

FORECASTING THE SOUTH AFRICAN ECONOMY WITH GIBBS SAMPLED BVECMs

RANGAN GUPTA*

Abstract

The paper uses the Gibbs sampling technique to estimate a heteroscedastic Bayesian Vector Error Correction Model (BVECM) of the South African economy for the period 1970:1-2000:4, and then forecasts GDP, consumption, investment, short and long term interest rates, and the CPI over the period of 2001:1 to 2005:4. We find that a tight prior produces relatively more accurate forecasts than a loose one. The out-of-sample-forecast accuracy resulting from the Gibbs sampled BVECM is compared with those generated from a Classical VECM and a homoscedastic BVECM. The homoscedastic BVECM is found to produce the most accurate out of sample forecasts.

Keywords: VECM and BVECM, Forecast Accuracy, BVECM Forecasts, VECM Forecasts, Gibbs Sampling

1. INTRODUCTION

This paper uses the Gibbs sampling technique to estimate a Bayesian Vector Error Correction Model (BVECM) of the South African economy for the period 1970: 1-2000:4 to forecast GDP, consumption, investment, short and long term interest rates, and the CPI over 2001:1 to 2005:4. The out-of-sample-forecast accuracy resulting from the BVECM is compared with those generated from a BVECM estimated using Theil's (1971) mixed estimation method and a standard Classical VECM.

The motivation for this analysis emanates from a recent paper by Gupta (2006). In this study, the author compares the ability of Vector Autoregressive (VAR) Models and Vector Error Correction Models (VECM), both Classical and Bayesian in nature, in forecasting the South African economy. For this purpose, the study estimates these models using quarterly data on the same set of variables over the same sample period, as that of ours, and then, in turn, compares the out-of-sample forecast errors generated by these models over the period of 2001:1 to 2005:4. The BVECM model, with the most tight priors, was found to outperform the other models, in terms of forecasting consumption, CPI, GDP investment and the 91 days Treasury bill rate. The Classical VECM, however, performed the best, in terms of forecasting the measure of the long-term interest rate. Our paper, thus, simply tries to investigate whether we are able to produce lower forecasting errors over the same forecasting horizon and for the same set of variables, when the BVECM is re-estimated using Gibbs sampling.

* Associate Professor, University Of Pretoria, Department of Economics, Pretoria, 0002, South Africa. Phone: +27 12 420 3460, Fax: +27 12 362 5207, Email: Rangan.Gupta@up.ac.za. I would like to thank all the participants of the Brownbag Seminar Series at the Department of Economics, University of Pretoria for many helpful comments and, especially, Mr. Dave Liu for the diligent research assistance.

Traditionally, Bayesian Vector Autoregressive (BVAR) and Bayesian Vector Error Correction (BVEC) models have been estimated using Theil's (1971) mixed estimation technique, where the error structure is assumed to follow the Gauss-Markov assumptions. However, in this paper, the estimation of the BVECMs using Gibbs sampling allows us to incorporate outliers or non-constant variances. Note, with heteroscedastic errors, we could have used the generalised least squares method to implement Theil's (1971) mixed estimation. But it is often the case that with the Gibbs sampling methodology, complex estimation problems, such as these, are simplified considerably by conditioning on unknown parameters, by assuming that these values are known.¹ Moreover, as Kadiyala and Karlsson (1997) point out, amongst different ways of carrying out the Monte Carlo integration, the Gibbs sampling algorithms are less adversely affected by model size, and hence perform no worse than other important sampling methods.² The Vector Autoregressive (VAR) model, though 'a theoretical', is particularly useful for forecasting purposes. An unrestricted VAR model, as suggested by Sims (1980), can be written as follows:

$$y_t = C + A(L)y_t + \varepsilon_t \quad (1)$$

where y is a $(n \times 1)$ vector of variables being forecasted; $A(L)$ is a $(n \times n)$ polynomial matrix in the backshift operator L with lag length p , *i.e.* $A(L) = A_1L + A_2L^2 + \dots + A_pL^p$; C is a $(n \times 1)$ vector of constant terms, and ε is a $(n \times 1)$ vector of error terms. Note that, in our case, we assume that $\varepsilon \sim N(0, \sigma^2 V)$, $V = \text{diag}(v_1, v_2, \dots, v_n)$ where V is a $n \times n$ matrix, with the relative variance terms v_1, v_2, \dots, v_n assumed to be fixed but unknown. But, for the BVARs and BVECMs, estimated with Theil's (1971) mixed estimation procedure, and the Classical VECM, ε is an $(n \times 1)$ vector of white noise error terms.

Focusing on the practical case, of y_t being a vector of n time series that are integrated³ to the order of 1 ($I(1)$),⁴ the ECM counterpart of the VAR, given by (1), is captured by a VECM as follows:⁵

$$\Delta y_t = \pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t \quad (2)$$

where $\pi = -\left(I - \sum_{i=1}^p A_i\right)$ and $\Gamma_i = -\sum_{j=i+1}^p A_j$.

The Engle-Granger (1987) Representation Theorem asserts that if the coefficient matrix π (the cointegrating space) has reduced rank $r < n$, then there exist matrices CC and β each with rank r such that $\pi = \alpha\beta'$ and $\beta'y_t$ is $I(0)$. Note r is the number of cointegrating relations (the cointegrating rank) and each column of β is the cointegrating

¹ See section 2 for further details.

² See Kadiyala and Karlsson (1997) and the references cited therein for further details.

³ A series is said to be integrated of order q , if it requires q differencing to transform it to a zero-mean, purely non-deterministic stationary process.

⁴ LeSage (1990) and references cited therein for further details regarding most macroeconomic time series being $I(1)$.

