(1)}}), \sigma^2_n\right),\ x = x_{(1)}, \ldots, x_{(K)} \tag{6}$$

$$\sigma^2_f \sim \text{InvGamma}(\alpha_f, \beta_f) \tag{7}$$

$$\sigma^2_s \sim \text{InvGamma}(\alpha_s, \beta_s) \tag{8}$$

$$\sigma^2_\gamma \sim \text{InvGamma}(\alpha_\gamma, \beta_\gamma) \tag{9}$$

$$\sigma^2_n \sim \text{InvGamma}(\alpha_n, \beta_n) \tag{10}$$

$$\eta^2 \sim \text{InvGamma}(\alpha_l, \beta_l) \tag{11}$$

2.4 Estimation

Samples from the joint posterior

$$p(\mathbf{V}, \sigma_n^2, \sigma_f^2, \sigma_s^2, \sigma_\gamma^2, \eta^2 | \mathbf{N}^*_{t_{(2)}+}, \mathbf{V}^*)$$

are drawn using a Markov chain Monte Carlo (MCMC) sampler (Metropolis, Rosenbluth, Rosenbluth, Teller & Teller, 1953; Hastings, 1970). Full conditional densities for the variance parameters can be calculated analytically. Full conditionals for the vital rate parameters are not as tractable since each vital rate enters the likelihood through the non-linear function $\mathscr{M}(\mathbf{V})$. Therefore, these parameters are updated, one age-by-year-at-a-time, using Metropolis-Hastings steps.

3 Simulation Study

A simulation study can be used to illustrate model outputs and estimate actual coverage of the posterior intervals.

3.1 Method

True, underlying fertility rates, survival proportions, migration counts and population counts were contrived for an hypothetical population with four five-year age groups plus an open ended group (0–4, 5–9, 10–14, 15+) over four five-year periods (1960–1964, 1965–1969, 1970-1974, 1975–1980). These are shown in Tables 1 and 2.

Values for the hyperparameters $\alpha_f, \beta_f, \ldots$ were set by specifying the measurement uncertainty of the corresponding vital rate parameters with 95 percent prior belief. This was done by choosing the accuracy at two quantiles of the prior distribution; the median and the 95$^{\text{th}}$ percentile. For example, to determine α_f and β_f, it was specified that median(σ_f^2) corresponds to age-specific fertility rates measured with an accuracy of ± 10 percent. The 95$^{\text{th}}$ percentile was then set to be twice the value of the median. The See Appendix A for details. The hyperparameters used in the simulation are reported in Table 3.

With these values set, the following steps were implemented: for $i = 1, \ldots, N$

1. Use the values in Tables 1 and 2 to generate "measurements-with-error" of each vital rate by drawing a sample of $f_{x_{(1)}t_{(1)}}, \ldots f_{x_{(K)}t_{(J)}}, \ldots, s_{x_{(K+1)}t_{(J)}}, \ldots, g_{x_{(K)}t_{(J)}}, \ldots, n_{x_{(K)}t_{(1)}}$ from their repsective prior distributions (see Section 2.3.2), with variances given in Table 3.

2. Generate "measurements-with-error" of population counts at times $t_{(2)}, \ldots t_{(J)}$ by sampling $n^*_{x_{(1)}t_{(1)}}, \ldots n^*_{x_{(K)}t_{(J)}}$ from the likelihood described in Section 2.3.1. Use the values in Tables 2 and 3 for the means and variance, repsectively.

3. Use the Bayesian model to generate samples from the joint posterior of vital rates. The inputs are:

   - *Fixed* hyper-parameters: $\alpha_f, \beta_f, \ldots$ from Table 3.
   - *Fixed* vital rate means: $f^*_{xt}, \ldots, n^*_{xt_{(1)}}$ generated in Step 1.
   - Measured population counts at times $t_{(2)}, \ldots t_{(J)}$ generated in Step 2.

Table 1. True vital rates used in the simulation study.

	Time period			
x	[1960, 1965)	[1965, 1970)	[1970, 1975)	[1975, 1980)
		Fertility Rate		
0	0.00	0.00	0.00	0.00
5	2.00	2.00	2.00	2.00
10	1.50	1.50	1.50	1.50
15	0.00	0.00	0.00	0.00
		Survival Proportion		
0	0.90	0.90	0.90	0.90
5	0.95	0.95	0.95	0.95
10	0.85	0.85	0.85	0.85
15	0.80	0.80	0.80	0.80
20+	0.10	0.10	0.10	0.10
		Migration Proportion		
0	-0.03	-0.05	0.03	0.05
5	-0.05	-0.10	0.05	0.10
10	-0.06	-0.11	0.06	0.11
15	-0.01	-0.01	0.01	0.01

Table 2. True population counts used in simulation study.

x	1960	1965	1970	1975	1980
0	7643	8760	9698	11760	14910
5	6205	7015	7763	9522	11927
10	4242	5026	5388	6912	8879
15	3171	3609	4140	4855	6346

- Start values for all parameters of interest. For vital rates, these are set to the fixed vital rate means; for variances, these are set to the medians of the prior distributions.

For each age-specific vital rate, determine the proportion of times the underlying truth was contained in 95 percent posterior intervals produced by the N simulations and compare with the nominal value.

3.2 Results

3.2.1 Quantiles from a single replication

Ninety five percent posterior probability intervals based on a single replication of the simulation study are shown in Figures 2–5. These plots illustrate the model outputs one would obtain when using this method in practice, although the true values would, of course, be un-

Table 3. Specification of hyper-priors for simulation study.

