

Landon-Lane, John

**Working Paper**

## The impact on forecasts and impulse responses of restricting drift in a Vector Autoregression

Working Paper, No. 2001-14

**Provided in Cooperation with:**

Department of Economics, Rutgers University

*Suggested Citation:* Landon-Lane, John (2001) : The impact on forecasts and impulse responses of restricting drift in a Vector Autoregression, Working Paper, No. 2001-14, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:

<https://hdl.handle.net/10419/79177>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# The Impact on Forecasts and Impulse Responses of Restricting Drift in a Vector Autoregression

John Landon-Lane<sup>\*†</sup>

Rutgers University and The University of New South Wales

November 1, 2001

## Abstract

Inference on the long-run properties of a Vector Autoregression (VAR) consisting wholly of  $I(1)$  variables are made using Bayesian methods. In particular, the implications on the forecast and impulse response function distributions of directly estimating and restricting the drift parameters of variables in a VAR are studied. It is found that the forecast and impulse response distributions are sensitive to the prior distribution placed over the drift parameters.

**Keywords:** Vector Autoregression, Impulse Response Function, Long-Run Forecasts, Markov chain Monte Carlo, Gibbs-Sampler

**JEL Classification:** C32, C11, C13

---

<sup>\*</sup>Contact address: Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA. E-mail: lane@econ.rutgers.edu

<sup>†</sup>I would like to thank members of the Time Series / Applied Macroeconomic group at UNSW, in particular Ron Bewley, Glenn Otto and Minxian Yang for their comments. I would also like to thank John Geweke for very helpful comments. Support from the *Australian Research Council* is gratefully acknowledged.

# 1 Introduction

Vector Autoregressions (VAR) have become a very popular tool used in forecasting and policy analysis since the influential work of Sims (1980).<sup>1</sup> In VAR's that contain non-stationary variables it has been shown there is the possibility that linear relationships between the levels of the variables can improve forecasting performance<sup>2</sup>. In these cases the VAR can be represented as a vector error correction (VEC) model which have their foundations in Sargan (1964) and Davidson, Hendrey, Srba and Yeo (1978).<sup>3</sup>

Along with model and parameter uncertainty, the imprecision of the estimates of the drift parameters in a Vector Autoregression (VAR) containing only series that are integrated of order one (I(1)) has long been acknowledged as a source of poor forecasting performance in the VAR (Clements and Hendry 1998). An imprecise estimate of the drift that is apparent in a variable is more serious the longer the forecast horizon since forecasts, from such a model, converge asymptotically to linear trends with slope equal to the estimated drift.

A number of commonly modelled economic variables, such as inflation, interest rates, and exchange rates are generally thought to not exhibit non-zero drift. Since other variables such as GDP and money supply exhibit strong drift, many VAR's that have been used in modeling and forecasting contain a mixture of drifting and non-drifting variables, the "mixed-drift" case. In these cases it has been shown, in Bewley (2000) and others, that it is possible to spuriously estimate non-zero drift for variables that, *a priori*, would not be expected to have drift. This, clearly, has rather dramatic impacts on the forecasting performance of the VAR.

---

<sup>1</sup>see Hamilton (1994) for a short review of VAR's and their uses in forecasting and policy analysis.

<sup>2</sup>Theoretical support for this can be found in Engle and Yoo (1987) and empirical support for this can be found in LeSage (1990) and Bewley and Parry (1991) among others

<sup>3</sup>see Banerjee, Dolado, Galbraith and Hendrey (1993) for a review of the error correction and co-integration literature

Along with macroeconomic forecasting, VAR models are extensively used by policy makers. In particular, the orthogonalized impulse response function implied by the VAR is an important tool in macroeconomic modeling<sup>4</sup>. While the drift parameters do not play a direct role in the calculation of the impulse response function, it is shown that correctly estimating the drift can have an impact on estimates of the parameters of the VAR which in turn impacts upon the estimates of the orthogonalized impulse response function. In essence, the impulse response function is a long-run detrended forecast from a VAR and given that it is common to report impulse response functions for up to 24 quarters or 48 months, it is clear that small differences in parameter estimates could have important effects on the long run behaviour of the various “forecasts” of the VAR. Therefore, it is desirable, when estimating a VAR with variables of mixed drift, to be able to restrict the drift for some variables in a VAR to be zero when this is appropriate.

A natural way to incorporate prior beliefs on the drift parameters of a VAR is via Bayesian methods. One problem with making inferences on the long-run properties of a VAR, such as the value of the drift parameters, is that there usually is very little information in the data about the long-run. In such a case, it would be reasonable to expect that the prior beliefs could have some influence on the estimated long-run properties of the VAR.