⁵ See Dickey *et al.* (1991) and Johansen (1995) for further technical details.

vector, and the elements of α are known as the adjustment parameters in the VECM. α is also known as the loading matrix and has a dimension $n \times r$. Since it is not possible to use conventional OLS to estimate α and β , Johansen's (1988) full information maximum likelihood estimation is used to determine the cointegrating rank of π , using the r most significant cointegrating vectors to form β , from which a corresponding α is derived. Note that the specification in (2) is in line with the Engle and Granger (1987) Representation Theorem.

The VAR model, generally, uses equal lag length for all the variables of the model. One drawback of VAR models is that many parameters are needed to be estimated, some of which may be insignificant. This problem of overparameterization, resulting in multicollinearity and a loss of degrees of freedom, leads to inefficient estimates and possibly large out-of-sample forecasting errors. One must remember that in the VECMs, besides the parameters corresponding to the lagged values of the variables, the parameters corresponding to the error correction terms are also estimated, possibly enhancing the problem of overparameterization further.

A rational approach⁶ to overcoming this overparameterization, as described in Litterman (1981), Doan *et al.* (1984), Todd (1984), Litterman (1986) and Spencer (1993), is to use a Bayesian VAR (BVAR) model. Instead of eliminating longer lags, the Bayesian method imposes restrictions on these coefficients by assuming that they are more likely to be near zero than the coefficients on shorter lags. However, if there are strong effects from less important variables, the data can override this assumption. The restrictions are imposed by specifying normal prior distributions with zero means and small standard deviations for all coefficients with the standard deviation decreasing as the lags increase. The exception to this is, however, the coefficient on the first own lag of a variable, which has a mean of unity. Litterman (1981) used a diffuse prior for the constant. This is popularly referred to as the 'Minnesota prior' due to its development at the University of Minnesota and the Federal Reserve Bank at Minneapolis. Note that, as described in (2), an identical approach can be taken to implement a Bayesian variant of the Classical VECM based on the Minnesota prior.

Formally, as discussed above, the Minnesota prior means take the following form:

$$\beta_i \sim N(1, \sigma_{\beta_i}^2) \quad \text{and} \quad \beta_j \sim N(0, \sigma_{\beta_j}^2) \quad (3)$$

where β_i denotes the coefficients associated with the lagged dependent variables in each equation of the VAR, while β_j represents any other coefficient. In the belief that lagged dependent variables are important explanatory variables, the prior means corresponding to them are set to unity. However, for all the other coefficients, β_j 's, in a particular equation of the VAR, a prior mean of zero is assigned, to suggest that these variables are less important to the model.

The prior variances $\sigma_{\beta_i}^2$ and $\sigma_{\beta_j}^2$, specify uncertainty about the prior means $\bar{\beta}_i = 1$, and $\bar{\beta}_j = 0$, respectively. Because of the overparameterization of the VAR, Doan *et al.* (1984) suggested a formula to generate standard deviations as a function of small numbers of hyperparameters: w , d , and a weighting matrix/(i, j). This approach allows the

⁶ Two other solutions, often adapted, are simply to exclude the insignificant lags based on statistical tests, or to use a near VAR, which specifies an unequal number of lags for the different equations.

forecaster to specify individual prior variances for a large number of coefficients based on only a few hyperparameters. The specification of the standard deviation of the distribution of the prior imposed on variable j in equation i at lag m , for all i, j and m , defined as $S(i, j, m)$, can be specified as follows:

$$S(i, j, m) = [w \times g(m) \times f(i, j)] \frac{\hat{\sigma}_j}{\hat{\sigma}_i} \quad (4)$$

with $f(i, j) = 1$, if $i = j$ and k_{ij} otherwise, with $(0 \leq k_{ij} \leq 1)$, $g(m) = m^{-d}$, $d > 0$. Note that $\hat{\sigma}_i$ is the estimated standard error of the univariate autoregression for variable i . The ratio $\hat{\sigma}_j / \hat{\sigma}_i$ scales the variables so as to account for differences in the units of measurement and, hence, causes specification of the prior without consideration of the magnitudes of the variables. The term w indicates the overall tightness and is also the standard deviation on the first own lag, with the prior getting tighter as we reduce the value. The parameter $g(m)$ measures the tightness on lag m with respect to lag 1, and is assumed to have a harmonic shape with a decay factor of d , which tightens the prior on increasing lags. The parameter $f(i, j)$ represents the tightness of variable j in equation i relative to variable i , and by increasing the interaction, *i.e.* the value of k_{ij} , we can loosen the prior.⁷

Besides, the Minnesota prior on the β 's, we impose a diffuse prior on σ , *i.e.* $\sigma \sim \left(\frac{1}{\sigma}\right)$.

The parameters to be estimated are β , σ and the relative variance terms (v_1, v_2, \dots, v_n) . Note the estimation of $p + 1$ number of β coefficients, given that the number of lags equal to p , σ and n number of v_i terms using n data observations might seem very problematic.

However, following Geweke (1993), we assign an independent $\frac{\chi^2(r)}{r}$ prior distribution

to the v_i terms, which depends on a single hyperparameter r . Specifically we assume:

$\frac{r}{v_i} \sim \frac{\chi^2(r)}{r}$ and $r \sim \Gamma(m, k)$ to take account of heteroscedasticity and outliers. This allows

us to estimate the additional n parameters associated with the v_i 's by just adding a single parameter r to the estimation procedure for our model. The details regarding the prior assigned to the V_i terms can be motivated by considering that the mean of the prior, is

equal to unity, and the prior variance is $\frac{2}{r}$. This implies that as r becomes very large, the

prior reflects the special case of $\varepsilon_i \sim N(0, \sigma^2 I_n)$. Note that the role of $V_i \neq I_n$ is to robustify

against outliers and observations containing large variances by down-weighting these observations. Large r values are, thus, associated with a prior belief that there are no

outliers and non-constant variances. The values assigned to the hyperparameter r is derived by imposing a prior distribution of $\Gamma(m, k)$ to this parameter, which has a mean

of $\frac{m}{k}$ and a variance of $\frac{m}{k^2}$. So using $m = 8$ and $k = 2$ would assign a prior to r centered

on a small $r = 4$ and variance of $r = 2$. Given that $\frac{r}{r-2}$ and $\frac{r}{r+2}$ is the mean and mode of the prior, respectively, small values of the hyperparameter r , causes v_{ij} to take on a skewed form where the mean and mode are quite different.⁸

⁷ For an illustration, see Dua and Ray (1995).

