Modeling Mortality with a Bayesian Vector Autoregression

Carolyn Ndigwako Njenga & Michael Sherris

Modeling Mortality with a Bayesian Vector Autoregression

Carolyn Ndigwako Njenga
Australian School of Business
University of New South Wales, Sydney, NSW, 2052
Australia
Email: ndigwako@hotmail.com

Michael Sherris
Australian School of Business
University of New South Wales, Sydney, NSW, 2052
Email: m.sherris@unsw.edu.au

4th March 2011

Abstract

Mortality risk models have been developed to capture trends and common factors driving mortality improvement. Multiple factor models take many forms and are often developed and fitted to older ages. In order to capture trends from young ages it is necessary to take into account the richer age structure of mortality improvement from young ages to middle and then into older ages. The Heligman and Pollard (1980) model is a parametric model which captures the main features of period mortality tables and has parameters that are interpreted according to age range and effect on rates. Although time series techniques have been applied to model parameters in various parametric mortality models, there has been limited analysis of parameter risk using Bayesian techniques. This paper uses a Bayesian Vector Autoregressive (BVAR) model for the parameters of the Heligman-Pollard model and fits the model to Australian data. As VAR models allow for dependence between the parameters of the Heligman-Pollard model they are flexible and better reflect trends in the data, giving better forecasts of the parameters. Forecasts can readily incorporate parameter uncertainty using the models. Bayesian Vector Autoregressive (BVAR) models are shown to significantly improve the forecast accuracy of VAR models for mortality rates based on Australian data. The Bayesian model allows for parameter uncertainty, shown to be a significant component of total risk.

Keywords: mortality, parameter risk, vector auto-regression, Bayesian, Heligman-Pollard model

JEL Classifications: J11, C11, G22

1   Introduction

Mortality rates around the world have been improving significantly over the 20th century. In Australia significant improvements have occurred particular since the 1970's. Longevity risk has been the focus of much recent research because of the need for financial products to manage the risk that individuals will outlive their financial resources. This is important because Australian retirees are increasingly accountable for managing their own longevity risk (Henry, 2009). The baby boomers born between 1946 and 1964 have begun to enter retirement.

Institutions and social schemes such as Social Security in USA that offer living benefits may face fiscal stress regardless of implementing policy changes such as increasing the eligibility age for normal retirement benefits from 65 to 67 by 2025 (Lee and Skinner (1999)). In order for product providers of longevity insurance to meet their guarantees of payments to individuals it is necessary to assess the risk that they take on and the resulting capital requirements. This risk should not only take into account trend risk but also parameter risk in the models used. The classical Lee-Carter model has many parameters and only one common factor for mortality improvements.

In contrast, parametric mortality models smooth mortality rates and trends, reducing the number of parameters and parameter risk. They can be used to compare mortality in different countries and at different times (Congdon, 1993). The Heligman and Pollard (1980) model is a parametric model that fits the whole age range of mortality rates. Parameters provide information about mortality for differing age ranges. Parameters are modeled as a time series to capture trends and risks in mortality rates. Forfar and Smith (1987), Rogers (1986), McNown and Rogers (1989) and Thompson, Bell, Long, and Miller (1989) use time series techniques to model the time evolution of mortality model parameters. Dependence between lifetimes at a given point in time is influenced by common environmental factors and this creates parameter risk (Denuit & Frostig, 2009).

Similar modeling techniques have been applied to multivariate economic data. The modeling techniques have direct application to mortality data and parametric mortality models. Reichmuth and Sarferaz (2008) and Pedroza (2006) apply Bayesian techniques to non-parametric mortality models. Dellaportas, Smith, and Stavropoulos (2001) and Sharrow, Clark, Collinson, Kahn, and Tollman (2010) apply Bayesian techniques to the fitting of the parameters of the Heligman and Pollard model. They do not apply this approach to the time evolution of the parameters. This is required to quantify parameter risk in a dynamic mortality model. Denuit and Frostig (2007, 2008, 2009) quantify parameter risk for the classical Lee and Carter (1992) model which allows for dependence in lifetimes due to exposure to common factors.

Parametric mortality models for mortality rates capture the trends and volatility of

large body of data with a small number of parameters. The parameters are easy to interpret and easy to manipulate for analysis (Rogers, 1986; Congdon, 1993). Parametric models allow comparisons of mortality rates over time and by region. Non-parametric models such as Lee and Carter (1992) have been popular. Econometric models provide a more general framework for modeling mortality trends and volatility.

In this paper, the parameters in the Heligman-Pollard model are modeled as a multivariate time series system in an age-period mortality model. The correlation between the Heligman and Pollard model parameters is captured in the econometric methods allowing interaction between changes in mortality for differing ages. A vector autoregression (VAR) model is used for the relationship between the past (lagged) values and current values of the Heligman-Pollard parameters. The classical unrestricted VAR model does not account for uncertainty in its coefficients. A Bayesian VAR model accounts for the uncertainty in the VAR coefficients allowing quantification of parameter risk. These models will be referred to as the HP-VAR and HP-BVAR models respectively.

Models, such as the Lee-Carter model, are used to model old age mortality and perform best when estimated using data from middle ages to older ages. Models for older ages often ignore trends at younger ages by excluding the mortality rates of those age 64 and under. Improvement in the age structure of mortality that is due to improvement in mortality at younger ages contains information useful for older age mortality because of the systematic improvements that impact all ages to a greater or lesser extent.

This paper applies Bayesian techniques to capture parameter risk in HP-VAR and HP-BVAR mortality models. Australian mortality is analysed using population data from the Human Mortality Database to model longevity at the population level. The paper extends the methodology of McNown and Rogers (1989), that combines parametric mortality models with time series models, by modeling mortality with a HP-BVAR. The results demonstrate that the Bayesian Vector Autoregressive model provides improved forecasts of the model parameters with a 50 (1946-1995) year look-back horizon used to forecast parameters for 12 years (1996-2007). Parameter risk is quantified and shown to be a significant component of total mortality uncertainty. Although not the focus of the paper, the results have important implications for the determination of capital requirements for product providers of life annuities and pensions.

2   Data

The data set used is Australian population data obtained from the Human Mortality Database (HMD, *Human Mortality Database*, 2010). The HMD data for ages above 80 are not the actual death rates (Wilmoth, Andreev, Jdanov, & Glei, 2007). At the older ages

the volatility of mortality rates increases. The maximum age used for modeling is 89 in order to increase the reliability of the model estimation. For Australian deaths before 1964 the Human Mortality Database data was only provided in 5 year age groups and was split into annual data using cubic splines and smoothed as described in McNeil, Trussell, and Turner (1977). Until 1971 the population data was not adjusted for net undercounts and the data from 1971 onwards is of better quality and complete.

Mortality in Australia, as for many developed economies, has been declining steadily over the recent past as shown in Figure 2, prepared using the method in Peng (2008). The vertical axis on the main plot represents ages with age zero (0) at the bottom increasing to age one hundred and ten and over (110+) at the top. The data values are divided into three categories and represented as low values (Yellow or Very Light Grey), medium values (Orange or moderately grey) and high values (Red or Darker Grey). Missing data for the very old ages is represented by white areas at the top of the main plot.
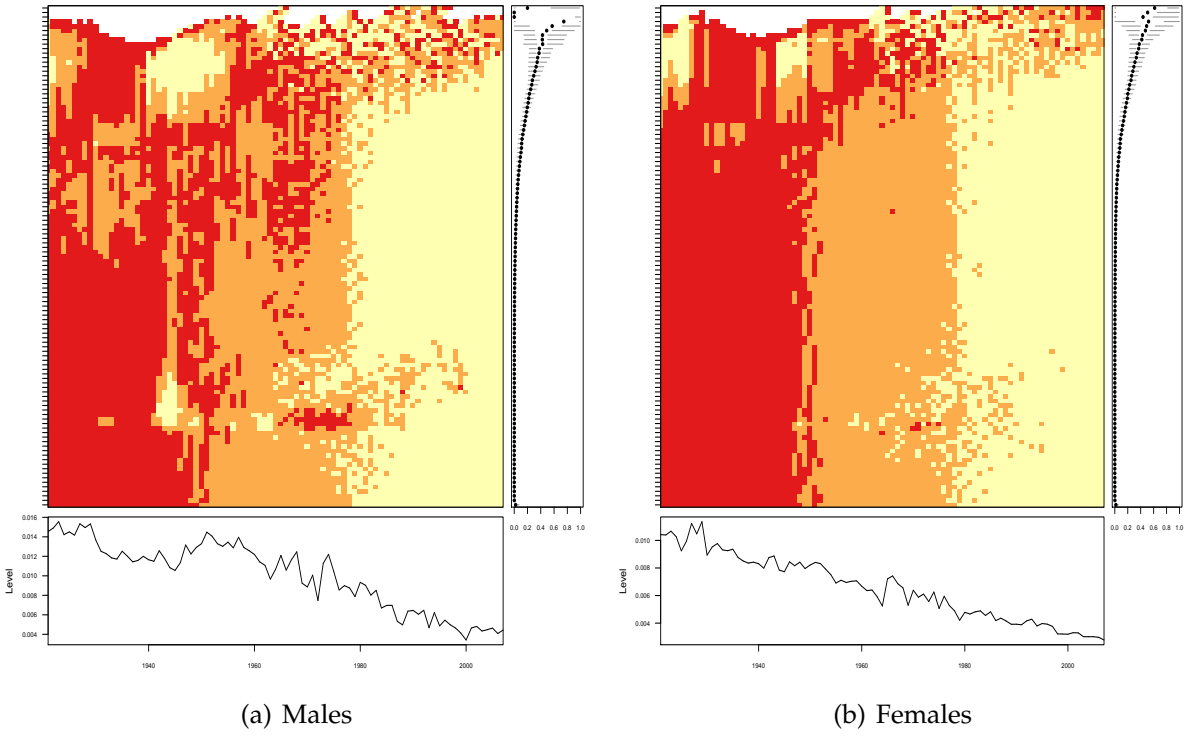

(a) Males   (b) Females

Figure 1: Declining Mortality Rates for Australia from 1921 to 2007 for ages 0 to 110+ from bottom to top. Red=High mortality, Orange=Moderate mortality, Yellow=Low mortality, White=Missing Data. The panel on the right of the main plots has boxplots of the data in each time series. In the bottom panel are median values across all the time series of mortality rates for each time point.

There are more yellow areas (light shade) on the right hand side of the plot and more red areas (dark shade) on the left hand side of the plot. Mortality rates have declined significantly. From the box plots in the panel on the right side of the main
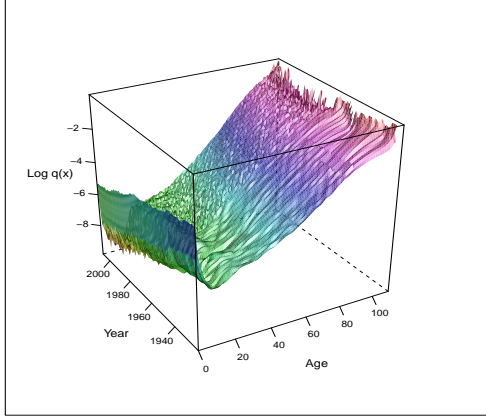
Figure 2: Log q_x for Australian Males
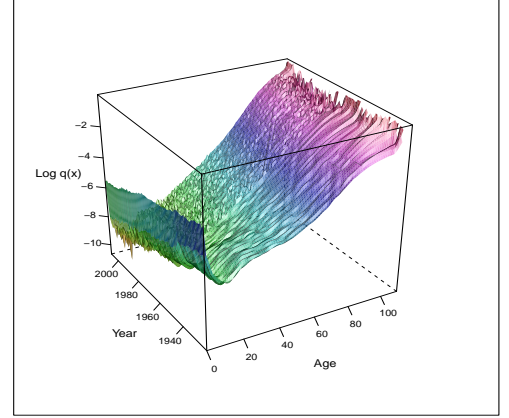

Figure 3: Log q_x for Australian Females

plot, the variation in mortality rates is highest for the older ages, the dots are increasing and the whiskers (lines on either side of the dot) are longer for the older ages. The plot of levels on the bottom is the median values for all the mortality rate time series and quantifies the strong downward trend in mortality rates. There has been a decline in the rate of decrease in mortality rates in more recent years, from the late 1990s. An explanation for this is that mortality trend improvements are subject to a law of "diminishing returns" (Wong-Fupuy and Haberman (2004)). Capturing these changes in trends is an important feature of a mortality model.

Mortality curves reflect four main features (Heligman and Pollard (1980); Rogers (1986)). Initially a high death rate in the first year of life that declines in childhood then rises in late teenage years to form a hump. This hump is more pronounced in males due to accidents and less pronounced and later in females due to maternal causes. At older ages the death rates increases at a more or less constant rate.

Figures 2 and 3 plot the Australian mortality data across time and age.

Figure 4 shows the typical pattern across age for mortality rates using the Heligman-Pollard model.

Mortality rates have maintained this shape over long time periods. Parametric models can capture these mortality rate curves and allow efficient dynamic models to be fitted to the data (Congdon 1993).

3 Notation

The probability of death between times t and $t + 1$ for individuals aged x is denoted $q_{x,t}^{true}$ at time t. $E(x, t)$ denotes the number of observed exposed-to-risk between ages x and $x + t$ at time t and $D(x, t)$ the number of observed deaths in those exposed aged between x and $x + t$ at time t. $q_{x,t}^{true}$ is estimated by the crude mortality rate in year
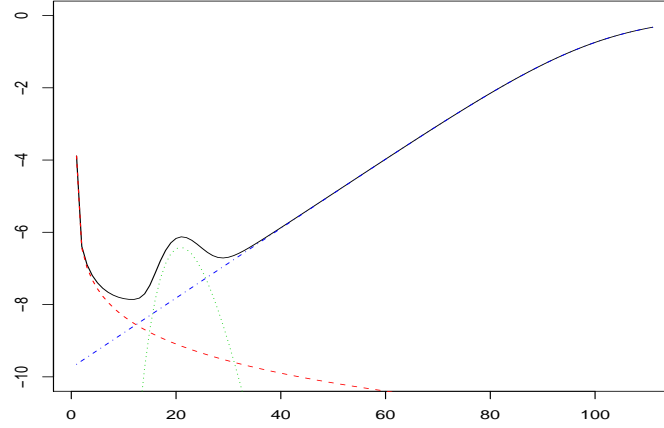
Figure 4: Heligman and Pollard Model. Red broken line = childhood mortality. Green dotted line = adult mortality. Blue dot-dashed line = old age mortality.

t, $q_{x,t}$, using the observed deaths and observed exposures. The crude death rates are denoted and defined as $m(x,t) = \frac{D(x,t)}{E(x,t)}$. μ_x denotes the instantaneous rate (force) of mortality. Assuming a constant force of mortality for integer age, x, and time, t, and a stationary population, the force of mortality, $\mu_{x,t}$, is approximated by the age specific central death rates, $m(x,t)$. For a given set of parameters, θ_t:

$$m(x,t) \simeq \mu_x(\theta_t) = \mu_{x+\xi,t+\tau}(\theta_t) \qquad 0 \leq \xi, \tau < 1$$

Crude death rates, $m(x,t)$, are related to the probability of death by $q_{x,t} = 1 - \exp(-m(x,t))$.

