Forecasting Vital Rates and Population with Bayesian Vector Autoregressions

Javier Meseguer*
Social Security Administration

February 2010

Abstract

Demographic forecast applications relevant to the pension and life insurance industries pose a particular challenge, as they require very long predictive horizons. The significant uncertainty involved calls for a multivariate approach that is flexible, can accommodate complex interactions among the series being modeled and takes into account uncertainty from multiple sources. This paper explores the application of Bayesian vector autoregressive processes. Long-range forecasts of age-specific mortality and fertility rates are generated using two variants of the framework: One with an informative prior steady-state and another with a diffuse prior. We use the Lee-Carter approach as a benchmark for comparison and produce population forecasts via the cohort component method.

1 Introduction

Traditionally, demographers have approached population forecasting by developing assumptions about changes over time in the vital rates (mortality, fertility and net migration). These so-called deterministic scenarios are derived through a combination of historical trend extrapolation and expert judgment. For instance, in its annual report, the Board of Trustees to the U.S. Social Security Administration presents three sets of alternative long-range population projections encompassing a 75-year horizon. The intermediate or Alternative 2 scenario represents the most likely outcome in the Trustees' judgment. In addition, two alternative scenarios are reported,

*The analyses, opinions and findings of this paper represent the views of the author. They are not necessarily those of the Social Security Administration. Please address correspondence to Javier Meseguer, 500 E Street, SW, Washington, DC 20254. Telephone: (202) 456-6244. Email: javier.meseguer@ssa.gov.

based on assumptions about plausible variation in the assumed vital rates: a more optimistic scenario (low-cost or Alternative 1) and a more pessimistic one (high-cost or Alternative 3).

The alternative scenarios offer useful information about the plausible outcomes expected by an expert or group of experts. However, given the lack of an underlying statistical model, the scenarios are incompatible with a probabilistic representation of uncertainty. This drawback has possibly contributed towards an increasing acceptance of time-series methods by demographers and other social scientists.

One of the challenges particular to demographic applications is the need to generate forecasts over unusually long horizons. For example, in order to evaluate the solvency of a pension plan one must consider how the age structure of the population evolves over time. As a result, it is important to develop multivariate models that offer a great deal of flexibility in the possible range of outcomes; can accommodate complex interactions among the series being modeled and take into consideration multiple sources of uncertainty. This paper explores the use of Bayesian vector autoregressive (BVAR) models as suitable candidates.

Approaching inference from a Bayesian perspective offers certain advantages over the traditional sampling theory framework. First, Bayesian inference derives from an exact finite-sample probability distribution, regardless of sample size. Hence, there is no need to rely on asymptotic theory approximations. Second, by providing a formal mechanism for the inclusion of available prior information, it is possible to accommodate model restrictions that do a better job at reproducing the stylized facts in the data. Finally, since the Bayesian paradigm treats a model's parameters as random variables, the generated predictive distributions account for variation due to both the sampling and parameter outcomes.

The remainder of this paper is structured as follows. In Section 2 we briefly describe the data, comprising fertility rates for seven age groups, as well as gender-specific mortality rates for 22 age groups. For the purpose of generating population projections via the cohort component method, net migration is treated exogenously, taking the Alternative 2 or intermediate assumptions from the Social Security Trustees as given.

In Section 3 we discuss the stochastic approach to mortality and fertility forecasting pioneered by Ronald Lee and Lawrence Carter. The resulting projections are used as a benchmark for comparison through out the paper.

The alternative BVAR models are the subject of Section 4. Shrinkage Minnesota-type priors on the autoregressive coefficients provide a means to successfully minimize the problem of over-parameterization that is common in vector autoregressions.

Meanwhile, the mean-deviation reduced-form of the process facilitates a more intuitive interpretation. In particular, it allows specifying a prior density on the steady-state or unconditional mean.

Section 5 presents the resulting long-run projections of fertility and mortality. Two different versions of the BVAR approach are entertained. One variant considers an uninformative or diffuse prior density on the steady-state. The other centers the steady-state prior around the Alternative 2 ultimate assumptions. The generated fertility and mortality forecasts are combined to produce age-specific population projections.

2 The Vital Rates

The set of vital rates used in this paper were constructed on the basis of data provided by the Social Security Administration's Office of the Chief Actuary. They include both the historical series and the projections for the Alternative 2 scenario in the 2009 Trustees Report. The year 2008 is taken as the cut-off point for the sample data in all cases, while the objective is to generate annual forecasts from 2009 to 2100.

In generating deterministic scenarios, the Trustees develop fertility assumptions for women by single year of age, from 14 to 49. The available historical series begins in 1917. For modeling purposes, the annual birth rates are collected into 7 age groups (ages 15-19, ages 20-24,..., ages 45-49). Notice that we omit fertility at age 14. This is of little practical consequence. It is simply done to maintain consistency with the remaining vital rates, which in most cases are aggregated into five-year age groups. The Trustees' mortality assumptions, on the other hand, are developed by single year of age, gender and cause of death. For each gender, we construct an all-cause mortality series that involves 22 age groups (age 0, ages 1-4, ages 5-9,.., ages 95-99, and ages 100+). Historical mortality dates as far back as the year 1900.

The central panel of Figure 1 displays the age profiles of male and female logarithmic mortality for the years 1925, 1960 and 2008. Clearly, the age structure of mortality has shown a remarkable degree of regularity over time. For any given period, all-cause logarithmic mortality declines smoothly from birth up to about ages 10-14. From ages 14 on up, the mortality rate rises steadily, although there is an increasingly prominent hump at ages 15-19, especially for males, often referred to as the "accident hump." This occurrence is associated with motor-vehicle fatalities.

Not only has mortality across the ages maintained a fairly stable shape, but it has also generally followed a downward secular trend. Life expectancy at birth for males and females, along with the projections in the 2009 Trustees report are shown in the middle left panel of Figure 1. Barring unusual shocks like the 1918 Spanish influenza pandemic, improvement in life expectancy over the last century has been linear. Most of the gains in life expectancy before 1950 were achieved through large reductions in the death rates associated with the younger ages, whereas survival improvement after age 65 took center stage in the latter half of the century.

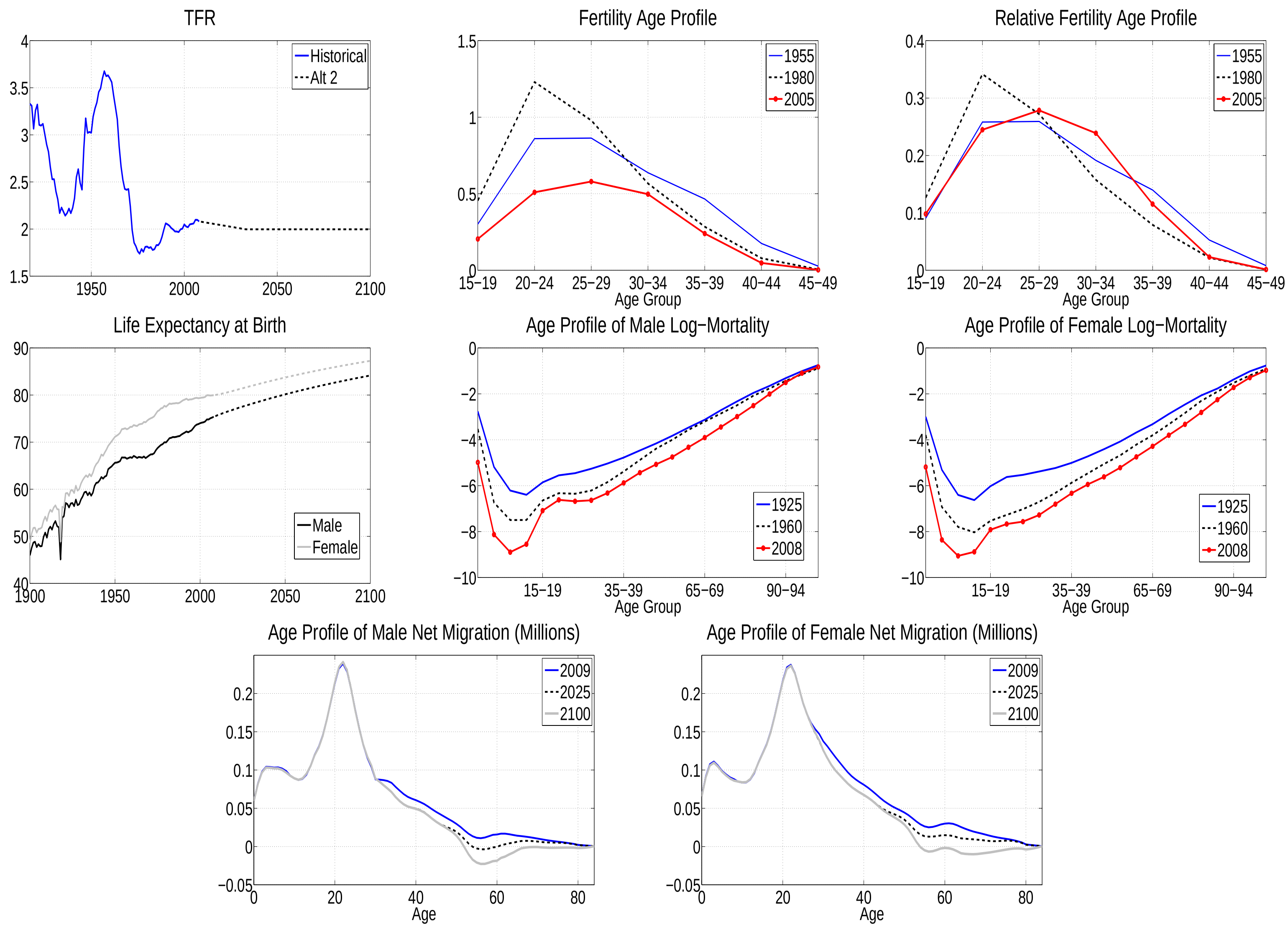
Unlike the case of mortality, fertility has historically displayed a great deal of volatility. A commonly used summary measure of a population's birth experience is the total fertility rate (TFR). It is defined as the average number of children born to a woman if she were to exhibit each age-specific birth rate throughout the entire childbearing period. Mathematically, the TFR is calculated as the sum of all the age-specific fertilities. A value of about 2.1 births per woman is considered replacement level, leading to zero population growth in the absence of net migration and assuming constant mortality.

The upper left panel of Figure 1 displays the U.S. TFR beginning in 1917, as well as the Alternative 2 scenario for the 2009 Trustees Report. Clearly, the most significant feature in the series is its variability. The TFR declined from a high of 3.3 at the end of World War I to a low of 2.1 during the Great Depression. Following World War II, the baby boom ensued (from 1946 to 1964), with the TFR reaching a value of 3.7 at its peak in 1957. By 1976, the TFR had attained its lowest historical level of 1.7, ascending again through much of the 1980s. Since 1990 it has remained fairly stable at about replacement level. As pointed out by Lee (1993), some of the factors believed to have contributed to the secular decline in fertility (such as reduced child mortality and the availability of contraceptive technology) have mostly run their course. This suggests that past fertility levels may be unlikely to be observed in the future.

Although overall fertility levels have experienced wide variation, the age pattern has had a relatively stable shape over long periods of time. The top center panel of Figure 1 illustrates the age profile of fertility for the years 1955, 1980 and 2005. Dividing by the TFR at each particular period yields the corresponding relative fertility rates. Clearly, when the effect of changing overall fertility levels is removed from the age profile, the result is greater regularity in the movement of the age-specific birth-rates.

In general, the birth rates for women in their teens and twenties have moved closely together in the same direction through most of the historical period. On the other hand, up until about the mid-1970s, the fertility rates at older ages (women in their forties and mid-thirties), trended downward regardless of the overall fertility level. The most recent experience is of particular interest. While birth rates for women over thirty years of age have steadily increased since the 1980s, fertility for teenagers and women in their early twenties has actually declined over the last 15 years or so. This tempo effect has been attributed in large part to a fundamental shift in the behavior of women entering the labor force, choosing to postpone motherhood

Figure 1: Selected features of the vital rates.



early in their careers. The mean age of a mother at first birth in the United States has steadily increased over the last several decades.

The last component of the vital rates (net migration) is not directly modeled in this paper. This is in part due to the lack of reliable historical data (particularly when it comes to illegal immigration). Moreover, future legal immigration is solely a function of legislation. For the purpose of generating population forecasts, we take the Alternative 2 scenario as the future realization of the series. The Trustees develop net migration assumptions by gender and single year of age, from birth to age 84. The bottom panel of Figure 1 illustrates the Alternative 2 male and female age profiles (expressed in millions of people) over the years 2009, 2025 and 2100.

3 A Benchmark Model

Possibly the best known stochastic model for mortality and fertility forecasting is some variation of the framework pioneered by Ronald Lee and Lawrence Carter. The so called Lee-Carter model (Lee and Carter; 1992), postulates the logarithm of a set of age-specific mortality rates as the linear function of a time-varying index of the general level of mortality

$$\log(y_{i,t}) = \alpha_i + \beta_i k_t + \varepsilon_{i,t}, \tag{1}$$

for $i = 1, \ldots, m$ and $t = 1, \ldots T$, where the term $y_{i,t}$ denotes mortality at age i and time t.

The age-specific set of parameters α_i describes the average shape of the log-mortality rates, while the slope coefficients β_i determine both the direction and magnitude by which mortality at every age varies with the index k_t. Subject to appropriate identifying constraints, least square estimates of the parameters can be obtained by applying singular value decomposition to the matrix of centered logarithmic death rates. Then, the estimates of the mortality index are treated as a univariate time-series that is projected into the future. In most applications, a random-walk with drift is found to provide a good empirical fit for k_t

$$k_t = k_{t-1} + \mu + e_t, \qquad e_t \sim N(0, \sigma_e^2). \tag{2}$$

Finally, the resulting forecasts of the mortality index are substituted back into equation (1), in order to recover a probability distribution of the future death rates for each age category.