Variance parameter	Specification	Median, 95$^{\text{th}}$ percentile	Inverse gamma shape, scale
σ_f^2 (Fertility rate)	Accuracy of f_{xt} measurements at median = ± 10 percent.	0.0538^2, 0.076^2	10.2116, 0.0286
σ_s^2 (Survival proportion)	Accuracy of $s_{xt}/(1-s_{xt})$ measurements at median = ± 10 percent.	0.0538^2, 0.076^2	10.2116, 0.0286
σ_γ^2 (Migration proportion)	Accuracy of γ_{xt} measurements at median = ± 0.2.	0.102^2, 0.1443^2	10.2051, 0.1028
σ_n^2 (Population at $t_{(1)}$)	Accuracy of $n_{xt_{(1)}}$ measurements at median = ± 10 percent.	0.0538^2, 0.076^2	10.2116, 0.0286
η^2 (Population at $t_{(2)}+$)	Accuracy of $n_{xt_{(2)+}}$ measurements at median = ± 10 percent.	0.0538^2, 0.076^2	10.2116, 0.0286

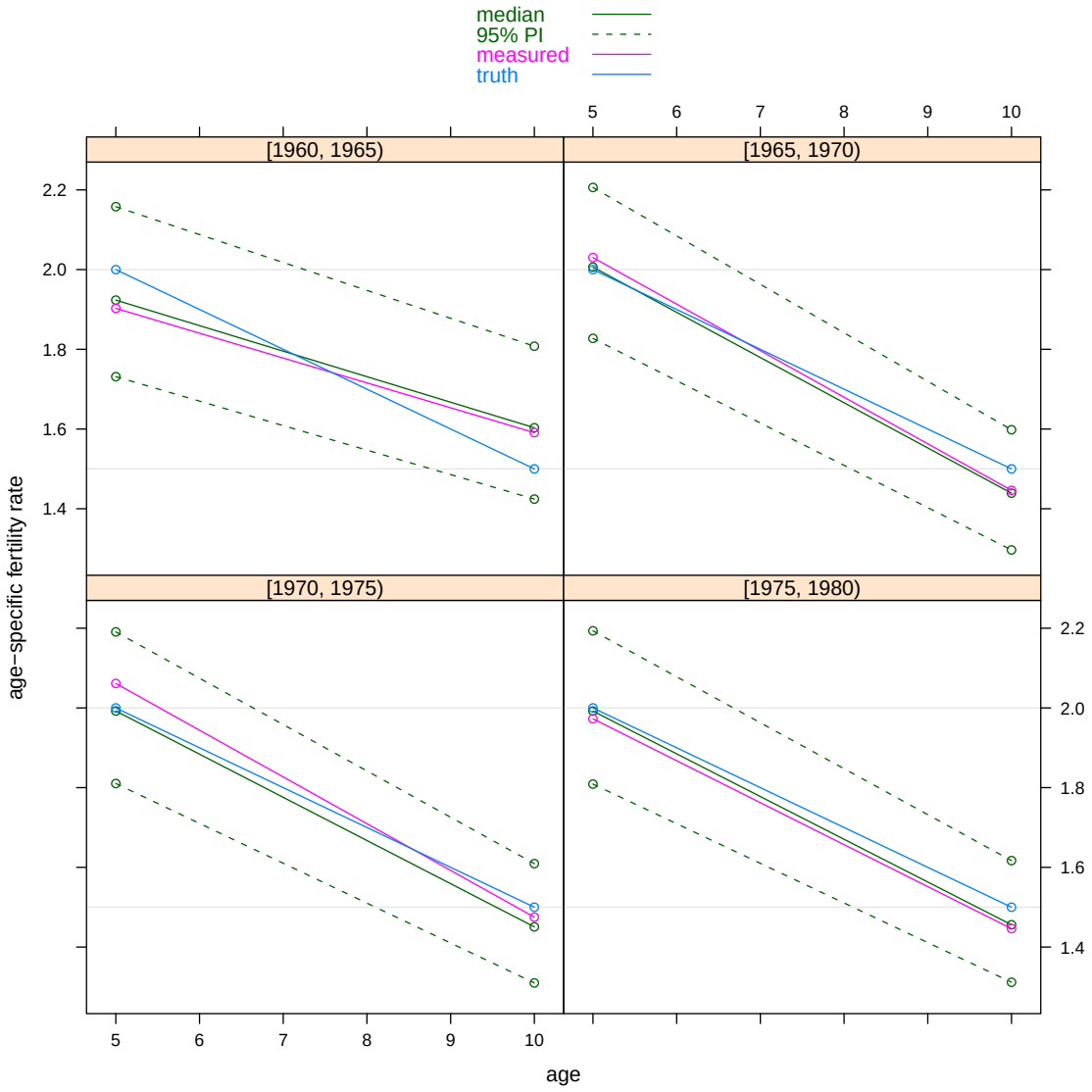

Figure 2. Ninety five percent marginal posterior probability intervals for age-specific fertility rate taken from a single replication of the simulation study, with true and measured values overlaid.

known. Posterior samples of summary measures such as TFR and life expectancy at birth can also be constructed from the vital rate posterior samples. These are shown in Figures 6 and 7. Also obtainable, but not shown, are interval estimates of age-specific population counts at times $t_{(2)}, \ldots t_{(J)}$.

