Institute of Actuaries of Australia
BIENNIAL CONVENTION 2011
BEYOND THE MANDATE
10 – 13 APRIL 2011 • HILTON SYDNEY

Modeling Mortality with a Bayesian Vector Autoregression

Carolyn Njenga & Prof Michael Sherris

Outline

1 Introduction

- Australian Mortality
- Model and Parameter Risk
- Motivation

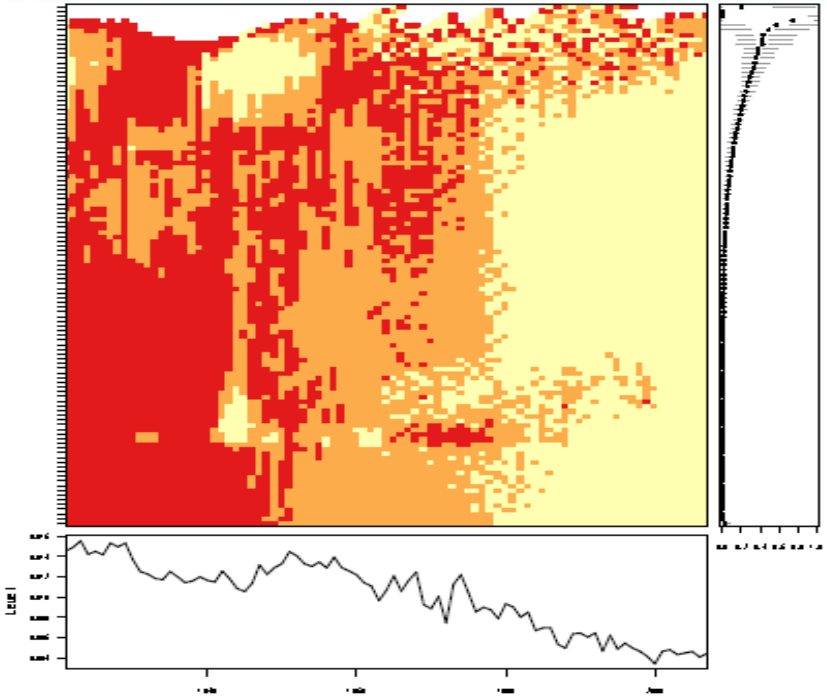
2 Brief Literature Review

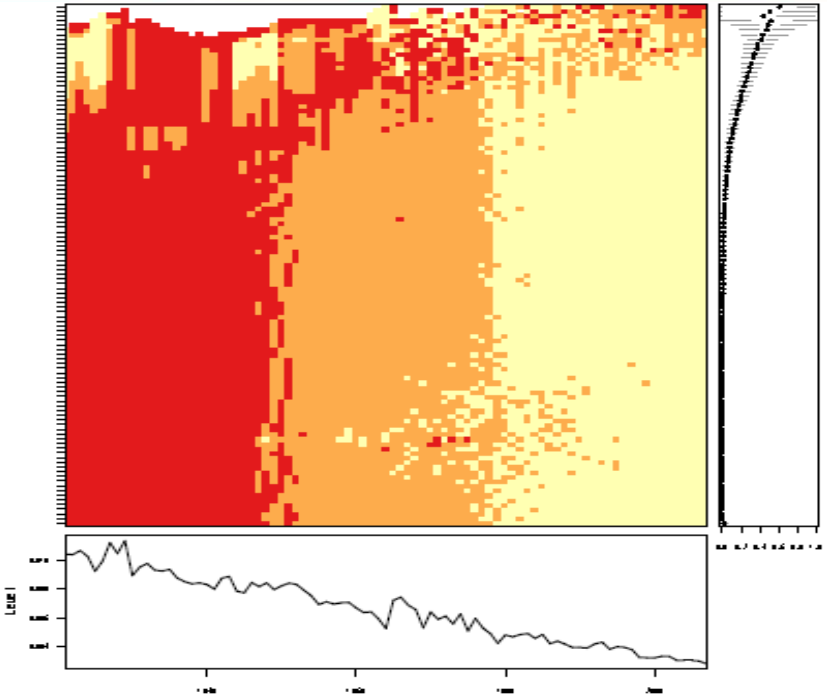
3 HP-BVAR Model

- Methodology
- Results

4 Conclusion

AUSTRALIAN MORTALITY





Declining Mortality Rates (Australia) from 1921 to 2007 for ages 0 to 110+ from bottom to top for males (left) & females (right). Red=High mortality, Orange=Moderate mortality, Yellow=Low mortality, White=Missing Data. Panel on right of the main plots has boxplots of data in each time series. The bottom panel has median values across time series of mortality rates for each time point.

Aim: Model probability of death, q_x

Principle of Parsimony (Jeffreys, 1961)

An inverse relationship between prior degree of belief in model and number of parameters in model.

Model Risk:

Find model that best satisfies some criterion e.g. BIC using the best estimate of its <u>parameter set</u>, θ. Best fit often $\Rightarrow$ overparameterization.