There are two important modifications of the Lee-Carter framework that tend to significantly improve forecast accuracy. One correction involves setting α_i to the logarithm of the age-specific mortality rates corresponding to the last period in the sample. This enables the generated projections to match the initial conditions in the jump-off year. The second improvement deals with the fact that the OLS estimates of the parameters minimize mean square error in the logarithm of the death rates. As a result, the total number of deaths predicted by the model is not guaranteed to equal the observed death counts in the sample. The proposed solution is a second-stage re-estimation of the mortality index k_t, to match some aggregate quantity in the sample like observed deaths or life expectancy at birth.[1]

The Lee-Carter model yields a parsimonious stochastic approach to mortality forecasting that is easy to implement and often produces reasonable forecasts for all-cause age-specific mortality. There are however, certain limitations. In particular, a linear trend in the mortality index k_t entails a constant geometric rate of decline for each age-specific mortality

$$\frac{d\log(y_{i,t})}{dt} = \beta_i \frac{dk_t}{dt}. \tag{3}$$

Moreover, regardless of which univariate time series process is used to forecast k_t, the rates of mortality decline for different ages (say i and j) maintain a constant ratio to one another over time

$$\frac{d\log(y_{i,t})/dt}{d\log(y_{j,t})/dt} = \frac{\beta_i\ dk_t/dt}{\beta_j\ dk_t/dt} = \frac{\beta_i}{\beta_j}. \tag{4}$$

In addition, since the Lee-Carter model is estimated from a sampling theory perspective, it incorporates uncertainty through a single source (the sampling uncertainty derived from forecasting k_t). Specifically, the generated forecasts are conditional on the point estimates of the parameters. As a result, any uncertainty in the estimation of α_i, β_i and the drift parameter μ is not captured by the projections.

Borrowing from the Lee-Carter analytical framework, Lee (1993) proposes a similar technique to generate long-term projections of fertility. In this case, the birth rates themselves $f_{i,t}$ are postulated to be a function of a time-varying index of the general level of fertility f_t

$$f_{i,t} = \alpha_i + \beta_i f_t + \varepsilon_{i,t}. \tag{5}$$

Unique identification of the parameters is obtained through standard restrictions, requiring the sums of f_t and β_i to respectively equal zero and one. These constraints lead to a convenient interpretation of the fertility index in each period as the deviation of the total fertility rate from its long-term average A

$$\mathrm{TFR}_t = f_t + A + E_t, \tag{6}$$

with $A = \sum_i \alpha_i$ and $E_t = \sum_i \varepsilon_{i,t}$. Upon estimation of the parameters, the fitted total fertility rate $F_t = f_t + A$ (or alternatively the fertility index f_t) provides a univariate

[1] For further details about the Lee-Carter approach see for instance Lee (2000) and Lee and Miller (2001).

time-series that can be projected into the future using standard Box-Jenkins methods. Lee (1993) finds an ARMA(1,1) specification to yield an adequate fit

$$F_t = c + \phi F_{t-1} + e_t + \gamma e_{t-1}, \qquad e_t \sim N(0, \sigma_e^2). \tag{7}$$

Finally, the forecasts of the age-specific fertility rates $f_{i,t}$ are recovered via equation (5) from those of F_t.

Notice that the linear specification in equation (5) cannot prevent the occurrence of negative birth rates. As in the case of mortality, Lee (1993) initially considered modeling the logarithm of $f_{i,t}$, which effectively imposes a lower bound of zero. However, he found this approach yielding a poor fit of the TFR over the sample period. The problem with the alternative logarithmic transformation derives from the influence of the declining trend in fertility at the older ages. As a result, Lee (1993) advocates modeling the fertility rates themselves and simply discarding any negative simulated paths. As with the mortality model, two important adjustments generally improve forecast performance. First, it is recommended that the intercept estimates α_i be set to the observed age-specific birth rates in the jump-off year of the projections. In addition, a second stage re-estimation of the fitted total fertility rate F_t matches fitted and actual aggregate births.

In forecasting fertility, the biggest departure from the Lee-Carter approach to mortality is by far the explicit incorporation of expert opinion. Lee (1993) found that various ARMA(p,q) specifications could fit the data equally well, but produce quite different point forecasts. Even small sustained differences in the TFR can have a huge impact over time on the resulting population projections. Thus, Lee (1993) thought it best to constrain the central tendency of the fitted TFR to yield some prescribed ultimate value F^*. We follow this approach by setting $F^* = 2$. A TFR of two children per woman represents the intermediate or Alternative 2 assumption by the Board of Trustees to the U.S. Social Security Administration. For a covariance stationary ARMA(1,1) process, the unconditional mean of its forecasts can be forced to have the prescribed value by estimating the alternative specification

$$F_t = 2(1 - \phi) + \phi F_{t-1} + e_t + \gamma e_{t-1}. \tag{8}$$

Using this approach in combination with the Lee-Carter mortality model, Lee and Tuljapurkar (1994) generate long-term stochastic forecasts of the U.S. population. Again, as in the case of mortality, the fertility forecasts fail to incorporate any estimation uncertainty for α_i, β_i and the parameters in the ARMA process.

4 Bayesian Vector Autoregressions

Originally introduced by Sims (1980), the vector autoregressive specification generalizes a univariate autoregression, so that the current value of the i^{th} time-series is not just determined by its own lags, but also by the past values of all other variables in the system. Formally, the so-called reduced form of a VAR(p) model encompassing m time-series can be represented by

$$\mathbf{\Pi}(L)\boldsymbol{y}_t = \mathbf{\Phi}\boldsymbol{d}_t + \boldsymbol{\varepsilon}_t, \qquad \text{for } t = 1, \ldots, T, \tag{9}$$

where $\boldsymbol{y}_t$ is an $m \times 1$ dimensional vector of the variables in the model at time t, $\boldsymbol{d}_t$ is a $q \times 1$ vector of exogenous variables or deterministic components, L is the back-shift operator with $\mathbf{\Pi}(L) = \boldsymbol{I}_m - \mathbf{\Pi}_1 L - \cdots - \mathbf{\Pi}_p L$, and $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{\Sigma})$. The p matrices $\mathbf{\Pi}_i$ of dimension $m \times m$ contain the autoregressive coefficients, while in the special case $\boldsymbol{d}_t = 1$, the set of parameters $\mathbf{\Phi}$ reduces to an $m \times 1$ vector of intercepts

$$\boldsymbol{y}_t = \mathbf{\Phi} + \mathbf{\Pi}_1 \boldsymbol{y}_{t-1} + \mathbf{\Pi}_2 \boldsymbol{y}_{t-2} + \cdots + \mathbf{\Pi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t. \tag{10}$$

VAR processes can be very useful in capturing dynamic correlation patterns among multiple data series. However, their versatility can also come at a steep cost. Specifically, the number of coefficients in a VAR model increases geometrically with the number of time-series in the system. For instance, a VAR(p) process with m equations and no deterministic components other than an intercept comprises a total of $pm^2 + m$ coefficients, in addition to the $m(m+1)/2$ distinct elements in the disturbance covariance matrix $\mathbf{\Sigma}$. As the number of parameters increases relative to the number of available observations, the result is imprecise coefficient estimates (large associated standard errors). This problem is typically referred to as overfitting and often leads to poor forecast performance.

One way to overcome over-parametrization in a model is by imposing restrictions on the range of values some or all of the parameters can have. In this context, the Bayesian approach offers a natural framework. Let $L(\mathbf{Y}|\mathbf{\Pi}, \mathbf{\Phi}, \mathbf{\Sigma})$ denote the likelihood function of a vector autoregression, where $\mathbf{Y}$ represents a set of m observed time-series. From a Bayesian perspective, a complete model also requires the specification of a prior density on the parameters $\pi(\mathbf{\Pi}, \mathbf{\Phi}, \mathbf{\Sigma})$. Bayesian inference then proceeds by conditioning on the observables through Bayes's theorem, in order to derive the so-called posterior density[2]

$$\pi(\mathbf{\Pi}, \mathbf{\Phi}, \mathbf{\Sigma}|\mathbf{Y}) \propto \pi(\mathbf{\Pi}, \mathbf{\Phi}, \mathbf{\Sigma})\ L(\mathbf{Y}|\mathbf{\Pi}, \mathbf{\Phi}, \mathbf{\Sigma}). \tag{11}$$

The prior density represents any knowledge the researcher has about the model's parameters prior to observing the data. Thus, it provides the means to accommodate any available non-sample information. This information might be guided by statistical theory considerations, empirical evidence from previous analysis or simply the

[2] A verbal statement of Bayes's theorem is simply that the posterior density is proportional to the product of the prior and likelihood function.

judgment of the researcher. On the other hand, Bayes's theorem can be thought of as the mechanism through which the sample information contained in the likelihood function modifies the knowledge embodied in the prior. Once the posterior density $\pi(\mathbf{\Pi}, \mathbf{\Phi}, \mathbf{\Sigma}|\mathbf{Y})$ is derived, it can be used to: (1) produce optimal point estimates of the parameters; (2) construct credible intervals and regions of the parameter space corresponding to some posterior probability; and (3) estimate any feature of the predictive density associated with future observations.

Sims (1980) was well aware of the potential for over-parametrization in VAR models and was first in suggesting using the Bayesian framework. The key to this effort was finding sensible prior restrictions to impose on the autoregressive coefficients, as a means to increase precision. This led to the development of the well-known random-walk or Minnesota prior described by Litterman (1986), as well as subsequent extensions of this framework proposed by Doan, Litterman and Sims (1984), Kadiyala and Karlsson (1993, 1997) and Sims and Zha (1998).[3]

Given an arbitrarily long lag p, Bayesian vector autoregressive (BVAR) models under the Minnesota prior impose stochastic constraints on the autoregressive coefficients, based on two premises. First, it is assumed that recent values are likely to contain more useful information about a variable's future direction than more distant ones. This restriction is implemented in practice by increasing the prior probability that the coefficients in the model are closer to zero as a function of lag length. In addition, it is assumed that the movement over time of the variables being modeled can be well approximated by a random-walk around an unknown deterministic component.

Formally, consider the reduced form of a VAR(p) model in terms of its univariate components

$$\begin{aligned}\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{m,t} \end{bmatrix} = \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \vdots \\ \Phi_m \end{bmatrix} + \begin{bmatrix} a_{111} & a_{121} & \cdots & a_{1m1} \\ a_{211} & a_{221} & \cdots & a_{2m1} \\ \vdots & & \ddots & \vdots \\ a_{m11} & a_{m21} & \cdots & a_{mm1} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \vdots \\ y_{m,t-1} \end{bmatrix} + \cdots \\ \cdots + \begin{bmatrix} a_{11p} & a_{12p} & \cdots & a_{1mp} \\ a_{21p} & a_{22p} & \cdots & a_{2mp} \\ \vdots & & \ddots & \vdots \\ a_{m1p} & a_{m2p} & \cdots & a_{mmp} \end{bmatrix} \begin{bmatrix} y_{1,t-p} \\ y_{2,t-p} \\ \vdots \\ y_{m,t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{m,t} \end{bmatrix}\end{aligned} \tag{12}$$

Let a_{ijp} denote the autoregressive coefficient associated with variable j in equation i at lag p, and let σ_{ijp} represent its corresponding prior standard deviation. Under

[3]See Geweke and Whiteman (2006) for a lively chronological account of these developments.

the Minnesota prior, each individual density of a_{ijp} is independent normal with mean equal to one for the first lag of the dependent variable in each equation, and zero for all other coefficients and all other lags

$$a_{ijp} \sim \begin{cases} N(1, \sigma_{ijp}^2), & \text{for } i = j \text{ and } p = 1. \\ \\ N(0, \sigma_{ijp}^2), & \text{otherwise.} \end{cases} \tag{13}$$

This is equivalent to setting the prior means of the parameters in equation (10) to $\mathbf{\Pi}_1 = \boldsymbol{I}_m$ and $\mathbf{\Pi}_l = \mathbf{0}_m$, for $l = 2, \ldots, p$. The result is to center the prior for the i^{th} equation around the random-walk specification

$$y_{i,t} = y_{i,t-1} + c_i + \varepsilon_{i,t}. \tag{14}$$

Moreover, the parameters corresponding to any exogenous variables or intercepts are also assumed independent normal.

$$\Phi_i \sim N(0, \sigma_{\Phi_i}^2). \tag{15}$$

All of the Minnesota prior standard deviations are determined by a set of four so-called hyperparameters (denoted by λ_1, λ_2, λ_3 and λ_4). When it comes to the autoregressive coefficients, two cases can be distinguished: (1) the prior standard deviations on the lags of the dependent variable and (2) the standard deviations associated with the autoregressive parameters of the remaining variables

$$\sigma_{ijp} = \begin{cases} \lambda_1 p^{-\lambda_3}, & \text{for } j = i. \\ \\ \lambda_1 \lambda_2(i,j) p^{-\lambda_3} \left(s_i / s_j\right), & \text{for } j \neq i. \end{cases} \tag{16}$$

Likewise, the prior standard deviations of the deterministic components (either intercepts or any exogenous variables corresponding to the set of parameters $\mathbf{\Phi}$), are specified as follows

$$\sigma_{\Phi_i} = \lambda_4 s_i. \tag{17}$$

The term (s_i/s_j) serves as a scale factor that accounts for variation due to differences in the measurement units of the variables. An "empirical Bayes" approach is followed in practice, estimating s_i from the data as the standard error in a p-lag autoregression of variable i.

The hyperparameter λ_1 represents the prior standard deviation on the first lag of variable i in equation i. Decreasing its value has the effect of shrinking the diagonal elements of $\mathbf{\Pi}_1$ toward one and all other elements toward zero. For this reason, λ_1 is often referred to as the "overall tightness" of the prior, as it determines how tightly the random-walk specification is imposed in practice.