⁸ See Chapter 6 of LeSage (1999) for further details.

The heteroscedastic Bayesian VECM is estimated using Gibbs sampling, by beginning with arbitrary values for β , σ , v_i and r . We then compute the mean and variance of β using its posterior density conditional on the initial values of σ , v_i and r , given by $\beta | (\sigma, V) \sim N[H(X'V^{-1}y + \sigma^2R'T^{-1}c), \sigma^2H]$ with $H = (X'V^{-1}X + R'T^{-1}R)^{-1}$, where X denotes the lagged values of y and $\beta \sim N(c, T)$, with c and T defined according to the Minnesota prior, outlined in (3) and (4). Now we use the computed mean and variance of β to draw a multivariate normal random vector, β^1 , based on which we calculate the conditional posterior density for σ given by $\left[\sum_{i=1}^n (e_i^2/v_i)/\sigma^2 \right] (\beta, V) \sim \chi^2(n)$, $e_i = y_i - X_i'\beta$, and then, in turn, use this value along with a random $\chi^2(n)$ draw to determine σ^1 . Then using β^1 and σ^1 , we calculate the posterior distribution of V conditional on (β, σ) , $\left[\sum_{i=1}^n (\sigma^{-2}e_i^2 + r)/v_i \right] (\beta, V) \sim \chi^2(r+1)$, and use the value with an n -vector of random $\chi^2(r+1)$ draws to determine v_i , $i = 1, \dots, n$. Finally, we draw a $\Gamma(m, k)$ value to update the initial value of r . These steps constitute a single pass of the Gibbs sampler. We make a large number of passes, in our case 10,000 with the initial 1000 discarded to prevent issues of initialisation, to build up a sample β^j , σ^j , v_i^j and r^j of j values from which we can approximate the posterior distributions for our parameters.

The Bayesian variants of the VARs and VECMs with homoscedastic errors are, on the other hand, estimated using Theil's (1971) mixed estimation technique. Specifically, suppose we denote a single equation of the VAR model as: $y_1 = X\beta + \varepsilon_1$, with $\text{Var}(\varepsilon_1) = \sigma^2I$, then the stochastic prior restrictions for this single equation can be written as:

$$\begin{bmatrix} m_{111} \\ m_{112} \\ \cdot \\ \cdot \\ \cdot \\ m_{mk} \end{bmatrix} = \begin{bmatrix} \sigma/\sigma_{111} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \sigma/\sigma_{112} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 & \sigma/\sigma_{mk} \end{bmatrix} \begin{bmatrix} a_{111} \\ a_{112} \\ \cdot \\ \cdot \\ \cdot \\ a_{mk} \end{bmatrix} + \begin{bmatrix} u_{111} \\ u_{112} \\ \cdot \\ \cdot \\ \cdot \\ u_{mk} \end{bmatrix} \quad (5)$$

Note, $\text{Var}(u) = \sigma^2I$ and the prior mean m_{ijk} and σ_{ijk} the form shown in (3) and (4), respectively, for the Minnesota prior. With (5) written as:

$$r = R\beta + u \quad (6)$$

the estimates for a typical equation are derived using (7) as follows:

$$\hat{\beta} = (X'X + R'R)^{-1} (X'y_1 + R'r) \quad (7)$$

Essentially then, the method involves supplementing the data with prior information on the distribution of the coefficients. The number of observations and degrees of freedom are increased by one in an artificial way, for each restriction imposed on the parameter estimates. The loss of degrees of freedom due to over-parameterization associated with a classical VAR model is, therefore, not a concern in the BVARs and BVECMs.

To the best of our knowledge, this is the first attempt to simultaneously analyze the role of Bayesian VECMs estimated using Theil's (1971) mixed estimation technique and Gibbs sampling in making economy-wide forecasts, given that the studies of LeSage (1990), LeSage and Pan (1995), Dowd and LeSage (1997), and LeSage and Krivelyova (1999) were only regional in nature, and based on the mixed estimation method. Finally unlike most studies in the forecasting literature based on the BVARs, we also check for the robustness of our analysis by specifying alternative values of the hyperparameters for the Bayesian priors, available in the literature. Our study can, thus, be viewed as an attempt to extend the existing literature on forecasting, and especially Gupta (2006), by incorporating the role of the Classical and Bayesian variants of VECMs, under alternative assumptions of the error structure.

The rest of the paper is organized as follows. Besides the introduction and the conclusions, section 2 presents the basic idea of Gibbs sampling. Section 3 sets out the BVECM, estimated using Gibbs sampling, for the South African Economy, while section 4 compares the accuracy of the out-of-sample forecasts generated from alternative models, namely the Classical VECM and the BVECM estimated using Theil's (1971) mixed estimation method, with that of the Gibbs sampled BVECM.