3.2.2 Estimated Coverage

The estimated coverage of the the posterior intervals is shown in Figures 8–11. The error bars represent the inaccuracy due to using a finite N and are Bayesian posterior intervals based

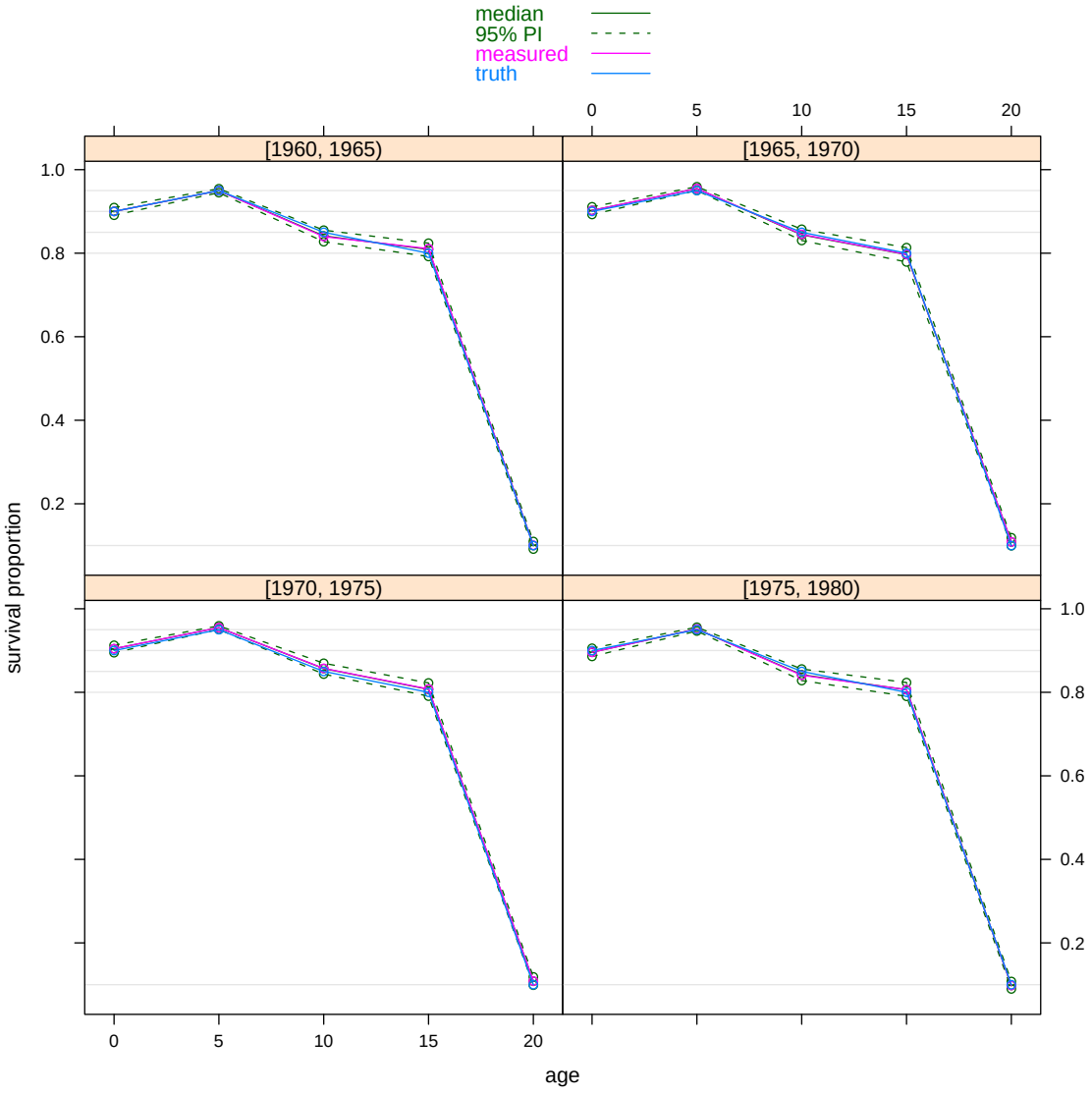

Figure 3. Ninety five percent marginal posterior probability intervals for age-specific survival proportion taken from a single replication of the simulation study, with true and measured values overlaid.