Incorporating prior beliefs into a VAR via Bayesian methods has become a standard method in the macroeconomic forecasting literature. Bayesian VAR's (BVAR's) aim to improve forecasting performance by reducing the effect of parameter uncertainty. It is well known (for example see Doan, Litterman and Sims (1984), Todd (1984) and Litterman (1986)) that priors that shrink the VAR to a model consisting of purely random walks have proved successful at forecasting reasonably large systems. More recently, a number of other priors, developed from Litterman priors, have been introduced to incorporate the possibility of co-integration into the BVAR. For example, see Sims and Zha (1998) and Waggoner and Zha

---

<sup>4</sup>see Sims (1992) and Sims (1980) for a discussions and uses of the orthogonalized impulse response function.

(1999) for a discussion of these extensions.

The approach advocated in this paper, unlike Bewley (2000) who only does mixed estimation, is to use Markov chain Monte Carlo methods to directly estimate and restrict the drift parameters in a VAR, thus allowing restrictions on the drift parameters to be easily implemented, while still allowing for the use of existing well established prior distributions for the parameters that make up the VAR. In particular, the SUR representation of a VAR is utilized in designing a Gibbs-sampling scheme to estimate the VAR with restrictions on the drift parameters. In the second part of the paper, the sensitivity of the forecast and impulse response function distributions to restrictions on the values of drift parameters are studied. The general method to directly estimate the drift parameters of a VAR consisting wholly of I(1) variables is introduced in Section 2 and an illustration of the method is given in Section 3.

## 2 Model and Estimation

Consider an  $n$  dimensional vector,  $y_t$ , consisting wholly of I(1) variables. Suppose that this vector of I(1) variables can be represented by a VAR of order  $p$ . That is,

$$y_t = a + \sum_{i=1}^p A_i y_{t-i} + \epsilon_t \quad (1)$$

where  $a$  is a  $n \times 1$  vector of constants and  $A_i$  is an  $n \times n$  coefficient matrix for each  $i = 1, \dots, p$ .

One can follow Clements and Hendry (1998) and rewrite (1) as

$$[\Delta y_t - \delta] = \gamma [\alpha' y_{t-1} - \mu] + \sum_{i=1}^{p-1} B_i [\Delta y_{t-i} - \delta] + \epsilon_t, \quad (2)$$

where  $\delta$  is an  $n \times 1$  vector of drift parameters,  $\alpha'$  is an  $r \times m$ ,  $0 \leq r \leq n$ , matrix of cointegrating vectors,  $\mu$  is their unconditional mean and

$$B_j = - \sum_{i=j+1}^p A_i.$$

It is simple then to show that the constant vector,  $a$ , in (1) is given by

$$a = (I - \sum_{i=1}^{p-1} B_i)\delta - \gamma\mu. \quad (3)$$

Thus, unless  $\mathbf{B}_i = 0$  for all  $i$  and  $\gamma = 0$ , it is possible to have all elements of  $a$  non-zero even if some elements of  $\delta$  are zero. Thus using the estimated constants in forecasting as estimates of the drift parameters can lead to misleading results. Therefore, one needs to be able to estimate the drift parameters,  $\delta$ , directly. Even when  $\delta$  can be directly estimated non-linearly there is still the issue that in the data there may appear to be a drift present in the variable. For example, if one looked at US short term interest rates for the late 60s, 70s and early 80s (Figure 1, Section 3) there would appear to be a strong drift present. However, *a priori*, one would not expect an interest rate to contain a non-zero drift, especially in the long run. Therefore, it is also necessary to be able to place restrictions over the drift parameters in a VAR.

Abstracting from the case where there are co-integrating relationships evident in the data<sup>5</sup>, the analysis will concentrate on the VAR in differences(DVAR) given by

$$[\Delta y_t - \delta] = \sum_{i=1}^{p-1} B_i[\Delta y_{t-i} - \delta] + \epsilon_t. \quad (4)$$

That is, the variables in the DVAR are I(1) but not co-integrated.

There are a number of ways in which the parameters,  $\delta$  and  $B_i$  can be estimated. One could use non-linear estimation procedures, such as maximum likelihood, on (2). However, in the case of maximum likelihood the size of the parameter space can get quite large very quickly. For example, a VAR(5) for a vector of dimension 4 contains 72 unknown parameters. In this case it would be quite difficult to impose linear restrictions on the drift parameters,  $\delta$ . Alternatively, one could follow Bewley (2000) and apply the Bewley (1979) transformation to the first difference version of (1) yielding an exactly identified simultaneous equation system

---

<sup>5</sup>see Section 4 for a short discussion of how to extend the method to the case of co-integration

that is linear in  $\delta$ . The transformed system can then be estimated using 2SLS and linear restrictions on  $\delta$  can easily be implemented but then 3SLS or iterated 3SLS would need to be used .