The number of deaths is large so its distribution can be estimated by a Normal distribution with expectation and variance that are both equal to $\mathbb{E}[m(x,t)]E(x,t)$, assuming a Poisson distribution for number of deaths.

Denoting the estimated values of $m_{x,t}$ by $\hat{m}_{x,t}$. The mortality model residuals are $e_{x,t} = m_{x,t} - \hat{m}_{x,t}$. The residuals $e_{x,t}$ are also approximately Normal and the standardized mortality rates are:

$$z_{x,t} = \frac{D_{x,t} - E_{x,t}\hat{m}_{x,t}}{\sqrt{E_{x,t}\hat{m}_{x,t}}} \simeq \frac{m(x,t) - \mathbb{E}[m(x,t)]}{\sqrt{\mathbb{E}[m(x,t)]/E(x,t)}} \tag{1}$$

These standardized residuals are approximately independent and identically distributed $N(0,1)$ variables. Positive and negative residuals should be randomly distributed for the underlying model to be valid (Cairns, Blake, Dowd, Coughlan, & Epstein, 2007).

3.1 Heligman and Pollard Model

Denote by $\theta_t' = (A_t, B_t, C_t, D_t, E_t, F_t, G_t, H_t)$ model parameters at time t. The probability of death $q_{x,t}$ at time point t is assumed given by the Heligman-Pollard model:

$$q_x(\theta_t) = q_{x,t} = A_t^{(x+B_t)^{C_t}} + D_t \exp[-E_t(\log\{\frac{x}{F_t}\})^2] + \frac{G_t H_t^x}{1 + G_t H_t^x} \tag{2}$$

The parameters A, B, C, D, G are all in $[0,1]$, E, F in $[0,\infty)$ and H in $[0,15]$.

Each of the parameters has a demographic interpretation. The first component, $A^{(x+B)^C}$, is a rapidly declining exponential to reflect the fall in mortality during the early childhood years. The middle term in the model reflects accident mortality for males and accident and maternal mortality for females. The third term in the model, GH^x is a Gompertz exponential that reflects the rise in mortality in adults due to the aging of the body.

Parameters are estimated for q_x for $x=0,1,\ldots,85$ for the Heligman-Pollard model for the years 1946-2007 by minimizing the weighted sum of squared errors between observed, q_x and the fitted, $\hat{q}_x$:

$$S^2 = \sum_{x=0}^{85} \frac{1}{q_x^2}(\hat{q}_x - q_x)^2 \tag{3}$$

The weights in Equation 3 are used in Heligman and Pollard (1980) and are based on the assumption that the coefficient of variation is constant across age. The variance of $q_{x,t}$, denoted by $\sigma_{x,t}^2$, is assumed directly proportional to $q_{x,t}^2$. Dellaportas et al. (2001) argue that Equation 2 is too restrictive as it assumes a deterministic relationship between $q_{x,t}$ and θ_t. Consequently not all the variation in $q_{x,t}$ is explained by θ_t. This requires assumptions about the distribution of $q_{x,t}$. Dellaportas et al. (2001) consider different probability distributions and apply Bayesian techniques to estimate θ_t.

Denote by $\Theta = (\theta_{t_0}, \ldots, \theta_{t_T})$ the matrix of the estimated parameters at times $t_0, \ldots, t_T$:

$$\Theta = \begin{pmatrix} A_{t_0} & A_{t_1} & \ldots & A_{t_T} \\ B_{t_0} & B_{t_1} & \ldots & B_{t_T} \\ & \vdots & \ddots & \\ H_{t_0} & H_{t_1} & \ldots & H_{t_T} \end{pmatrix} \tag{4}$$

For old age mortality, the third term of Equation 2 is sufficient for estimating $q_{x,t}$. In particular:

$$\frac{q_{x,t}}{p_{x,t}} = G_t H_t^x \tag{5}$$

Assume that $p_{x,t} \to 1$. The trends in the relative rate of increase in mortality represent the rate at which G is declining (Heligman & Pollard, 1980). As shown in Wong-Fupuy

and Haberman (2004) the relative rate of change of mortality increase is:

$$\frac{q'_{x,t}}{q_{x,t}} = \ln(H)$$

A small $\ln(H)$ implies that there has been a small decline in G while a larger $\ln(H)$ implies that there has been a larger decline in G.

4   Econometric Models

Denote by y_t the univariate time series that are in θ_t. Time series are classified as either stationary or non-stationary (integrated). When time series show non-stationarity in the mean or have a trending behavior the process is integrated. A stationary process is modeled using an equation with fixed coefficients estimated from past data.

Australian Mortality rates by age exhibit non-stationary behavior at many ages. Sherris and Njenga (2009) provide a more detailed analysis. When mortality rates are estimated using a parametric relationship they are affected by shocks or unexpected changes in the parameters of the model. If y_t is a stationary time series then the shocks gradually die away and their effect reduces with time. If y_t is a non-stationary time series then the effect of the shock will persist infinitely.

For an AR(1) time series y_t:

$$y_t = \kappa + \delta t + \phi u_{t-1} + \epsilon_t \qquad \epsilon_t \sim WN(0, \sigma^2) \tag{6}$$

$\kappa + \delta t$ is the deterministic linear trend and $\phi u_{t-1} + \epsilon_t$ is AR(1). When $|\phi| < 1$ the process y_t is stationary, denoted I(0), about the deterministic linear trend (or trend stationary). When $\phi = 1$, the AR(1) can be written as $u_t = u_{t-1} + \epsilon_t = u_0 + \sum_{i=1}^{t} \epsilon_i$, which is a stochastic trend. In this case, y_t is I(1) with drift. Stationary (I(0)) processes are short memory processes. After long lags observations at different times are independent. Integrated (I(1)) processes are long memory processes. After long lags observations are not independent so differencing data leads to a loss of information.

The Augmented Dickey Fuller (ADF) Test is used to test for integration. The null hypothesis is that $\phi = 1(H_0 : \phi = 1 \Rightarrow y_t \sim I(1)$) or that the time series is difference stationary against the alternative that it is trend stationary ($H_a : |\phi| < 1 \Rightarrow y_t \sim I(0)$).

An autoregressive-moving average model, ARMA(p,q), is defined by a pth order stochastic difference equation:

$$y_t = \delta + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \tag{7}$$

where $\epsilon_t \, WN(0, \sigma^2)$

ARMA(p,q) models are based on the assumption that the process y_t is stationary. If the data is trending then Box and Jenkins (1976) propose differencing the data d times to eliminate the trend and hence suitable for modeling using ARMA(p,q). Such a model is an ARIMA(p,d,q) (Hamilton, 1994).

In McNown and Rogers (1989) a parameterized time series is modeled as described in Thompson et al. (1989) using the techniques of Box and Jenkins (1976). A unit root in the time series of the parameters required differencing the time series to achieve stationarity (McNown & Rogers, 1989). The parameters for an ARIMA models were then extrapolated to obtain a series of Heligman-Pollard curves with time varying parameters. Because of the assumption of independence of the Heligman-Pollard parameters in the process given in Rogers (1986) the forecasts are not accurate and will be inconsistent (Lee, 1992). Lee (1992) also notes the absence of confidence intervals which is attributed to problems that arise from the independence assumption.

Sims (1980) developed a Vector Autoregression model with p lags, VAR(p) for expressing a set of variables as a weighted linear combination of each variable's past values and the past values of the other variables in the set. The VAR(p) models are more flexible than AR models and have a rich structure that captures more features of the mortality system.

θ_t, $t = 1, 2, \ldots, T$ is a column vector of n variables. For the Heligman-Pollard parameters θ_t is $\theta_t = (A_t, B_t, \ldots, H_t)'$ with $n = 8$. The unrestricted VAR(p) model is:

$$\theta_t = c + \sum_{l=1}^{p} \Omega_l \theta_{t-l} + \epsilon_t \tag{8}$$

$c = (c_1, \ldots, c_n)'$ is a $n \times 1$ vector of unknown constants. Ω_l is an unknown $n \times n$ matrix of coefficients of θ_{t-l} at lag l. $\epsilon_t, t = 1, \ldots, T$ are independent identically distributed errors distributed as $N_n(0, \Sigma)$ as $\mathbb{E}(\epsilon_t) = 0$, $\mathbb{E}(\epsilon_t \epsilon_t') = \Sigma$ and $\mathbb{E}(\epsilon_t \epsilon_{t-l}) = 0$ for $l \neq 0$. ϵ_t measures the degree to which the contemporaneous vector θ_t is determined by the VAR(p). ϵ_t is a variable that is influenced by the number of lags, p in the VAR model and the choice of coefficients, $v = (c, \Omega_1, \ldots, \Omega_p)$, that give weights in the linear combination that forms the VAR(p).

To illustrate, for $n = 3$ and $p = 2$, the VAR(2) can be written:

$$\theta_t = c + \Omega_1 \theta_{t-1} + \Omega_2 \theta_{t-2} + \epsilon_t \tag{9}$$

Equation 9 is written in matrix form as:

$$
\begin{bmatrix} \theta_{1t} \\ \theta_{2t} \\ \theta_{3t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} \omega_{11}^1 & \omega_{12}^1 & \omega_{13}^1 \\ \omega_{21}^1 & \omega_{22}^1 & \omega_{23}^1 \\ \omega_{31}^1 & \omega_{32}^1 & \omega_{33}^1 \end{bmatrix} \begin{bmatrix} \theta_{1t-1} \\ \theta_{2t-1} \\ \theta_{3t-1} \end{bmatrix} + \begin{bmatrix} \omega_{11}^2 & \omega_{12}^2 & \omega_{13}^2 \\ \omega_{21}^2 & \omega_{22}^2 & \omega_{23}^2 \\ \omega_{31}^2 & \omega_{32}^2 & \omega_{33}^2 \end{bmatrix} \begin{bmatrix} \theta_{1t-2} \\ \theta_{2t-2} \\ \theta_{3t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{bmatrix} \tag{10}
$$

Equation 10 can be written equation-by-equation as:

$$
\begin{aligned}
\theta_{1t} &= c_1 + \omega_{11}^1\theta_{1t-1} + \omega_{12}^1\theta_{2t-1} + \omega_{13}^1\theta_{3t-1} + \omega_{11}^2\theta_{1t-2} + \omega_{12}^2\theta_{2t-2} + \omega_{13}^2\theta_{3t-2} + \epsilon_{1t} \\
\theta_{2t} &= c_2 + \omega_{21}^1\theta_{1t-1} + \omega_{22}^1\theta_{2t-1} + \omega_{23}^1\theta_{3t-1} + \omega_{21}^2\theta_{1t-2} + \omega_{22}^2\theta_{2t-2} + \omega_{23}^2\theta_{3t-2} + \epsilon_{2t} \\
\theta_{3t} &= c_3 + \omega_{31}^1\theta_{1t-1} + \omega_{32}^1\theta_{2t-1} + \omega_{33}^1\theta_{3t-1} + \omega_{31}^2\theta_{1t-2} + \omega_{32}^2\theta_{2t-2} + \omega_{33}^2\theta_{3t-2} + \epsilon_{3t}
\end{aligned} \tag{11}
$$

The VAR(p) can be written in concise matrix notation which is useful for specifying the prior distribution of a VAR(p). Let $\theta=(\theta_p,\theta_{p+1},\ldots,\theta_T)$, $v=(c,\Omega_1,\ldots,\Omega_p)$, $X=(X_0,\ldots,X_{T-1})$, $X_t=(1,\theta'_t,\ldots,\theta'_{t-p+1})$, $\xi=(\epsilon_p,\ldots,\epsilon_T)$. Then, the VAR(p) is:

$$
\theta = Xv + \xi \tag{12}
$$

The coefficients of the VAR(p) are estimated using the OLS estimator:

$$
v^{OLS} = (X'X)^{-1}X'\theta \quad \text{or} \quad v_i^{OLS} = (X'X)^{-1}X'\theta_i \tag{13}
$$

where θ_i is a $T \times 1$ vector with the t-th element θ_{it}.

The vector autoregressive (VAR) model extends univariate autoregressive models to dynamic multivariate and provides better forecasts than univariate time series models (Zivot & Wang, 2006). VAR models are used to describe and forecast multivariate time series for stationary time series.

For non-stationary time series a Vector Error Correction term is added to form a vector error correction model (VECM) and it is necessary to test for the existence of a stationary linear combination of the non-stationary terms (cointegration).

The unknown relationship that generates the multivariate time series is the evolution of q_x over time. When two variables simultaneously influence each other they are referred to as endogenous variables. Since all the parameters are explained within the mortality system the parameters of the Heligman-Pollard model are all treated as endogenous variables.

From a VAR model it is possible to analyse the impacts that the variables have on each other over time and how the variables respond to other unobservable factors. It is also possible to assess if knowing about the past values of some of the parameters tell us about the future values of other parameters. Key questions that may be asked include whether the parameters that model the mortality schedule are related to each other, if the current and past values of a parameter affect other parameters, how a

change in one variable relates to a change in another variable, if the change in the other variable is immediate or does it occur with a delay and, finally, how shocks to the mortality system affect the variables.

4.1 Model Estimation

The first step in estimating a VAR is selecting the lag length, p. This decision is based on minimizing a selection criterion (Lütkepohl, 1991). The Bayesian Information Criterion (BIC(p)) and Hannan-Quinn Criterion (HQ(p)) penalize VAR(p) models with large (p) and are:

$$BIC(p) = \ln |\tilde{\Sigma}(p)| + \frac{\ln T}{T} p n^2 \qquad HQ(p) = \ln |\tilde{\Sigma}(p)| + \frac{2 \ln \ln T}{T} p n^2 \qquad (14)$$

where

$$\tilde{\Sigma}(p) = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t'$$

If all or some of the time series are I(1) while others are I(0) it is necessary to consider a cointegrated VAR since it is possible for the I(1) variables to be cointegrated. Cointegration relations are not directly apparent from a VAR(p) in levels such as Equation 8. It must be transformed into Vector Error Correction Model (VECM) by taking the first difference of Y_t so that:

$$\Delta \theta_t = c + \Pi \theta_{t-1} + \sum_{k=1}^{p-1} \Gamma_k \Delta \theta_{t-k} + \epsilon_t, \qquad k = 1, \dots, p-1 \qquad (15)$$

where $\Pi = \sum \Omega_i - \mathbf{I}_n$ and $\Gamma_k = -\sum_{i=k+1}^{p} \Omega_i$.