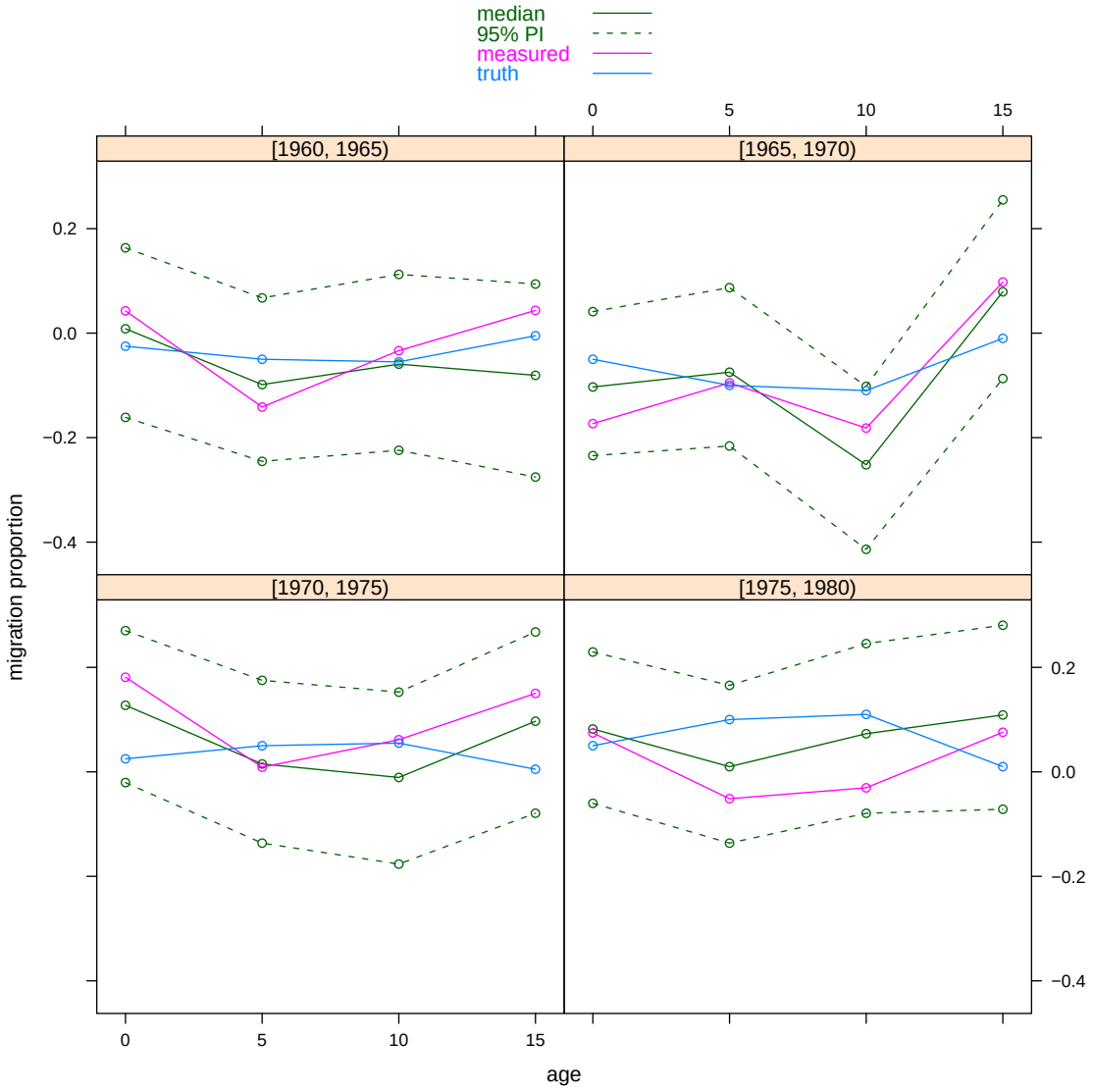

Figure 4. Ninety five percent marginal posterior probability intervals for age-specific migration proportion taken from a single replication of the simulation study, with true and measured values overlaid.

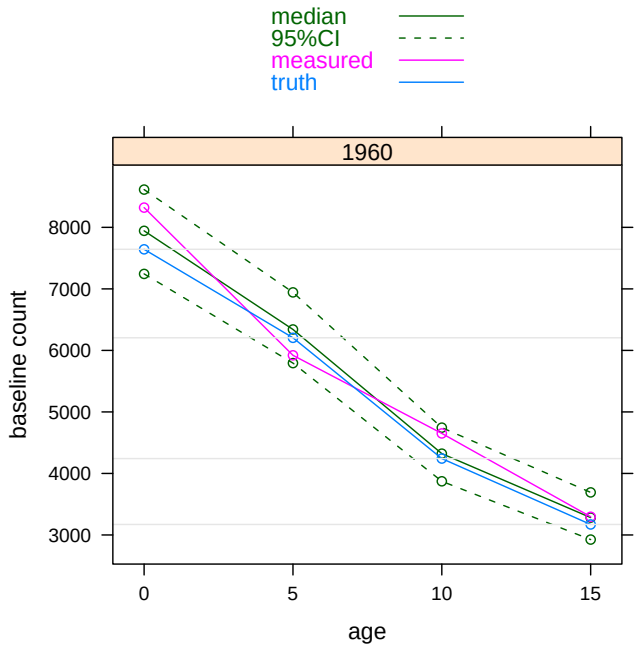

Figure 5. Ninety five percent marginal posterior probability intervals for population count at $t_{(1)}$ taken from a single replication of the simulation study, with true and measured values overlaid.

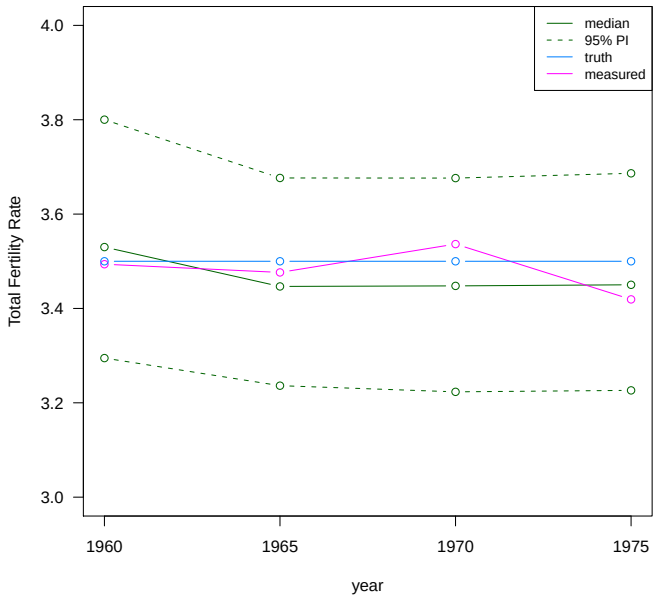

Figure 6. Ninety five percent marginal posterior probability intervals for total fertility rate taken from a single replication of the simulation study, with true and measured values overlaid.

