Long-Range Forecasts of Mortality and Life Expectancy Using Bayesian Vector Autoregressions

Javier Meseguer
Social Security Administration
Office of Policy - Division of Economic Research

This report provides a summary of ongoing research undertaken by the author. The analyses, opinions and findings contained in this report represent the views of the author and are not necessarily those of the Social Security Administration. Any errors or omissions are solely the author's responsibility.

Introduction

Using time-series statistical techniques to produce population forecasts has become standards practice among demographers and other social scientists. Unlike the traditional deterministic "scenarios" methodology, stochastic models are capable of generating projections that are consistent with a probabilistic representation of uncertainty. In this context, the approach to mortality forecasting pioneered by Ronald Lee, Lawrence Carter and many others following on their footsteps has received much attention in recent years. This report explores the use of Bayesian vector autoregressive (BVAR) models as an alternative to the Lee-Carter method.

The Bayesian framework offers a number of significant advantages over the traditional sampling theory approach. First, Bayesian inference derives from an exact finite-sample probability distribution, regardless of sample size. Hence, there is no need to rely on asymptotic theory approximations that can be poor in small samples.

Second, by providing a formal mechanism for the inclusion of available prior information, it is possible to introduce model restrictions that better accommodate the main empirical regularities characterizing all-cause mortality. This, in turn, can translate into improved forecast performance.

Finally, since the Bayesian paradigm treats a model's parameters as random variables, the generated predictive distributions account for variation due to both the sample and parameter outcomes. This ability to coherently incorporate parameter uncertainty can be particularly relevant in the context of demographic applications, where for example, very long forecast horizons are required to evaluate the solvency of a pension plan.

Time-Series Specifications for Mortality Forecasting

The so called Lee-Carter model,[8] postulates the logarithm of a set of age-specific mortality rates as the linear function of a time-varying index of the general level of mortality

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

for $i = 1,\ldots,m$, and $t = 1,\ldots,T$, where the term $y_{i,t}$ denotes mortality at age i and time t. The age-specific set of parameters α_i describes the average shape of the logarithmic mortality rates, while the slope coefficients β_i determine both the direction and magnitude by which mortality at every age varies with the index k_t.

Subject to appropriate identifying constraints, least square estimates of the parameters in the Lee-Carter model can be obtained by applying singular value decomposition to the matrix of centered logarithmic death rates. Then, the estimates of the mortality index are treated as a univariate time-series that is projected into the future. In most applications, a random-walk with drift is found to provide a good empirical fit for k_t. Finally, the resulting forecasts of the mortality index are "plugged" back into equation (1), in order to recover a probability distribution of the future death rates for each age category.

Several time-series econometric specifications have been suggested as alternatives to the Lee-Carter method.[4] One such possibility is to model the first difference of logarithmic mortality as a p^{th}-order autoregressive process AR(p)

$$\Delta\log(y_{i,t}) = c_i + \sum_{s=1}^{p} \phi_{s,i}\Delta\log(y_{i,t-s}) + e_{i,t}. \qquad (2)$$

In this case, future changes in the individual mortality rates are determined by their own past values plus a random disturbance term $e_{i,t}$. The m age-specific series are estimated within a system of seemingly unrelated regression equations (SURE), in order to accommodate the significant degree of contemporaneous correlation characterizing all-cause mortality data. A variant of this framework is for instance used by the Congressional Budget

Office's stochastic model of long-term trust fund finances.[3]

A further extension of the model in equation (2) is a quasi-vector or QVAR(p) autoregression.[4] The latter includes an index of the general level of logarithmic mortality as part of each univariate autoregression, where the index itself is a function of all the age-specific death rates. Of course, an even more general specification involves a full vector autoregressive (VAR) process, in which the lagged values of all m mortality series appear in the right side of each equation

$$\Delta \log(y_{i,t}) = c_i + \sum_{i=1}^{m} \sum_{s=1}^{p} \phi_{s,i} \Delta \log(y_{i,t-s}) + e_{i,t}. \quad (3)$$

However, estimation of an unrestricted VAR model is not feasible in this context, without a priori imposing severe restrictions on the autoregressive coefficients.

One potential problem with unrestricted VAR processes is the fact that the number of parameters increases geometrically with the set of variables being modeled. Hence, as m increases, the number of autoregressive coefficients involved is eventually guaranteed to outstrip the number of available observations. For instance, a VAR(p) model with p = 1 lag and m = 21 mortality age groups comprises 462 autoregressive coefficients. This problem is referred to in the literature as overfitting (when too many parameters are estimated relative to the number of data points). The result is usually imprecise parameter estimates that often translate into poor out-of-sample forecast performance.

Bayesian vector autoregressive (BVAR) models can successfully overcome the overfitting problem. This is accomplished through the concept of shrinkage, by imposing a priori stochastic restrictions on the autoregressive coefficients, so that their value is determined by a handful of so-called hyperparameters. The following sections describe a Bayesian prior that provides the researcher with a great deal of control over the relative influence that the lags of every series have on each equation.