The "lag weight" hyperparameter λ_2 affects the prior standard deviation of variable j in equation i, relative to λ_1. Setting $0 < \lambda_2 < 1$ shrinks the off-diagonal elements of $\mathbf{\Pi}_l$ toward zero for $l = 1, \ldots, p$. This implies that for a given equation, variation in the dependent variable is explained to a greater extent by its own lags than by any other variable. By contrast, setting $\lambda_2 = 1$ treats the lags of the dependent and all other variables equally. The notation $\lambda_2(i, j)$ emphasizes the possibility of specifying alternative degrees of shrinkage for different variables in a given equation.

The "lag decay" hyperparameter $\lambda_3 > 0$ in the Minnesota prior standard deviations has the effect of shrinking all the autoregressive coefficients toward their prior means as a function of lag length p. A value $\lambda_3 = 1$ implies a harmonic rate of decay. An increase in λ_3 shrinks higher-order lags toward zero more quickly, giving greater importance to more recent lags than to more distant ones.

Finally, the hyperparameter λ_4 determines the prior standard deviations associated with exogenous variables or intercept coefficients in the set of parameters $\mathbf{\Phi}$. Decreasing λ_4 tightens the prior on these parameters around their mean. Typically, the Minnesota prior is made non-informative on the intercept parameters by selecting a zero mean and an arbitrarily large value for λ_4.

4.1 Mean-Adjusted BVAR Models and the Steady-State

As referenced in Appendix A.1, it is well known that for a covariance stationary VAR model, the expectation of its long-horizon forecasts converges to the unconditional mean or steady-state of the process ($\boldsymbol{\mu}$ in equation 27). In other words, when it exists, $\boldsymbol{\mu}$ represents the future expected value of the time-series in the indefinite future. Villani (2006) considers Bayesian estimation for the following alternative re-parametrization of the VAR model in mean-adjusted form

$$\mathbf{\Pi}(L)(\boldsymbol{y}_t - \mathbf{\Psi}\boldsymbol{d}_t) = \boldsymbol{\varepsilon}_t, \qquad \text{for } t = 1, \ldots, T, \tag{18}$$

which again, for $\boldsymbol{d}_t = 1$ reduces to

$$\boldsymbol{y}_t = \mathbf{\Psi} + \mathbf{\Pi}_1(\boldsymbol{y}_t - \mathbf{\Psi}) + \mathbf{\Pi}_2(\boldsymbol{y}_t - \mathbf{\Psi}) + \cdots + \mathbf{\Pi}_p(\boldsymbol{y}_t - \mathbf{\Psi}) + \boldsymbol{\varepsilon}_t. \tag{19}$$

The reduced-form parametrization in equation (10) is linear in $\mathbf{\Phi}$, $\mathbf{\Pi}_1$,..., $\mathbf{\Pi}_p$, while the steady-state $\boldsymbol{\mu}$ is a complex non-linear function of these parameters. By contrast, the mean-adjusted model in equation (19) is non-linear in the coefficients. However, in this case, the unconditional mean of the process $\boldsymbol{\mu} = \mathbf{\Psi}$ is an explicit component of the specification, facilitating the introduction of a prior on the steady-state.

Clearly, the choice between these two alternative VAR parametrization boils down to whether one prefers specifying a prior density for a set of intercepts $\mathbf{\Phi}$ or for a set of steady-state parameters $\mathbf{\Psi}$. We favor the latter option for good reasons. First, quantifying any prior judgment about the intercepts is difficult. In fact, this is why most applications of BVAR models under the Minnesota prior tend to be uninformative for the intercept coefficients. On the other hand, the steady-state parameters in the mean-adjusted form have an intuitive interpretation. It is certainly much easier to provide an educated guess about the likely range of say future average fertility for a specific age group, than it is to provide a similar opinion about the intercept coefficient in the corresponding VAR equation.

A second advantage of the mean-adjusted parametrization is that even a mildly informative steady-state prior might prevent the generated long-run projections from ending up differing substantially from prior belief. Of course, at the other end of the spectrum of preferences, one may wish to exclude any prior opinion at all. In this case, the steady-state prior can be made uninformative by selecting an arbitrarily large standard deviation. This way, the prior density remains relatively flat over the relevant region of the parameter space, making any two intervals of similar size within such region roughly equally likely a priori. Both scenarios are explored in this paper.

Early computation of BVAR models under the Minnesota prior initially used a quasi-Bayesian equation-by-equation mixed estimation strategy involving a number of simplifying assumptions. The most troubling of these was to require a fixed known diagonal residual covariance matrix $\mathbf{\Sigma}$, with diagonal elements s_i^2 determined from the data. Kadiyala and Karlsson (1993, 1997) have extended this framework to full-Bayesian system-wide estimation, allowing for a random non-diagonal residual covariance matrix. A particularly flexible approach is the so-called normal-diffuse prior specification, which has been incorporated by Villani (2006) in the mean-adjusted version of the BVAR model.

Formally, let $\mathbf{\Pi} = \left[\begin{array}{ccc} \mathbf{\Pi}_1 & \cdots & \mathbf{\Pi}_p \end{array} \right]'$. Assuming prior independence between $\mathbf{\Sigma}$, $\mathbf{\Psi}$ and $\mathbf{\Pi}$, the normal-diffuse prior takes the form

$$\pi(\mathbf{\Sigma}) \propto |\mathbf{\Sigma}|^{-(m+1)/2}, \qquad \mathrm{vec}(\mathbf{\Pi}) \sim N(\boldsymbol{\theta}_\Pi, \mathbf{\Omega}_\Pi), \qquad \mathrm{vec}(\mathbf{\Psi}) \sim N(\boldsymbol{\theta}_\Psi, \mathbf{\Omega}_\Psi). \tag{20}$$

The prior corresponding to the disturbance covariance matrix $\mathbf{\Sigma}$ is the usual diffuse specification in multivariate normal linear models, known as Jeffrey's prior. The autoregressive coefficients and steady-state parameters are multivariate normal. The restrictions implied by the Minnesota prior can be incorporated by setting $\boldsymbol{\theta}_\Pi$ and $\mathbf{\Omega}_\Pi$ to the individual elements defined in expressions (13) and (16), respectively. Meanwhile, prior belief about the future long-run expectation of the variables in the system can be used to specify the prior means and variances corresponding to the unconditional mean $\mathbf{\Psi}$. Alternatively, the steady-state prior can be made uninformative by selecting large individual variances $\mathbf{\Omega}_\Psi$.

At the heart of all Bayesian inference is estimation of the posterior density, which in turn requires the evaluation of an often multi-dimensional integral. MCMC

(Markov Chain Monte Carlo) techniques have revealed as indispensable tools to sample from nonstandard distributions, having a profound effect on the advancement of Bayesian and non-Bayesian inference alike.[4]

In the context of BVAR models, the flexibility of the normal-diffuse prior framework comes at the expense of an analytically intractable posterior. The lack of a closed-form solution is of no consequence, as long as a representative sample of posterior parameters can be drawn. Kadiyala and Karlsson (1997) evaluate the efficiency of alternative computational algorithms, favoring a Gibbs sampling approach. The latter yields a joint posterior sample of parameter draws by sampling iteratively from the set of known conditional distributions. Specifically, a three-block Gibbs sampler is required to cycle over $\mathbf{\Sigma}$, $\mathbf{\Pi}$ and $\mathbf{\Psi}$. For ease of reference, the full conditional posterior densities of the parameters derived by Villani (2006) are reproduced in Appendix A.2.

In summary, the nature of the prior information guiding the BVAR models specified in this paper are driven by multiple considerations. A diffuse prior for the disturbance covariance matrix $\mathbf{\Sigma}$ seems appropriate, given the lack of prior opinion and obvious difficulties in quantifying it. The Minnesota-type restrictions on the autoregressive coefficients $\mathbf{\Pi}$ are designed to avoid overfitting. Diffuse or informative priors on the steady-state $\mathbf{\Psi}$ can be argued on their own merits. The former reflect a preference for forecasts that are determined by the assumed data generating process and observed data only. The latter favor the notion of expert opinion additionally guiding the resulting forecasts. We provide an illustration of both cases.

Finally, one last prior restriction imposed on the estimated models is covariance stationarity. This constraint is motivated by two reasons. First, the mean-adjusted BVAR parametrization is locally non-identified. Specifically, any posterior draw of $\mathbf{\Pi}$ failing the stationarity condition described in Appendix A.1 implies that the unconditional mean does not exist and thus, $\mathbf{\Psi}$ cannot be identified. Formally, the stationarity constraint enters the prior multiplicatively through the following indicator function

$$\mathrm{I}(\mathbf{\Pi}^*) = \begin{cases} 1, & \text{if } \ |\eta_{\max}| < 1. \\ \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

with $\mathbf{\Pi}^*$ representing the matrix defined in (25). Computationally this restriction simply requires rejecting any parameter draws in the Gibbs chain violating the stability of the model. A further reason justifying the stationarity constraint is the prior belief that it is implausible for the time-series being modeled to display explosive behavior. From a practical standpoint, this is a key assumption in the context of the

[4]An introduction and tutorial to these methods can be found in Brooks (1998), while more detailed treatment is provided for instance in Chen et al. (2000).

lengthy forecast horizons entertained. Hence, even under the alternative BVAR parameterization in equation (10), where identification is not an issue, it is still sensible to impose stationarity.

5 Vital Rate and Population Forecasts

We compare vital rate and population projections from three separate specifications over a forecast horizon of 92 years (2009-2100). The first model is based on the framework for mortality and fertility forecasting discussed in Section 3, which is jointly referred to as the Lee-Carter model from this point forward. The second and third specifications are BVAR models, both fitted to the age-specific mortality and fertility rates.

The 2 BVAR models entertained are identical in the number of lags and the selected values of their corresponding hyperparameters. They differ only with respect to the prior specification of the steady-state coefficients. In one case, the steady-state prior density is made uninformative by setting the parameter means to zero and the standard deviations to ten. This variant of the process is referred to as the "diffuse" BVAR model through out the paper.

The second Bayesian vector autoregression is specified so as to incorporate some degree of judgment about the long-run mean of the generated forecasts. For this "informative" BVAR model, the steady-state prior means are set to the Alternative 2 ultimate assumptions in the 2009 Trustees Report. Formalizing such prior judgment further requires the specification of standard deviations. Their magnitude determines in turn the level of prior opinion uncertainty. In order to quantify our opinion about the range of possible variation in the steady-state, we settle on a fixed rule that takes into account both historical experience and expert judgment. In particular, for each series, we use the standard deviation corresponding to the cumulative mean of the joint path followed by the historical data and the Alternative 2 projections.

From a computational standpoint, all of the reported forecast estimates for the Lee-Carter model are based on a total of 10,000 simulations. The BVAR models were estimated by running 12,000 iterations of the three-block Gibbs sampler referenced in Section 4 and Appendix A.2. The initial 2,000 iterations were discarded as part of the burn-in phase of the MCMC chain, leaving a sample of 10,000 posterior draws for inference. Convergence of the sampler was monitored informally by looking at trace plots for selected coefficients.[5]

[5] See Cowles and Carlin (1996) for a review of convergence diagnosis in MCMC samplers.

5.1 Fertility Projections

Fertility rates for 7 age groups over the historical period from 1917 to 2008 are modeled within both the Lee-Carter and BVAR frameworks. For the former, the intercept parameters α_i are set to the observed birth rates in 2008, while fitted total fertility is matched to the actual number of aggregate births. In addition, the TFR follows a constrained ARMA(1,1) with an unconditional mean of 2. The resulting parameter estimates are shown in Table 1, along with the values of the age-specific coefficients α_i and β_i.

Bayesian vector autoregressions with $p = 2$ lags are estimated using the logarithm of fertility instead. In this case, the standard random-walk assumption of the Minnesota prior does not provide a suitable approximation to the individual series. Therefore, it is replaced with the prior assumption that each variable follows a stationary univariate first-order autoregression with a coefficient value of 0.8. Even with this change, the logarithmic transformation of the fertility rates tends to induce explosiveness, with a large portion of the parameter draws being rejected due to non-stationarity. This problem is solved by setting the overall tightness hyperparameter to a small value ($\lambda_1 = 0.05$). As a consequence, the proportions of non-stationary draws discarded are only 5.4% and 1.8% for the diffuse and informative BVAR models, respectively. The lag-decay hyperparameter is set to $\lambda_3 = 1$, while the lag-weight of the coefficients for other variables in each equation takes the value $\lambda_2 = 0.5$.

Notice that each of the 7 equations in the BVAR system contains 14 autoregressive coefficients. Hence, we do not bother reporting any posterior summaries for these parameters. Table 1, however, presents the prior values assigned to the steady-state densities. In all cases, for the diffuse version of the BVAR process the prior means are given a value of 0, while the standard deviations are set to 10. On the other hand, the steady-state prior densities for the informative BVAR model are centered around the logarithm of the corresponding Alternative 2 ultimate fertility assumptions. Likewise, the standard deviations are those of the cumulative means resulting from the joint path followed by the historical data and the Alternative 2 projections.

For each of the 7 fertility age groups and the total fertility rate, the top and central panels in Figure 2 display the forecasts generated by the models over the period from 2009 to 2100. The graphs illustrate the means and 90% credible intervals of the posterior predictive densities from the BVAR models (solid lines). The corresponding quantities from the Lee-Carter approach are shown as dashed lines. For reference, the Alternative 2 scenario is also included.

One of the consequences of modeling the logarithm of fertility in the BVAR models is that the steady-state prior densities associated with the actual birth rates are log-normal. Exponentiating the posterior draws of $\mathbf{\Psi}$ allows us to recover posterior

Table 1: Selected parameter estimates and prior coefficients for the fertility models.