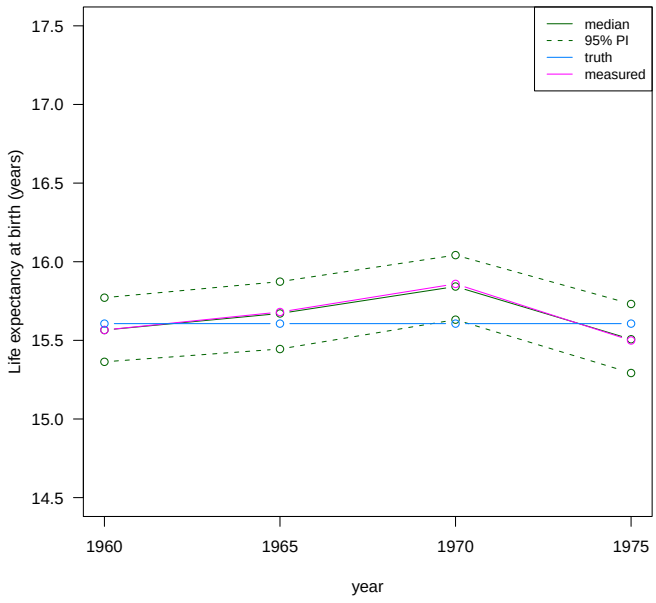

Figure 7. Ninety five percent marginal posterior probability intervals for life expectancy at birth taken from a single replication of the simulation study, with true and measured values overlaid.

on a uniform prior for the probability that the true value is contained within the interval. Specifically, if Y is the number of replications in which the posterior interval contained the true parameter value for any age-time specific vital rate parameter,

$$\pi \sim \text{Beta}(1,1) \stackrel{d}{=} \text{Unif}(0,1), \quad Y \sim \text{Bin}(N,\pi)$$
$$\pi|Y \sim \text{Beta}(Y+1, N-Y+1).$$

The endpoints of the error bars are the 2.5^{th} and 97.5^{th} percentiles of $\pi|Y$. The circle is Y/N.

3.3 Discussion

Results from the single replication illustrate how the posterior probability intervals succeed in capturing the true parameter values approximately 95 percent of the time. They also show the relative accuracy with which each parameter has been estimated; age-specific fertility rates have wider probability intervals than age-specific survival proportions, for example.

If the model is well calibrated in the sense that the 95 percent posterior probability intervals contain the true value 95 percent of the time, approximately 95 percent of the intervals from the simulation replications should contain the truth for each vital rate parameter. The results suggest that this was the case for the particular hypothetical population used.