Empirical Regularities in All-Cause Age-Specific Mortality

Mortality for U.S. males and females combined and 21 age categories (age 0, ages 1-4, ages 5-9,…, ages 90-94, and ages 95+), is used to fit the BVAR models estimated in this report. The data set was constructed from the period life tables published by the Social Security Administration's Office of the Chief Actuary.[1] The sample is restricted to the time frame (1928-2001), so as to exclude the excessive volatility displayed by the series in earlier years (particularly during the 1918 Spanish influenza pandemic). Figures 1, 2 and 3 illustrate a number of features that are common to all-cause age-specific mortality data.

Figure 1: Surface of the logarithmic age profile of mortality (1928-2001).

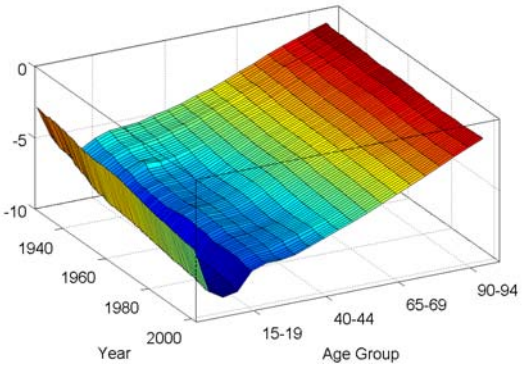


Figure 2: Contours of the logarithmic age profile of mortality (1928-2001).

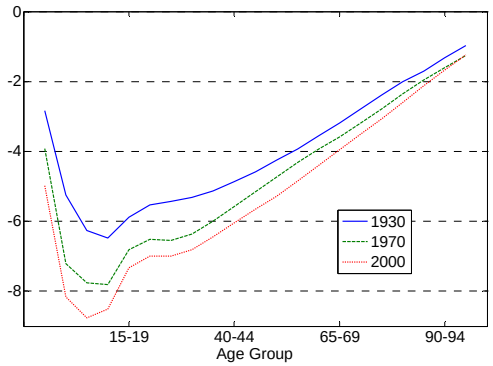
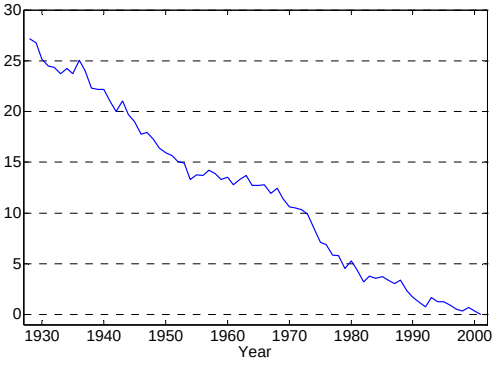


Figure 3: Lee-Carter index of logarithmic mortality (1928-2001).



One evident characteristic in the data is the regularity in the shape of logarithmic mortality over the ages, resembling an "elongated v." For any given period, mortality declines smoothly from birth until about ages 10-14 and then, it increases almost linearly for the remaining ages until death. Moreover, for the portion of the sample roughly involving the second half of the 20th century, the death rates experience a sharp increase associated with motor-vehicle fatalities in the 15-19 age group, often referred to as "the accident hump." This can be seen in the surface and contours of the age profile of logarithmic mortality in Figures 1 and 2, respectively.

A second feature of the age profile in all-cause mortality is a downward trend. That is, while mortality across the age groups maintains a fairly regular shape, it also shifts downward over time. This is illustrated in Figure 3, which displays the estimated Lee-Carter mortality index. Evidently, the general level of logarithmic mortality follows a downward trend that is approximately linear over the time period entertained. This empirical regularity is confirmed by the high degree of cross-correlation in the levels of logarithmic mortality, as the death rates move closely together.

Table 1 presents the 21×21 sample correlation matrix for the first difference of logarithmic mortality. Notice how in general, the degree of correlation among the age groups tends to decrease with the order of temporal adjacency. For instance, the sample correlation between age 0 and ages 1-4 is 0.61 (second entry in the first column of Table 1). On the other hand, the correlation in $\Delta\log(y_{i,t})$ between age 0 and ages 95+ is only 0.19 (last entry in the first column of Table 1). This simply implies that age groups that are temporally closer to one another have greater ability at predicting each others movements than those that are far apart. BVAR models can be calibrated to exploit this feature, by having each lag carry a weight that is proportional to the correlation structure in the data.