The variables in θ_t are at most I(1), $\Delta \theta_t$ are I(0). $\Pi \theta_t$ is the only term that may contain I(1) variables. Since $\Delta \theta_t$ is I(0) then $\Pi \theta_t$ must also be I(0). Π is the long run impact matrix and its rank determines the number of common stochastic trends. Due to the unit root(s) Π is a singular matrix and thus it is not of full rank. Let rank(Π)=r. If r=0 then $\Pi = 0$ and θ_t is I(1) with no cointegration. The VECM is therefore simply a first differences VAR(p-1) in this case. If $0 < r < n$, θ_t is I(1) with r linearly independent cointegrating vectors and $n - r$ common stochastic trends.

Denote the VECM by $H(r)$. When $r = 0$ there is no cointegration and when $r = n$ the corresponding VAR(p) is made up of stationary variables. Testing for cointegration involves a nested model:

$$H(0) \subset \cdots \subset H(r) \subset \cdots \subset H(n)$$

Rank(Π) determines the number of cointegrating relations in θ_t. Likelihood ratio (LR) statistics for determining r are determined by the estimated eigenvalues of Π,

$\lambda_i^{eigen}, i = 1, \ldots, n.$

Johansen's LR statistic test is formulated to test the nested hypothesis:

$$H_0(r) : r = r_0; \qquad H_a(r) : r > r_0 \tag{16}$$

The likelihood ratio trace statistic, LR_{trace} is:

$$LR_{trace}(r_0) = -T \sum_{i=r_0}^{n} \ln(1 - \lambda_i^{eigen}) \tag{17}$$

4.2   Bayesian VAR

VAR models are often overparameterized since they impose no theoretical restrictions to guide the specification of the model (Litterman, 1986; Zivot & Wang, 2006; Sims & Zha, 1998; Robertson & Tallman, 1999; Brandt & Freeman, 2006; Baltagi, 2002). A VAR requires the estimation of $n + pn^2$ coefficients (parameters). The estimates of the coefficients c and Ω_l using an unrestricted VAR(p) model are considered to be fixed quantities. These estimates of coefficients do not accurately reflect the underlying relationship because some of the estimated coefficients of the VAR model are non-zero purely by chance. Restrictions may be imposed to reduce the number of parameters being estimated.

Bayesian inference assumes the data is fixed but the population parameters are random. It requires knowledge of the distribution of the random parameters. In the case of the Bayesian VAR there is uncertainty regarding the distribution of the coefficient matrices. This is reflected in the prior and resulting posterior distribution of the coefficients. One way of doing this is by giving the non-zero coefficients of recent observations more weight (Robertson & Tallman, 1999). Litterman (1986) develops a Bayesian method that views the coefficients c and Ω_l as random variables rather than viewing them as fixed quantities like in the unrestricted VAR(p).

Litterman (1986) specified the form of the prior distributions by giving them specific mean values and measuring the variation (the "tightness" of the distributions) around these given prior mean values for a set of hyperparameters. This is known as Litterman's Prior or the Minnesota Prior because it was part of Litterman's work at the Federal Reserve Bank of Minneapolis and the University of Minnesota (Robertson & Tallman, 1999). This method was extended by Sims and Zha (1998) to give the prior that shall be used in the HP-BVAR model described in this paper.

What follows is the formulation of the Bayesian VAR starting with the unrestricted VAR(p) from Equation 8 building it into a Bayesian VAR with Litterman's Prior and a Bayesian VAR with Sims-Zha's Prior.

Each Ω_l, the individual elements, ω_{ij}^l, are independent normally distributed ran-

dom variables and each variable follows a random walk with a drift that may be nonzero. The random walk assumption is implemented by giving the following means for the lagged coefficient matrices. At $l = 1$, the mean of the coefficient matrix, Ω_1, is the identity matrix (the prior mean for the coefficient of the $j = i$-th variable in equation i is 1) and at $l \neq 1$, the mean of the coefficient matrix is the zero matrix:

$$
\mathbb{E}[\Omega_l] = \begin{cases}
\begin{bmatrix} \omega_{11}^1 & \omega_{12}^1 & \omega_{13}^1 \\ \omega_{21}^1 & \omega_{22}^1 & \omega_{23}^1 \\ \omega_{31}^1 & \omega_{32}^1 & \omega_{33}^1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, & \text{if } l = 1; \\[2em]
\begin{bmatrix} \omega_{11}^l & \omega_{12}^l & \omega_{13}^l \\ \omega_{21}^l & \omega_{22}^l & \omega_{23}^l \\ \omega_{31}^l & \omega_{32}^l & \omega_{33}^l \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, & \text{if } l \neq 1;
\end{cases} \tag{18}
$$

Denote the variance of Ω_l as Ψ_l such that $\sqrt{\mathbb{V}[\omega_{ij}^l]} = \sqrt{\psi_{lij}}$. The random walk assumption imposed on the VAR(p) is loosened by specifying the standard deviations of the individual elements, ω_{ij}^l, of the coefficient matrices. The prior standard deviations are measures of confidence in the prior means. A high confidence that the estimated coefficients match the prior mean is indicated by a small prior standard deviation.

First, consider the elements along the diagonal of Ω_l, that is, ω_{ij}^l when $i = j$. Litterman (1986) defined a hyperparameter, λ_1^h, to be the prior standard deviation of ω_{ii}^l which controls the extent to which the process is a random walk. As λ_1^h gets smaller, the random walk assumption becomes stronger since there is little variation around the prior mean of ω_{ii}^l which was set at 1 as in Equation 18 and ω_{ij}^l when $i \neq j$ will tend to zero.

Litterman (1986) next considered each equation in the VAR(p) system to impose further restrictions on the standard deviations of the prior means.

Using the example with $n = 3$ and $p = 2$ any row in Equation 12 can be written as:

$$
\theta_{it} = c_i + \omega_{i1}^1 \theta_{1t-1} + \omega_{i2}^1 \theta_{2t-1} + \omega_{i3}^1 \theta_{3t-1} + \omega_{i1}^2 \theta_{1t-2} + \omega_{i2}^2 \theta_{2t-2} + \omega_{i3}^2 \theta_{3t-2} + \epsilon_{it} \tag{19}
$$

The variation in the i-th variable comes from its past values, the past values of the other variables in the system and the estimated constant. Litterman (1986) took the prior mean of the constant c_i as zero and the standard deviation of c_i to be a weight or fraction, denoted by the hyperparameter, $\lambda_4^h > 0$, of the variation in the i-th variable, σ_i and set the standard deviation of c_i to be $\sigma_i \lambda_4^h$. A small λ_4^h implies that there is little variation in c_i from zero.

Let σ_i be the standard deviation of the i-th variable and σ_j be the standard deviation of the j-th variable. The ratio, $\frac{\sigma_i}{\sigma_j}$ affects the value of the coefficient of θ_{jt-1} in equation

i. If $\sigma_i < \sigma_j$ then less weight will be given to the coefficient of θ_{jt-1} in the *i*-th equation and this coefficient will tend to zero. The contribution of the variable's own lags to the variable's variation and the contribution of the other variable's lags to the variable's variation is measured by a hyperparameter $\lambda_2^h, (0, 1]$. $\lambda_2^h = 1$ implies that the contribution to the variation from the variable's own lagged values and the variation from the other variables lagged values are the same. A small λ_2^h implies that ω_{ij}^l when $i \neq j$ will tend to zero. The coefficients at lags $l > 1$ have a prior mean of zero. Another hyperparameter, $\lambda_3^h > 0$, is used in the form $l^{\lambda_3^h}$. If $\lambda_3^h > 1$, as the number of lags, $l > 1$ gets larger $l^{\lambda_3^h}$ also gets larger.

Finally specifying Litterman (1986)'s prior standard deviations for the ω_{ij}^l in Ω_l as:

$$\sqrt{\psi_{lij}} = \begin{cases} \dfrac{\lambda_1^h}{l^{\lambda_3^h}}, & \text{if } i = j; \\ \dfrac{\sigma_i \lambda_1^h \lambda_2^h}{\sigma_j l^{\lambda_3^h}}, & \text{if } i \neq j; \end{cases} \tag{20}$$

gives Litterman's Prior.

To illustrate this, consider the system of equations that forms the VAR(p) with $n = 3$ and $p = 2$. The diagonal of the prior covariance matrix for the *i*-th equation is given by the standard deviations in parenthesis below the coefficients as:

$$\begin{aligned}
\theta_{1t} &= \underset{(\sigma_1 \lambda_4^h)}{c_1} + \underset{(\lambda_1^h)}{\omega_{11}^1 \theta_{1t-1}} + \underset{(\frac{\sigma_1 \lambda_1^h \lambda_2^h}{\sigma_2})}{\omega_{12}^1 \theta_{2t-1}} + \underset{(\frac{\sigma_1 \lambda_1^h \lambda_2^h}{\sigma_3})}{\omega_{13}^1 \theta_{3t-1}} + \underset{(\frac{\lambda_1^h}{2^{\lambda_3^h}})}{\omega_{11}^2 \theta_{1t-2}} + \underset{(\frac{\sigma_1 \lambda_1^h \lambda_2^h}{\sigma_2 2^{\lambda_3^h}})}{\omega_{12}^2 \theta_{2t-2}} + \underset{(\frac{\sigma_1 \lambda_1^h \lambda_2^h}{\sigma_3 2^{\lambda_3^h}})}{\omega_{13}^2 \theta_{3t-2}} + \epsilon_{1t} \quad (21) \\[2em]
\theta_{2t} &= \underset{(\sigma_2 \lambda_4^h)}{c_2} + \underset{(\frac{\sigma_2 \lambda_1^h \lambda_2^h}{\sigma_1})}{\omega_{21}^1 \theta_{1t-1}} + \underset{(\lambda_1^h)}{\omega_{22}^1 \theta_{2t-1}} + \underset{(\frac{\sigma_2 \lambda_1^h \lambda_2^h}{\sigma_3})}{\omega_{23}^1 \theta_{3t-1}} + \underset{(\frac{\sigma_2 \lambda_1^h \lambda_2^h}{\sigma_1 2^{\lambda_3^h}})}{\omega_{21}^2 \theta_{1t-2}} + \underset{(\frac{\lambda_1^h}{2^{\lambda_3^h}})}{\omega_{22}^2 \theta_{2t-2}} + \underset{(\frac{\sigma_2 \lambda_1^h \lambda_2^h}{\sigma_3 2^{\lambda_3^h}})}{\omega_{23}^2 \theta_{3t-2}} + \epsilon_{2t} \\[2em]
\theta_{3t} &= \underset{(\sigma_3 \lambda_4^h)}{c_3} + \underset{(\frac{\sigma_3 \lambda_1^h \lambda_2^h}{\sigma_1})}{\omega_{31}^1 \theta_{1t-1}} + \underset{(\frac{\sigma_3 \lambda_1^h \lambda_2^h}{\sigma_2})}{\omega_{32}^1 \theta_{2t-1}} + \underset{(\lambda_1^h)}{\omega_{33}^1 \theta_{3t-1}} + \underset{(\frac{\sigma_3 \lambda_1^h \lambda_2^h}{\sigma_1 2^{\lambda_3^h}})}{\omega_{31}^2 \theta_{1t-2}} + \underset{(\frac{\sigma_3 \lambda_1^h \lambda_2^h}{\sigma_2 2^{\lambda_3^h}})}{\omega_{32}^2 \theta_{2t-2}} + \underset{(\frac{\lambda_1^h}{2^{\lambda_3^h}})}{\omega_{33}^2 \theta_{3t-2}} + \epsilon_{3t}
\end{aligned}$$

The prior means, $\mathbb{E}[v']$, and the prior covariance of the coefficients, $\mathbb{V}[v']$, are summarised as:

$$v' = \mathbb{E}[v'] = \mathbb{E} \begin{bmatrix} c' \\ \Omega_1' \\ \Omega_2' \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \overline{v}_{i=1} & \overline{v}_{i=2} & \overline{v}_{i=3} \end{bmatrix} \tag{22}$$

Denote the diagonals of the *i*-th equation's prior covariance matrix under the Litterman Prior by $\overline{\mathfrak{G}}_i$ (whose elements are the standard deviations in parenthesis in Equa-

tion 22 e.g. $\overline{\mathfrak{G}}_1 = (\sigma_1\lambda_4^h), (\lambda_1^h), (\frac{\sigma_1\lambda_1^h\lambda_2^h}{\sigma_2}), (\frac{\sigma_1\lambda_1^h\lambda_2^h}{\sigma_3}), (\frac{\lambda_1^h}{2^{\lambda_3^h}}), (\frac{\sigma_1\lambda_1^h\lambda_2^h}{\sigma_2 2^{\lambda_3^h}}), (\frac{\sigma_1\lambda_1^h\lambda_2^h}{\sigma_3 2^{\lambda_3^h}}))$. The posterior mean of v, denoted v^{Lit} is used as a point estimator for v (Lütkepohl, 2005). This is calculated for the i-th equation as:

$$v_i^{Lit} = (\overline{\mathfrak{G}}_i^{-1} + \sigma_i^2 X'X)^{-1}(\overline{\mathfrak{G}}_i^{-1}\overline{v}_i + \sigma_i^2 X'\theta_i) \tag{23}$$

Litterman's Prior estimates the coefficients for one equation at a time.

Kadiyala and Karlsson (1997) analyse other options of priors that are used in estimating the coefficients of a VAR(p). Sims and Zha (1998) generalized the result into one prior. Sims and Zha (1998) replaced Litterman's Prior with a form of the Normal-Wishart prior as follows.

First, the prior distribution of the coefficients given Σ is assumed to be normal. The prior distribution of the covariance matrix, Σ is inverse Wishart (Drèze & Richard, 1983) with $\overline{S}$ the diagonal scale matrix in the inverse Wishart prior distribution of Σ.