2. A NOTE ON THE BASICS OF GIBBS SAMPLING

Gibbs sampling dates to the work on image analysis of Geman and Geman (1984) and is related to the EM algorithm⁹ which dealt with maximum likelihood and Bayesian estimation with missing information. Assume a parameter vector $\theta = (\theta_1, \theta_2)$, a prior $p(\theta)$, and likelihood $l(\theta|y)$ that produces a posterior distribution $p(\theta|y) = cp(\theta)l(\theta|y)$, with y the data and c a normalising constant. It is often the situation, that the posterior distribution over all parameters are extremely difficult to work with. On the other hand, if we can partition our parameter space into two sets θ_1, θ_2 , and had initial estimates for $\theta_1(\hat{\theta}_1)$ we could estimate θ_2 conditional on θ_1 using $p(\theta_2|y, \hat{\theta}_1)$. Denoting the estimate $\hat{\theta}_2$, derived by using the posterior mean or mode of $p(\theta_2|y, \hat{\theta}_1)$, and consider that we are now able to construct a new estimate of θ_1 based on the conditional distribution $p(\theta_1|y, \hat{\theta}_2)$, which can be used to construct another value for $\hat{\theta}_2$, and so on.

In general, for the case of k parameters the algorithm can be summarised as follows:

Initialise θ_0

Repeat {

Sample $\theta_1^{r+1} \sim p[\theta_1|y, (\theta_2^r, \dots, \theta_k^r)]$

= Sample $\theta_2^{r+1} \sim p[\theta_2|y, (\theta_1^{r+1}, \theta_3^r, \dots, \theta_k^r)]$

Sample $\theta_k^{r+1} \sim p[\theta_k|y, (\theta_1^{r+1}, \theta_2^{r+1}, \dots, \theta_{k-1}^{r+1})]$ }

Geman and Geman (1984) showed that the stochastic process θ^t from this approach to sampling the complete sequence of conditional distributions represents a Markov chain with correct equilibrium distribution. In fact, Gibbs sampling is closely related to Markov Chain Monte Carlo (MCMC) methods. An important issue in using Gibbs sampling is convergence of the sampler to the posterior distribution. Based on theory, we know that

⁹ See Dempster *et al.* (1977) for details.

the sampler converges in the limit as the number of draws tends to infinity. However, in any applied problem one must determine how many draws to make with the sampler.¹⁰ There is some evidence that linear regression models exhibit rapid convergence. It should be noted that once convergence is achieved, we need to carry on making passes to build up a sample from the posterior distribution which we can use to make inferences about the parameters. Note convergence, generally, tends to be a function of how complicated the set of conditional distributions are.

3. A GIBBS SAMPLED BVECM MODEL FOR THE SOUTH AFRICAN ECONOMY

Along the lines of Litterman (1986), Ni and Sun (2005), Gupta (2006) and Gupta and Sichei (2006), we estimate a BVECM model using Gibbs sampling for the South African economy for the period of 1970:1 to 2000:4, based on quarterly data. We then compute out-of-sample one- through eight-quarters-ahead forecasts for the period of 2001:1 to 2005:4, and then compare the accuracy of the forecast relative to the forecasts generated by a Classical VECM, and a BVECM estimated using Theil's (1971) mixed estimation technique. The variables included are real GDP, consumption, investment, 91 days Treasury Bill rate, 10 years and longer government bond rate, and the CPI. All data are seasonally adjusted in order to, *inter alia*, address the fact, as pointed out by Hamilton (1994:362), that the Minnesota prior is not well suited for seasonal data. All data are obtained from the Quarterly Bulletin of the Reserve Bank of South Africa. Note the real variables correspond to the values of the variables at year 2000's prices.

The model¹¹ is estimated with a constant and four lags of each variable, as in Dua and Ray (1995).¹² Moreover, given the fact that one cointegrating relationship was found, the model also included one error-correction term.¹³ All variables, except for the measures of the short- and long-term interest rates, have been measured in natural logarithms. Note Sims *et al.* (1990) indicate that with the Bayesian approach entirely based on the likelihood function, the associated inference does not need to take special account of nonstationarity. Given this, the variables have been specified in levels.¹⁴

¹⁰ For a comprehensive list of *ad hoc* convergence tests, refer to LeSage (1999).

¹ Based on 10,000 random draws for the 26 parameters in each equation, hosts of alternative convergence tests, outlined in LeSage (1999), were carried out to ensure that the sampler converges in the limit. Given the large number of parameters in the model, the convergence tests have not been reported to save space. However, all the results can be made available upon request from the author.

¹² Hafer and Sheehan (1989) find that the accuracy of the forecasts from the VAR is sensitive to the choice of lags. Their results indicated that shorter-lagged models are more accurate, in terms of forecasts, than longer lag models. Therefore, alternative lag structures for the Gibbs sampled BVECM were also examined. When we reduce the lag length to 3 and then to 2, we find marginal improvements in the accuracy of all six variables, but the rank of ordering, resulting from the alternative forecasts remained unchanged.

¹³ The cointegrating relationships are based on the trace statistics compared to the critical values at the 95 per cent level. From the results of the test, we observed that the null hypothesis of $r < 1$ was rejected at the 95 per cent level because the trace statistic of 65.333 is less than the associated critical value of 69.819.

¹⁴ However, using the Augmented Dickey Fuller, the Phillips-Perron tests, all the 6 variables were found to be, first-order difference stationary, *i.e.* integrated of order 1 ($I(1)$).