Bayesian Vector Autoregressions and the Minnesota Prior

Let $L(\theta \mid Y)$ denote a probability model (typically in the form of a likelihood function), where Y represents the observable data and θ is a set of unknown parameters to be estimated. From a Bayesian perspective, a complete model also requires the specification of a prior density on the parameters $\pi(\theta)$. Bayesian inference then proceeds by conditioning on the observables through Bayes's theorem, in order to derive the so-called posterior density

$$\pi(\theta \mid Y) = \frac{\pi(\theta)\, L(\theta \mid Y)}{f(Y)} \propto \pi(\theta)\, L(\theta \mid Y) \qquad (4)$$

The prior density represents any knowledge the researcher has about θ prior to observing the data and provides the means to accommodate any available non-sample information. This information might be guided by statistical theory considerations, empirical evidence from previous analysis or simply, the subjective judgment of the researcher. On the other hand, Bayes's theorem can be thought of as the mechanism through which the sample information contained in the likelihood function modifies the knowledge about θ embodied in the prior. Once the posterior density is derived, it can be used to produce optimal point estimates of the parameters and to construct credible intervals and regions of the parameter space corresponding to some posterior probability.

One problem with vector autoregressions is overfitting, due to the large number of parameters involved. A VAR(p) model with m equations comprises a total of pm^2+m coefficients, in addition to the $m(m+1)/2$ distinct elements in the residual covariance matrix. Overfitting is successfully overcome by the random-walk or Minnesota prior, which derives its name from the research undertaken at the University of Minnesota and the Federal Reserve Bank of Minneapolis.[5] Originally developed as an alternative to the predictions generated by large-scale structural macroeconomic models, BVAR systems under the Minnesota prior have accumulated an impressive forecast record, becoming one of the most widely applied Bayesian time series forecasting approaches to date.[6]

Given an arbitrarily long lag p, the Minnesota prior imposes inexact (stochastic) prior constraints on the model's parameters, based on two premises:

1. The movement over time of the variables being modeled is assumed to be well approximated by a random-walk around an unknown deterministic component.

2. Recent values are likely to contain more useful information about a variable's future direction than more distant ones.

The first prior restriction is well suited to the purposes of this paper, given that most applications

of the Lee-Carter method find that a random-walk with drift fits well the general level of mortality. The second constraint is implemented in practice by increasing the prior probability that the coefficients in the model are closer to zero as a function of lag length.

Formally, let a_{ijp} represent the parameter associated with variable j in equation i at lag p, and let σ_{ijp} denote its corresponding prior standard deviation. The Minnesota prior specifies the individual priors of a_{ijp} as independent normal distributions with mean equal to one for the first lag of the dependent variable in each equation, and zero for all other coefficients and all other lags

$$a_{ijp} \sim \begin{cases} N(1, \sigma_{ijp}^2), & \text{for } i = j, \text{ and } p = 1. \\ N(0, \sigma_{ijp}^2), & \text{otherwise.} \end{cases} \quad (5)$$

This, of course, centers the prior density for the i^{th} equation around the random-walk specification

$$y_{i,t} = c_i + y_{i,t-1} + e_{i,t}. \quad (6)$$

The standard deviations of the autoregressive coefficients in the Minnesota prior are defined in terms of four hyperparameters (denoted by λ_1, λ_2, λ_3 and λ_4), as follows

$$\sigma_{ijp} = \begin{cases} \dfrac{\lambda_1}{p^{\lambda_3}}, & \text{for } i = j. \\ \dfrac{\lambda_1 \lambda_2(i,j)}{p^{\lambda_3}} \left(\dfrac{s_i}{s_j} \right), & \text{for } i \neq j. \\ \lambda_4 s_i, & \text{for } c_i. \end{cases} \quad (7)$$

Notice that for a given equation, three cases can be distinguished: (1) the prior standard deviations on the lags of the dependent variable; (2) those associated with the autoregressive coefficients of the remaining variables; and (3) the standard deviations of the intercept parameters c_i.

Hyperparameter λ_1 represents the prior standard deviation on the first lag of variable i in equation i. For a given equation, decreasing its value has the effect of shrinking the prior density on the coefficient of the first lag of the dependent variable toward one, and the prior densities on all other parameters toward zero. For this reason, λ_1 is often referred to as the "overall tightness" of the Minnesota prior, as it determines how tightly the random-walk specification is imposed a priori.

The second hyperparameter (λ_2) affects the prior standard deviation of variable j in equation i, relative to λ_1. Setting $0 < \lambda_2 < 1$ shrinks the prior densities of all variables other than the dependent variable toward zero. This implies that for a given equation, variation in the dependent variable is explained to a greater extent by its own lags than by any other variable. By contrast, setting $\lambda_2 = 1$ treats the lags of the dependent and all other variables equally.

Two limiting cases are of particular interest. When $\lambda_2 \rightarrow 0$ each equation follows the functional specification in equation (2), which is a univariate autoregression. In addition, as λ_1 also approaches zero, each equation follows a random-walk with drift instead.

It is important to emphasize that the notation $\lambda_2(i,j)$ indicates the possibility of specifying alternative degrees of shrinkage (values of λ_2) for different variables. In other words, for a given equation, it is possible to allow the lag of each series to assume a different prior weight with which to impact the current value of the dependent variable. Such degree of flexibility can in fact be relied upon to exploit the temporal contiguity relationships in the correlation structure of the data in Table 1. This is accomplished by assigning a smaller value to $\lambda_2(i,j)$ the lower the correlation is between the i^{th} and j^{th} mortality series.