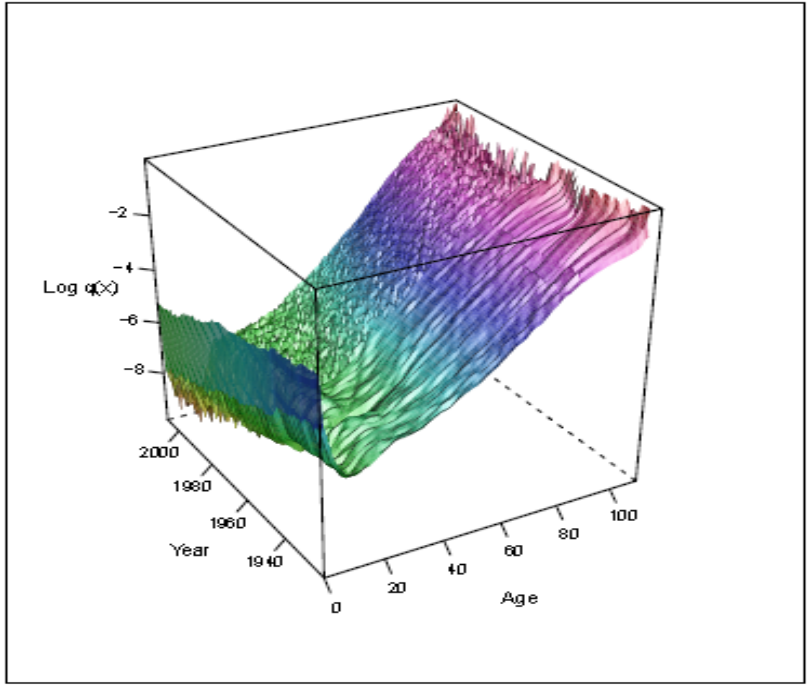
Parameter Risk:

Assume model is correct and θ has posterior distribution $f(\theta|\text{past data})$ then derive posterior distribution for q_x given the past data.

Not much done on parameter risk.

Significant in tail distribution of q_x

Shape of logarithm of probability of death, ln q_x



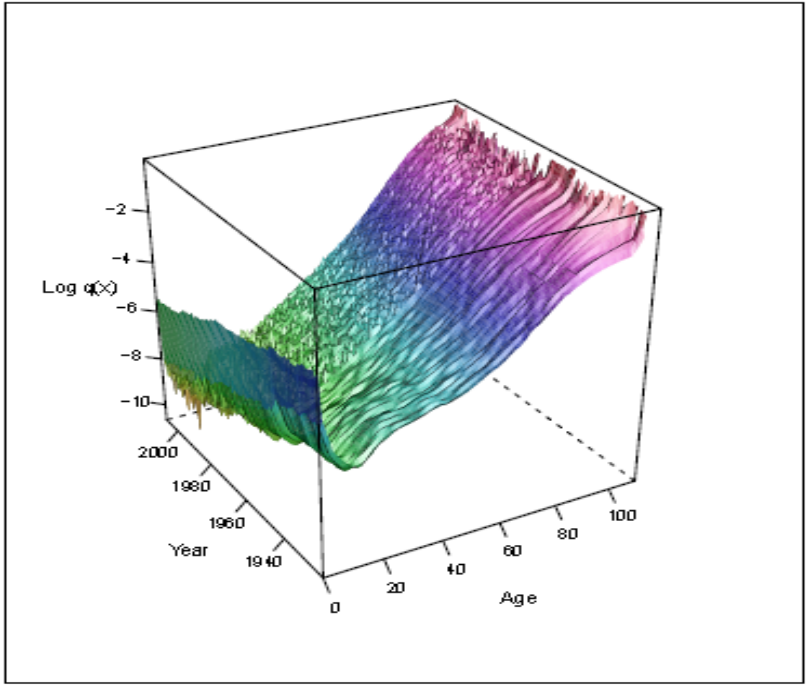


Figure: Log q_x for Australian Males (left) and Females (right). Shape remains similar.

Motivation for using Heligman and Pollard (1980):