In the context of macroeconomic forecasting, it has long been acknowledged that imposing prior information improves forecasting performance (see Litterman (1986)). Priors for the parameters  $A_i$  and  $B_i$  are well developed and have known properties. The Bewley (1979) transformation transforms the parameters in (1) in a non-linear fashion. For example, the Bewley transformed version of the first difference form of (1) is

$$\Delta y_t = \delta + \sum_{i=0}^{p-2} C_i \Delta^2 y_{t-i} + \zeta_t \quad (5)$$

where

$$C_i = \left[ I - \sum_{m=1}^{p-1} B_m \right]^{-1} \sum_{k=j+1}^{p-1} B_k$$

for  $j = 0, \dots, p-2$ . The question of what would be the appropriate priors for the coefficient matrices  $C_i$  is still an open question. The aim of this paper, therefore, is to develop a sampling scheme that is able to estimate (2), and hence be able to use existing well-defined priors for the coefficient matrices, while placing restrictions on the drift parameters,  $\delta$ .

The sampling algorithm that is developed here makes use of the fact that the VAR representation can be reformulated into a seemingly unrelated regressions (SUR) form. The SUR form is easy to estimate using standard Bayesian methods and allows for cross equation equality restrictions on the drift parameters to be easily imposed.

Consider the following model specification. Conditional on  $\delta$ , the model, (2), can be written as

$$\tilde{\mathbf{y}} = \tilde{\mathbf{Z}}\beta + \epsilon \quad (6)$$

where

$$\tilde{\mathbf{y}} = (\hat{\Delta \mathbf{y}}'_1, \dots, \hat{\Delta \mathbf{y}}'_n)',$$

$$\tilde{\mathbf{Z}} = I_n \otimes \mathbf{X}, \quad \mathbf{X} = [\hat{\Delta \mathbf{y}}_{1,-1} \quad \dots \quad \hat{\Delta \mathbf{y}}_{n,-(p-1)}],$$

and

$$\hat{\Delta}\mathbf{y}_i = \begin{pmatrix} \Delta y_{i1} - \delta_i \\ \vdots \\ \Delta y_{iT} - \delta_i \end{pmatrix}.$$

Note that  $\hat{\Delta}y_{i,-j}$  represents the  $T \times 1$  vector of observations  $(\hat{\Delta}y_{i,1-j}, \dots, \hat{\Delta}y_{i,T-j})'$ . Also note that  $\tilde{\mathbf{y}}$  and  $\tilde{\mathbf{Z}}$  are functions of  $\mathbf{y}$  and  $\delta$ .

The distributional assumption for  $\epsilon$  is that

$$\epsilon \sim N(\mathbf{0}, \mathbf{H}^{-1} \otimes I_T), \quad (7)$$

where  $\mathbf{H}$  is an  $n \times n$  precision matrix. Finally note that the coefficient vector  $\beta$  is the row-wise vectorisation of the concatenated coefficient matrix  $[B_1, \dots, B_{p-1}]$ .

Thus, conditional on  $\delta$ , we have the standard SUR formulation of a VAR. The data density in this case is

$$p(\mathbf{y}|\delta, \beta, \mathbf{H}) = (2\pi)^{-Tm/2} |\mathbf{H}|^{T/2} \exp\left(-\frac{1}{2}(\tilde{\mathbf{y}} - \tilde{\mathbf{Z}}\beta)'(\mathbf{H} \otimes I_T)(\tilde{\mathbf{y}} - \tilde{\mathbf{Z}}\beta)\right). \quad (8)$$

Define  $\mathbf{S}$  to be the  $n \times n$  matrix defined by

$$\mathbf{S} = [s_{ij}], \quad s_{ij} = (\tilde{y}_i - \tilde{\mathbf{X}}_i\beta_i)'(\tilde{y}_j - \tilde{\mathbf{X}}_j\beta_j), \quad (9)$$

where  $\tilde{y}_k$  and  $\beta_k$  are the obvious partitions of  $\tilde{\mathbf{y}}$  and  $\beta$  pertaining to the  $k^{th}$  equation of the VAR. Then (8) can be simplified to

$$p(\mathbf{y}|\delta, \beta, \mathbf{H}) = (2\pi)^{-Tm/2} |\mathbf{H}|^{T/2} \exp\left(-\frac{1}{2}tr\mathbf{S}\mathbf{H}\right). \quad (10)$$