The term p^{λ_3} in the Minnesota prior standard deviations has the effect of shrinking all the coefficients toward their prior means as a function of lag length p. In this case, the hyperparameter λ_3 represents the rate of lag-decay, where a value $\lambda_3 = 1$ implies a harmonic rate of decay. An increase in λ_3 shrinks higher-order lags toward zero more quickly, giving greater importance to more recent lags than to more distant ones.

Finally, hyperparameter λ_4 determines the prior standard deviations associated with the intercept parameters c_i, as well as any exogenous variables introduced in the VAR system. Decreasing λ_4 tightens the prior densities on these parameters around their mean. Typically, in most applications, the Minnesota prior is made noninformative on the intercept coefficients by selecting an arbitrarily large value for λ_4. In addition, the term (s_i/s_j) serves as a scale factor that accounts for variation due to differences in the measurement units of the variables. An "empirical Bayes" approach is followed in practice, estimating s_i from the data as

the standard error in a *p*-lag autoregression of variable *i*.

Estimation and Inference in BVAR Models

Estimation of BVAR systems can be carried out under different prior distributional assumptions. The models implemented in this report correspond to the normal-diffuse variety[7,6] The latter combines a diffuse prior on the residual covariance matrix and a multivariate normal density on the autoregressive coefficients, with first and second moments respectively defined in equations (5) and (7). This framework offers additional flexibility over other approaches (such as the asymmetric prior treatment of different equations), but comes at the expense of lacking posterior analytical tractability. Fortunately, Markov chain Monte Carlo (MCMC) techniques are well suited for estimation purposes.[2]

All the estimates presented are based on samples of 30,000 posterior parameter draws, which were obtained via Gibbs sampling, after discarding an initial number of 2,000 "burn-in" iterations. The Gibbs sampler is an MCMC algorithm that exploits the tractability of the posterior conditional densities in the specified BVAR systems. Specifically, the posterior conditional density of the regression coefficients is known to be multivariate normal, while the inverse of the residual covariance matrix has a wishart conditional posterior distribution.

Notice that when working with differenced data as it is the case in this study, the prior mean for the first lag of the dependent variable in equation (5) should be set to zero. This implies a first lag coefficient of unity in the levels of logarithmic mortality (which is a random-walk specification). Furthermore, in order to avoid the possibility of explosive behavior when generating mortality projections over very long forecast horizons, the estimated BVAR processes are constrained to the stationary region of the parameter space.

Two alternative Bayesian vector autoregressive specifications are entertained for posterior inference, denoted as BVAR(1)-I and BVAR(1)-II, respectively. Both models involve an overall tightness of $\lambda_1 = 0.3$ and a prior on the intercept coefficients that is made noninformative by setting $\lambda_4 = 10^5$. The first model effectively replicates the functional form in equation (2), that consists of a system of first-order univariate autoregressions with a common residual covariance matrix. This is achieved by selecting $\lambda_2(i,j) = 0.001$, for $i \neq j$. On the other hand, the second model accommodates the temporal contiguity regularities observed in the correlation structure of the data. In particular, letting $\Omega_{i,j}$ denote the corresponding element of the sample correlation matrix in Table 1, the prior lag weights are set to $\lambda_2(i,j) = 0.8 \times \Omega_{i,j}$, for $i \neq j$. In addition, both models involve a harmonic rate of lag decay ($\lambda_3 = 1$), which happens to make no difference, since the selected lag length is $p = 1$.

Posterior Predictive Validation and Model Choice

Thoughtful statistical analysis should address the two related issues of model assessment and model choice. The former investigates how well a model fits the data. The latter deals with selecting the "best" model, given a collection of competing specifications $\{M_1, M_2, \ldots, M_k\}$. From a Bayesian perspective, it is actually possible to determine the posterior probability associated with each model. Therefore, the Bayesian solution to the model selection problem is to choose the model with highest posterior probability.

Formally, the posterior probability associated with model M_j is given by

$$\text{pr}(M_j \mid Y) = \frac{\text{pr}(M_j)\, f(Y \mid M_j)}{\sum_{j=1}^{k} \text{pr}(M_j)\, f(Y \mid M_j)}, \qquad (8)$$

Where $\text{pr}(M_j)$ represents the prior probability assigned to model M_j, and $f(Y \mid M_j)$ is the marginal likelihood previously appearing in denominator of equation (4). The latter serves as the normalizing constant that makes the posterior density proper, by forcing it to integrate to one

$$f(Y \mid M_j) = \int \pi(\theta \mid M_j)\, L(\theta \mid M_j)\, d\theta. \qquad (9)$$

Of course, when equal probabilities are assigned to each model then, the specification with the highest posterior probability is also the one with the largest marginal likelihood value.