	Lee-Carter		Diffuse Prior $\mathbf{\Psi}$		Informative Prior $\mathbf{\Psi}$	
Age Group	α_i	β_i	Mean	Std. Dev.	Mean	Std. Dev.
15-19	0.1694	0.1107	0	10	-1.7755	0.1185
20-24	0.4478	0.3645	0	10	-0.8034	0.1289
25-29	0.5327	0.2419	0	10	-0.6298	0.0893
30-34	0.5198	0.1327	0	10	-0.6543	0.0852
35-39	0.2696	0.1047	0	10	-1.3108	0.2091
40-44	0.0548	0.0411	0	10	-2.9041	0.3718
45-49	0.0033	0.0044	0	10	-5.7138	0.6496

Lee-Carter - F_t: ARMA(1,1)		BVAR	
ϕ	0.9645	p	2
γ	0.4903	λ_1	0.05
σ_e	0.1012	$\lambda_2(i,i)$	1
-	-	$\lambda_2(i,j)$	0.5
-	-	λ_3	1
-	-	Prior Mean diag($\mathbf{\Pi}_1$)	0.8

samples for the variables in their original form. We show the prior densities corresponding to the informative BVAR specification in the bottom panels of Figure 2, along with estimates of the posterior steady-state densities for the diffuse and informative versions of the model. The latter were approximated using a kernel density estimator.

It is important to recognize that the Lee-Carter approach basically models the TFR and then allocates the resulting fertility across the ages, on the basis of their intercept and slope coefficients. A BVAR process, on the other hand, models directly the age-specific fertilities and then recovers the TFR as the sum of all the paths. By design, the mean forecast of the TFR for the Lee-Carter model is constrained to the Alternative 2 assumption. For instance, its value in the year 2100 is 2.01 children per woman. The individual steady-state priors for the informative BVAR model induce a similar point forecast of 2.01 children per woman and narrower bounds than the Lee-Carter model. Not surprisingly, the diffuse BVAR version produces somewhat wider interval forecasts. In addition, its mean slightly overshoots the Alternative 2 value (a mean TFR of 2.13 children per woman in the year 2100).

In general, for the first three age groups (age 15-29), the posterior steady-state densities corresponding to the diffuse BVAR specification are heavily skewed to the left and rather spread out in comparison with the informative priors. As a result, the

Figure 2: Fertility forecasts and steady-state densities.

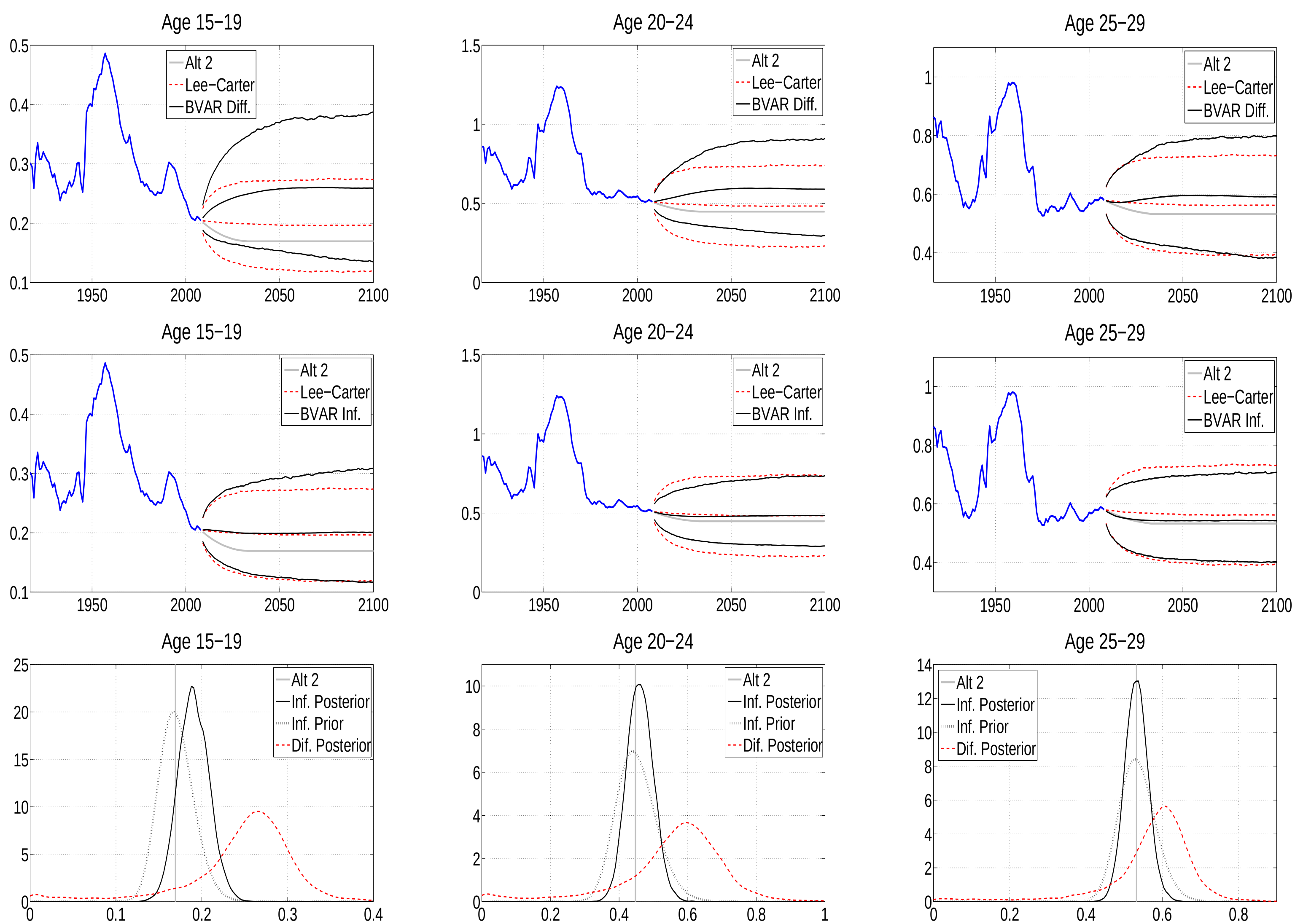


Figure 2: (Continued).

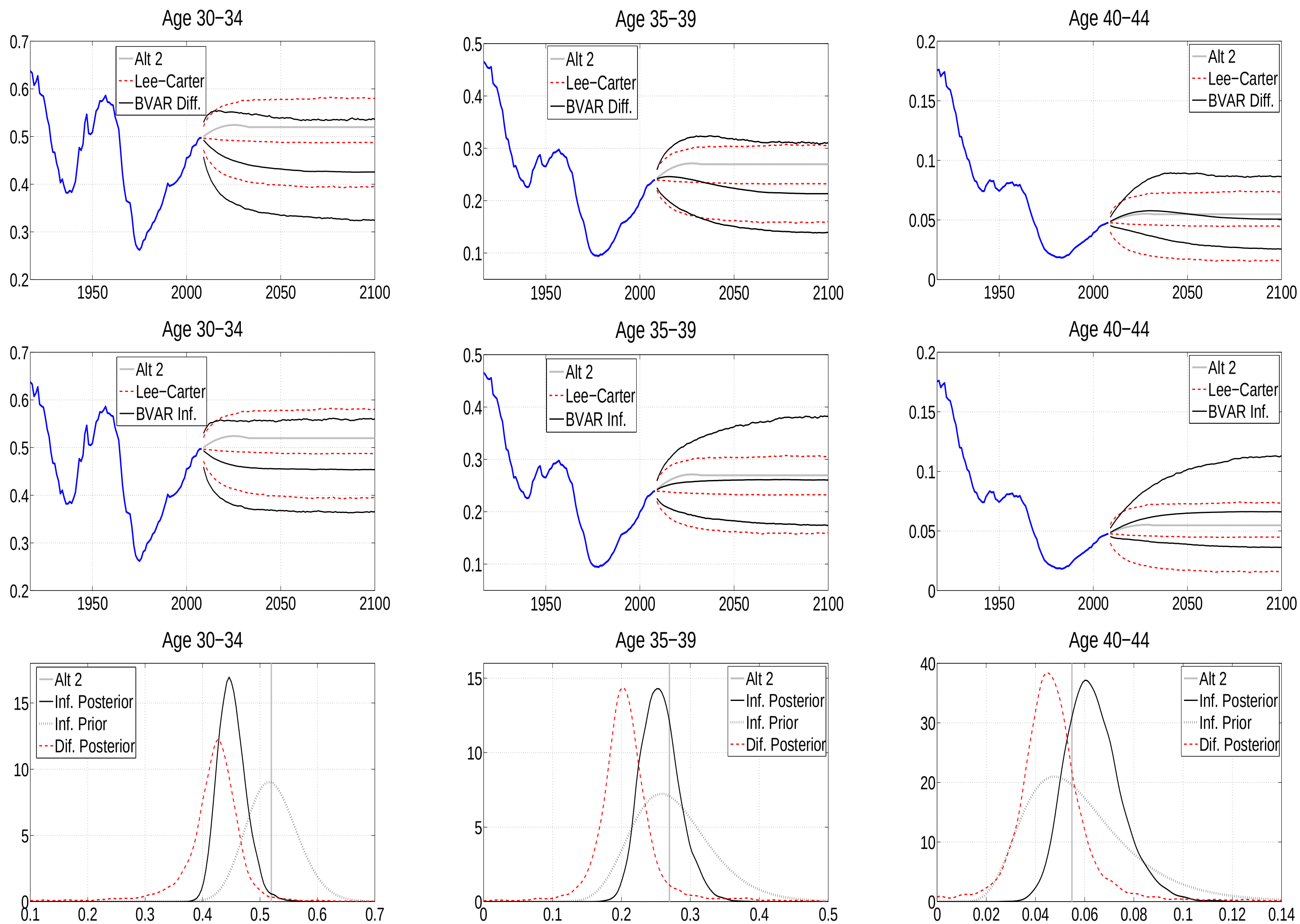
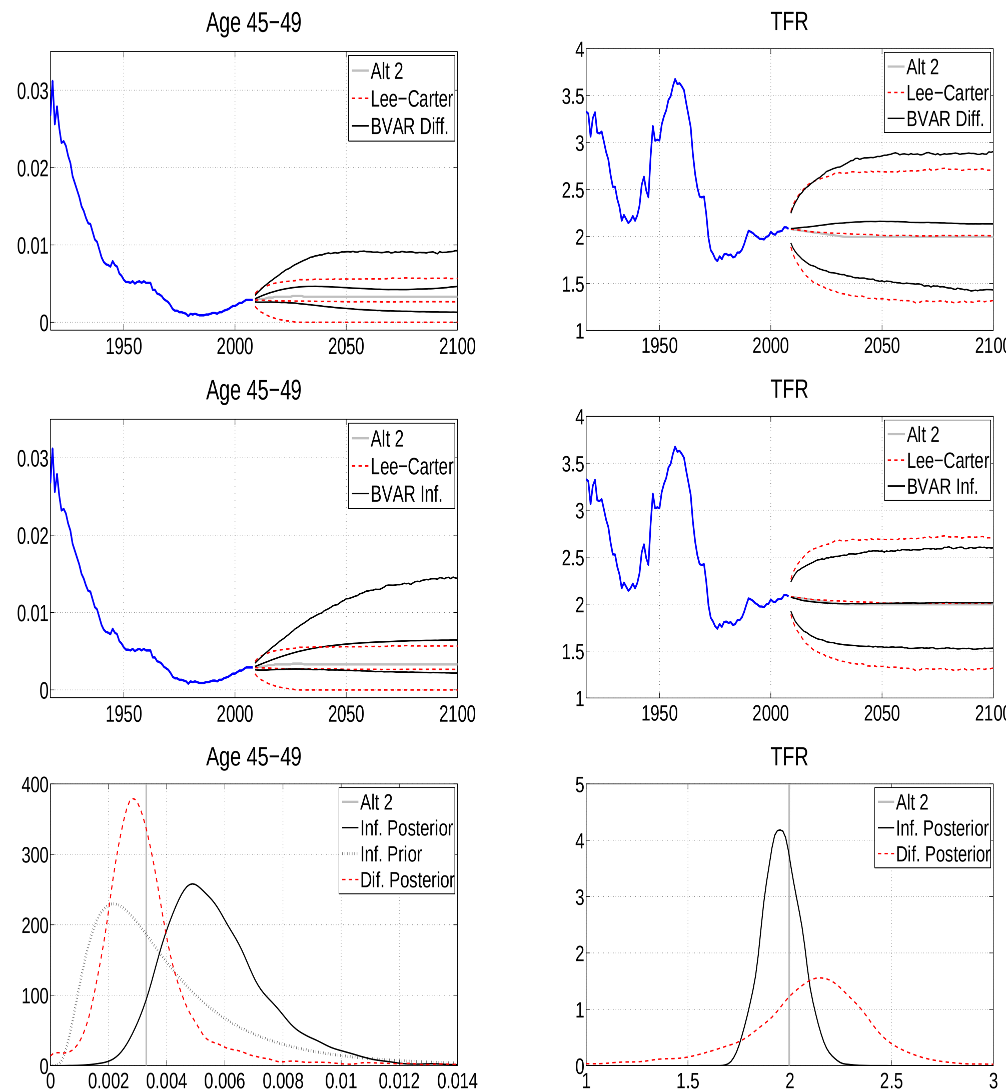


Figure 2: (Continued).



mean projections of the informative model tend to move in the direction of the intermediate assumptions and come closer to them. However, for the oldest ages (age groups 40-44 and 45-49), the diffuse posterior densities of the steady-state are actually more precise than their informative counterparts. This is due to the particular characteristics of the log-normal distribution. Specifically, the steady-state normal prior with zero mean and variance equal 100 is very non-informative when it comes to the logarithm of age-specific fertility. However, the log-normal prior density corresponding to the fertility rates places a significant amount of mass within a small interval around zero, while being roughly flat over the remaining range. As a result, this prior density can actually be quite informative in some cases. It is possible to define a more suitable diffuse prior by setting a non-zero mean or re-scaling the data, although the issue is not pursued any further.