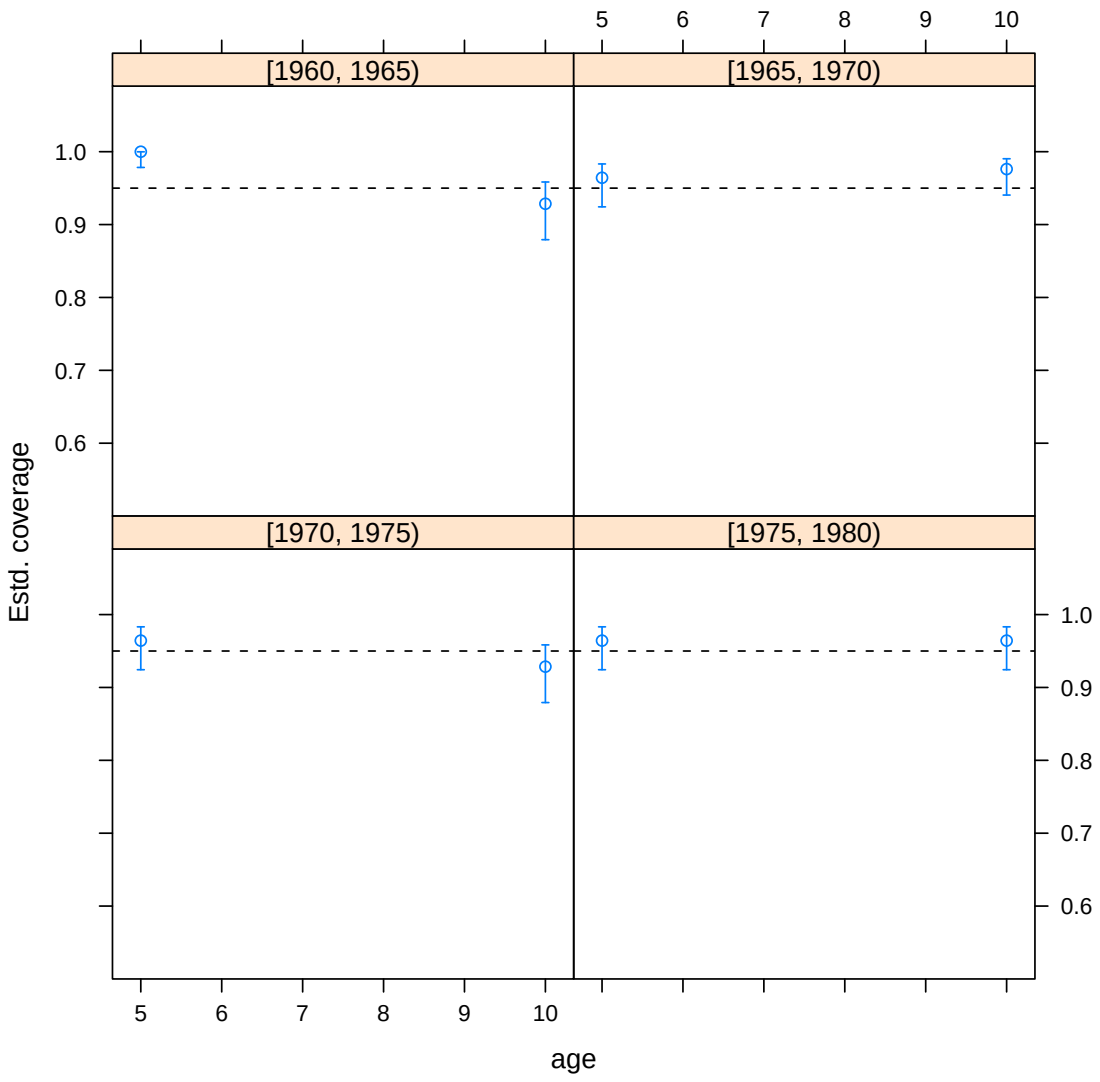

Figure 8. Estimated coverage with approximate Monte Carlo error for fertility rate, $N = 168$.

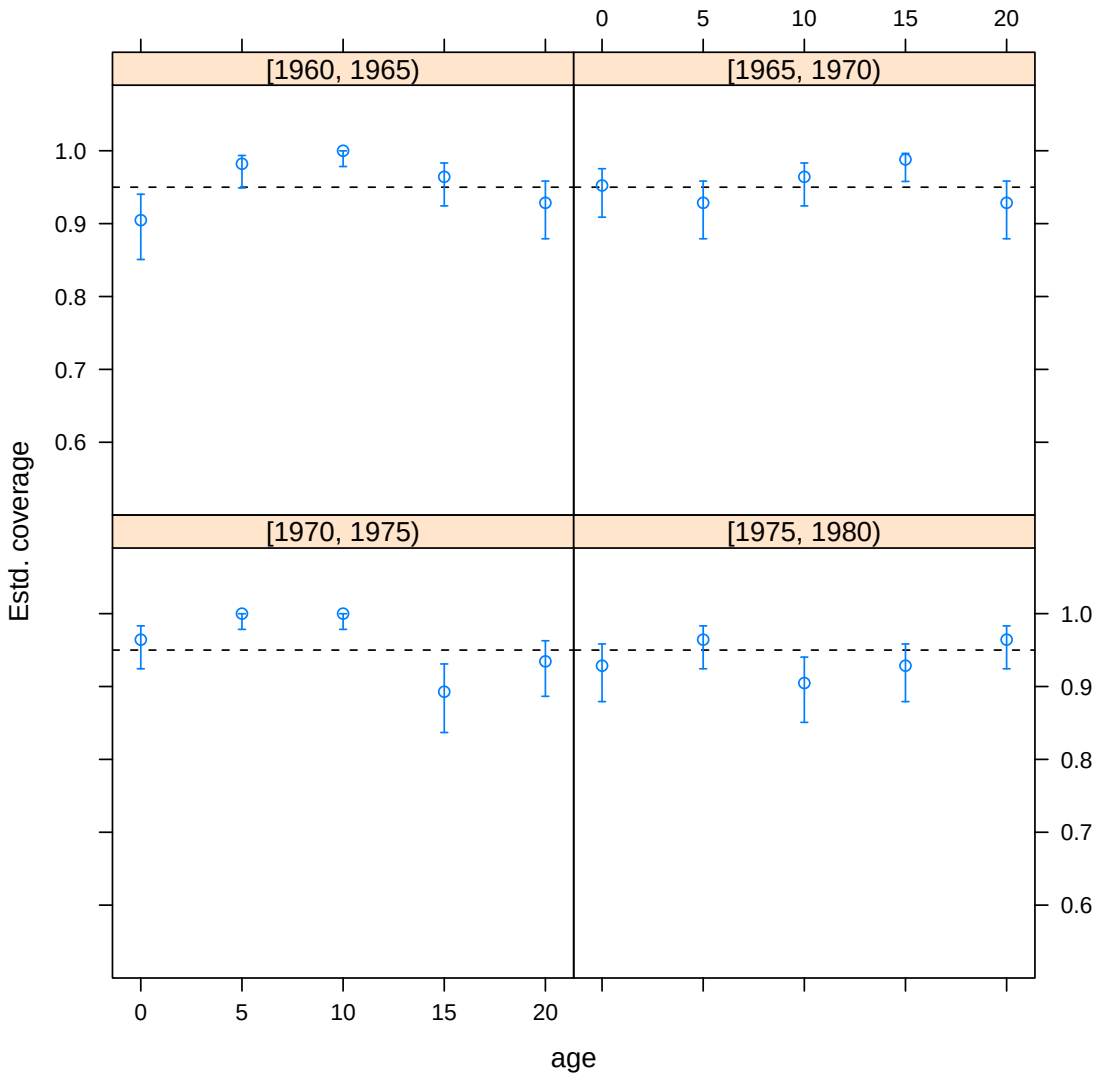

Figure 9. Estimated coverage with approximate Monte Carlo error for survival proportion, N = 168.