The accurate estimation of marginal likelihoods is often one of the most challenging computational problems in Bayesian applications. The approach followed in this report exploits a very insightful connection between the marginal likelihood of a model and its predictive density. Let $Y_T^0 = \{y_1^0, y_2^0, \ldots, y_T^0\}$ represent a sample of T observations (where the fact that it is observed is explicitly emphasized with a superscript), and let $Y_F = \{y_{T+1}, y_{T+2}, \ldots, y_F\}$ denote a future (unobserved) sample. The model's predictive density for Y_F is given by

$$P(Y_F \mid Y_T^0) = \int P(Y_F \mid Y_T^0, \theta)\, \pi(\theta \mid Y_T^0)\, d\theta. \quad (10)$$

Once the future sample is observed, the quantity $P(Y_F^0 \mid Y_T^0)$ is known as the predictive likelihood and can be shown to represent the multiplicative factor used to update the marginal likelihood, as new data becomes available.[6] This implies that a model's marginal likelihood can be decomposed in terms of the product of its predictive likelihood factors

$$f(Y_T^0 \mid \mathrm{M}_j) = \prod_{t=1}^{T} P(y_t^0 \mid y_{t-1}^0, y_{t-2}^0, ..., y_0^0, \mathrm{M}_j). \quad (11)$$

Using this approach, the logarithm of the marginal likelihood estimates for the BVAR(1)-I and BVAR(1)-II models are respectively 4,276.5 and 4,478.7. Consequently, the estimates suggest a significant improvement in predictive performance for the second model, as a result of accommodating the temporal contiguity regularities present in the correlation structure of the data (Table 1).

The predictive decomposition in equation (11) emphasizes the link between model selection and the out-of-sample prediction record of the specified models. It formally embodies the maxim that "a model is as good as its predictions." Moreover, by following the predictive route, it is possible to specifically target criteria for model choice that are consistent with the intended use of the models. For instance, the variables directly fit to the BVAR systems in this application are the first differences of logarithmic mortality. However, what matters ultimately is the models' performance in forecasting the mortality rates themselves (their levels), as well as nonlinear functions of the death rates, such as life expectancy at various ages.

Posterior predictive validation represents the Bayesian equivalent of classical goodness-of-fit diagnosis. Let $g(y_t)$ denote any function of the age-specific mortalities. Clearly, given a sample of posterior draws, for every available observation $g(y_t^0)$ there is an associated predictive density $P(y_t \mid y^0_{t-1},\ldots,y^0_{t-p},\mathrm{M}_j)$ from which synthetic draws of $g(y_t)$ can be generated. This provides a way to identify any discrepancies between the observed data and the posited model, as well as the means to specify meaningful model selection criteria. In addition, from a computational perspective, the analysis can be carried out in a straightforward fashion as a simple by-product of the posterior simulation.

For all 21 age-specific mortalities, life expectancy at birth and life expectancy at ages 19, 35, 65 and 80, Table 2 presents some useful results for the purposes of model validation and model choice. The first two columns in Table 2 display the sum of squared residuals for the two models, defined as

$$\mathrm{SSR} = \sum_{t=1}^{T} \left(g(y_t^0) - E[g(y_t)] \right)^2. \quad (12)$$

For ease of presentation, the third column in Table 2 shows the ratio of SSR between the two models. In addition, the last two columns in Table 2 report the percentage of observations in the sample that lie within the 90% credible intervals generated by the models.

Evidently, with the exception of mortality at ages 50-54, the BVAR(1)-II specification achieves a point forecast error reduction for all other projected quantities ranging anywhere between 2% to 22%. This finding is fully consistent with the previously reported marginal likelihood estimates favoring the BVAR(1)-II model. It supports the notion that the lags of other series different from the dependent variable carry useful predictive information. Furthermore, both models produce forecast intervals with an actual probability content that is close to the nominal 90% coverage. Hence, the BVAR systems seem to provide a reasonably good fit to the observed data.

Long-Range Forecasts

Long-rage projections (2002-2075) corresponding to the predictive densities of the BVAR(1)-II model were generated, based on a sample of 30,000 posterior draws. For selected forecast periods, Table 3 presents the 5th, 50th and 95th quantiles of life expectancy at various ages, as well as the width of the resulting interval predictions. Similar results corresponding to the Lee-Carter model are also included for comparison. For the purposes of illustration, Figures 4 through 6 display historical and projected values of life expectancy at birth and ages 35 and 65, while Figures 7 through 9 show equivalent forecasts for mortality at ages 30-34, 60-64 and 75-79.

In general, the point forecasts generated by the BVAR specification are fairly close to those of the Lee-Carter method, with the latter model typically yielding slightly greater increases in mortality improvement over time. For instance, the Lee-Carter projected median life expectancy at birth in the years 2015 and 2075 exceeds the BVAR approach by roughly three and a half weeks and seven months, respectively.

Figure 4: Life Expectancy at birth.

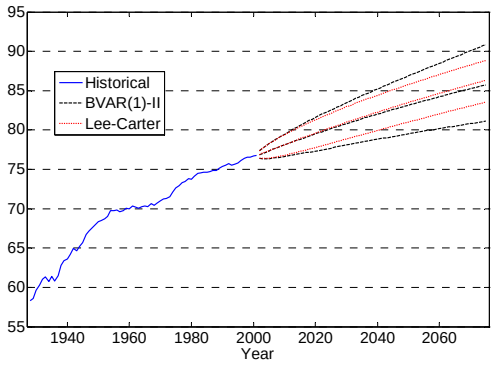


Figure 5: Life Expectancy at age 35.