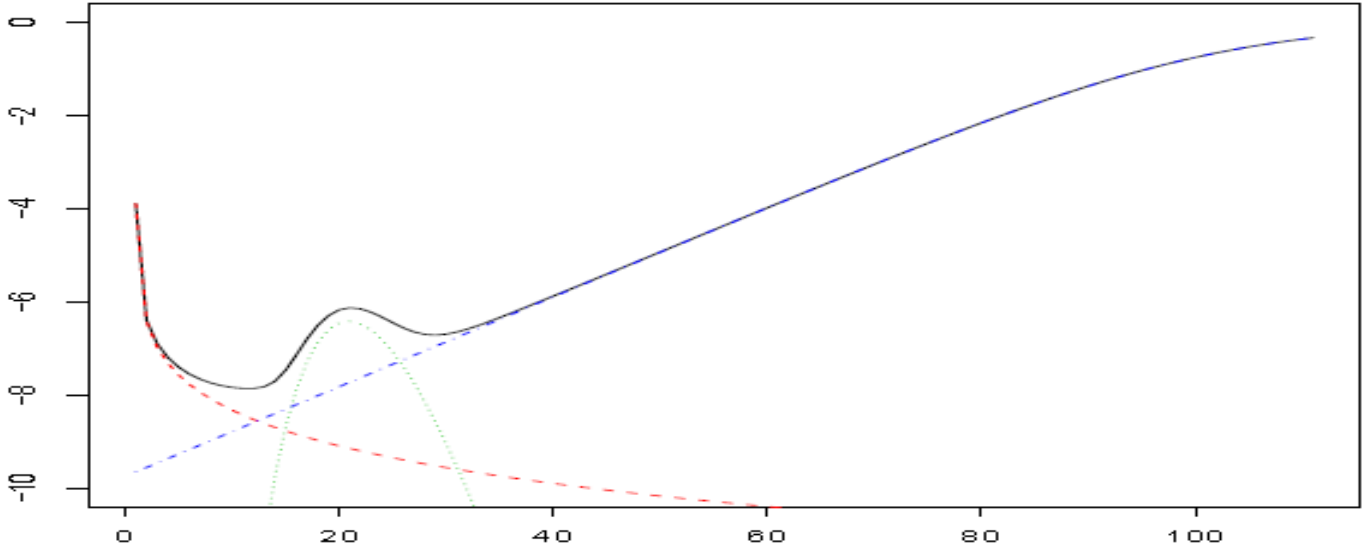


Fig: Preserves shape of mortality. Change values of parameters to reflect shift in mortality levels.

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

$\theta'_t = (A_t, B_t, C_t, D_t, E_t, F_t, G_t, H_t)$ is the set of parameters at time t.

Advantages of Heligman and Pollard (1980)'s First Law of Mortality

- Each of the parameters has a straightforward interpretation.
- Model mortality at a fixed point in time for all ages.
- Extract the parameters at a sequence of points in time.
- Model the parameters as a time series.
- Used as basis for age-period tables. e.g. Forfar & Smith (1987), McNown and Rogers (1989), Dellaportas, Smith and Stavropoulos (2001) and Njenga and Sherris (2009).

BRIEF LITERATURE REVIEW

McNown and Rogers (1989):

- Fitted Heligman-Pollard Mortality model to annual series of age-specific death rates in the United States then Extrapolated the fitted parameters using univariate time-series methods.
- The time series of the parameters exhibited highly non-stationary behaviour and this made it difficult to select a proper model for extrapolation.

Gap in this work

- Uses univariate time series models.
- Assumes independence of parameters.

Dellaportas, Smith, and Stavropoulos (2001):

- Heligman-Pollard first law is too restrictive. Assume $p(\theta)$ is the prior distribution for θ then obtain Posterior Distribution of Heligman Pollard Parameters $p(\theta|q_x) \propto p(\theta)p(q_x)$
- Use Markov Chain Monte Carlo methods to fit the parameters and update θ at each iteration.

Gap in this work

- Use Bayesian techniques to estimate the parameters but do not model the time evolution of the parameters.

HP-BVAR MODEL

Description of the HP-BVAR Model

Model Time Evolution of Parameters of Heligman-Pollard Model Using a Bayesian VAR model (in next section).

Highlights:

- Based on Heligman and Pollard (1980) Model⇒Maintains Shape, easy to interpret etc
- Multivariate Time Series⇒Captures Correlation in the Parameters
- Bayesian⇒Uses prior information to update parameter estimates

METHODOLOGY

Unrestricted VAR(p): No uncertainty in $\upsilon = (c, \Omega_1, \ldots, \Omega_p)$

$$\theta_t = c + \sum_{l=1}^{p} \Omega_l \theta_{t-l} + \epsilon_t \quad \text{or} \quad \theta = X\upsilon + \xi \tag{2}$$

Equation i:

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

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.

Cointegration: Stationary (I(0))Linear Combination of Non-stationary (I(1)) variables.

VAR(p) assumes θ_t are I(0).

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).

Bayesian VAR - Basic Concept

- VAR models impose no theoretical restrictions to guide model specification $\Rightarrow$ not parsimonious (Litterman, 1986; Zivot & Wang,2006; Sims & Zha, 1998; Robertson & Tallman, 1999; Brandt & Freeman, 2006; Baltagi, 2002).
- v is considered to be a fixed quantity in unrestricted VAR(p).
- VAR is often overparameterized.
- Bayesian VAR (BVAR) has uncertainty regarding the distribution of v, reflected in the prior & resulting posterior distribution of v.
- B-VAR has prior which assumes some coefficients to zero, changes from zero only if the coefficient has information.

Sims and Zha (1998) (Normal-Wishart) Prior:
Assume a prior distribution for $v = (c, \Omega_1, \Omega_2)$

$$\overline{v}^{PRIOR} = \mathbb{E}[v] = \mathbb{E}\begin{bmatrix} c & \Omega_1 & \Omega_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{3}$$

Assumes each variable follows a random walk.

Gives a distribution for the coefficients.

Under Normal-Wishart Prior, posterior is also Normal-Wishart:

Bayesian VAR - Sims and Zha (1998)

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 imposes 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.