5.2 Mortality Projections

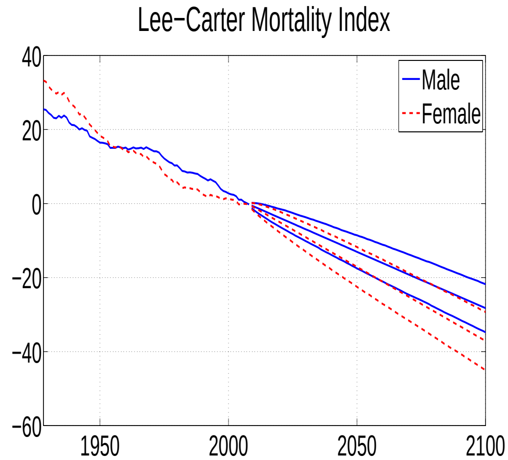
Separate models for male and female mortality are estimated, each comprising 22 age groups. In order to avoid excessive volatility (particularly the severe variation in mortality due to the 1918 Spanish influenza pandemic), we settle on a sample covering the historical period from 1928 to 2008. The Lee-Carter approach models the logarithm of the individual rates. As it is customary, the intercept parameters α_i are set to the observed logarithmic death rates in 2008, while a second-stage re-estimation of the mortality index matches life expectancy at birth in the sample. Tables 2 and 3 display the parameter estimates. Clearly, the value of the estimated drift coefficients μ indicates greater mortality improvement for females than for males. This is is also evident in Figure 3, which displays estimates of the mortality index k_t, along with mean and 90% interval projections.

The Bayesian vector autoregressions are fitted in terms of first differences in logarithmic mortality $\Delta \log(y_{i,t}) = \log(y_{i,t}) - \log(y_{i,t-1})$, using $p = 1$ lag. This entails the estimation of 484 autoregressive coefficients per model. For logarithmic mortality, the random-walk assumption in the Minnesota prior provides a reasonable approximation. In light of the differenced transformation in the data, we set the prior mean of the first lag of the dependent variable in each equation to zero.[6] In all cases, the overall tightness hyperparameter is assigned a value $\lambda_1 = 0.2$. The value of the lag-decay hyperparameter λ_3 is irrelevant, since only one lag is specified. None of the draws of the Gibbs sampler had to be discarded due to lack of covariance stationarity.

One of the regularities in the data is the fact that the individual mortality rates tend to move closely together. A look at the 22×22 sample correlation matrix of $\Delta \log(\mathbf{Y})$ reveals an expected pattern, in that the degree of correlation tends to

[6]Consider a first order autoregression using differenced data: $\Delta x_t = a + b\Delta x_{t-1} + e_t$. In this case, a coefficient value $b = 0$ is consistent with a random-walk in levels: $x_t = a + x_{t-1} + e_t$.

Figure 3: Lee-Carter mortality index projections.



decrease with the order of temporal adjacency among the age groups. For instance, the sample correlation for male mortality between age 20-24 and age 25-29 is 0.85. On the other hand, the correlation between age 20-24 and age 90-94 is only 0.18. This feature in the correlation structure of the data implies that age groups that are closer to one another have greater ability at predicting each others movements. We incorporate such information by setting the lag-weight hyperparameter proportional to the correlation matrix $\lambda_2(i, j) = 0.8\times$ Corr$[\Delta \log(\mathbf{y}_i), \Delta \log(\mathbf{y}_j)]$, for $i \neq j$. In other words, for a given equation, the autoregressive coefficients for age groups that are far from the dependent variable age group have smaller prior variances. As a result, they experience greater shrinkage toward their zero means and have less influence over the future movement of the series.

Tables 2 and 3 show the values assigned to the prior steady-state densities of the BVAR models. Again, for the diffuse process the prior means are set to zero and the standard deviations to ten. On the other hand, the steady-state prior densities for the informative BVAR model are centered around the average difference in logarithmic mortality followed by the Alternative 2 projections. Similarly, the prior standard deviations are those of the cumulative means of differenced logarithmic mortality along the historical data and the Alternative 2 scenario.

For four selected male mortality age groups (ages 15-19, 30-34, 55-59 and 80-84), Figure 5 displays the means and 90% credible intervals of the posterior predictive densities over the period from 2009 to 2100. The corresponding quantities from the Lee-Carter approach are shown as dashed lines. Figure 6 presents similar projections for female mortality. We do not bother displaying the steady-state posterior densities, given that they represent the long-run means of differenced log mortality. It is also not possible to recover the steady-state densities of the mortality rates from the posterior

Table 2: Selected parameter estimates and prior coefficients for the male mortality models.

	Lee-Carter		Diffuse Prior $\mathbf{\Psi}$		Informative Prior $\mathbf{\Psi}$	
Age Group	α_i	β_i	Mean	Std. Dev.	Mean	Std. Dev.
0	-4.9850	0.0996	0	10	-0.0171	0.0083
1-4	-8.1296	0.1211	0	10	-0.0145	0.0149
5-9	-8.8945	0.1080	0	10	-0.0151	0.0081
10-14	-8.5490	0.0870	0	10	-0.0143	0.0076
15-19	-7.0854	0.0430	0	10	-0.0081	0.0098
20-24	-6.6182	0.0423	0	10	-0.0078	0.0093
25-29	-6.6776	0.0463	0	10	-0.0084	0.0102
30-34	-6.6362	0.0477	0	10	-0.0089	0.0101
35-39	-6.3195	0.0481	0	10	-0.0091	0.0065
40-44	-5.8767	0.0453	0	10	-0.0091	0.0093
45-49	-5.4323	0.0430	0	10	-0.0095	0.0075
50-54	-5.0708	0.0416	0	10	-0.0101	0.0073
55-59	-4.7490	0.0376	0	10	-0.0113	0.0096
60-64	-4.3261	0.0334	0	10	-0.0119	0.0058
65-69	-3.9058	0.0295	0	10	-0.0085	0.0064
70-74	-3.4451	0.0272	0	10	-0.0083	0.0053
75-79	-2.9897	0.0269	0	10	-0.0082	0.0070
80-84	-2.5137	0.0229	0	10	-0.0079	0.0062
85-89	-2.0075	0.0164	0	10	-0.0053	0.0052
90-94	-1.5035	0.0127	0	10	-0.0050	0.0058
95-99	-1.0979	0.0108	0	10	-0.0048	0.0070
100+	-0.8308	0.0097	0	10	-0.0044	0.0072

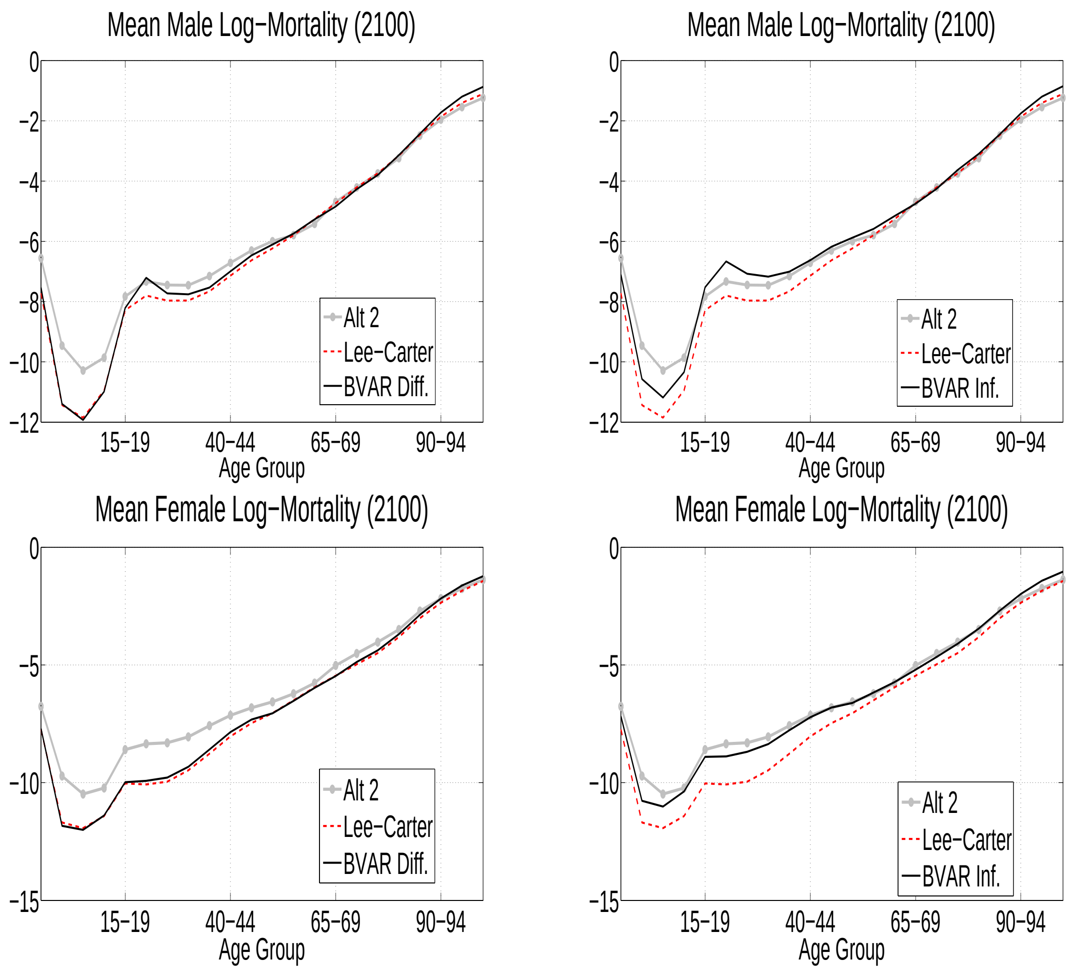
Lee-Carter - k_t: Random-Walk		BVAR	
μ	-0.3052	p	1
σ_e	0.4186	λ_1	0.2
-	-	$\lambda_2(i,i)$	1
-	-	$\lambda_2(i,j)$	$0.8\times$ Corr$[\Delta \log(\mathbf{y}_i), \Delta \log(\mathbf{y}_j)]$
-	-	λ_3	1
-	-	Prior Mean diag($\mathbf{\Pi}_1$)	0

Table 3: Selected parameter estimates and prior coefficients for the female mortality models.

	Lee-Carter		Diffuse Prior $\mathbf{\Psi}$		Informative Prior $\mathbf{\Psi}$	
Age Group	α_i	β_i	Mean	Std. Dev.	Mean	Std. Dev.
0	-5.1886	0.0703	0	10	-0.0171	0.0085
1-4	-8.3606	0.0925	0	10	-0.0148	0.0152
5-9	-9.0552	0.0795	0	10	-0.0155	0.0105
10-14	-8.8809	0.0699	0	10	-0.0147	0.0103
15-19	-7.9154	0.0581	0	10	-0.0074	0.0137
20-24	-7.6694	0.0662	0	10	-0.0074	0.0146
25-29	-7.5642	0.0659	0	10	-0.0081	0.0129
30-34	-7.2747	0.0604	0	10	-0.0085	0.0120
35-39	-6.8016	0.0540	0	10	-0.0085	0.0101
40-44	-6.3305	0.0466	0	10	-0.0088	0.0092
45-49	-5.9437	0.0417	0	10	-0.0095	0.0065
50-54	-5.6164	0.0390	0	10	-0.0103	0.0060
55-59	-5.2150	0.0350	0	10	-0.0110	0.0078
60-64	-4.7423	0.0329	0	10	-0.0112	0.0057
65-69	-4.2857	0.0319	0	10	-0.0080	0.0065
70-74	-3.8002	0.0319	0	10	-0.0078	0.0058
75-79	-3.3277	0.0318	0	10	-0.0076	0.0079
80-84	-2.8094	0.0274	0	10	-0.0075	0.0066
85-89	-2.2565	0.0205	0	10	-0.0049	0.0064
90-94	-1.7245	0.0172	0	10	-0.0049	0.0065
95-99	-1.2938	0.0150	0	10	-0.0049	0.0070
100+	-0.9732	0.0124	0	10	-0.0044	0.0062

Lee-Carter - k_t: Random-Walk		BVAR	
μ	-0.40	p	1
σ_e	0.4970	λ_1	0.2
-	-	$\lambda_2(i,i)$	1
-	-	$\lambda_2(i,j)$	$0.8\times$ Corr$[\Delta\log(\mathbf{y}_i), \Delta\log(\mathbf{y}_j)]$
-	-	λ_3	1
-	-	Prior Mean diag($\mathbf{\Pi}_1$)	0

Figure 4: Mean forecasts of logarithmic mortality in the year 2100.



draws of $\boldsymbol{\Psi}$. Clearly, the most distinguishable feature in the graphs is how much narrower the Lee-Carter interval forecasts are relative to those of the BVAR models, whether under an informative or a diffuse steady-state prior.

The logarithm of the mean forecasts in the year 2100 over the age profile are shown in Figure 4. For males at young ages (0-19), the Lee-Carter and diffuse BVAR models yield lower point predictions than the Trustees assumptions. This is also the case for the informative approach, although to a lower degree. However, the informative model predicts higher mortality than the Alternative 2 scenario at ages 20-49, while the opposite tends to apply to the mean forecasts of the diffuse and Lee-carter approaches. For females, the Lee-Carter and diffuse BVAR models forecast significantly lower mortality than the Trustees assumptions almost across the board. In this case, the informative BVAR model comes much closer the Alternative 2 scenario.

Figure 5: Male mortality forecasts for various age groups.