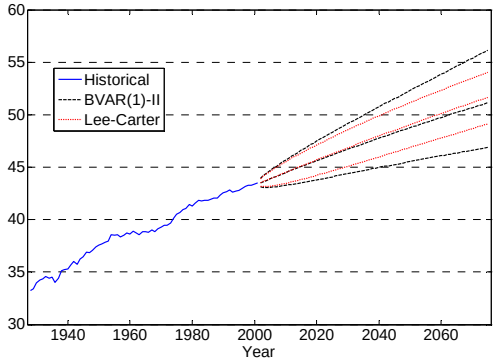


Figure 6: Life Expectancy at age 65.

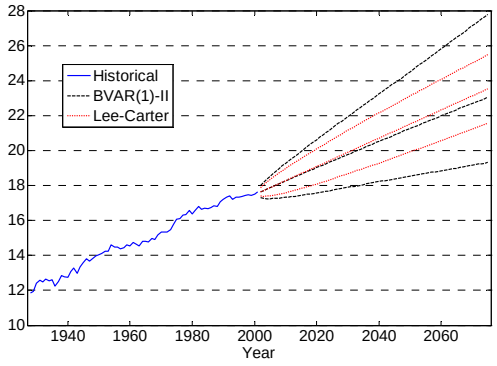


Figure 7: Mortality at ages 30-34.

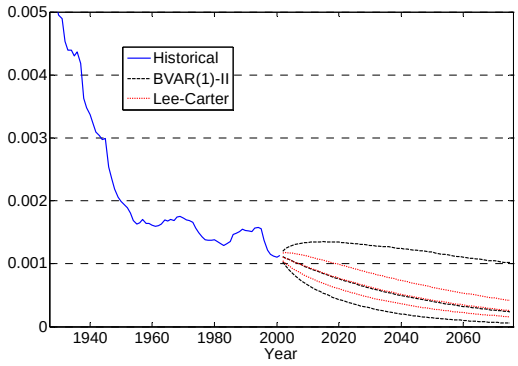


Figure 8: Mortality at ages 60-64.

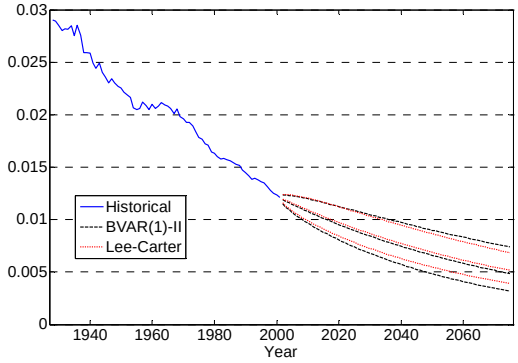
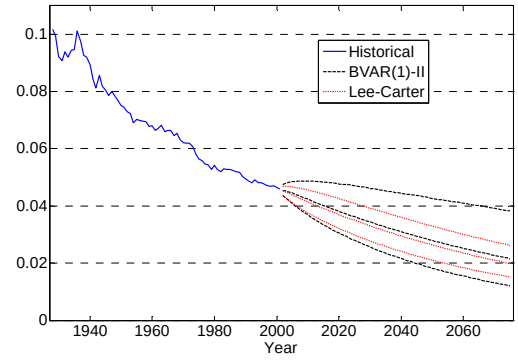


Figure 9: Mortality at ages 75-79.



The credible intervals of the BVAR model, on the other hand, are wider than the corresponding confidence intervals in the Lee-Carter method. This tends to be the case in particular for the older age groups. In other words, the difference among the intervals widens not only as a function of the forecast horizon, but as age increases. For example, the Bayesian interval projections in 2075 for life expectancy at birth, at age 65 and at age 80 are respectively 1.8, 2.2 and 3 times wider than those of the Lee-Carter model.

At this point, it is important to emphasize the methodological distinctions between the two models. Regarding the sources of uncertainty, the Bayesian interval forecasts incorporate variation due to both the sample and parameter outcomes. The Lee-Carter approach does not account for parameter uncertainty. From an empirical stance, the latter model has sometimes been criticized for producing implausibly narrow intervals.[9] Ongoing research by the author of this report has also found a tendency for the Lee-Carter method to generate

mortality projections that are "too narrow" for the oldest age groups. Indeed, it is this type of criticism that has led to a number of extensions of the Lee-Carter model, in order to accommodate greater uncertainty in some of the parameter estimates.[8,9] Generally, however, there are both practical and conceptual problems to the notion of parameter uncertainty within the classical inference perspective. This is due to the fact that the sampling distribution of a classical estimator has support on the sample space, not on the parameter outcomes. Hence, the BVAR approach presented in this report seems promising in addressing some of the shortcomings of the Lee-Carter method.