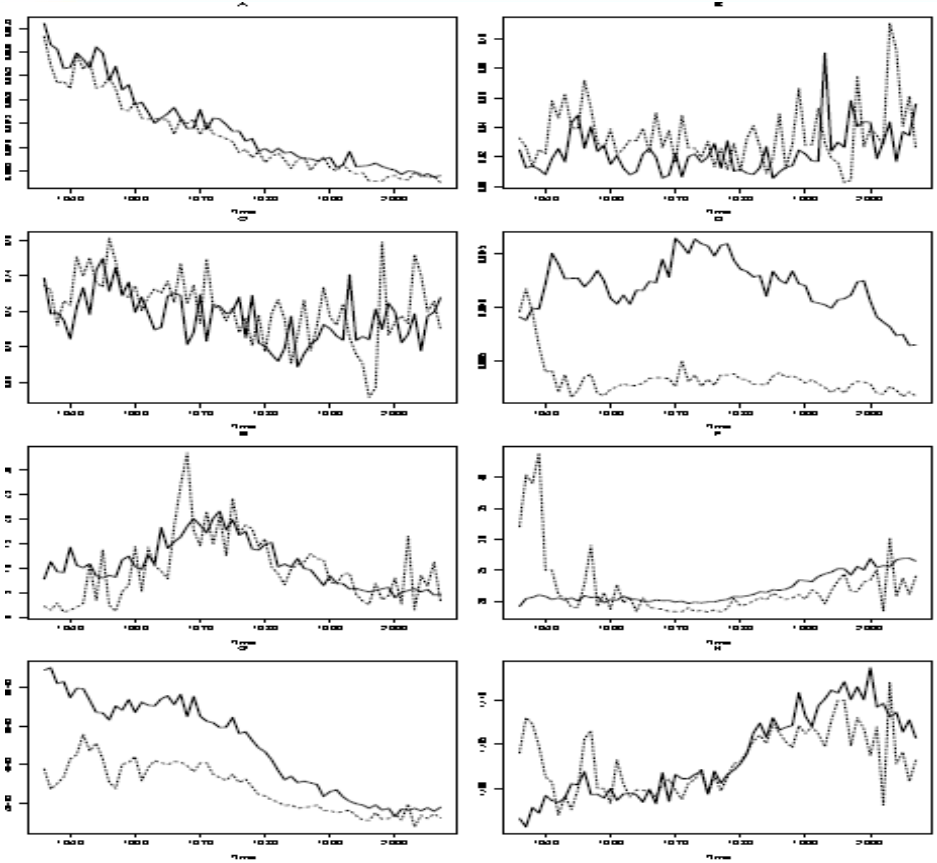
In particular the BVAR model allows the standard deviations of the coefficients to decrease as the lags increase.

Allows for non-stationary & cointegrated variables.

Summary of Methodology

- Estimate parameters of Heligman-Pollard Model, θ
- Test for unit-roots/stationarity
- Determine optimal lag length for VAR(p)
- If there are unit roots test for cointegration
- Use B-VAR(p) to calculate posterior density of θ i.e. $f(\theta|\text{past data})$
- Estimate $\hat{q}_{x,\tau}$

RESULTS



	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	****

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	****

Correlation & Significance Males (top) & Females:
upper diagonal part contains correlation coefficient estimates;
lower diagonal part has corresponding p-values.
Parameters with significant correlation have p-values <0.001.

Figure on left: HP Parameters for Australia Male (Solid Line) & Females (Dotted Line) from 1921 to 2007.
Stationarity: Test for difference stationarity find all I(1) except B_f.

- Using various information criteria, p = 2 $\Rightarrow$ fit a VAR(2).

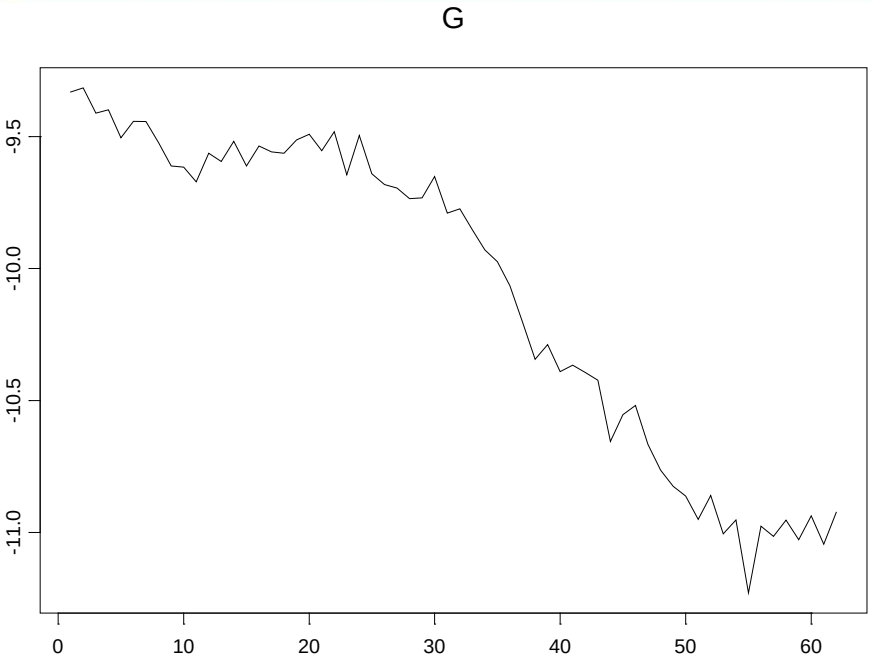
- There is 1 cointegration relation at 95% confidence level for both Males and Females