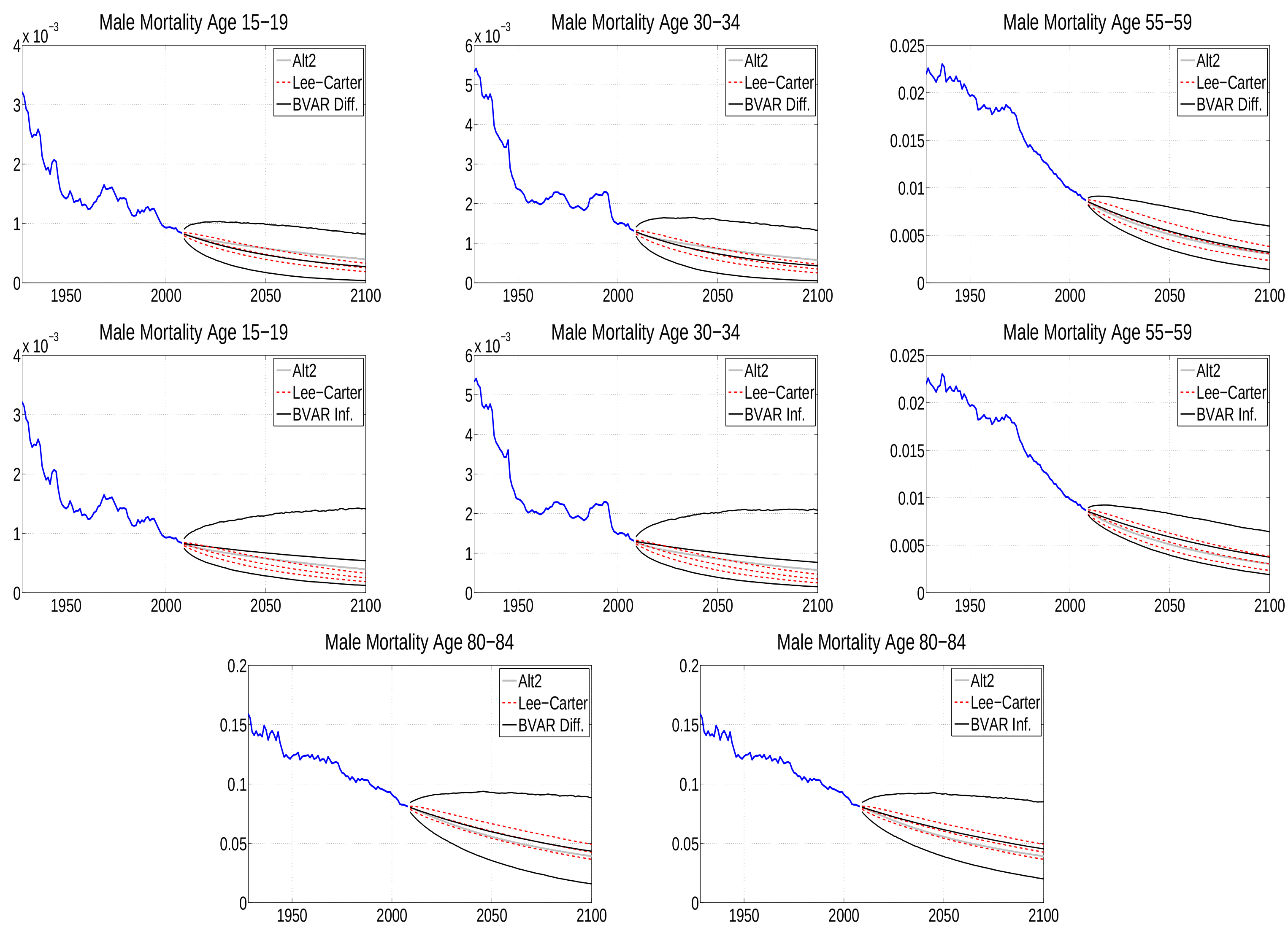


Figure 6: Female mortality forecasts for various age groups.

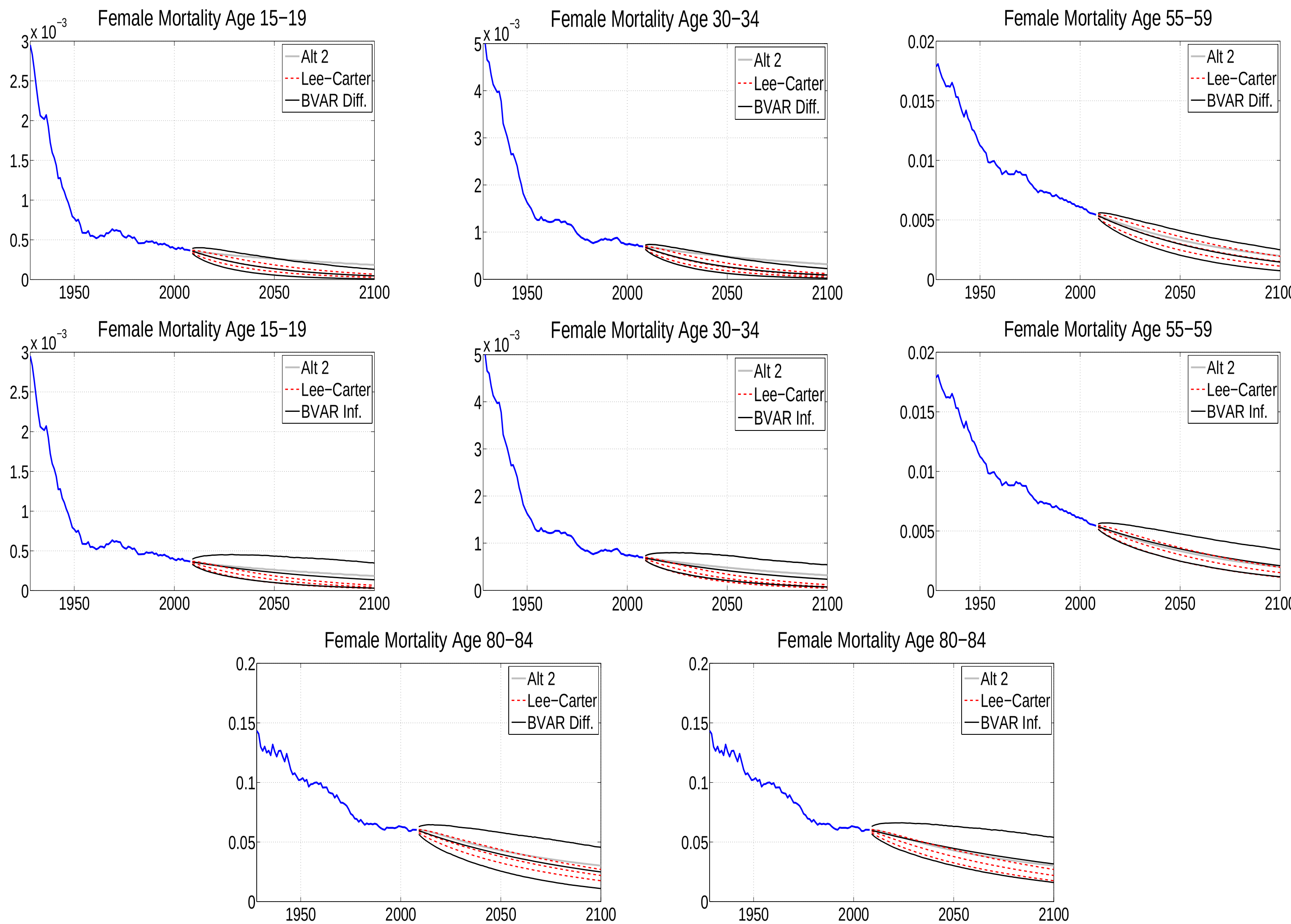
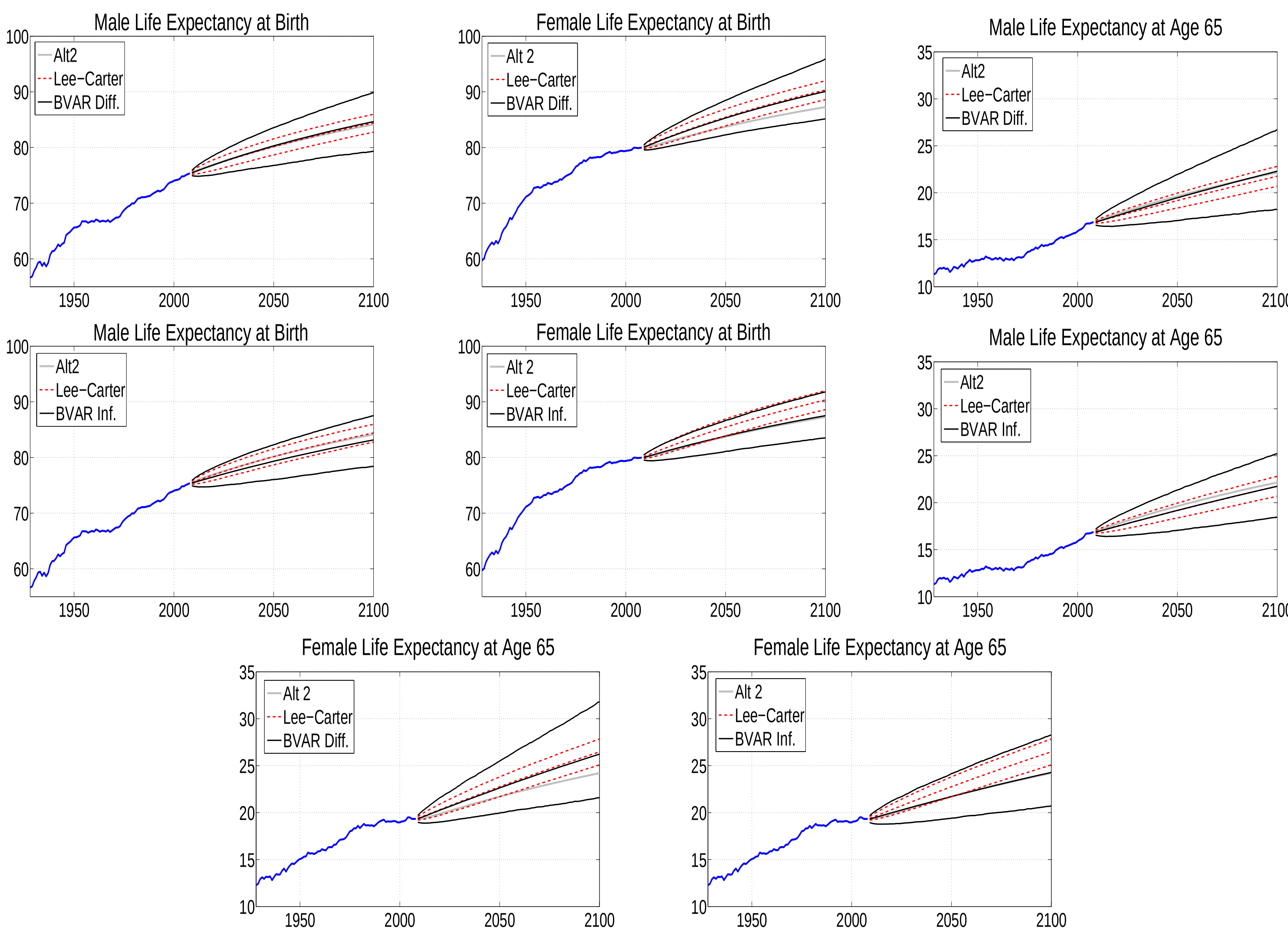


Figure 7: Forecasts of life expectancy at birth and at age 65 for males and females.



A metric that takes into account the aggregate effect of mortality projections is life expectancy $e_{i,t}$. It is formally defined as the average number of remaining years an individual age i at time t is expected to live, and it is a highly non-linear function of the age-specific mortality rates. Figure 7 presents male and female forecasts of life expectancy at birth and at age 65. Again, the 90% confidence interval forecasts corresponding to the Lee-Carter model appear quite narrow relative to the BVAR approach.

For males in the year 2100, the Alternative 2 assumption of life expectancy at birth is 84.1 years. The mean forecasts of the Lee-Carter and diffuse BVAR specifications are respectively 84.4 and 84.6. However, lower and upper bounds associated with the diffuse BVAR model range from 79.3 to 89.9 years of life. This range narrows to between 82.8 and 86 years with the Lee-Carter approach. For the informative BVAR model, the mean forecast of male life expectancy at birth in the year 2100 is 83.1, so it underpredicts the Alternative 2 value by exactly one year.

Finally, when it comes to female mortality, both the Lee-Carter and diffuse BVAR approaches forecast significantly better mortality improvement than assumed under the Alternative 2 scenario. In the year 2100, the Alternative 2 assumption of life expectancy at birth is 87.3 years. The informative BVAR model generates a mean forecast of 87.5. Meanwhile, the mean projections corresponding to the Lee-Carter and diffuse BVAR models are 90.3 and 90.1, respectively.

5.3 Population Projections

The generated vital rate forecasts are used to produce probabilistic population projections via the cohort component method. Following the description in Hollman et al. (2000), the cohort component method can be defined by the following equations

$$\begin{cases} P_t(a) = B_{t-1,t} - D_{t-1,t}(a) + M_{t-1,t}(a), & \text{for } a = 0. \\ P_t(a) = P_{t-1}(a-1) - D_{t-1,t}(a) + M_{t-1,t}(a), & \text{otherwise.} \end{cases} \tag{22}$$

where $P_t(a)$ denotes the population age a at time t, while $B_{t-1,t}$, $D_{t-1,t}$ and $M_{t-1,t}$ respectively represent births, deaths and net migration during the time interval from $t-1$ to t. Given the intermediate assumptions on net migration and the forecast paths of age-specific mortality and fertility, these equations provide a mechanism to recursively update forward the given population in the jump-off year.