References

1. Bell, F. C. and M. L. Miller (2005). "Life Tables for the United States Social Security Area 1900-2100." *Social Security Administration*, Actuarial Study No. 120, Pub. No. 11-11536. Available at http://www.ssa.gov/OACT/NOTES/s2000s.html.

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

3. Congressional Budget Office (CBO) (2001). "Uncertainty in Social Security's Long-Term Finances: A Stochastic Analysis." Available at http://www.cbo.gov/ftpdocs/32xx/doc3235

4. Denton, F. T., C. H. Feaver and B. G. Spencer (2005). "Time Series Analysis and Stochastic Forecasting: An Econometric Study of Mortality and Life Expectancy." *Journal of Population Economics* 18, 203-227.

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

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

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

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

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

Table 1: Sample correlation of first difference in logarithmic mortality.

Age	0	1-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95+
0	1.00																				
1-4	0.61	1.00																			
5-9	0.47	0.71	1.00																		
10-14	0.36	0.56	0.73	1.00																	
15-19	0.27	0.57	0.64	0.78	1.00																
20-24	0.29	0.47	0.55	0.67	0.80	1.00															
25-29	0.27	0.49	0.55	0.62	0.78	0.84	1.00														
30-34	0.28	0.47	0.50	0.66	0.74	0.73	0.86	1.00													
35-39	0.25	0.45	0.55	0.63	0.69	0.65	0.83	0.88	1.00												
40-44	0.34	0.46	0.43	0.58	0.67	0.62	0.71	0.78	0.83	1.00											
45-49	0.39	0.41	0.47	0.53	0.57	0.47	0.58	0.61	0.70	0.79	1.00										
50-54	0.40	0.46	0.45	0.38	0.48	0.42	0.51	0.51	0.62	0.73	0.75	1.00									
55-59	0.29	0.24	0.43	0.45	0.48	0.43	0.43	0.44	0.48	0.62	0.74	0.71	1.00								
60-64	0.36	0.27	0.30	0.34	0.34	0.25	0.30	0.36	0.41	0.53	0.67	0.76	0.76	1.00							
65-69	0.32	0.27	0.31	0.34	0.38	0.31	0.34	0.30	0.38	0.56	0.59	0.70	0.73	0.75	1.00						
70-74	0.25	0.19	0.24	0.26	0.27	0.22	0.20	0.21	0.28	0.44	0.50	0.60	0.71	0.79	0.80	1.00					
75-79	0.29	0.27	0.35	0.39	0.34	0.29	0.31	0.27	0.42	0.49	0.56	0.67	0.72	0.73	0.84	0.82	1.00				
80-84	0.28	0.25	0.26	0.30	0.32	0.28	0.25	0.23	0.32	0.49	0.50	0.62	0.69	0.71	0.80	0.87	0.87	1.00			
85-89	0.20	0.10	0.21	0.27	0.23	0.16	0.14	0.15	0.31	0.40	0.46	0.54	0.60	0.65	0.76	0.80	0.86	0.88	1.00		
90-94	0.20	0.10	0.21	0.28	0.24	0.18	0.15	0.15	0.31	0.41	0.46	0.51	0.60	0.62	0.73	0.76	0.85	0.87	0.97	1.00	
95+	0.19	0.12	0.23	0.30	0.25	0.22	0.19	0.17	0.31	0.42	0.46	0.49	0.60	0.58	0.70	0.69	0.82	0.83	0.90	0.97	1.00

Table 2: Posterior predictive validation and model choice.

	Sum of Squared Residuals			*Coverage of 90% Intervals*	
$g(y_t^0)$	BVAR(1)-I	BVAR(1)-II	$\frac{\text{BVAR(1)-II}}{\text{BVAR(1)-I}}$	BVAR(1)-I	BVAR(1)-II
Age 0	0.107×10^{-3}	0.961×10^{-2}	0.895	93.06	94.44
Ages 1-4	0.122×10^{-5}	0.101×10^{-5}	0.828	95.83	97.22
Ages 5-9	0.946×10^{-7}	0.731×10^{-7}	0.773	95.83	98.61
Ages 10-14	0.639×10^{-7}	0.505×10^{-7}	0.790	97.22	95.83
Ages 15-19	0.298×10^{-6}	0.238×10^{-6}	0.799	95.83	97.22
Ages 20-24	0.963×10^{-6}	0.807×10^{-6}	0.838	95.83	95.83
Ages 25-29	0.737×10^{-6}	0.625×10^{-6}	0.848	97.22	97.22
Ages 30-34	0.657×10^{-6}	0.598×10^{-6}	0.909	93.06	94.44
Ages 35-39	0.923×10^{-6}	0.876×10^{-6}	0.949	94.44	95.83
Ages 40-44	0.120×10^{-5}	0.112×10^{-5}	0.934	95.83	94.44
Ages 45-49	0.158×10^{-5}	0.149×10^{-5}	0.944	95.83	98.61
Ages 50-54	0.351×10^{-5}	0.352×10^{-5}	1.002	94.44	94.44
Ages 55-59	0.644×10^{-5}	0.612×10^{-5}	0.951	93.06	94.44
Ages 60-64	0.126×10^{-4}	0.100×10^{-4}	0.789	94.44	95.83
Ages 65-69	0.249×10^{-4}	0.211×10^{-4}	0.845	95.83	94.44
Ages 70-74	0.654×10^{-4}	0.534×10^{-4}	0.816	97.22	97.22
Ages 75-79	0.231×10^{-3}	0.174×10^{-3}	0.752	94.44	94.44
Ages 80-84	0.588×10^{-3}	0.514×10^{-3}	0.873	95.83	95.83
Ages 85-89	0.163×10^{-2}	0.151×10^{-2}	0.928	95.83	95.83
Ages 90-94	0.310×10^{-2}	0.287×10^{-2}	0.928	95.83	95.83
Ages 95-99	0.474×10^{-2}	0.434×10^{-2}	0.915	95.83	95.83
Life Exp. Age 0	0.683×10^{1}	0.664×10^{1}	0.973	92.96	94.37
Life Exp. Age 19	0.454×10^{1}	0.435×10^{1}	0.957	94.37	94.37
Life Exp. Age 35	0.360×10^{1}	0.333×10^{1}	0.927	94.37	95.77
Life Exp. Age 65	0.201×10^{1}	0.178×10^{1}	0.888	92.96	95.77
Life Exp. Age 80	0.136×10^{1}	0.127×10^{1}	0.931	95.77	97.18