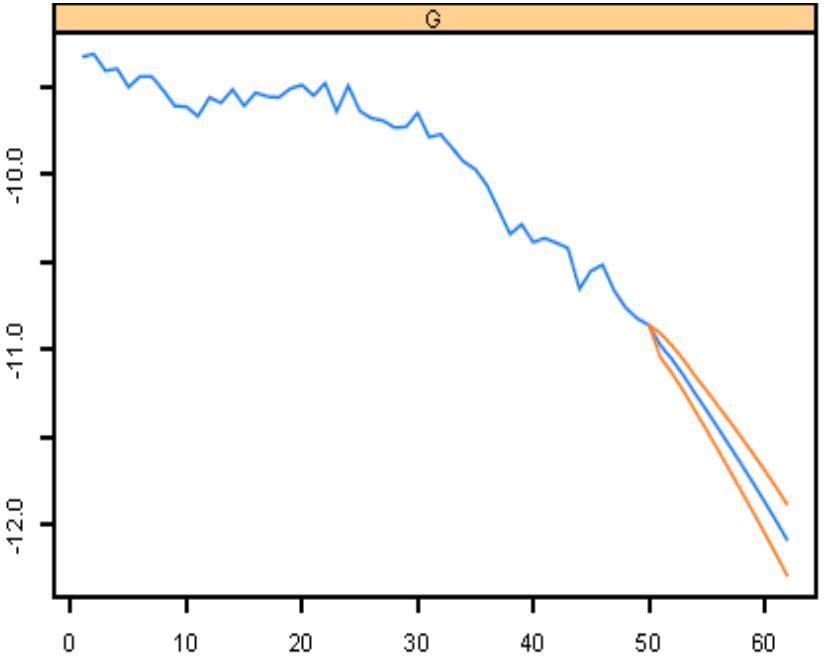
- Fit a BVAR(2) with allowance for unit roots and cointegration.

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.

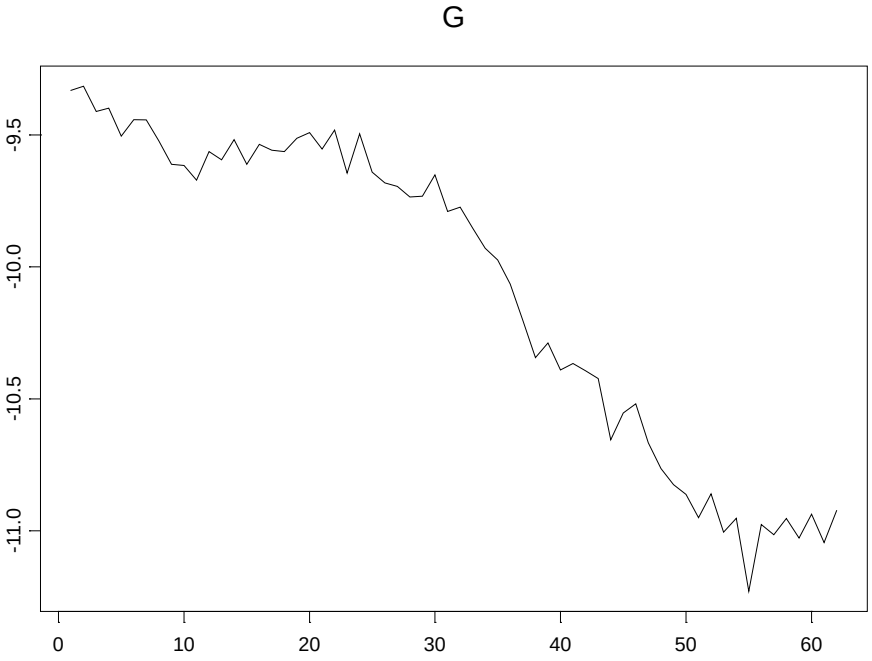
For longevity risk the uncertainty at the older ages is of most interest.