The so called, 'optimal' Bayesian prior is selected on the basis of the Mean Absolute Percentage Error (MAPE) values of the out-of-sample forecasts. Specifically, the six-variable BVECM is estimated for an initial prior for the period of 1971:1 to 2000:4 and, then we forecast for 2001:1 through 2005:4. Since we use four lags, the initial four quarters of the sample, 1970:1 to 1970:4, are used to feed the lags. We generate dynamic forecasts, as would naturally be achieved in actual forecasting practice. During each quarter of the forecast period, the models are estimated in order to update the estimate of the coefficient before producing 8-quarters-ahead forecasts. This iterative estimation and 8-step-ahead forecast procedure was carried out for 20 quarters, with the first forecast beginning in 2001:1. This experiment produced a total of 20 one-quarter-ahead forecasts, 20-two-quarters ahead forecasts, and so on, upto 20 8-step-ahead forecasts. We use the algorithm in the Econometric Toolbox of MATLAB,¹⁵ for this purpose. The MAPEs¹⁶ for the 20, quarter 1 through quarter 8 forecasts were then calculated for the six variables of the model. The average of the MAPE statistic values for one- to eight-quarters-ahead forecasts for the period 2001:1 to 2005:4 are then examined. Thereafter, we change the prior and a new set of MAPE values is generated. The combination of the parameter values, in the prior, that produces the lowest average MAPE values is selected, as the 'optimal' Bayesian prior. Following Doan (2000) and Dua *et al.* (1999), we choose 0.1 and 0.2 for the overall tightness (w) and 1 and 2 for the harmonic lag decay parameter (d). Moreover, as in Dua and Ray (1995), we also report our results for a combination of $w = 0.3$ and $d = 0.5$ - Moreover, a symmetric interaction function fit, j is assumed with $k_{ij} = 0.5$, as in Dua and Smyth (1995) and LeSage (1990). Finally, following LeSage (1999), the value of r is set equal to 4 to account for a heteroscedastic error structure.¹⁷

4. EVALUATION OF FORECAST ACCURACY

To evaluate the accuracy of forecasts generated by the BVECMs, we need to perform alternative forecasts. To make the MAPEs comparable with the BVECMs, we report the same set of statistics for the out-of-sample forecasts generated from a Classical VECM and the 'optimal' BVECM ($w = 0.1, d = 2$) obtained in Gupta (2006). In Tables 1 through 6, we compare the MAPEs of one- to eight-quarters-ahead out-of-sample-forecasts for the period of 2001:1 to 2005:4, generated by the the Classical VECM and the 'optimal' BVECM in Gupta (2006) with that of the 5 alternative Gibbs sampled BVECMs. The conclusions from these tables are as follows:

¹⁵ All statistical analysis was performed using MATLAB, version R2006a.

¹⁶ Note that if A_{t+n} denotes the actual value of a specific variable in period $t+n$ and ${}_tF_{t+n}$ the forecast made in period t for $t+n$, the MAPE statistic can be denned as

$$\left[\frac{1}{N} \sum abs \left(\frac{A_{t+n} - {}_tF_{t+n}}{A_{t+n}} \right) \right] \times 100, \text{ where } abs \text{ stands for the absolute value. For } n=1, \text{ the}$$

summation runs from 2001:1 to 2005:4, and for $n = 2$, the same covers the period of 2001:2 to 2005:4 and so on.

¹⁷ The Gibbs sampled estimates of the BVECMs with a high value of r ($= 100$) yields very similar estimates and out-of-sample errors as with the models estimated using Theil's (1971) mixed estimation method.

Table 1. MAPE(2001:1-2005:4): Final consumption expenditure by households in logs (Indices 2000 = 100)

QA	VECM	BVECM ($w = 0.1, d = 2$)	BVECM_G				
			($w = 0.3, d = 0.5$)	($w = 0.2, d = 1$)	($w = 0.2, d = 2$)	($w = 0.1, d = 1$)	($w = 0.1, d = 2$)
1	0.0001	0.0007	0.0006	0.0008	0.0018	0.0028	0.0004
2	0.0031	0.0024	0.0033	0.0180	0.0020	0.0053	0.0013
3	0.0031	0.0029	0.0063	0.0010	0.0037	0.0063	0.0013
4	0.0030	0.0025	0.0072	0.0004	0.0059	0.0063	0.0006
5	0.0038	0.0032	0.0102	0.0009	0.0072	0.0083	0.0007
6	0.0045	0.0035	0.0139	0.0010	0.0088	0.0097	0.0007
7	0.0056	0.0043	0.0172	0.0015	0.0103	0.0114	0.0009
8	0.0068	0.0057	0.0213	0.0007	0.0010	0.0139	0.0019
AVE	0.0038	0.0032	0.0100	0.0030	0.0051	0.0080	0.0010

MAPE: Mean Absolute Percentage Error; QA: Quarter Ahead; BVECM_G: Gibbs Sampled BVECM.

Table 2. MAPE(2001:1-2005:4): CPI in logs (Indices 2000 = 100)

QA	VECM	BVECM($w = 0.1, d = 2$)	BVECM_G				
			($w = 0.3, d = 0.5$)	($w = 0.2, d = 1$)	($w = 0.2, d = 2$)	($w = 0.1, d = 1$)	($w = 0.1, d = 2$)
1	0.0099	0.0111	0.0091	0.0027	0.0081	0.0152	0.0144
2	0.0140	0.0146	0.0138	0.0061	0.0116	0.0254	0.0236
3	0.0248	0.0244	0.0264	0.0082	0.0227	0.0450	0.0415
4	0.0236	0.0236	0.0265	0.0008	0.0224	0.0495	0.0477
5	0.0070	0.0071	0.0122	0.0225	0.0065	0.0419	0.0394
6	0.0143	0.0125	0.0066	0.0475	0.0124	0.0290	0.0280
7	0.0348	0.0318	0.0244	0.0722	0.0310	0.0176	0.0174
8	0.0594	0.0546	0.0475	0.0996	0.0533	0.0014	0.0029
AVE	0.0235	0.0225	0.0208	0.0325	0.0210	0.0281	0.0269

MAPE: Mean Absolute Percentage Error; QA: Quarter Ahead; BVECM_G: Gibbs Sampled BVECM.