Table 3: Long range projections of life expectancy at various ages (5^{th}, 50^{th} and 95^{th} quantiles)

Life Expectancy at Birth

	BVAR(1)-II				Lee-Carter			
Year	5^{th}	50^{th}	95^{th}	95^{th} -5^{th}	5^{th}	50^{th}	95^{th}	95^{th} -5^{th}
2010	76.56	77.93	79.25	2.69	76.71	77.98	79.18	2.47
2015	76.89	78.66	80.35	3.46	77.16	78.73	80.19	3.04
2025	77.61	80.03	82.35	4.74	78.20	80.16	81.97	3.77
2050	79.43	83.05	86.70	7.27	80.90	83.38	85.64	4.74
2075	81.04	85.64	90.71	9.67	83.46	86.22	88.72	5.26

Life Expectancy at Age 19

	BVAR(1)-II				Lee-Carter			
Year	5^{th}	50^{th}	95^{th}	95^{th} -5^{th}	5^{th}	50^{th}	95^{th}	95^{th} -5^{th}
2010	58.38	59.65	60.88	2.50	58.60	59.69	60.75	2.15
2015	58.63	60.29	61.88	3.25	58.98	60.35	61.66	2.68
2025	59.22	61.50	63.74	4.52	59.88	61.63	63.29	3.41
2050	60.84	64.30	67.89	7.06	62.31	64.62	66.79	4.48
2075	62.27	66.78	71.81	9.54	64.69	67.34	69.79	5.10

Life Expectancy at Age 35

	BVAR(1)-II				Lee-Carter			
Year	5^{th}	50^{th}	95^{th}	95^{th} -5^{th}	5^{th}	50^{th}	95^{th}	95^{th} -5^{th}
2010	43.23	44.35	45.48	2.24	43.40	44.38	45.35	1.95
2015	43.46	44.93	46.41	2.95	43.73	44.98	46.19	2.45
2025	44.01	46.05	48.15	4.15	44.56	46.16	47.71	3.15
2050	45.48	48.67	52.17	6.69	46.79	48.95	51.03	4.24
2075	46.82	51.04	55.99	9.16	49.03	51.56	53.93	4.90

Life Expectancy at Age 65

	BVAR(1)-II				Lee-Carter			
Year	5^{th}	50^{th}	95^{th}	95^{th} -5^{th}	5^{th}	50^{th}	95^{th}	95^{th} -5^{th}
2010	17.30	18.18	19.12	1.82	17.55	18.19	18.85	1.30
2015	17.41	18.57	19.81	2.39	17.77	18.60	19.43	1.66
2025	17.68	19.32	21.14	3.46	18.32	19.41	20.52	2.20
2050	18.50	21.19	24.39	5.89	19.86	21.44	23.03	3.17
2075	19.27	22.97	27.64	8.36	21.49	23.45	25.38	3.88

Life Expectancy at Age 80

	BVAR(1)-II				Lee-Carter			
Year	5^{th}	50^{th}	95^{th}	95^{th} -5^{th}	5^{th}	50^{th}	95^{th}	95^{th} -5^{th}
2010	7.77	8.57	9.44	1.67	8.19	8.58	8.97	0.78
2015	7.73	8.79	9.97	2.24	8.32	8.82	9.33	1.01
2025	7.72	9.22	10.96	3.24	8.65	9.32	10.01	1.36
2050	7.84	10.35	13.46	5.61	9.59	10.61	11.67	2.07
075	7.99	11.49	16.08	8.08	10.64	11.96	13.33	2.69