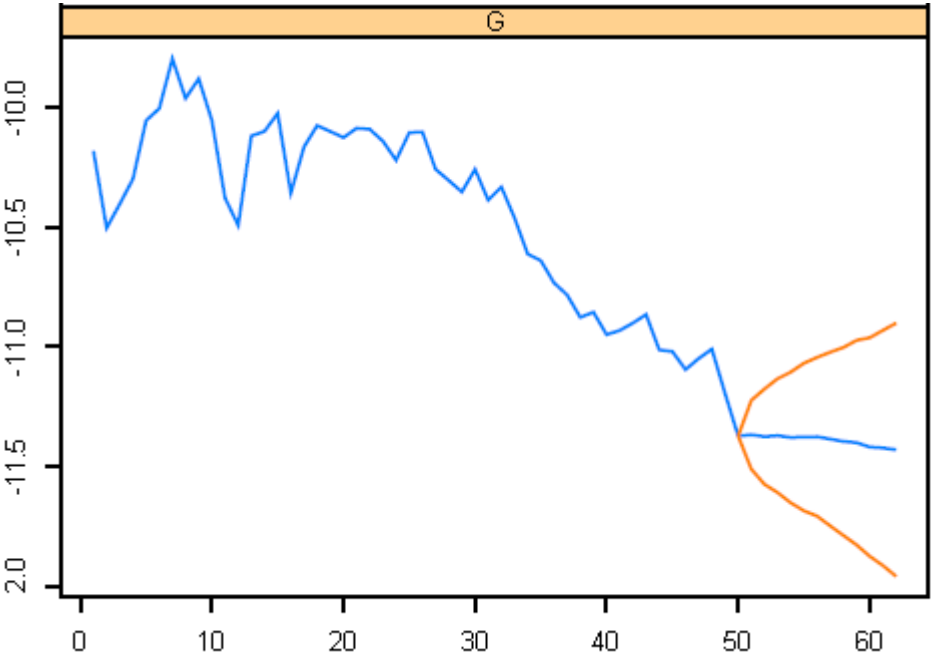
Parameter G is the base level of old age mortality.





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 & over the same horizon has not reached a long run distribution for the parameters.

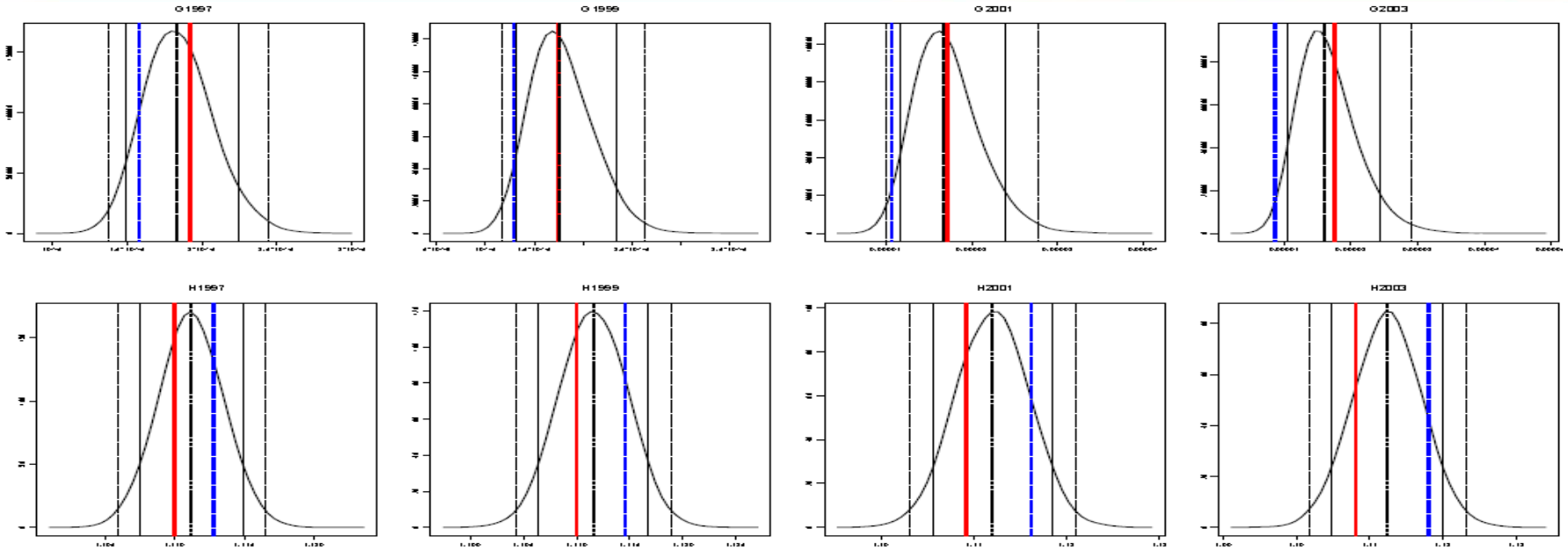
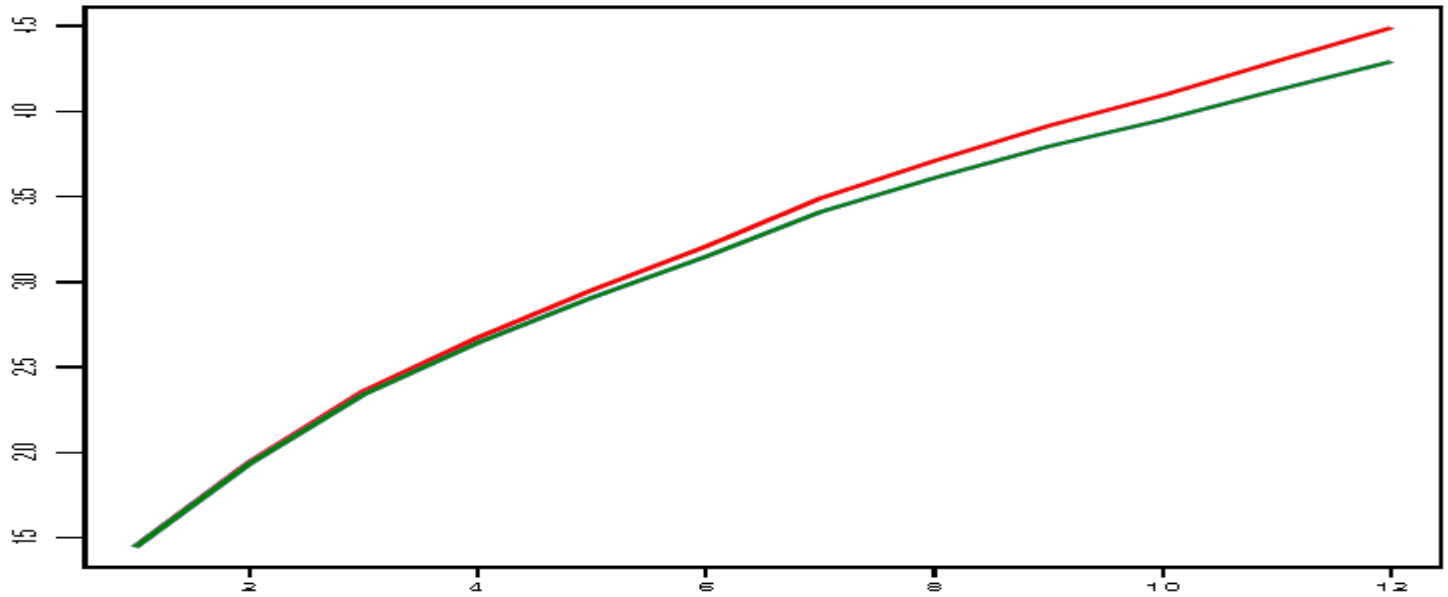