Table 3. MAPE(2001:1-2005:4): Real GDP in logs (Indices 2000 = 100)

QA	VECM	BVECM($w = 0.1, d = 2$)	BVECM_G				
			($w = 0.3, d = 0.5$)	($w = 0.2, d = 1$)	($w = 0.2, d = 2$)	($w = 0.1, d = 1$)	($w = 0.1, d = 2$)
1	0.0002	0.0006	0.0009	0.0030	0.0002	0.0000	0.0005
2	0.0015	0.0017	0.0037	0.0030	0.0015	0.0017	0.0004
3	0.0041	0.0037	0.0073	0.0055	0.0032	0.0048	0.0019
4	0.0042	0.0037	0.0080	0.0037	0.0030	0.0050	0.0018
5	0.0035	0.0028	0.0080	0.0030	0.0016	0.0045	0.0005
6	0.0018	0.0008	0.0075	0.0011	0.0007	0.0029	0.0015
7	0.0011	0.0004	0.0082	0.0010	0.0024	0.0021	0.0030
8	0.0013	0.0005	0.0100	0.0013	0.0028	0.0025	0.0032
AVE	0.0022	0.0018	0.0067	0.0027	0.0019	0.0029	0.0016

MAPE: Mean Absolute Percentage Error; QA: Quarter Ahead; BVECM_G: Gibbs Sampled BVECM.

(i) Optimal BVECM versus VECM: Since, in this case, we replicate the results of Gupta (2006), as expected we find that, except for the 10 years and longer government bond rate, the BVECM model with the most tight prior ($w = 0.1, d = 2$) outperforms the VECM model.¹⁸

¹⁸ We were able to duplicate the results of Gupta (2006) for the four other alternative combinations of the hyperparameters of the prior. Since the BVECM with $w = 0.1, d = 2$ outperformed all the other models, except for the VECM for the long-term interest rate, we just chose to report the above mentioned BVECM. Note as in Gupta (2006), we found that this BVECM, when compared to the alternative BVECMs, did only second best, in terms of the out-of-sample forecasting errors, for the consumption expenditures of the households.

Table 4. MAPE(2001:1-2005:4): 10 years and over government bond rate

QA	VECM	BVECM($w = 0.1, d = 2$)	BVECM_G				
			($w = 0.3, d = 0.5$)	($w = 0.2, d = 1$)	($w = 0.2, d = 2$)	($w = 0.1, d = 1$)	($w = 0.1, d = 2$)
1	0.4915	0.5183	0.4073	0.4151	0.4350	0.4345	0.4326
2	0.6522	0.7076	0.4704	0.5205	0.5259	0.4805	0.4858
3	0.9582	0.9739	0.6221	0.6585	0.6616	0.6671	0.6241
4	0.4117	0.5236	0.0322	0.0867	0.1045	0.1361	0.0685
5	0.0648	0.0634	0.4619	0.4855	0.4387	0.3680	0.4688
6	0.3251	0.4883	0.1804	0.2625	0.1825	0.0937	0.2118
7	0.4070	0.5860	0.1805	0.3334	0.2347	0.1033	0.2592
8	0.8089	1.0466	0.1414	0.1074	0.0119	0.1885	0.0082
AVE	0.5149	0.6135	0.3120	0.3587	0.3244	0.3090	0.3199

MAPE: Mean Absolute Percentage Error; QA: Quarter Ahead; BVECM_G: Gibbs Sampled BVECM.

Table 5. MAPE(2001:1-2005:4): Investment expenditure in logs (Indices 2000 = 100)

QA	VECM	BVECM($w = 0.1, d = 2$)	BVECM_G				
			($w = 0.3, d = 0.5$)	($w = 0.2, d = 1$)	($w = 0.2, d = 2$)	($w = 0.1, d = 1$)	($w = 0.1, d = 2$)
1	0.0549	0.0460	0.0560	0.0532	0.0482	0.0493	0.0481
2	0.0541	0.0577	0.0569	0.0568	0.0563	0.0586	0.0604
3	0.0416	0.0430	0.0481	0.0465	0.0424	0.0507	0.0501
4	0.0458	0.0432	0.0637	0.0470	0.0466	0.0624	0.0567
5	0.0507	0.0362	0.0767	0.0470	0.0436	0.0645	0.0574
6	0.0815	0.0608	0.1122	0.0708	0.0711	0.0985	0.0894
7	0.0591	0.0388	0.1046	0.0580	0.0529	0.0852	0.0756
8	0.0565	0.0316	0.1102	0.0541	0.0492	0.0871	0.0759
AVE	0.0555	0.0447	0.0786	0.0542	0.0513	0.0695	0.0642

MAPE: Mean Absolute Percentage Error; QA: Quarter Ahead; BVECM_G: Gibbs Sampled BVECM.

Table 6. MAPE(2001:1-2005:4): 91 days treasury bill rate

QA	VECM	BVECM($w = 0.1, d = 2$)	BVECM_G				
			($w = 0.3, d = 0.5$)	($w = 0.2, d = 1$)	($w = 0.2, d = 2$)	($w = 0.1, d = 1$)	($w = 0.1, d = 2$)
1	0.2592	0.0984	0.2163	0.1842	0.1148	0.1384	0.1189
2	0.2781	0.1147	0.3531	0.299	0.224	0.2588	0.2458
3	0.1809	0.3508	0.0907	0.0288	0.1097	0.0872	0.0896
4	0.2150	0.5387	0.1098	0.0724	0.1626	0.1583	0.1509
5	0.2782	0.1126	0.5491	0.4152	0.3376	0.3114	0.3429
6	1.0232	0.5894	1.1869	1.1563	1.0615	1.0262	1.0661
7	1.2847	0.8317	1.4501	1.5098	1.361	1.3168	1.3662
8	1.6072	1.0886	1.7072	1.8576	1.665	1.6052	1.6687
AVE	0.6408	0.4656	0.7079	0.6904	0.6295	0.6128	0.6311

MAPE: Mean Absolute Percentage Error; QA: Quarter Ahead; BVECM_G: Gibbs Sampled BVECM.