The relation between the Minnesota prior and the Normal-Wishart prior is as follows. Let η be the $(1+np)n \times 1$ vector of stacked intercepts and coefficients in the VAR(p), e.g.:

$$\eta = vec(v) = (c_1, c_2, c_3, \omega_{11}^1, \omega_{21}^1, \omega_{31}^1, \ldots, \omega_{13}^2, \omega_{23}^2, \omega_{33}^2)' \tag{24}$$

Similarly, $\eta^{OLS} = vec(v^{OLS})$. Also, let ϑ be the $nT \times 1$ stacking of the T observations of the first variable, then the T observations of the second variable and so on:

$$\vartheta = vec(\theta) = (\theta_{11}, \theta_{12}, \ldots, \theta_{1T}, \theta_{21}, \ldots, \theta_{3T})' \tag{25}$$

In a similar way, $e = vec(\xi)$. Then, the VAR(p) takes the form:

$$\vartheta = (I_n \otimes X)\eta + e \tag{26}$$

where $e \sim N(0, \sigma \otimes I_T)$. $\otimes$ is the Kronecker Product that for two matrices
$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$
and
$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{bmatrix}$

Denote the sampling density of ϑ as $p(\vartheta|\eta, \Sigma)$. This sampling density can be broken up into two parts. Firstly, the distribution for η given Σ:

$$\eta|\Sigma, \vartheta \sim N(\eta^{OLS}, \Sigma \otimes (X'X)^{-1}) \tag{27}$$

Secondly, $p(\vartheta|\eta, \Sigma)$ also has a part where Σ^{-1} has a Wishart distribution:

$$\Sigma^{-1}|\vartheta \sim W([(\vartheta - Xv^{OLS})'(\vartheta - Xv^{OLS})]^{-1}, T - (1 + np) - n - 1) \qquad (28)$$

Under the Normal-Wishart Prior, the posterior distribution is also Normal-Wishart and the posterior mean or point estimate of v is:

$$v^{NW} = (\overline{\mathcal{H}}^{-1} + X'X)^{-1}(\overline{\mathcal{H}}^{-1}\overline{v} + X'\theta) \qquad (29)$$

The estimate of the error covariance matrix, Σ is:

$$\Sigma^{OLS} = T^{-1}(\theta'\theta - v'^{NW}(X'X + \overline{\mathcal{H}}^{-1})v^{NW}) + \overline{v}'\overline{\mathcal{H}}^{-1}\overline{v} + \overline{S} \qquad (30)$$

$\overline{S}$ is a diagonal matrix with elements $((\frac{\sigma_1}{\lambda_0^h})^2, (\frac{\sigma_2}{\lambda_0^h})^2, (\frac{\sigma_3}{\lambda_0^h})^2)$

Sims-Zha's prior standard deviations for the ω_{ij}^l are specified as:

$$\sqrt{\psi_{lij}} = \frac{\lambda_0^h \lambda_1^h}{\sigma_j l^{\lambda_3^h}} \qquad (31)$$

$\overline{\mathcal{H}}$ has diagonal elements $(\lambda_0^h \lambda_4^h)^2, (\frac{\lambda_0^h \lambda_1^h}{\sigma_1})^2, (\frac{\lambda_0^h \lambda_1^h}{\sigma_2})^2, (\frac{\lambda_0^h \lambda_1^h}{\sigma_3})^2, (\frac{\lambda_0^h \lambda_1^h}{\sigma_1 2^{\lambda_3^h}})^2, (\frac{\lambda_0^h \lambda_1^h}{\sigma_2 2^{\lambda_3^h}})^2, (\frac{\lambda_0^h \lambda_1^h}{\sigma_3 2^{\lambda_3^h}})^2)$

$\overline{\mathcal{H}} \otimes \overline{S}$ gives a matrix with elements equal to the variances under Litterman's Prior when $\lambda_2^h = 1$. The Normal-Wishart Prior makes λ_2^h redundant because the scale matrix of the inverse Wishart distribution, $\overline{S}$ has the constant weight, λ_0^h in its denominator and prevents the prior from distinguishing between the lags of the i-th variable and those of the other variables in the i-th equation.

The basic VAR(p) in levels model assumes that the series θ_t are stationary (Hamilton, 1994). Many economic time series are non-stationary and in some cases cointegrated. In order to incorporate this into the Bayesian VAR models, Sims and Zha (1998) added two hyperparameters that incorporated non-stationarity and cointegration into the priors. The first hyperparameter is based on Doan, Litterman, and Sims (1984) which restricts the sum of the coefficients on the lagged values of the dependent variable in each equation to 1 (i.e. $\sum_{l=1}^p \Omega_l = I_n$) and the sum of the coefficients on each of the other variables to 0 in a BVAR(p) in first differences. This restriction is referred to as the sum of coefficients prior.

From Equation 15 this implies that the BVAR in first differences can be written as:

$$\Delta\theta_t = c + \sum_{k=1}^{p-1} \Gamma_k \Delta\theta_{t-k} + \epsilon_t, \qquad k = 1, \dots, p-1 \qquad (32)$$

since $\Pi\theta_{t-1} = \sum_{l=1}^p \Omega_l - I_n = 0$ because $\sum_{l=1}^p \Omega_l = I_n$ under the sum of coefficients

prior.

Eliminating the Π term implies that there is no cointegration but there are non-stationary variables in the system. The sum of coefficients prior is controlled by using the hyperparameter, $\mu_5^h \geq 0$. Sims and Zha (1998) assume that a good forecast of θ_i at some point $t = \tau$ is the average of the lagged values of θ_i, denoted as $\bar{\theta}_i$, but $\bar{\theta}_i$ does not help to predict the values of θ_j for $j \neq i$. They introduced n dummy variables, $\bar{\theta}_i^0$ at the start of each data set where $\bar{\theta}_i^0$ is the mean of the first p values of the i-th variable. μ_5^h assigns weights to the $\bar{\theta}_i^0$ such that a large $\bar{\theta}_i^0$ implies a high likelihood of $\Pi = 0$ such that in the long run the values of $\theta_i^0 \rightarrow \bar{\theta}_i^0$ (Doan et al., 1984; Sims & Zha, 1998; Robertson & Tallman, 1999; Summers, 2001).

The weighted dummy observations for the case n=3 and $p = 2$ is:

$$
\begin{bmatrix}
0 & \mu_5^h\bar{\theta}_1^0 & 0 & 0 & \mu_5^h\bar{\theta}_1^0 & 0 & 0 \\
0 & 0 & \mu_5^h\bar{\theta}_2^0 & 0 & 0 & \mu_5^h\bar{\theta}_2^0 & 0 \\
0 & 0 & 0 & \mu_5^h\bar{\theta}_3^0 & 0 & 0 & \mu_5^h\bar{\theta}_3^0
\end{bmatrix}
\tag{33}
$$

The sum of coefficients prior implies that as $\mu_5^h \rightarrow \infty$, $\Pi \rightarrow 0$ the number of unit roots is the same as the number of variables in the system and there is no cointegration. The additional hyperparameter μ_6^h allows the BVAR model to incorporate cointegration by assuming there is one cointegrating relation and sets the value of the constant $c_i = 1$. The matrix of the weighted dummy observations for $n = 3$ and $p = 2$ is:

$$
\begin{bmatrix}
\mu_6^h & \mu_6^h\bar{\theta}_1^0 & \mu_6^h\bar{\theta}_2^0 & \mu_6^h\bar{\theta}_3^0 & \mu_6^h\bar{\theta}_1^0 & \mu_6^h\bar{\theta}_2^0 & \mu_6^h\bar{\theta}_3^0 \\
1 & \theta_{12} & \theta_{22} & \theta_{32} & \theta_{11} & \theta_{21} & \theta_{31} \\
1 & \theta_{13} & \theta_{23} & \theta_{33} & \theta_{12} & \theta_{22} & \theta_{32}
\end{bmatrix}
\tag{34}
$$

Sims-Zha's prior has a hyperparameter λ_0^h for the overall tightness of the standard deviation of the errors ϵ_{it} and their intercorrelations. This prior also allows for non-stationary time series and cointegrated relations.

This prior is suitable for modeling the parameters of the Heligman-Pollard model through time and this is the prior used in this paper to implement the HP-BVAR model.

$\lambda_0([0,1])$ controls the overall tightness of the prior on the error covariance matrix. As it increases the coefficients have increased variance in the structural form such that the model strays further from a random walk.

$\lambda_1(> 0)$ specifies how tight the random walk prior specification is. As $\lambda_1 \rightarrow 0$ the diagonal elements of the coefficient matrix for the first lag tend to one and all other elements tend to zero. This restriction is implemented only on the lagged matrices. As the value increases the random walk prior will not be enforced as strictly and the model will stray further from a random walk in the lags.

Increasing the value of $\lambda_3(> 0)$ shrinks the coefficients of higher order lags to zero and by allowing the parameters contained in these lags to vary less around their conditional mean of zero.

λ_4 controls the tightness of the prior on the constants and as $\lambda_4 \rightarrow 0$ the constants tend to zero. It is conditional on λ_0.

μ_5 controls the unit root prior. As it increases the likelihood that the model can be expressed in first differences also increases.

μ_6 controls the co-integration prior such that as it increases co-integration will be given more weight.

Choice of values for the hyperparameters is determined by obtaining values from previous studies or by evaluating the forecast performance of the model over a range of hyperparameters (Joiner, 2001).

The BVAR includes a requirement for the coefficients of long-term lags to be closer to zero than the short-term lags (Kadiyala & Karlsson, 1997). This is a way of imposing structure on the system. It provides a compromise between over-parameterization (in VAR modeling) and under-parameterization (in univariate modeling). It improves the accuracy of estimates and subsequent forecasts by introducing appropriate prior information into the model. In particular the BVAR model allows the standard deviations of the coefficients to decrease as the lags increase. Litterman (1986) incorporates a prior into the system by considering the behavior of each equation in the system on its own. Sims and Zha (1998); Waggoner and Zha (1999) incorporate prior information into the VAR model by considering the entire system.

5   Results

5.1   Parameters

The parameters of Heligman-Pollard Curves for 62 years from 1946 - 2007 are estimated with 50 years worth of observations from 1946-1995. VAR and Bayesian-VAR models are estimated for the fitted parameters. The fitted models are used to predict out-of-sample parameter forecasts for 1996-2007 which are compared to the observed estimated parameters. The estimated parameters of the Heligman-Pollard curves using data from 1946-2007 are shown in Figure 5. The male parameters, $\theta_{t,m}$ are the solid line while the female parameters, $\theta_{t,f}$ are the broken line.

A_m and A_f are both declining. Mortality of children at age 1 has been steadily declining. The gradient of the parameters is not as steep in recent years as it was between 1950 and 1970. The difference between mortality at age one and mortality at age 0, B_m and B_f, and the rate at which mortality decreases during childhood, C_m and C_f, show no distinct trend although the Bs are higher for more recent years while the

Cs are lower.

The intensity of young adult mortality is reflected in D_m and D_f. D_m, is consistently higher than D_f. Between 1960 and 1970 D_m increased and then started to decline from about 1980 although there was a short period of increase in the mid to late 1990s. An increase in D implies that mortality for young adults increased. E_m and E_f vary inversely with the spread of the young adult mortality hump. An increase in E means the hump is tight with little spread around the modal age while a decrease in E implies that hump is more spread around the modal age. Since the 1970s E_m and E_f have been declining. The spread of young adult mortality around the mortality hump is wide. F_m and F_f is the location or the modal age of the accident hump (and maternal hump for females). Before 1960, this location is higher for females than males but after 1965 F_f is lower than F_m.

G_m and G_f are the intercepts of the Gompertz curve at age 0. They represent the base level of senescent mortality. G_f is lower than G_m with a narrowing gap. H_m and H_f are the slope of the Gompertz curve. They represent the rate of increase of G_m and G_f. Until the late 1990s, late life mortality for both males and females had been declining at an increasing rate. However, from 2000 late life mortality for both males and females has been declining at a decreasing rate. This is consistent with the findings of Risk Management Solutions medical-based longevity risk model where the level of mortality improvement experienced in the last 30 years is estimated to tail off in 15 to 25 years (Risk Management Solutions, 2010a, 2010b). This is not easily captured when mortality models do not give more weight to recent observations and leads to underestimating of future mortality rates.

Testing for difference stationarity reveals that all the time series of the Heligman-Pollard Parameters are I(1) except for the female *B* parameter.

	A	B	C	D	E	F	G	H
A	*****	-0.128	0.572	0.204	0.268	-0.697	0.885	-0.824
B	0.322	*****	0.605	-0.299	-0.5	0.472	-0.374	0.368
C	<0.001	<0.001	*****	0.02	-0.014	-0.245	0.423	-0.416
D	0.111	0.018	0.876	*****	0.739	-0.663	0.372	-0.312
E	0.035	<0.001	0.911	<0.001	*****	-0.769	0.608	-0.614
F	<0.001	<0.001	0.055	<0.001	<0.001	*****	-0.874	0.84
G	<0.001	0.003	0.001	0.003	<0.001	<0.001	*****	-0.968
H	<0.001	0.003	0.001	0.013	<0.001	<0.001	<0.001	*****

Table 1: Correlation and Significance Males: upper diagonal part contains correlation coefficient estimates; lower diagonal part contains corresponding p-values. Parameters with significant correlation have p-values <0.001.