Figure 8 displays male population forecasts for the following three broad age groups: age 0-19, age 20-64 and age 65-100+. These roughly correspond to young dependents, workers and retirees. The graphs present means and 90% credible intervals from the posterior predictive densities over the 2009-2100 period. Figure 9 contains similar projections for the female population. Under the Alternative 2 scenario, the male and female population ages 0-19 in the year 2100 is respectively 62.6 and 59.8 million people. The corresponding mean projections for the informative BVAR model exceed these values by 1.9 millions for males and 2.1 millions for females. Likewise, the Lee-Carter mean forecasts over-predict the Alternative 2 population age 0-19 by 3.3 millions for males and 3.4 millions for females. By contrast, the diffuse BVAR model predicts and additional 17.3 million males and 16.8 million females over the Trustees assumptions.

Turning to the population age 20-64, the Alternative 2 scenario in the year 2100 involves 138.1 million males and 135.7 million females. The informative BVAR projections add an extra 260,000 and 1.5 million people, respectively. The Lee-Carter model predicts an additional 4.2 million males and 4.7 million women, while the counterpart figures for the diffuse vector autoregression are an extra 19.8 and 20.2 million people. Of course, the larger mean forecasts associated with the diffuse model are not surprising, given the higher projected average level of overall fertility. In terms of implied uncertainty, for age groups 0-19 and 20-64, the diffuse and informative BVAR models respectively generate the widest and narrowest intervals.

One striking feature of the graphs corresponding to the projected population age 65-100+ is how narrow the Lee-Carter interval forecasts are over most of the predicted horizon. In particular, for the female population, the Trustees assumption actually falls outside the Lee-Carter 90% confidence interval from the very first forecast period in 2009. It is not until 2088 that the Alternative 2 values begin to cross the lower bound of the intervals. This occurrence is in turn a direct result of the model's very narrow mortality interval projections. Consider that for individuals 65 and older in 2009, new births do not add members to their population until the year 2074. Hence, until that time, their projected numbers are exclusively determined by net migration and by survivors from the previous years. By contrast, while the diffuse model predicts mean female mortality improvement similar to that of the Lee-Carter approach, it nonetheless captures the Alternative 2 scenario as an outcome within a 90% posterior probability range.

Figure 10 displays total population projections (both male and female combined), along with aged and total dependency ratios. The Trustees assumption for 2100 is 514.8 million people. The mean forecasts corresponding to the informative BVAR and Lee-Carter models are respectively 520.2 and 539.1 millions. Meanwhile, the total population predicted by the diffuse model is 601.5 million people. All three models include the Alternative 2 assumption well within their uncertainty bands. A detailed break down of the mean projections by age group and gender for the years 2050 and 2100 can be glimpsed from the population pyramids in Figure 11. The graphs provide a stunning illustration of the impact that seemingly small increases in average fertility over the long-term can have on population levels at all ages.

Two widely used measures of population aging for the purpose of evaluating future solvency in public pension programs are the aged and total dependency ratios. The former is defined as the population aged 65 and over divided by the population aged 20-64 (roughly corresponding to the number of retirees supported by each worker). Its value for 2008 is 0.21, while the Trustees assumption in 2100 more than doubles it to 0.433. The mean projections generated by the informative BVAR and Lee-Carter models are 0.438 and 0.466, respectively. Surprisingly, the mean forecast corresponding to the diffuse BVAR process is 0.436, suggesting that in spite of the overall higher predicted population counts for these ages, the relative age structure for males and females combined is quite close to the Trustees assumptions.

Finally, an even broader measure of the financial burden placed on people of working age is the total dependency ratio. It involves the sum of the population 65 and over and under age 20, divided by the population aged 20-64. This ratio grows from a historical value of 0.67 in 2008 to 0.88 by the year 2100 under the Trustees assumptions. The mean prediction for the informative BVAR model is 0.891. The Lee-Carter and diffuse models yield 0.914 and 0.916, respectively.

Figure 8: Male population forecasts for three broad age groups (millions).

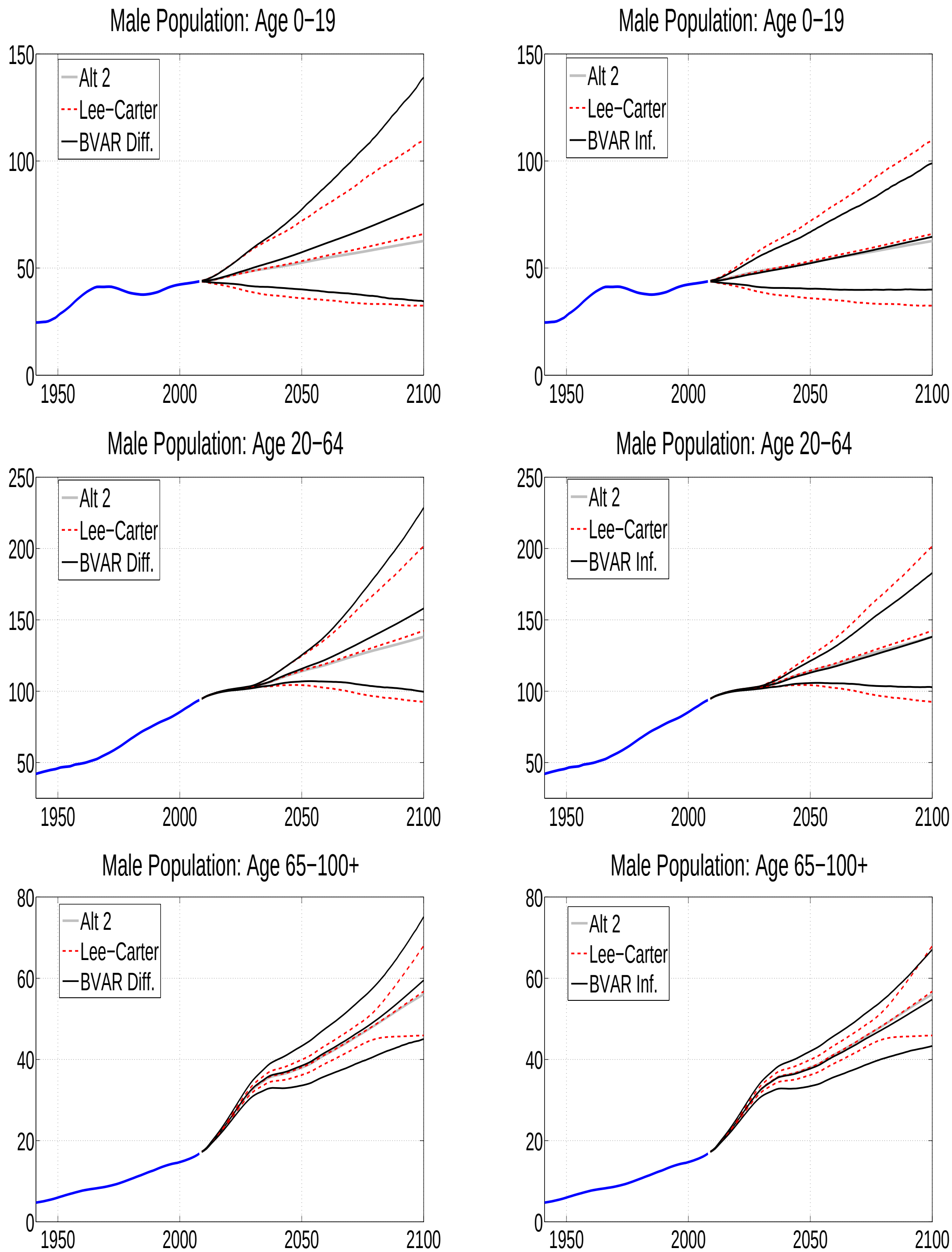


Figure 9: Female population forecasts for three broad age groups (millions).

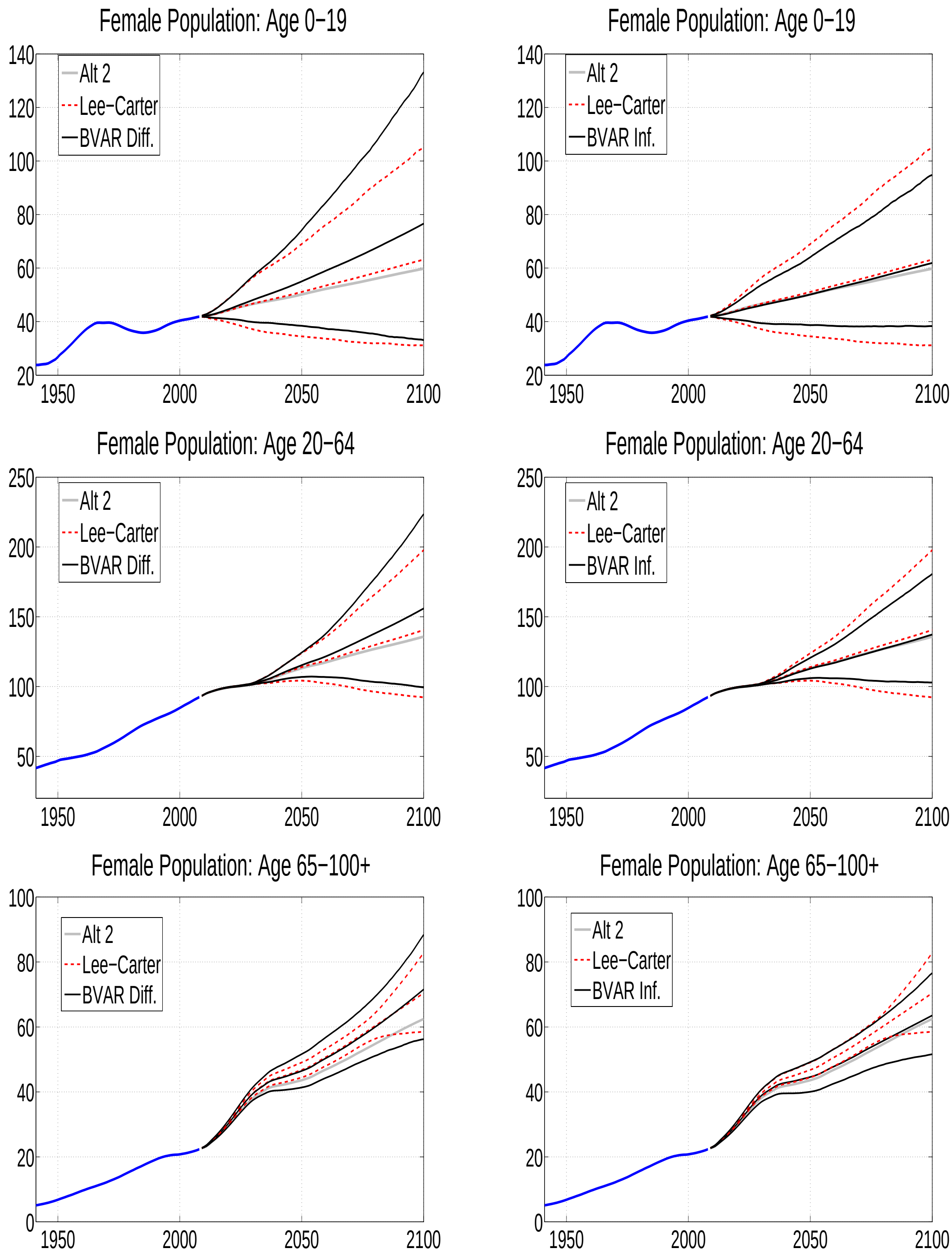


Figure 10: Forecasts of total population and dependency ratios.