(it) Gibbs Sampled BVECM¹⁹ (BVECM_G) versus VECM: In this case, it is always possible to come up with a BVECM_G model, based on alternative specification of the hyperparameters of the priors, that produces lower out-of-sample forecast errors, measured by the average MAPE, when compared to the VECM. Except for the CPI and the investment expenditure, the BVECM_G model with the most tight prior ($w = 0.1, d = 2$) does better than the VECM. However, amongst the alternative specifications of the

¹⁹ We also estimated Gibbs sampled BVARs for the alternative prior specifications, however, as in Gupta (2006), the BVARs were outperformed by the Classical VECM and the BVECMs estimated with Theil's (1971) mixed estimation as well as Gibbs sampling. Hence, the results have not been reported.

BVECM_Gs, the model performs only third best for the measure of long- and short-term interest rates. For the two interest rate measures BVECM_G with $w = 0.1$, $d = 1$ is optimal amongst the alternative BVECM_Gs.

(Hi) BVECM_G *versus* the 'optimal' BVECM: Except for the Treasury bill rate and the investment expenditures, we can always come up with a BVECM_G model, based on alternative specification of the hyperparameters of the priors, that produces lower average MAPE, when compared to the optimal BVECM, obtained in Gupta (2006). Amongst the BVECM_Gs, the model with the most tight priors, *i.e.* $w = 0.1$, $d=2$, outperforms the 'optimal' BVECM for the household consumption expenditures, GDP and 10 years and longer government bond rate.

(iv) Amongst the BVECM_Gs: Across the BVECM_Gs the models with tighter priors outperform the ones with loose priors except for the CPI. The optimal BVECM_G for the six variables are as follows: BVECM_G with $w = 0.1$, $d=2$, for consumption and GDP; $w = 0.3$, $d = 0.5$, for CPI; $w = 0.1$, $d= 1$, for 10 years and longer government bond rate and the treasury bill rate, and; $w = 0.2$, $d=2$, for the investment expenditures.

5. CONCLUSIONS AND AREAS FOR FURTHER RESEARCH

This paper uses the Gibbs sampling technique to estimate a BVECM of the South African economy for the period 1970:1-2000:4 to forecast GDP, consumption, investment, short and long term interest rates, and the CPI over 2001:1 to 2005:4. The out-of-sample-forecast accuracy resulting from the BVECM is compared with those generated from a BVECM estimated using Theil's (1971) mixed estimation method and a standard Classical VECM. Note, traditionally, BVAR and BVEC models have been estimated using Theil's (1971) mixed estimation technique, where the error structure is assumed to follow the Gauss-Markov assumptions. However, in this paper, the estimation of the BVECMs based on Gibbs sampling, allows us to incorporate outliers or non-constant variances.

In summary, we find that the BVECM_Gs with tighter priors are better suited for forecasting our six-variable model of the South African economy over the period of 2001:1 to 2005:4. However, it is difficult to find a unique BVECM_G consistently outperforming the VECM and the 'optimal' BVECM obtained in Gupta (2006). Unlike, the optimal BVECM which produced lower forecasting errors on average for 5 variables (consumption expenditure, GDP, 10 years and longer government bond rate and the Treasury bill rate) when compared to the VECM, the BVECM_G with the most tight priors can only outperform the latter in the case of four variables (consumption expenditure, CPI, GDP, investment expenditures and the Treasury bill rate). But it must be emphasised that it is always possible to come up with a BVECM_G, based on alternative prior specifications, that produces lower out-of-sample forecast errors in comparison to the Classical VECM. However, when compared to the 'optimal' BVECM, the BVECM_G, with the most tight priors, outperforms the former for only three variables (household consumption expenditures, GDP and 10 years and longer government bond rate). So if we treat the VECM as the benchmark model, then clearly the 'optimal' BVECM of Gupta (2006) is better suited for forecasting the six-variable South African economy over the period of 2001:1 to 2005:4, as it outperforms the Classical VECM for all the variables except for the measure of the long-term interest rate.

There are, however, limitations to using the Bayesian approach. Firstly, the forecast accuracy depends critically on the specification of the prior, and secondly, the selection of the prior based on some objective function for the out-of-sample forecasts may not be 'optimal' for the time period beyond the period chosen to produce the out-of-sample forecasts. Besides these, there are two other major concerns which are, however, general to any traditional statistically estimated model, for example the VARs and the VECMs, both Classical and Bayesian in nature, used for forecasting at the business cycle frequencies. Such procedures perform reasonably well as long as there are no structural changes experienced in the economy. Such changes, whether in or out of the sample, would then render the models inappropriate. Alternatively, these models are not immune to the 'Lucas Critique'.²⁰ Furthermore, the estimation procedures used here are linear in nature, and, hence, they fail to take into account nonlinearities in the data.

One and, perhaps, the best response to these objections has been the development of micro-founded Dynamic Stochastic General Equilibrium (DSGE) models, which are capable of handling both the problems arising out of the structural changes and the issues of nonlinearities.²¹ The current trend in the forecasting-literature is clearly dominated by the use of estimated versions of DSGE models, both via Classical and Bayesian methods, which, in turn, have also been found to produce better forecasts relative to the traditional forecasting models. In this regard, some studies worth mentioning are: Hansen and Prescott (1993), Ingram and Whiteman (1994), Rotemberg and Woodford (1995), Ireland (2001) and Zimmermann (2001), to name a few. Future research involving DSGE models to forecast the South African economy is, hence, clearly an area to delve into.