The parameters of the Heligman-Pollard model are found to be highly correlated which is consistent with Hartmann (1987). The correlation matrix of the parameters
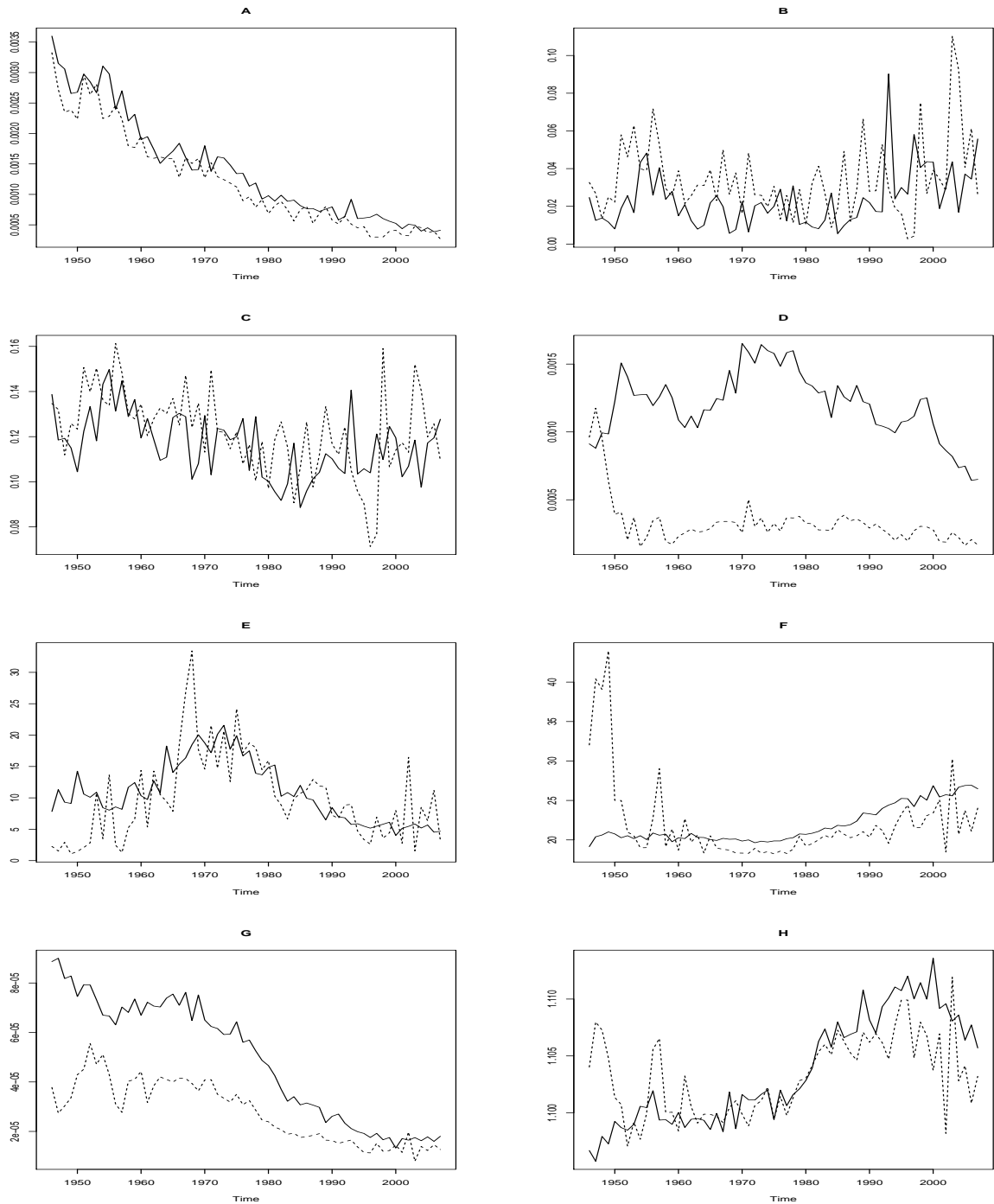
Figure 5: Heligman and Pollard Parameters for Australia Male and Females from 1946 to 2007. Solid Line=Males; Dotted Line=Females

are shown in Table 1 and 2 for males and females respectively.

The parameters with significant correlation have p-values <0.001. With the exception of D, there is significant correlation between parameters in different terms of Equation 2. A_m is significantly correlated with C_m, F_m, G_m and H_m; B_m with C_m, E_m and F_m, D_m with E_m and F_m, E_m with F_m,G_m and H_m, F_m with G_m and H_m, and G_m with H_m. For females the number of significant correlations are fewer but there is still sig-

19

	A	B	C	D	E	F	G	H
A	*****	0.104	0.567	0.512	-0.133	0.358	0.813	-0.401
B	0.42	*****	0.785	-0.084	-0.152	0.048	-0.056	0.07
C	<0.001	<0.001	*****	0.145	-0.022	0.079	0.462	-0.297
D	<0.001	0.517	0.262	*****	-0.127	0.714	0.138	0.204
E	0.304	0.237	0.863	0.325	*****	-0.583	0.265	-0.517
F	0.004	0.709	0.543	<0.001	<0.001	*****	-0.122	0.475
G	<0.001	0.665	<0.001	0.285	0.037	0.343	*****	-0.794
H	0.001	0.588	0.019	0.112	<0.001	<0.001	<0.001	*****

Table 2: Correlation and Significance Females: upper diagonal part contains correlation coefficient estimates; lower diagonal part contains corresponding p-values. Parameters with significant correlation have p-values <0.001.

nificant correlation between the parameters in the different terms in equation 2 since A_f is significantly correlated with C_f, D_f and G_f, C_f with B_f and G_f and F_f with H_f. This shows how old age mortality is correlated with mortality at younger ages.

5.2 VAR model for Parameters

A VAR(p) model is fitted to the logarithms of the time series of the Heligman-Pollard parameters for females and males using a model with an unrestricted constant as there is both presence of a constant and presence of a trend in many of the variables in the model. These are denoted by VAR(p)$_f$ and VAR(p)$_m$ for females and males respectively. $\theta_{t,f} = (A_{t,f}, B_{t,f}, C_{t,f}, D_{t,f}, E_{t,f}, F_{t,f}, G_{t,f}, H_{t,f})$ and $\theta_{t,m} = (A_{t,m}, B_{t,m}, C_{t,m}, D_{t,m}, E_{t,m}, F_{t,m}, G_{t,m}, H_{t,m})$ are the variables included in VAR$(p)_f$ and VAR$(p)_m$ which model the VAR(p) for females and males.

In order to minimize the criteria in Equation 14 the number of lags p is selected as 2. Since the time series of the variables (time series of the Heligman-Pollard model) in the VAR(p) have unit roots then it is also necessary to fit a VECM. Since a VAR(2) is suited to model the time series of the parameters, the corresponding VECM is of lag $p - 1 = 2 - 1 = 1$.

The next step involves testing for the existence of a stationary linear combination of the time series in the VECM.

Tables 3 and 4 show the results for the cointegration tests. In each case the LR_{test} rejects the null that $r_0 = 0$ at 1% significance and rejects the null that $r_0 = 1$ at 5% significance.

For forecasting the VECM is transformed back to a VAR.

A VAR(2) model is determined to be the appropriate model for the logarithms of the parameters of the Heligman-Pollard model. There are unit roots and cointegrated relations in the system. Using this information a Bayesian VAR is now estimated for

	Eigenvalue	Trace Stat	95% CV	99% CV
H(0)++	0.63	193.35	170.80	182.51
H(1)+	0.58	145.86	136.61	146.99
H(2)	0.47	104.52	104.94	114.36
H(3)	0.42	73.82	77.74	85.78
H(4)	0.36	47.30	54.64	61.24
H(5)	0.28	25.62	34.55	40.49
H(6)	0.17	9.65	18.17	23.46
H(7)	0.01	0.62	3.74	6.40

Table 3: Female Cointegration Tests

	Eigenvalue	Trace Stat	95% CV	99% CV
H(0)++**	0.79	221.13	170.80	182.51
H(1)+	0.61	146.81	136.61	146.99
H(2)	0.48	101.38	104.94	114.36
H(3)	0.40	69.91	77.74	85.78
H(4)	0.33	45.05	54.64	61.24
H(5)	0.26	25.49	34.55	40.49
H(6)	0.20	10.88	18.17	23.46
H(7)	0.00	0.06	3.74	6.40

Table 4: Male Cointegration Tests

Trace tests significant at the 5% level are flagged by ' +'.
Trace tests significant at the 1% level are flagged by '++'.
Max Eigenvalue tests significant at the 5% level are flagged by ' *'.
Max Eigenvalue tests significant at the 1% level are flagged by '**'.

the Australian data.

5.3 BVAR model

In modeling a B-VAR, cointegration is captured by the priors. There is no need for explicit modeling to be included. Since there are no values available from prior mortality studies a range of hyperparameters were considered. The best model performance was found to be for the set ($\lambda_0 = 0.9$, $\lambda_1 = 0.1$, $\lambda_2 = 1$, $\lambda_3 = 1$, $\lambda_4 = 0.05$, $\mu_5 = 5$, $\mu_6 = 5$). $\lambda_0 = 0.9$ - the model strays from a random walk. $\lambda_1 = 0.1$ - the diagonal elements of the coefficient matrix for the first lag tend to one and all other elements tend to zero. $\lambda_3 = 1$ - the parameters contained in higher lags vary less around their conditional mean of zero. $\lambda_4 = 0.05$ - constants tend to zero. $\mu_5 = 5$ - increased likelihood that the model can be expressed in first differences. $\mu_6 = 5$ - likelihood of co-integration is given more weight.
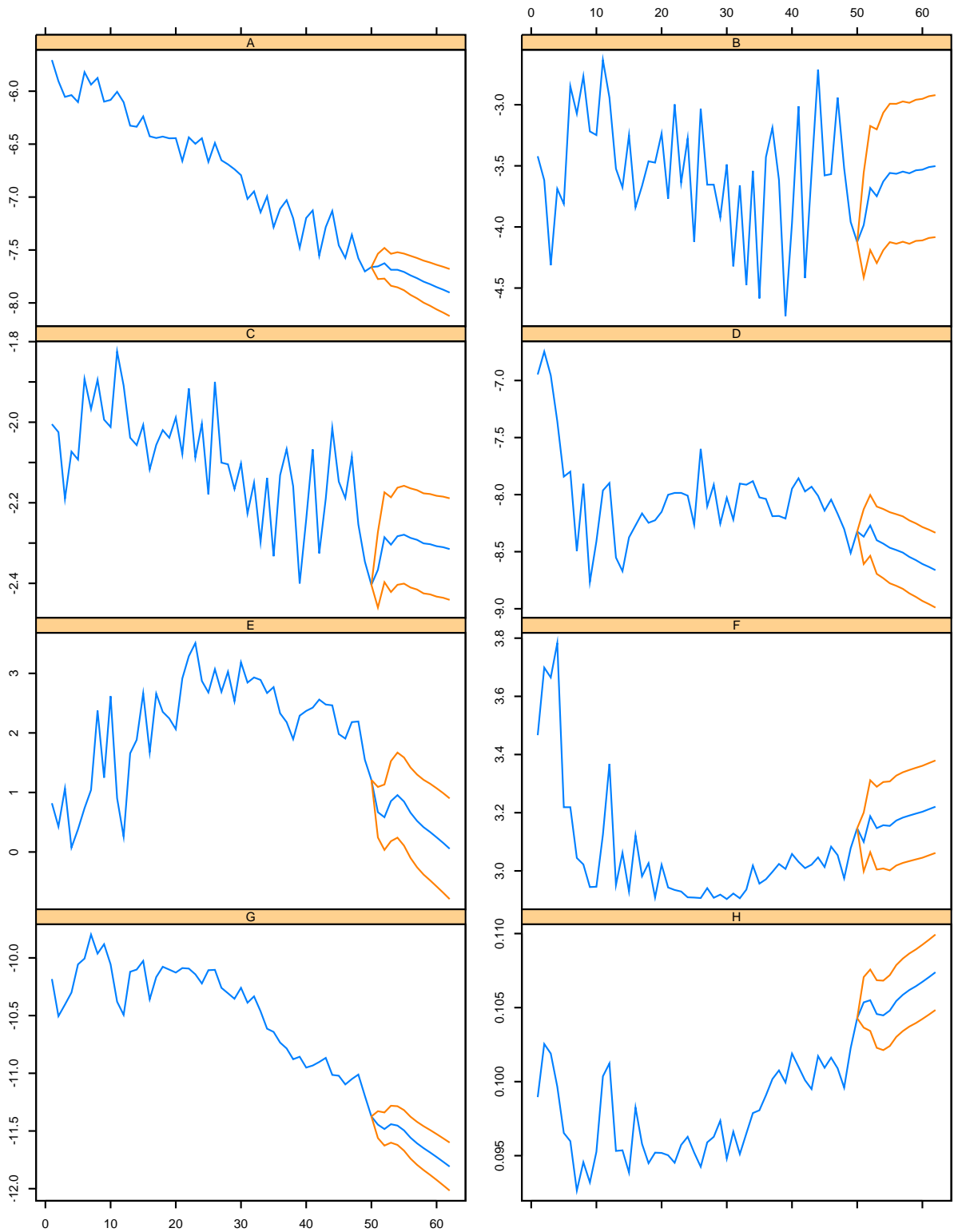
Figure 6: Fitted Values (1946-1995) and Projected Values (1996-2007) of $\ln \theta_{t,f}$ using VAR(2) Females

The predictions from fitting 50 years (1946-1995) of data are compared to the out-of-sample 12 years (1996-2007) of data that are available for the VAR model for females are displayed in Figure 6. The equivalent results for the BVAR(2) for females are shown in Figure 7. The VAR model shows an initial increase in parameter uncertainty that settles down to a long run distribution for each of the parameters. After an initial time period the parameter risk would be considered as having reached its maximum.

The results for the BVAR model show a significantly higher level of parameter risk and over the same horizon has not reached a long run distribution for the parameters. It is clear that parameter uncertainty is very significant and prediction intervals that take into account parameter risk are significantly higher than the case where this risk is not included. Figure 8 shows the equivalent BVAR(2) results for males confirming similar results as for females.

An advantage of the BVAR model is that parameters are estimated with greater accuracy as well as reflecting the true uncertainty in the predictive distribution. Table 5 shows this. The VAR model parameter estimates can lack intuitive interpretation whereas the BVAR model parameter estimates provide a more parsimonious model.

5.4 Mortality at Older Ages

For longevity risk the uncertainty at the older ages is of most interest. Figure 9 shows the observed and predicted parameter values for old age mortality using both the VAR and BVAR for males and females. The observed values are indicated by the green line, the VAR predictions by the red line and the B-VAR predictions by the blue line. $\ln(H)$ is the relative rate of old age mortality increase. This should reflect the decline in the rate of mortality increase after the mid-1990s. The plots show how the relative rate of change in mortality, given by $\ln(H)$, is over-estimated by the VAR whereas the Bayesian-VAR captures the reducing rate of change of old age mortality.