Fig: Parameter Risk in G and H (Males) for the predicted values for some odd years 1996-2007. Red (Thick Solid Line)=Observed, Black (Thick Broken Line)= BVAR Estimate, Blue (Thick Dot-Dashed Line)=VAR Estimate. 95% & 99% confidence intervals are indicated by the thin dotted and thin broken line respectively.

HP-BVAR has more accurate predictions.

Parameter (G) C.I. Ratio $= \dfrac{\text{Width Confidence Interval BVAR}}{\text{Width Confidence Interval VAR}}$



Red = 99% C.I. Width Ratio and Green = 90% C.I. Width Ratio.

Ratio greater than 1 $\Longrightarrow$ BVAR confidence interval is wider than VAR confidence interval.

99% Confidence intervals of old age predicted ln q_x

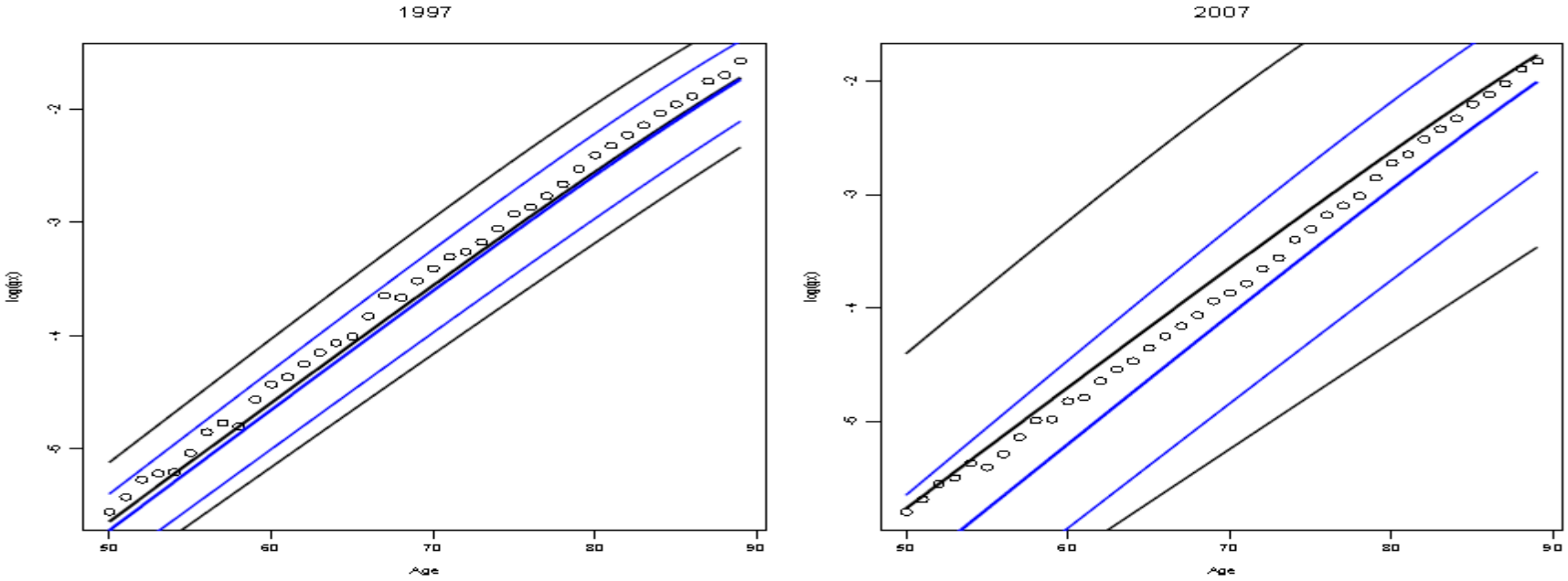


Fig: Predictions of ln q_x (Males) for selected years (1997 & 2007) VAR C.I. (BLUE) & BVAR C.I. (BLACK). C. I. Width increases with age. Thick line = Best Estimate. Thin Line = C. I.

HP-BVAR has wider confidence intervals and is more accurate than HP-VAR.

$$\text{RMSE} = \sqrt{\frac{1}{12}\sum_{\tau=1996}^{2007}(\hat{q}_{x,\tau} - q_{x,\tau})^2}$$