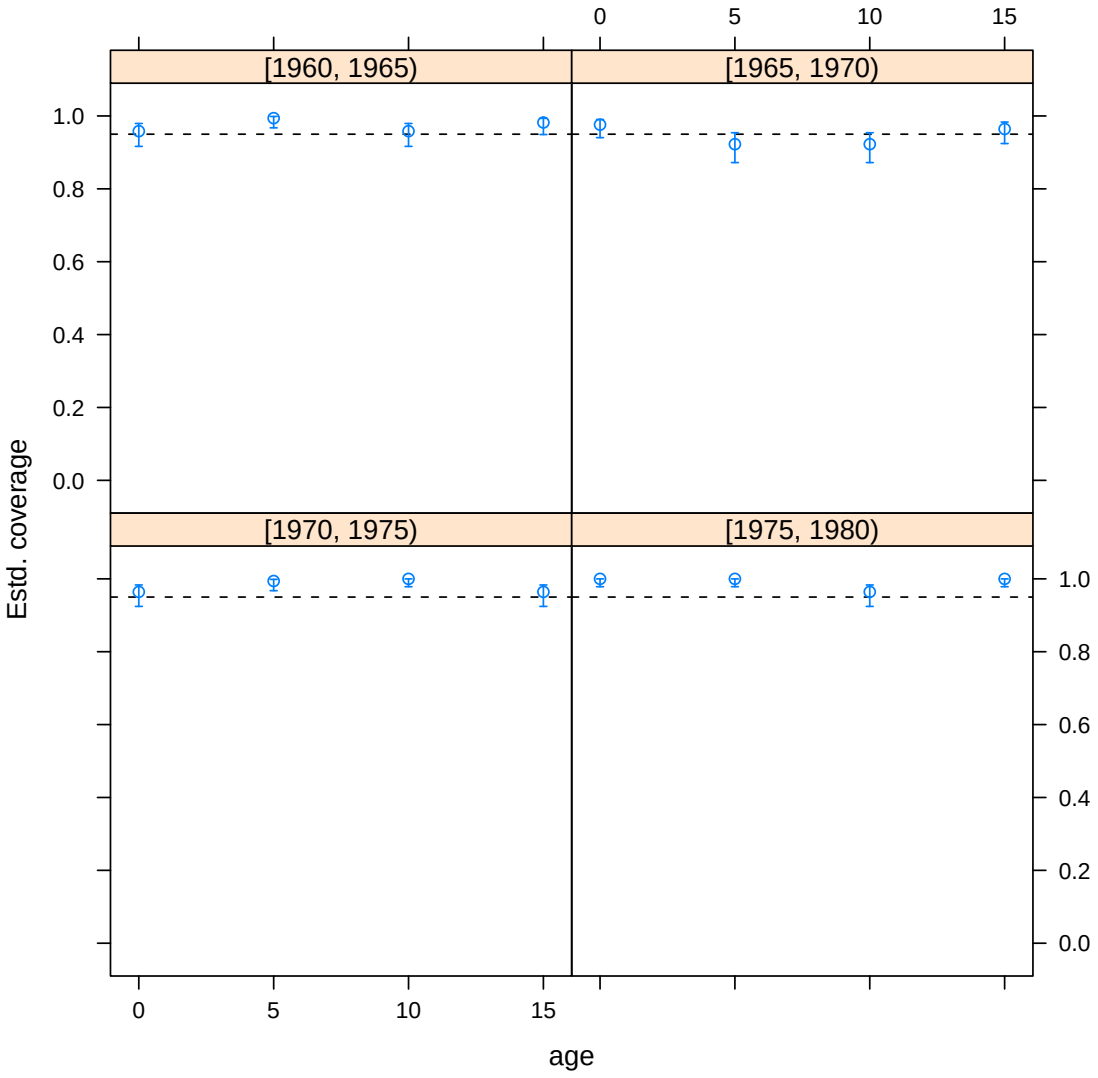

Figure 10. Estimated coverage with approximate Monte Carlo error for migration proportion, $N = 168$.

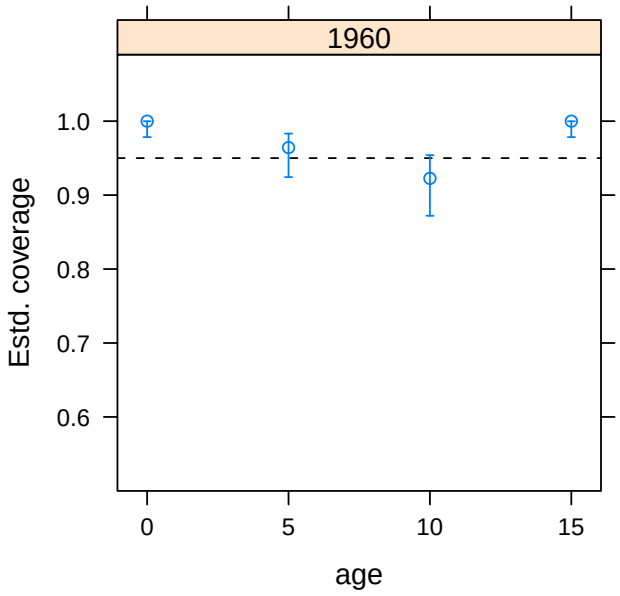

Figure 11. Estimated coverage with approximate Monte Carlo error for population count at $t_{(1)}$, $N = 168$.

4 Concluding Remarks

Much as CCMP projection produces deterministic estimates of population counts over the period of interest, the method proposed here requires little extra data yet produces probabilistic interval estimates not only of population counts, but also of fertility, mortality and migration parameters. Although it requires measurements of age-specific vital rates and population counts, measurement error is accounted for.

We have described a small simulation study that proved useful in studying the performance of the model. In future, this platform could be used to assess model performance under different data scenarios, as well as test sensitivity to different start values, for example.