Figures 10 and 11 show the predictive distributions from the BVAR model for the male older age mortality parameters in the Heligman-Pollard model for 1996-2007 . The actual parameter values are also shown. The VAR estimates are also shown for comparison. The plots show how the BVAR model provides more accurate predictions and also quantify uncertainty in parameter estimates consistent with the actual mortality experience.

		lnA.lag1	lnB.lag1	lnC.lag1	lnD.lag1	lnE.lag1	lnF.lag1	lnG.lag1	lnH.lag1	lnA.lag2	lnB.lag2	lnC.lag2	lnD.lag2	lnE.lag2	lnF.lag2	lnG.lag2	lnH.lag2
VAR_f	c																
lnA	2.7849	0.6859	-0.3366	0.8078	0.0069	-0.078	-0.0689	1.5703	96.2176	-0.0862	-0.1072	0.6698	-0.0316	-0.0626	0.2077	-1.4953	-132.0613
lnB	12.1821	0.8679	-0.8244	3.5019	0.2456	-0.2632	-0.3132	6.7306	390.6899	-0.7038	0.0177	1.2511	-0.0061	-0.1985	-0.064	-9.5697	-727.9425
lnC	1.3804	-0.057	-0.1265	0.7734	0.0169	-0.0278	-0.1019	1.6686	110.6891	0.0607	-0.1126	0.5251	-0.009	-0.0463	-0.0152	-1.851	-140.2451
lnD	-4.2349	-0.8283	0.0361	-0.1261	0.2589	0.0399	-0.0211	1.5385	144.1147	0.8145	-0.2725	0.1966	0.2621	0.0231	-0.0112	-1.2781	-121.3759
lnE	22.7082	-2.1337	-0.5828	4.6311	-1.0031	0.3188	0.1177	-0.1632	-35.5825	-0.1471	-1.0076	3.3633	1.4199	-0.2537	-3.1969	3.1735	238.6835
lnF	0.8106	0.1335	0.0162	-0.3199	0.2225	0.0036	-0.0328	0.164	27.5631	0.6323	-0.114	0.116	-0.1028	0.0183	0.352	-1.6556	-120.0886
lnG	-0.4342	-0.2967	-0.1608	1.1058	-0.1327	-0.0004	0.2405	1.5272	66.857	-0.2849	-0.0139	0.166	0.1297	-0.0561	-0.3044	0.7602	58.9427
lnH	-0.0061	0.0043	0.0021	-0.0152	0.0013	0.0002	-0.0033	-0.0168	-0.5172	0.004	-0.0004	-0.0005	-0.0021	0.0008	0.0061	-0.0014	-0.2494
BVAR_f	c																
lnA	-0.0002	0.9584	-0.028	-0.0954	-0.019	-0.0019	-0.056	0.0284	-0.2109	0.042	0.0252	0.0948	0.0197	-0.0107	0.0556	-0.028	0.1785
lnB	-0.0005	-0.1335	0.8887	-0.3707	-0.0761	0.0149	-0.1783	0.0856	2.9977	0.1343	0.0993	0.3668	0.0777	-0.0246	0.1763	-0.0845	-3.107
lnC	-0.0001	-0.0259	-0.0213	0.9241	-0.0158	0.0038	-0.0403	0.0193	0.4725	0.0262	0.019	0.0752	0.0161	-0.0071	0.0399	-0.0191	-0.4957
lnD	-0.0004	-0.0534	-0.0321	-0.1225	0.9227	-0.0035	-0.0468	0.0177	1.0263	0.0551	0.033	0.1257	0.078	0.0236	0.0437	-0.0168	-1.1237
lnE	0.0008	-0.0916	-0.0132	-0.0578	-0.224	0.8901	0.1687	0.0661	-7.0796	0.0896	0.0118	0.0541	0.2202	0.0571	-0.1663	-0.0672	7.1986
lnF	-0.0002	-0.0081	-0.0056	-0.0177	0.0201	0.0158	0.9193	-0.0045	1.0732	0.0086	0.0067	0.0193	-0.0195	-0.0038	0.0794	0.0048	-1.1083
lnG	0.0002	-0.0027	-0.0007	0.0063	-0.0083	-0.0197	0.0538	0.9942	0.6524	0.0024	-0.0006	-0.0076	0.008	-0.0045	-0.052	0.0054	-0.6035
lnH	0	0	0	-0.0001	0.0002	0.0003	-0.0008	0.0001	0.9852	0	0	0.0001	-0.0002	0	0.0008	-0.0001	0.0141
VAR_m	c																
lnA	-1.2646	0.5625	0.1854	-1.4227	-0.1851	-0.0484	2.1264	0.3902	23.7682	0.2857	-0.342	1.4633	0.5037	-0.0986	-0.1037	0.443	19.4773
lnB	3.4734	0.6575	1.9911	-10.3217	1.0621	-0.7431	5.554	0.5912	1.355	-0.6266	-2.0902	11.7959	0.7617	-0.2019	-1.8982	1.437	191.3919
lnC	1.6076	0.0616	0.3006	-1.4716	0.261	-0.1563	0.7405	0.3376	15.2666	-0.0963	-0.4365	2.4097	0.0391	-0.0801	-0.5485	0.2911	42.8929
lnD	-7.683	0.0963	0.1277	-0.7094	0.3414	0.139	1.3471	0.4867	21.1418	-0.0149	-0.0162	0.0519	0.1871	0.1442	0.0097	0.0012	21.0224
lnE	0.378	-0.3702	-0.1176	0.8857	-0.3909	0.4049	-0.8032	0.0253	23.7933	-0.1667	-0.0158	0.2298	0.1764	0.2594	1.6093	2.0437	133.1608
lnF	1.1445	0.0077	0.015	-0.0759	0.0102	-0.0518	0.2897	-0.0173	-4.3073	-0.006	-0.0275	0.1645	-0.0417	0.0268	0.1338	-0.2185	-13.2135
lnG	-0.7553	-0.1365	0.0437	-0.105	-0.0031	0.0582	-0.5758	1.3181	71.2391	0.1335	-0.1405	0.7253	-0.0777	-0.0907	0.26	0.2476	7.9704
lnH	0.0095	0.0014	-0.0014	0.0067	-0.0008	-0.0001	0.0017	-0.0148	-0.7626	-0.0009	0.0022	-0.0123	0.0005	0.001	0.0007	0.0032	0.316
BVAR_m	c																
lnA	-0.0002	0.9402	-0.014	-0.0751	-0.0489	0.0111	0.2096	-0.0109	-0.3581	0.0609	0.0122	0.0756	0.0493	-0.013	-0.2111	0.0115	0.2968
lnB	-0.0006	-0.1722	0.92	-0.4452	0.0779	0.0855	0.8295	-0.123	1.3494	0.1728	0.0689	0.4398	-0.0763	-0.0888	-0.8329	0.124	-1.4427
lnC	-0.0001	-0.0306	-0.0147	0.9164	0.0246	0.0203	0.146	-0.0209	0.2922	0.0309	0.0126	0.0828	-0.0242	-0.021	-0.1467	0.0211	-0.3153
lnD	0	0	0.0009	0.0026	0.9638	0.0084	0.0536	0.0495	-2.9827	0.0003	-0.0018	-0.0029	0.036	-0.0077	-0.0535	-0.0494	2.9749
lnE	0.0001	-0.01	0.0041	0.0326	-0.093	0.9262	-0.2281	0.0302	-1.8992	0.0108	-0.0037	-0.0314	0.0926	0.0727	0.2284	-0.0301	1.8935
lnF	0	0.0009	0.0008	0.0034	-0.0049	-0.0072	0.9585	-0.0005	0.0633	-0.0012	-0.0007	-0.0036	0.0049	0.007	0.0416	0.0004	-0.0564
lnG	0	0.0017	-0.0034	-0.019	0.0328	0.0132	-0.0597	0.9557	4.4882	-0.0005	0.004	0.0206	-0.0329	-0.0145	0.0593	0.0446	-4.5146
lnH	0	-0.0001	0.0001	0.0003	-0.0006	-0.0001	0.0004	0.0011	0.8953	0.0001	-0.0001	-0.0003	0.0006	0.0001	-0.0004	-0.0012	0.1049

Table 5: Coefficient Matrices of VAR_f and $BVAR_f$; VAR_m and $BVAR_m$ showing that some of the coefficients in the VAR are zero while in the VAR they are non-zero. A shortcoming of VAR models is some coefficients are estimated to be non-zero by chance.
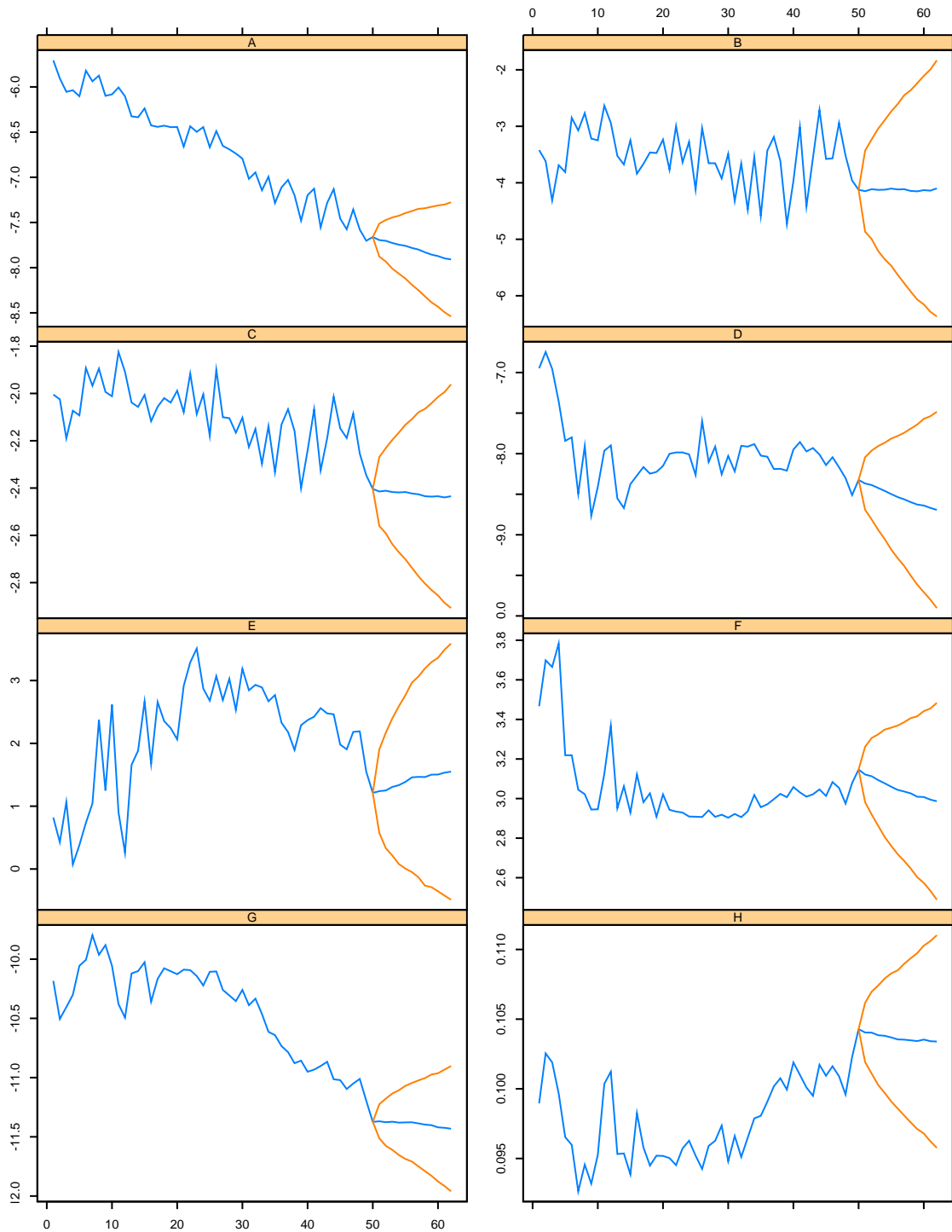
Figure 7: Fitted Values (1946-1995) and Projected Values (1996-2007) of $\ln\theta_{t,f}$ using BVAR(2) Females
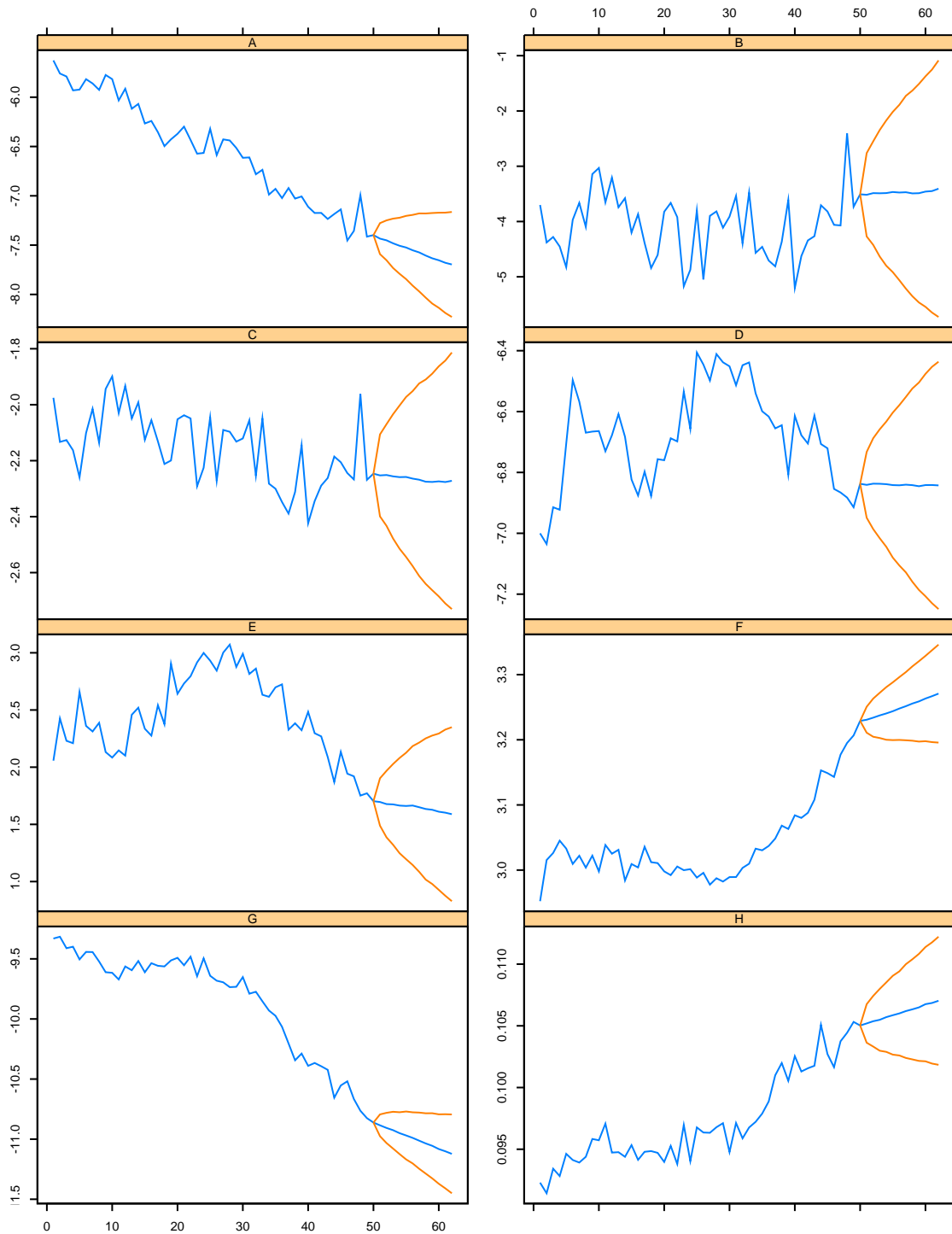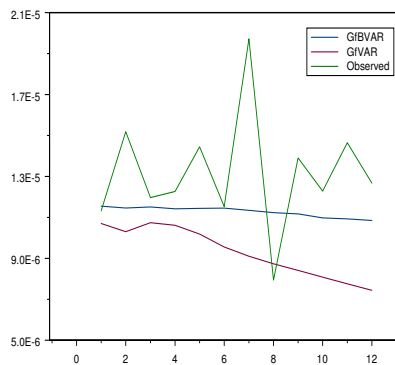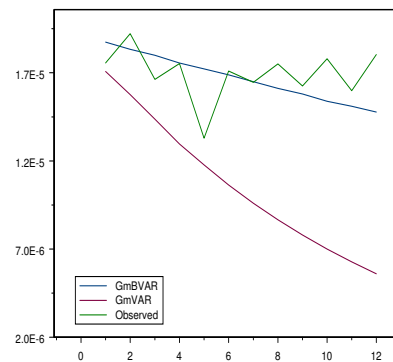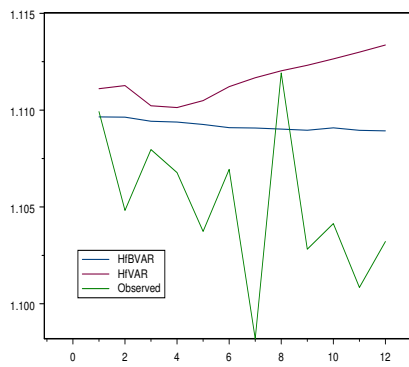
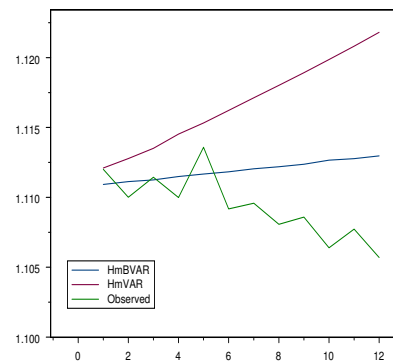Figure 8: Fitted Values (1946-1995) and Projected Values (1996-2007) of $\ln \theta_{t,m}$ using BVAR(2) Males