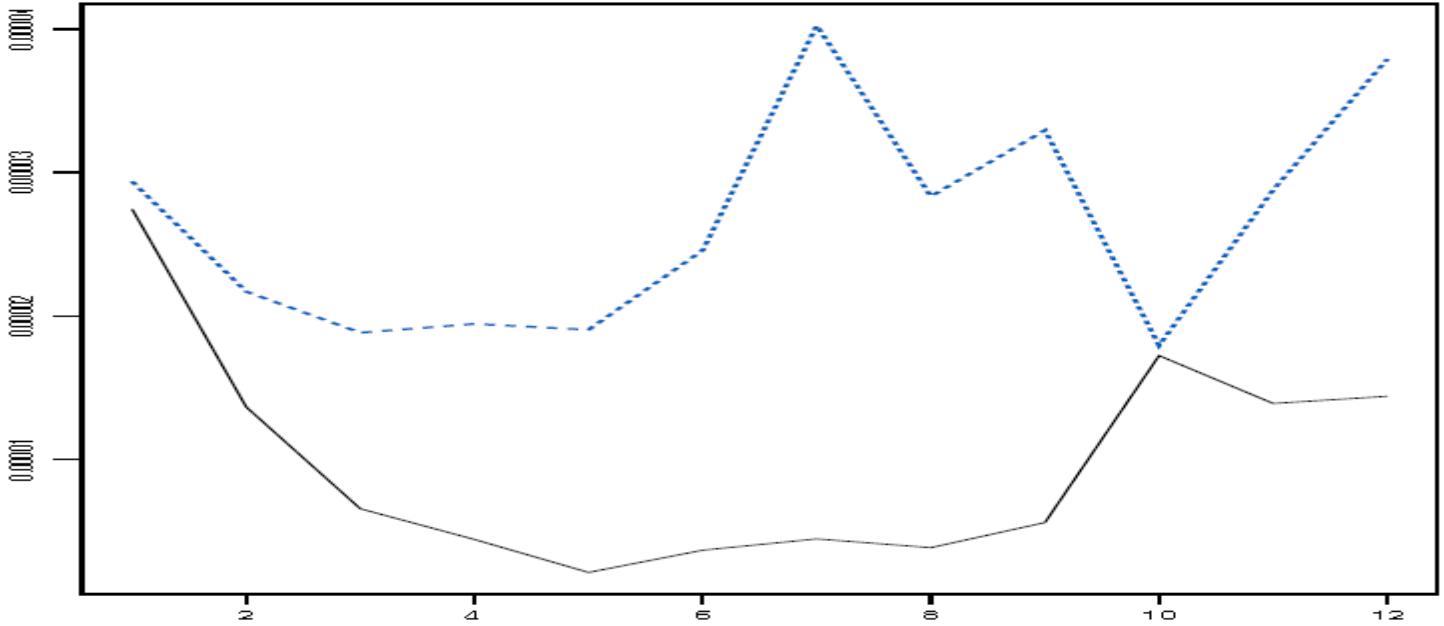


Fig: RMSE of Predicted q_x , x = 0, … , 90 from 1996-2007 using VAR (BLUE) and BVAR (BLACK).

Parameter uncertainty is very significant and prediction intervals that take into account parameter risk are significantly wider than the case where this risk is not included.

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.

The HP-BVAR model outperforms the HP-VAR model in terms of parsimony, goodness of fit and forecasting of mortality rates.

Further, more information about future mortality rate uncertainty is provided by the HP-BVAR model.

Conclusion

Model parameters of Heligman-Pollard Model including their correlation using econometric models.

Model with parameter risk gives more accurate predictions.

Model with parameter risk has wider confidence intervals.

Including parameter risk generates a parsimonious model.

Inherits Heligman and Pollard (1980) model problems e.g. female maternal/accident hump modal age (F_f) $\Rightarrow$ poor performance for females.

THANK YOU!