Although real demographic datasets are larger than those used in the simulation study, preliminary investigation suggests that our method scales to accommodate them. We note that complete sets of age-specific vital rate measurements may not be available. For example only ${}_1q_0$, ${}_4q_1$ and ${}_{45}q_{15}$ may be available for mortality. In such cases, a demographic model age-pattern (e.g. Brass, 1971) could be used to approximate a complete set of mortality measurements for input to the model. Similarly, we could use methods derived from the model of Coale & Trussell (1974) to approximate sets of age-specific fertility rate measurements. Migration could be handled similarly or estimated as zero with a large degree of error. The estimated accuracy of these modeled measurements is encoded in the prior variance parameters and contributes to the output through the width and location of the posterior probability intervals.

While we have focussed on female-only populations here, extension to two-sex populations will be the focus of future work.

A Derivation of Hyper-priors for Simulation Study

The parameters of the respective inverse gamma distributions can be determined by these specifications as follows:

FERTILITY AND POPULATION COUNTS Since fertility rates and population counts are measured on the log scale, this implies that, at the median level of accuracy,

$$\begin{aligned} &\Pr(f_{xt} - (0.1)f_{xt} \le f^*_{xt} \le f_{xt} + (0.1)f_{xt}) \ge 0.95 \\ \Rightarrow &\Pr(0.9 \le f^*_{xt}/f_{xt} \le 1.1) \ge 0.95 \end{aligned} \tag{12}$$

Then, since $\log f^*_{xt} \sim \text{Normal}(\log(f_{xt}), \sigma_f^2)$, (12) implies

$$\begin{aligned} &\Pr\big(\log(0.9) \le \log(f^*_{xt}) - \log(f_{xt}) \le \log(1.1)\big) \ge 0.95 \\ \Rightarrow &\Pr\big(\log(0.9) \le \log(f^*_{xt}) - \log(f_{xt}) \le -\log(0.9)\big) \ge 0.95 \\ \Rightarrow &\Pr\left(\frac{\log(0.9)}{\sigma_f^2} \le Z \le -\frac{\log(0.9)}{\sigma_f^2}\right) \ge 0.95, \quad Z \sim \text{Normal}(0,1) \end{aligned} \tag{13}$$

where (13) follows from $|\log(0.9)| \ge \log(1.1)$ and yields a symmetric interval. Thus, if $z_{0.025}$ is the $(0.025)^{th}$ quantile of a standard normal random variable, we should set

$$\text{median}(\sigma_f) = \frac{|\log(0.9)|}{z_{0.025}} \Rightarrow \text{median}(\sigma_f^2) = \left(\frac{\log(0.9)}{z_{0.025}}\right)^2.$$

The 0.95 quantile is set to be equal to twice the median. An Inverse Gamma distribution with these two quantiles can be found using a numerical search.

SURVIVAL Survival proportion (s_{xt}) is modeled on the logit scale, therefore additive proportionate errors need to be stated in terms of the odds of survival. Let

$$\mathscr{S}_{xt} \equiv \frac{s_{xt}}{1 - s_{xt}} \Rightarrow \log(\mathscr{S}_{xt}) = \text{logit}(s_{xt}).$$

Defining $\mathscr{S}^*_{xt}$ analagously for s^*_{xt}, and one can specify

$$\begin{aligned} \Pr\big(\mathscr{S}^*_{xt} - (0.1)\mathscr{S}^*_{xt} \le \mathscr{S}_{xt} \le \mathscr{S}^*_{xt} + (0.1)\mathscr{S}^*_{xt}\big) &\ge 0.95 \\ \Rightarrow \Pr\left(0.9 \le \frac{\mathscr{S}_{xt}}{\mathscr{S}^*_{xt}} \le 1.1\right) &\ge 0.95 \end{aligned}$$

whence

$$\begin{aligned} &\Pr\big(\log(0.9) \le \text{logit}(s_{xt}) - \text{logit}(s^*_{xt}) \le \log(1.1)\big) \ge 1 - \omega \\ &\Rightarrow \Pr\left(\frac{\log(0.9)}{\sigma_s} \le Z \le -\frac{\log(0.9)}{\sigma_s}\right) \ge 0.95, \quad Z \sim \text{Normal}(0,1) \end{aligned}$$

so that,

$$\text{median}(\sigma_s) = \frac{|\log(0.9)|}{z_\omega}$$

as before.

MIGRATION PROPORTION Migration proportion is not transformed. With a median accuracy of ±20 percent, median(σ_γ^2) is defined by

$$\begin{aligned} &\Pr\left(\gamma_{xt}^* - 0.2 \le \gamma_{xt} \le \gamma_{xt}^* + 0.2\right) \ge 0.95 \\ &\quad \Rightarrow \Pr\left(\frac{-0.2}{\sigma_\gamma} \le Z \le \frac{0.2}{\sigma_\gamma}\right) \ge 0.95, \qquad Z \sim \text{Normal}(0,1) \end{aligned}$$

thus

$$\text{median}(\sigma_\gamma) = \frac{0.2}{z_{0.025}}.$$

References

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, *22*(3), 547–581.

Brass, W. (1971). On the scale of mortality. In W. Brass (Ed.), *Biological Aspects of Demography* (pp. 69–110). London: Taylor and Francis Ltd.

Coale, A. J. & Trussell, J. (1974). Model fertility schedules: Variations in the age structure of childbearing in human populations. *Population Index*, *40*(2), 185–258.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, *57*(1), 97–109.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes*. Blackwell.