(a) ln(*G*) Females (b) ln(*G*) Males

(c) ln(*H*) Females (d) ln(*H*) Males

Figure 9: Estimated Old Age Parameters and the realised values from 1996-2007

VAR models produced inaccurate estimates when forecasted. The confidence bands for the estimated parameters are shown by the thin broken and thin dotted lines. The thick red solid line is the observed value of the parameter estimated for that year, the thick black broken line is the BVAR (with parameter risk) estimate of the parameter for that year and the thick blue dot-dashed line is the VAR (without parameter risk) estimate of the parameter for that year. The observed parameter value (thick red line) should fall within the confidence bands of the density of the distribution of the parameter, as estimated by the BVAR, 95% of time. The closer the observed value of the parameter for that year is to the BVAR estimate, the better the BVAR estimate. As the time from the last observed value used in the model, $T = 1995$, increases, the difference between the VAR estimate and the observed value increases. The plots show clearly that the BVAR projections outperform the VAR projections.

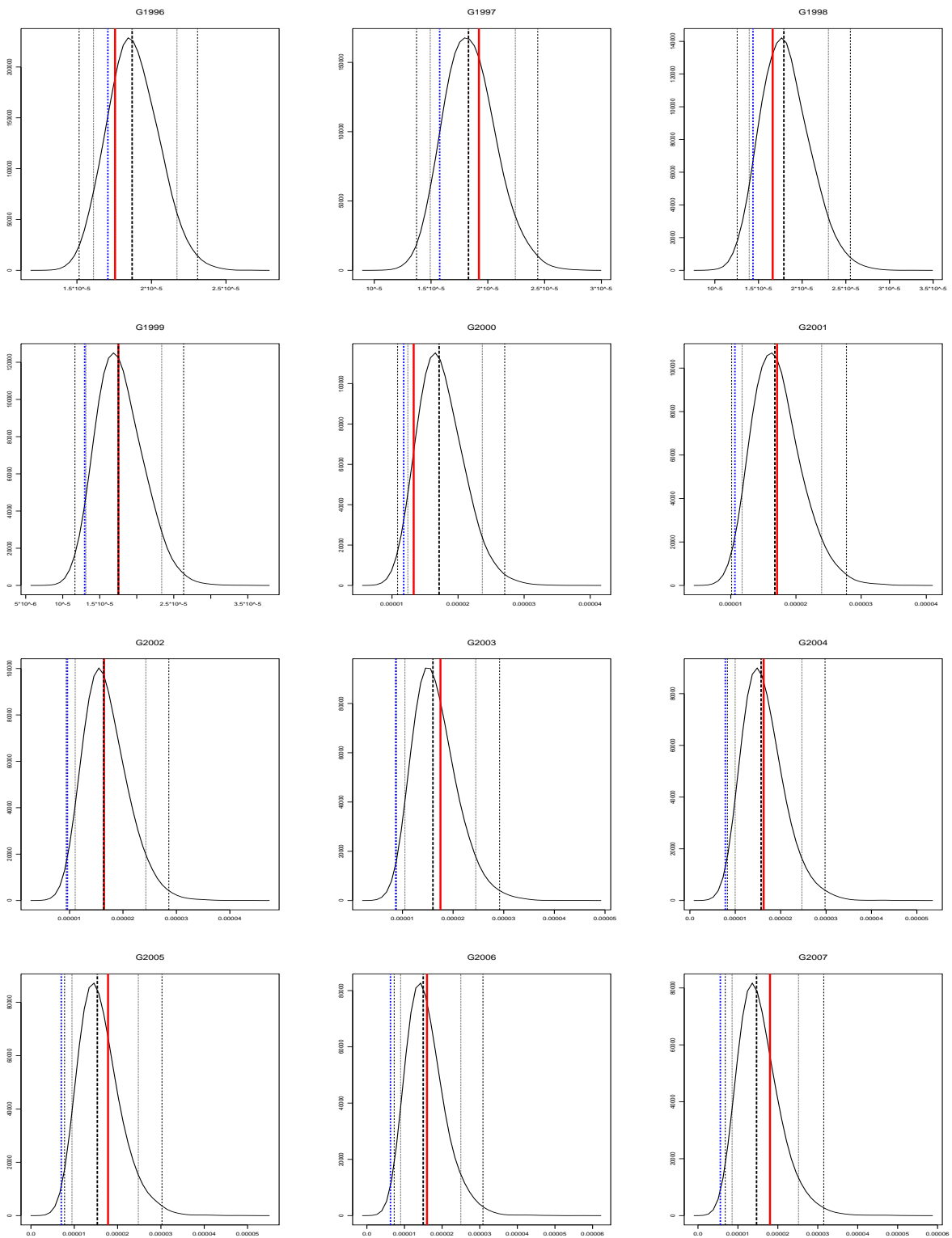The relative performance of the two models can be compared using the Root Mean

Figure 10: Parameter Risk in *G* (Males) for the predicted values for 1996-2007. Red (Thick Solid Line)=Observed, Black (Thick Broken Line)= BVAR Estimate, Blue (Thick Dot-Dashed Line)=VAR Estimate. The 95% and 99% confidence intervals are indicated by the thin dotted and thin broken line respectively.
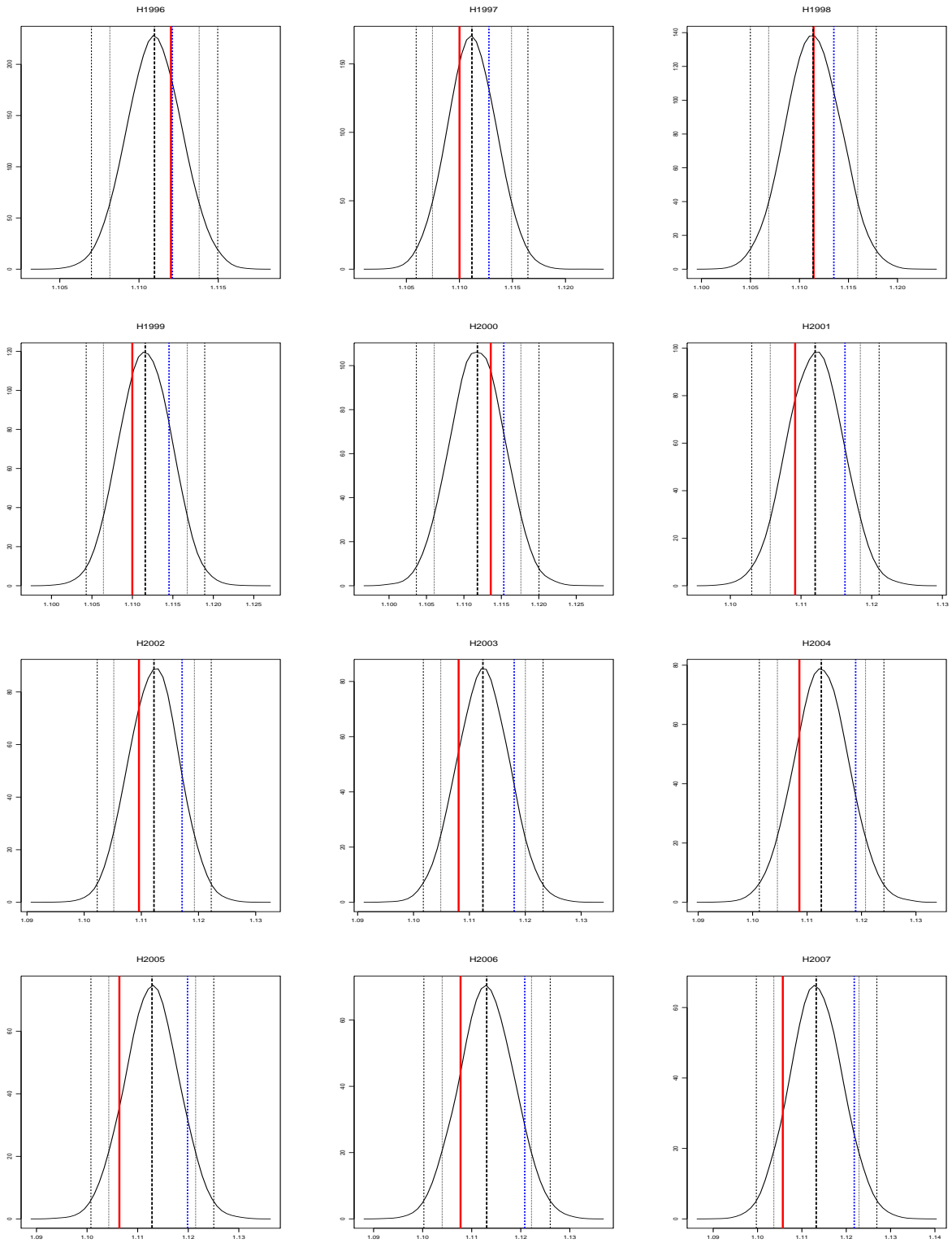
Figure 11: Parameter Risk in *H* (Males) for the predicted values for 1996-2007. Red (Thick Solid Line)=Observed, Black (Thick Broken Line)= BVAR Estimate, Blue (Thick Dot-Dashed Line)=VAR Estimate. The 95% and 99% confidence intervals are indicated by the thin dotted and thin broken line respectively.

Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (y_n^p - y_n)^2} \tag{35}$$

where for N observations, y_n is the observed value and y_n^p is the predicted value of observation n.

Table 6 shows the RMSE for the parameters for both males and females. The RMSE for the estimates from the B-VAR are smaller demonstrating the improved fit.

	VAR	B-VAR
G		
Female	2.39E-11	1.09E-11
Male	5.37E-11	3.25E-12
H		
Female	5.98E-05	2.96E-05
Male	8.00E-05	1.42E-05

Table 6: RMSE of predicted values 1996-2007 for parameters G and H using the VAR model (no parameter risk) and BVAR model (with parameter risk).

5.5 Mortality Rate Estimates

The performance of the models can also be considered in terms of the mortality rates that are forecasted from the fitted VAR and BVAR models for the Heligman-Pollard parameters. Denote the estimates of $q_{x,t}$ using the parameters predicted by the B-VAR as $\hat{q}_{x,t}^{B-VAR}$, the estimates predicted by the VAR as $\hat{q}_{x,t}^{VAR}$.

The sum of squared errors (SSE) for the mortality rates is given by:

$$SSE = \sum_{x,t} (q_{x,t}^p - q_{x,t})^2 \tag{36}$$

where the summation is taken over the entire projected age range and projected time period. The model with the smallest SSE provides a better fit to the data. Table 7 shows the SSE for male and female mortality rates for both the VAR and BVAR models.

	B-VAR	VAR
Females	0.01088	0.01041
Males	0.01045	0.02840

Table 7: Sum of squared errors of predicted values 1996-2007 for mortality rates using the VAR model (no parameter risk) and BVAR model (with parameter risk).