REFERENCES

- DEMPSTER, A. P., LAIRD, N. M. and BUBIN, D. B. (1977). Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39: 1-38.
- DICKEY, D. A., JANSEN, D. W and THORNTON, D. L. (1991). A primer on cointegration with an application to money and income. *Federal Reserve Bulletin*, Federal Reserve Bank of St. Louis, March/April: 58-78.
- DOAN, T. A. (2000). *RATS User's Guide, Version, 5.0*. Illinois: Estima.
- , LITTELMAN, R. B. and SIMS, C. A. (1984). Forecasting and conditional projections using realistic prior distributions. *Econometric Reviews*, 3: 1-100.
- DOWD, M. R. and LeSAGE, J. P. (1997). Analysis of spatial contiguity influences on state price level formation. *International Journal of Forecasting*, 13: 245-253.
- DUA, P. and RAY, S. C. (1995). A BVAR model for the Connecticut economy. *Journal of Forecasting*, 14: 167-180.
- and SMYTH, D. J. (1995). Forecasting U. S. home sales using BVAR models and survey data on households' buying attitude. *Journal of Forecasting*, 14: 167-180.
- , MILLER, S. M. and SMYTH, D. J. (1999). Using leading indicators to forecast U. S. home sales in a Bayesian vector autoregressive framework. *Journal of Real Estate Finance and Economics*, 18: 191-205.
- ENGLE, R. F. and GRANGER, C W. J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica*, 55: 251-276.

²⁰ See Lucas (1976) for details.

²¹ For a detailed review of the literature on the use of DSGE models for forecasting, see Zimmermann (2001).

- GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6: 721-741.
- GEWEKE, J. (1993). Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics*, 8, Supplement: Special Issue on Econometric Inference Using Simulation Techniques, S19-S40.
- GUPTA, R. (2006). Forecasting the South African economy with VARs and VECMs. *South African Journal of Economics*, 74, Forthcoming.
- and SICHEI, M. M. (2006). A BVAR model for the South African economy. *South African Journal of Economics*, 74: 391-409.
- HAFER, R. W and SHEEHAN, R. G. (1989). The sensitivity of VAR forecasts to alternative lag structures. *International Journal of Forecasting*, 5: 399-408.
- HAMILTON, J. D. (1994). *Time Series Analysis*, 2nd edition. Princeton: Princeton University Press.
- HANSEN, G. D. and PRESCOTT, E. C. (1993). Did technology shocks cause the 1990-91 recession? *American Economic Review Proceedings*, 83(2): 280-285.
- INGRAM, B. E and WHITEMAN, C. H. (1994). Supplanting the 'Minnesota prior: Forecasting macroeconomic time series using real business cycle model priors. *Journal of Monetary Economics*, 34 (3): 497-510.
- IRELAND, P. N. (2001). Technology shocks and the business cycle: An empirical investigation. *Journal of Economic Dynamics and Control*, 25: 703-719.
- JØHANSEN, S. (1988). Statistical analysis of co-integration vector. *Journal of Economic Dynamics and Control*, 12: 231-254.
- (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.
- KADIYALA, K. R. and KARLSSON, S. (1997). Numerical methods for estimation and inference in Bayesian VAR models. *Journal of Applied Econometrics*, 12: 99-132.
- LeSAGE, J. P. (1990). A comparison of the forecasting ability of ECM and VAR models. *Review of Economics and Statistics*, 72 (4): 664-671.
- (1999). *Applied Econometrics Using MATLAB*. Available at: <http://www.spatial-econometrics.com>.
- and KRIVELYOVA, A. (1999). A spatial prior for Bayesian autoregressive models. *Journal of Regional Science*, 39: 297-317.
- and PAN, Z. (1995). Using spatial contiguity as Bayesian prior information in regional

forecasting models.

International Regional Science Review, 18: 33-53.

LITTERMAN, R. B. (1981). A Bayesian procedure for forecasting with vector autoregressions.

Working Paper, Federal Reserve Bank of Minneapolis.

----- (1986). Forecasting with Bayesian vector autoregressions — Five years of experience. *Journal of Business and*

Economic Statistics, 4 (1): 25-38.

LUCAS, R. E., Jr. (1976). Econometric policy evaluation: A critique. *Carnegie Rochester Conference Series on Public Policy*,

1: 19-46.

NI, S. and SUN, D. (2005). Bayesian estimators for vector-autoregressive models. *Journal of Business and Economic*

Statistics, 23 (1): 105-117.

ROTEMBERG, J. and WOODFORD, M. (1995). Dynamic general equilibrium models with imperfectly competitive

product markets. In T F. Cooley (ed.), *Frontiers of Business Cycle Research*. New Jersey: Princeton University Press.

SIMS, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48: 1-48.

-----, STOCK, J. H. and WATSON, M. W (1990). Inference in linear time series models with some unit roots.

Econometrica, 58: 113-144.

SPENCER, D. E. (1993). Developing a Bayesian vector autoregression model. *International Journal of Forecasting*, 9,

407-421.

THEIL, H. (1971). *Principles of Econometrics*. New York: John Wiley.

TODD, R. M. (1984). Improving economic forecasting with Bayesian vector autoregression. *Quarterly Review*, Federal

Reserve Bank of Minneapolis, Fall: 18-29.

ZIMMERMANN, C. (2001). Forecasting with real business cycle models. *Indian Economic Review*, 36(1): 189-203.