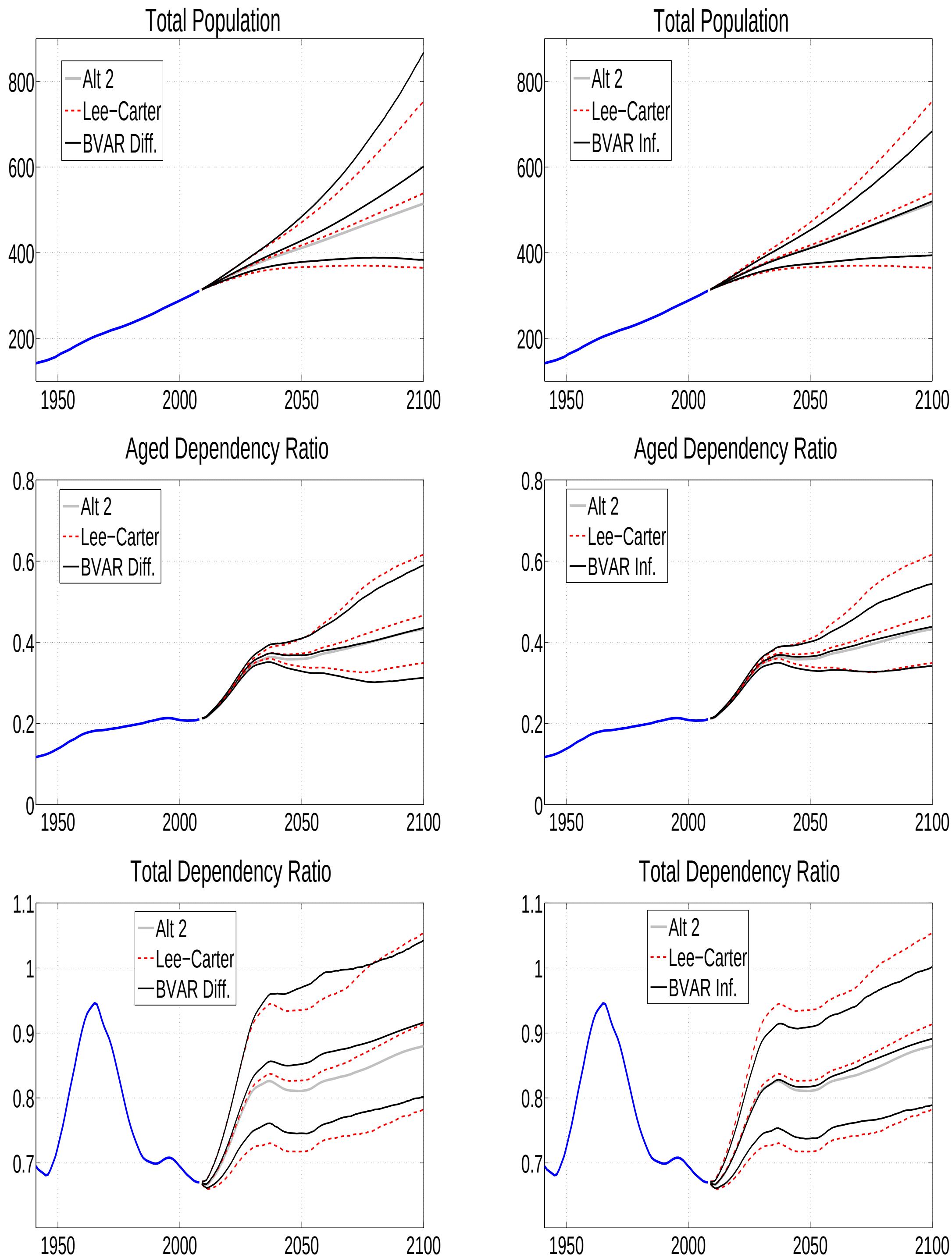
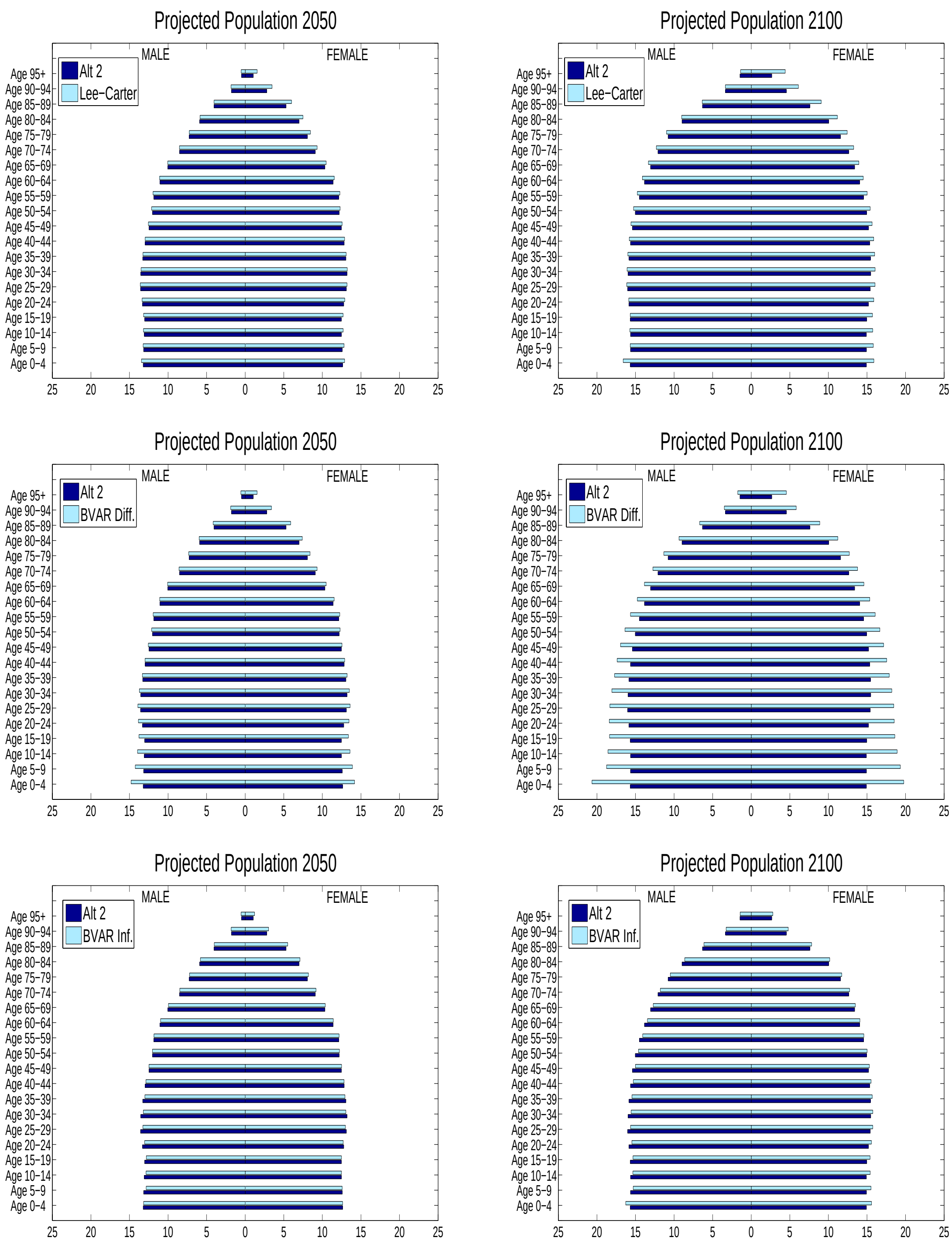


Figure 11: Population pyramids (mean forecasts in millions).



Appendix A

A.1. Stationarity in Vector Autoregressions

Following Hamilton (1994, chapter 10), the VAR(p) model in equation (10) can be rewritten as the first order system

$$\boldsymbol{y}_t^* = \boldsymbol{\Phi}^* + \boldsymbol{\Pi}^* \boldsymbol{y}_{t-1}^* + \boldsymbol{\varepsilon}_t^*, \tag{23}$$

where

$$\boldsymbol{y}_t^* = \begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{y}_{t-1} \\ \vdots \\ \boldsymbol{y}_{t-p+1} \end{bmatrix}, \; \boldsymbol{\Phi}^* = \begin{bmatrix} \boldsymbol{\Phi} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \; \boldsymbol{\varepsilon}_t^* = \begin{bmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \tag{24}$$

and

$$\boldsymbol{\Pi}^* = \begin{bmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{\Pi}_2 & \boldsymbol{\Pi}_3 & \cdots & \boldsymbol{\Pi}_{p-1} & \boldsymbol{\Pi}_p \\ \boldsymbol{I}_m & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_m & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{I}_m & \mathbf{0} \end{bmatrix}. \tag{25}$$

It can be shown that the process is covariance stationary if the eigenvalues of the matrix $\boldsymbol{\Pi}^*$ lie inside the unit circle

$$\left| \boldsymbol{I}_m \eta^p - \boldsymbol{\Pi}_1 \eta^{p-1} - \boldsymbol{\Pi}_2 \eta^{p-2} - \cdots - \boldsymbol{\Pi}_p \right| = 0. \tag{26}$$

Alternatively, covariance stationary requires the modulus of the largest eigenvalue of $\boldsymbol{\Pi}^*$ to be less than one ($|\eta_{\max}| < 1$). When this is the case, the model's long-horizon forecasts converge to the unconditional mean or steady-sate of the process

$$\boldsymbol{\mu} = (\boldsymbol{I}_m - \boldsymbol{\Pi}_1 - \boldsymbol{\Pi}_2 - \cdots - \boldsymbol{\Pi}_p)^{-1} \boldsymbol{\Phi}. \tag{27}$$

A.2. Conditional Posterior Densities in the BVAR Model

Let the symbol $\lfloor \rfloor$ denote the usual stacking of data vectors into a matrix comprising the entire sample. Given the prior densities in (20), Villani (2006) shows that the full conditional posteriors for the mean-adjusted BVAR model are given by

- Full conditional posterior of $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma} | \boldsymbol{\Pi}, \boldsymbol{\Psi}, \boldsymbol{Y} \sim IW(\boldsymbol{E}'\boldsymbol{E}, T), \tag{28}$$

$\boldsymbol{E} = \lfloor \boldsymbol{\Pi}(L)(\boldsymbol{y}_t - \boldsymbol{\Psi} \boldsymbol{d}_t) \rfloor$.

- Full conditional posterior of $\boldsymbol{\Pi}$

$$\operatorname{vec}(\boldsymbol{\Pi})|\boldsymbol{\Sigma},\boldsymbol{\Psi},\boldsymbol{Y} \sim N(\overline{\boldsymbol{\theta}}_{\Pi},\overline{\boldsymbol{\Omega}}_{\Pi}), \tag{29}$$

$\overline{\boldsymbol{\Omega}}_{\Pi}^{-1} = \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}_{\Psi}'\boldsymbol{X}_{\Psi} + \boldsymbol{\Omega}_{\Pi}^{-1}$, $\overline{\boldsymbol{\theta}}_{\Pi} = \overline{\boldsymbol{\Omega}}_{\Pi}\left[\operatorname{vec}(\boldsymbol{X}_{\Psi}'\boldsymbol{Z}_{\Psi}\boldsymbol{\Sigma}^{-1}) + \boldsymbol{\Omega}_{\Pi}^{-1}\boldsymbol{\theta}_{\Pi}\right]$, $\boldsymbol{Z}_{\Psi} = \lfloor \boldsymbol{y}_t - \boldsymbol{\Psi}\boldsymbol{d}_t \rfloor$ and $\boldsymbol{X}_{\Psi} = \lfloor \boldsymbol{y}_{t-1} - \boldsymbol{\Psi}\boldsymbol{d}_{t-1}, ..., \boldsymbol{y}_{t-p} - \boldsymbol{\Psi}\boldsymbol{d}_{t-p} \rfloor$.

- Full conditional posterior of $\boldsymbol{\Psi}$

$$\operatorname{vec}(\boldsymbol{\Psi})|\boldsymbol{\Sigma},\boldsymbol{\Pi},\boldsymbol{Y} \sim N(\overline{\boldsymbol{\theta}}_{\Psi},\overline{\boldsymbol{\Omega}}_{\Psi}), \tag{30}$$

$\boldsymbol{\Omega}_{\Psi}^{-1} = \boldsymbol{U}'\left(\boldsymbol{D}'\boldsymbol{D} \otimes \boldsymbol{\Sigma}^{-1}\right)\boldsymbol{U} + \boldsymbol{\Omega}_{\Psi}^{-1}, \overline{\boldsymbol{\theta}}_{\Psi} = \overline{\boldsymbol{\Omega}}_{\Psi}\left[\boldsymbol{U}'\operatorname{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\boldsymbol{D}) + \boldsymbol{\Omega}_{\Psi}^{-1}\boldsymbol{\theta}_{\Psi}\right], \boldsymbol{Z} = \lfloor \boldsymbol{\Pi}(L)\boldsymbol{y}_t \rfloor$, $\boldsymbol{U}' = \left(\boldsymbol{I}_{mq}, \boldsymbol{I}_q \otimes \boldsymbol{\Pi}_1', ..., \boldsymbol{I}_q \otimes \boldsymbol{\Pi}_p'\right)$ and $\boldsymbol{D} = \lfloor \boldsymbol{d}_t, -\boldsymbol{d}_{t-1}, ..., -\boldsymbol{d}_{t-p} \rfloor$.

References

Board of Trustees (2009). *Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds.*

Brooks, S. P. (1998). "Markov Chain Monte Carlo Method and Its Application." *The Statistician* **47**(1), 69-100.

Chen, M. H., Q. M. Shao and G. Ibrahim (2000). *Monte Carlo Methods in Bayesian Computation.* Springer-Verlag: New York.

Cowles, M. K. and B. P. Carlin (1996). "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review." *Journal of the American Statistical Association* **91**, 883-904.

Doan, T., R. Litterman and C. Sims (1984). "Forecasting and Conditional Projection Using Realistic Prior Distributions." *Econometric Reviews* **3**, 1-144.

Geweke, J. and C. Whiteman (2006). "Bayesian Forecasting." In *The Handbook of Economic Forecasting.* G. E. Clive, W. J. Granger and A. Timmerman, eds. North-Holland: Amsterdam.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press: New Jersey.

Hollman, F. W., T. J. Mulder and J. E. Kallan (2000). "Methodology and Assumptions for the Population Projections of the United States: 1999 to 2100." *Population Division Working Paper No. 38.* U.S. Census Bureau.

Kadiyala, K. R. and S. Karlsson (1993). "Forecasting with Generalized Bayesian Vector Autoregressions." *Journal of Forecasting* **12**, 365-378.

Kadiyala, K. R. and S. Karlsson (1997). "Numerical Methods for Estimation and Inference in Bayesian VAR-Models." *Journal of Applied Econometrics* **12**(2), 99-132.

Lee, R. D. (1993). "Modeling and Forecasting the Time-Series of U.S. Fertility: Age Distribution, Range and Ultimate level." *International Journal of Forecasting* **9**, 187-202.

Lee, R. D. (2000). "The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications." *North American Actuarial Journal* **4**(1), 80-91.

Lee, R. D. and L. R. Carter (1992). "Modeling and Forecasting U.S. Mortality." *Journal of the American Statistical Association* **87**, 659-671.

Lee, R. D. and T. Miller (2001). "Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality." *Demography* **38**(4), 537-549.

Lee, R. D. and S. D. Tuljapurkar (1994). "Stochastic Population Forecasts for the United States: Beyond High, Medium, and Low." *Journal of the American Statistical Association* **89**, 1175-1189.

Litterman, R. (1986). "Forecasting with Bayesian Vector Autoregressive Processes - Five Years of Experience." *Journal of Business and Economic Statistics* **4** (1), 25-38.

Sims, C. A. (1980). “Macroeconomics and Reality.” *Econometrica* **48**(1), 1-47.

Sims, C. A. and Tao A. Zha (1998). “Bayesian Methods for Dynamic Multivariate Models.” *International Economic Review* **39**(4), 949-968.

Villani, M. (2006). “Inference in Vector Autoregressive Models with an Informative Prior on the Steady State.” *Riksbank Research Paper* No. 19.