The RMSE for each predicted year for mortality rates is shown in figures 12(a) and

12(b) for females and males respectively. The root-mean-squared error (RMSE) in the HP-BVAR (black solid line) and the RMSE in the HP-VAR model (blue dotted line). There is a significant difference in the performance for males.
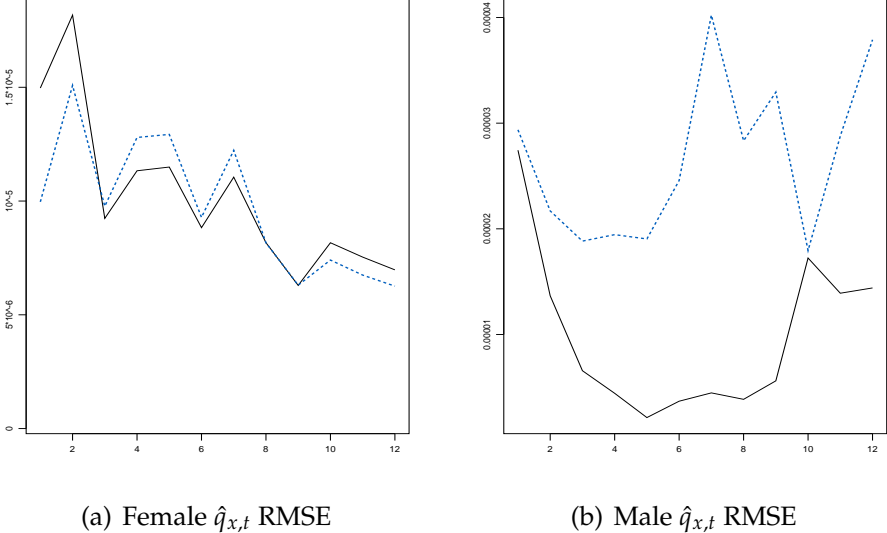

(a) Female $\hat{q}_{x,t}$ RMSE

(b) Male $\hat{q}_{x,t}$ RMSE

Figure 12: RMSE of predicted q_x for 12 years (1996-2007).

Figures 13 and 14 show the confidence intervals for $\ln q_x$. The solid black line and the dotted black line are the BVAR estimate and 99% confidence intervals respectively while the blue broken line and the blue dotted line are the VAR estimate and 99% confidence intervals. The plots confirm the superior performance of the BVAR model. They also demonstrate the increased uncertainty in prediction intervals for mortality rates that arises from quantifying parameter risk in a Bayesian model framework.
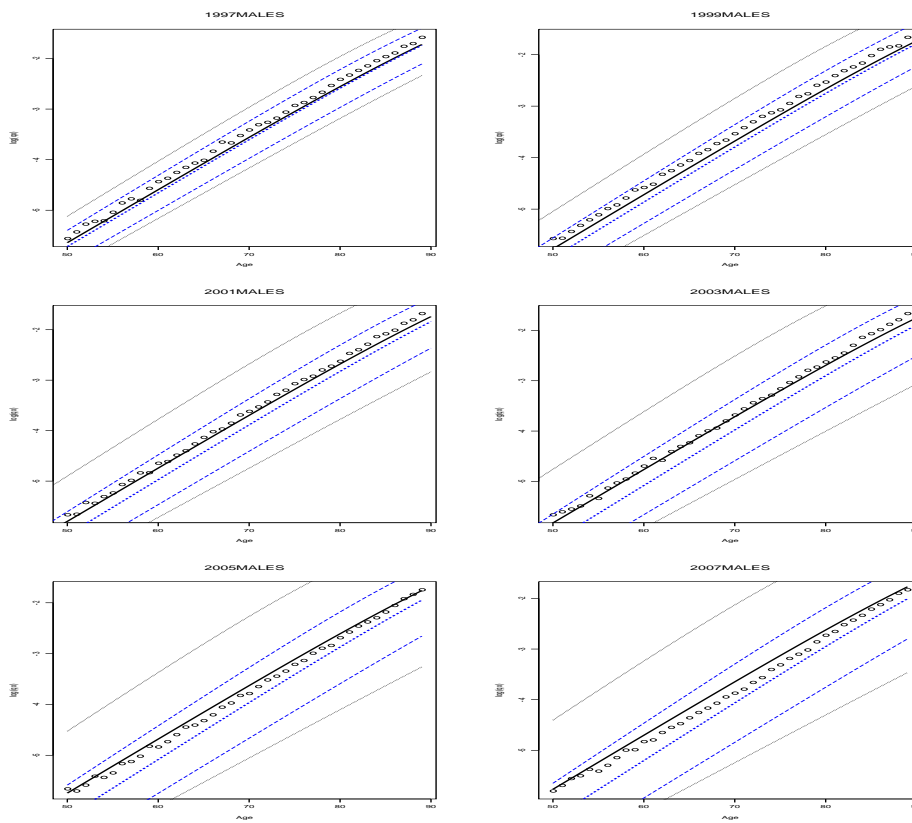
31

Figure 13: Old age Predictions of Male $\ln q_x$ by Year using HP-VAR and HP-BVAR and the observed $\ln q_x$ (small circles). The blue lines are for the HP-VAR model while the Black lines are for the HP-BVAR model. The HP-BVAR model has wider confidence intervals that reflect the parameter uncertainty.
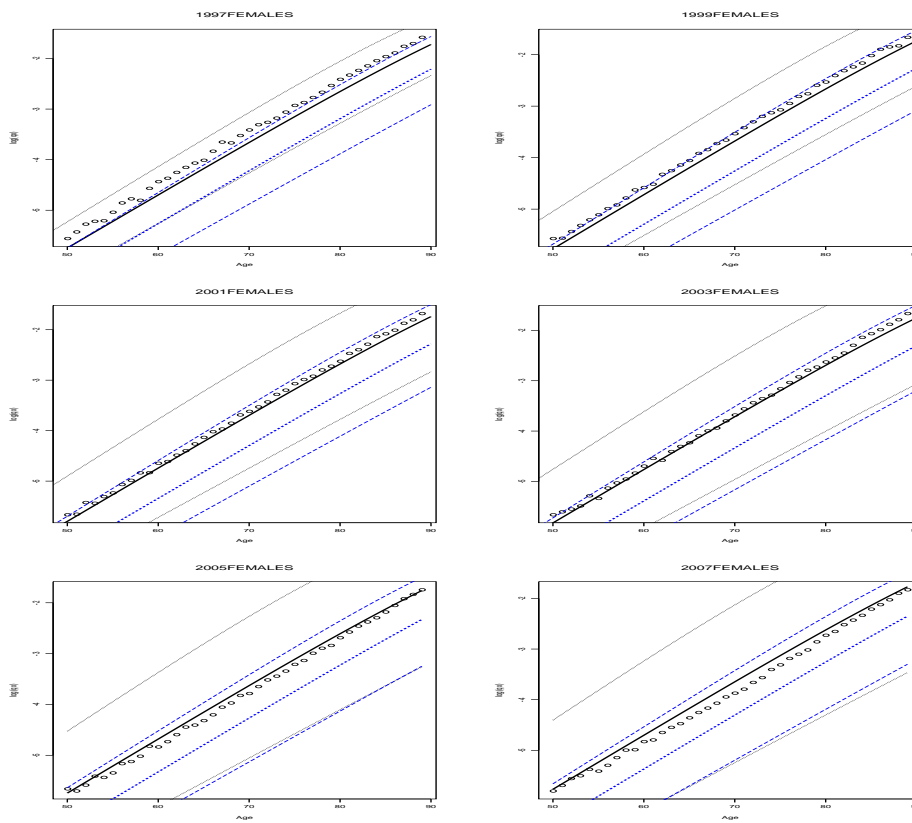
Figure 14: Old age Predictions of Female $\ln q_x$ by Year using HP-VAR and HP-BVAR and the observed $\ln q_x$ (small circles). The blue lines are for the HP-VAR model while the Black lines are for the HP-BVAR model. The HP-BVAR model has wider confidence intervals that reflect the parameter uncertainty.

6 Conclusions

Bayesian models allow a quantification of parameter risk in a predictive distribution. For mortality modeling, the full age range of mortality rates is often captured by a parametric model at any given time in order to provide a parsimonious fit to the data. To forecast future mortality, the parameters are fitted by a time series model and used for projecting expected rates and the volatility of the rates. There has been limited analysis and quantification of parameter uncertainty in mortality models.

This paper applies a Bayesian VAR model to the parameters of the Heligman-Pollard model fitted to Australian mortality data. The Heligman-Pollard model is a parsimonious representation of the mortality rates at any given time. Multivariate time series models then capture dependence across the whole age range by modeling the parameters including their correlation. The Bayesian VAR model provides a more flexible modeling framework for modeling mortality trends and risks than the more traditional VAR model.

HP-VAR and HP-BVAR models are implemented. Parameters of the HP-BVAR model are stochastic and this allows the quantification of parameter risk. The HP-VAR and HP-Bayesian VAR models are fitted and compared using Australian mortality data. Parameter uncertainty is shown to be an important factor to include in mortality models when assessing mortality risk. The HP-BVAR model is shown to outperform the HP-VAR model in terms of parsimony, goodness of fit and forecasting of mortality rates. At the same time, more information about future mortality rate uncertainty is provided by the model.

7 Acknowledgement

References

Baltagi, B. (2002). *Econometrics* (3rd ed.). Springer-Verlag, New York.

Box, G., & Jenkins, G. (1976). *Time Series Analysis: Forecasting and Control (Wiley Series in Probability and Statistics)*. Holden-Day San-Francisco.

Brandt, P. T., & Freeman, J. R. (2006). Advances in Bayesian Time Series Modeling and the Study of Politics: Theory Testing, Forecasting, and Policy Analysis. *Political Analysis*, *14*, 1-36.

Cairns, A. J., Blake, D. P., Dowd, K., Coughlan, G., & Epstein, D. (2007). A Quantitative Comparison of Stochastic Mortality Models Using Data from England & Wales and the United States. *SSRN eLibrary*. Available from SSRN: http://ssrn.com/abstract=1340389

Congdon, P. (1993). Statistical Graduation in Local Demographic Analysis and Projection. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *156*(2), 237-270.

Dellaportas, P., Smith, A. F. M., & Stavropoulos, P. (2001). Bayesian analysis of mortality data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *164*(2), 275-291.

Denuit, M., & Frostig, E. (2007). Association and heterogeneity of insured lifetimes in the Lee–Carter framework. *Scandinavian Actuarial Journal*, *2007*(1), 1 - 19.

Denuit, M., & Frostig, E. (2008). *First-Order Mortality Basis for Life Annuities.* Palgrave Macmillan. (Nature Publishing Group Metadata Repository [http://www.nature.com/oai/request] (Switzerland) ER)

Denuit, M., & Frostig, E. (2009). Life insurance mathematics with random life tables. *North American Actuarial Journal*, *13*(3), 339-355.

Doan, T., Litterman, R., & Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, *3*(1), 1 - 100.

Drèze, J. H., & Richard, J. F. (1983). Bayesian analysis of simultaneous equation systems. In (Vol. 1, p. 517-598).

Forfar, D. O., & Smith, D. M. (1987). The changing shape of English Life Tables. *Transactions of the Faculty of Actuaries*(40), 98-134.

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press.

Hartmann, M. (1987). Past and recent attempts to model mortality at all ages. *Journal of Official Statistics*, *3*(1), 19-36.

Heligman, L., & Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, *107*, 49-80.

Henry, K. (2009). *Australia's future tax system - The retirement income system: Report on strategic issues.*

Human Mortality Database. (2010, January). University of California, Berkeley (USA) Max Planck Institute for Demographic Research (Germany). Available from www.mortality.org or www.humanmortality.de

Joiner, A. (2001). Monetary Policy Effect in an Australian Bayesian VAR Model.

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

Lee, R. D. (1992). Stochastic demographic forecasting. *International Journal of Forecasting*, *8*(3), 315-327.

Lee, R. D., & Carter, L. R. (1992). Modeling and Forecasting U. S. Mortality. *Journal of*

the American Statistical Association, *87*(419), 659-671.

Lee, R. D., & Skinner, J. (1999). Will Aging Baby Boomers Bust the Federal Budget? *Journal of Economic Perspectives*, *13*(1), 115-140.

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

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis* (1st ed.). New York.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. New York.

McNeil, D. R., Trussell, T. J., & Turner, J. C. (1977). Spline Interpolation of Demographic Data. *Demography*, *14*(2), 245-252.

McNown, R., & Rogers, A. (1989). Forecasting Mortality: A Parameterized Time Series Approach. *Demography*, *26*(4), 645-660.

Pedroza, C. (2006). A Bayesian forecasting model: predicting U.S. male mortality. *Biostat*, *7*(4), 530-550.

Peng, R. (2008, 3 31). A Method for Visualizing Multivariate Time Series Data. *Journal of Statistical Software, Code Snippets*, *25*(1), 1–17. Available from `http://www.jstatsoft.org/v25/c01`

Reichmuth, W. H., & Sarferaz, S. (2008). *Bayesian Demographic Modeling and Forecasting: An Application to U.S. Mortality.*

Risk Management Solutions, I. (2010a). *Longevity Risk.* Available from `http://www.rms.com/Publications/Longevity_Risk_brochure.pdf`

Risk Management Solutions, I. (2010b). *RMS Unveils New Approach To Quantifying Longevity Risk.* Available from `http://www.rms.com/newspress/PR_071210_LongevityRisk.asp`

Robertson, J. C., & Tallman, E. W. (1999). Vector autoregressions: forecasting and reality. *Economic Review*(Q1), 4-18. Available from `http://ideas.repec.org/a/fip/fedaer/y1999iq1p4-18nv.84no.1.html`

Rogers, A. (1986). Parameterized Multistate Population Dynamics and Projections. *Journal of the American Statistical Association*, *81*(393), 48-61.

Sharrow, D. J., Clark, S. J., Collinson, M. A., Kahn, K., & Tollman, S. M. (2010). *The Age-Pattern of Increases in Mortality Affected by HIV: Bayesian Fit of the Heligman-Pollard Model to Data from the Agincourt HDSS Field Site in Rural Northeast South Africa.* University of Washington.

Sherris, M., & Njenga, C. N. (2009). Longevity Risk and the Econometric Analysis of Mortality Trends and Volatility. *SSRN eLibrary*. Available from SSRN: `http://ssrn.com/abstract=1458084`

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

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

Summers, P. M. (2001). Forecasting australia's economic performance during the asian

crisis. *International Journal of Forecasting*, *17*(3), 499-515.

Thompson, P. A., Bell, W. R., Long, J. F., & Miller, R. B. (1989). Multivariate Time Series Projections of Parameterized Age-Specific Fertility Rates. *Journal of the American Statistical Association*, *84*(407), 689-699.

Waggoner, D. F., & Zha, T. (1999). Conditional Forecasts in Dynamic Multivariate Models. *The Review of Economics and Statistics*, *81*(4), 639-651.

Wilmoth, J., Andreev, K., Jdanov, D., & Glei, D. (2007, May). *Methods Protocol for the Human Mortality Database* (Tech. Rep.). Human Mortality Database.

Wong-Fupuy, C., & Haberman, S. (2004). Projecting Mortality Trends: Recent Developments in the United Kingdom and the United States. *North American Actuarial Journal*, *8*(2), 56-83.

Zivot, E., & Wang, J. (2006). *Modelling Financial Time Series with S-plus* (Second ed.).