Conditionally conjugate priors are assigned to the parameters  $\delta$ ,  $\beta$ , and  $\mathbf{H}$ . In particular, independent Gaussian priors are assigned for  $\delta$  and  $\beta$  with

$$\beta \sim N(\underline{\beta}, \underline{\mathbf{H}}_\beta^{-1}) \quad \text{and} \quad \delta \sim N(\underline{\delta}, \underline{\mathbf{H}}_\delta^{-1}),$$



and a Wishart prior is assigned for  $\mathbf{H}$ , that is

$$\mathbf{H} \sim \text{Wish}(\underline{\mathbf{S}}^{-1}, \underline{\nu}).$$

Geweke (1999) shows that the conditional distributions for  $\beta$  and  $\mathbf{H}$  are

$$\beta|\delta, \mathbf{H}, \mathbf{y} \sim N(\bar{\beta}, \bar{\mathbf{H}}_{\beta}^{-1}),$$

where  $\bar{\mathbf{H}}_{\beta} = \underline{\mathbf{H}}_{\beta} + (\mathbf{H} \otimes \mathbf{X}'\mathbf{X})$  and  $\bar{\beta} = \bar{\mathbf{H}}_{\beta}^{-1} [\underline{\mathbf{H}}_{\beta}\underline{\beta} + (\mathbf{H} \otimes \mathbf{X}')\mathbf{y}]$ , and

$$\mathbf{H}|\delta, \beta, \mathbf{y} \sim W((\underline{\mathbf{S}} + \mathbf{S})^{-1}, \underline{\nu} + T).$$

Thus the conditional posterior distributions for  $\beta$  and  $\mathbf{H}$  are known and can be draws can be made from them directly using standard techniques.

In order to complete the Gibbs-sampling algorithm a posterior conditional distribution for  $\delta|\beta, \mathbf{H}, \mathbf{y}$  needs to be derived. This conditional distribution is derived using (2). For each  $t = 1, \dots, T$

$$\Delta y_t = \mathbf{E}\delta + \sum_{j=1}^{p-1} \mathbf{B}_j \Delta y_{t-j} + \epsilon_t \quad (11)$$

where  $\mathbf{E} = (I - \sum_{j=1}^{p-1} \mathbf{B}_j)$  and  $\epsilon_t \sim N(0, \mathbf{H}^{-1})$ . Then the conditional posterior density for  $\delta$  is

$$p(\delta|\beta, \mathbf{H}, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2}(\delta - \underline{\delta})' \underline{\mathbf{H}}_{\delta} (\delta - \underline{\delta}) \right\} \prod_{t=1}^T \exp \left\{ -\frac{1}{2} [v_t - E\delta]' \mathbf{H} [v_t - E\delta] \right\} \quad (12)$$

where  $v_t = (\Delta y_t - \sum_{j=1}^{p-1} \mathbf{B}_j \Delta y_{t-j})$ .

Completing the square, it can be shown that

$$\delta|\beta, \mathbf{H}, \mathbf{y} \sim N(\bar{\delta}, \bar{\mathbf{H}}_{\delta}^{-1}) \quad (13)$$

where

$$\bar{\mathbf{H}}_{\delta} = T \cdot \mathbf{E}' \mathbf{H} \mathbf{E} + \underline{\mathbf{H}}_{\delta}$$

and

$$\bar{\delta} = \overline{\mathbf{H}}_{\delta}^{-1} \left[ \underline{\mathbf{H}}_{\delta} \underline{\delta} + \mathbf{E}' \mathbf{H} \sum_{t=1}^T \left( \Delta y_t - \sum_{j=1}^{p-1} \mathbf{B}_j \Delta y_{t-j} \right) \right]$$

The general sampling procedure that will be used is a three block Gibbs sampler with the blocks being  $\delta$ ,  $\beta$  and  $\mathbf{H}$ . The three block Gibbs-Sampling algorithm is as follows. For each  $i = 1, \dots, (M-1)$ , given a draw from  $p(\delta, \beta, \mathbf{H}|\mathbf{y})$ ,  $(\delta^{(i)}, \beta^{(i)}, \mathbf{H}^{(i)})$ , a successive draw is made from  $p(\delta, \beta, \mathbf{H}|\mathbf{y})$  using the following algorithm:

$$\begin{aligned} \delta^{(i+1)} &\sim p(\delta|\beta^{(i)}, \mathbf{H}^{(i)}, \mathbf{y}) \\ \beta^{(i+1)} &\sim p(\beta|\delta^{(i+1)}, \mathbf{H}^{(i)}, \mathbf{y}) \\ \mathbf{H}^{(i+1)} &\sim p(\mathbf{H}|\delta^{(i+1)}, \beta^{(i+1)}, \mathbf{y}) \end{aligned} \tag{14}$$

where the condition posterior distributions for  $\delta$ ,  $\beta$  and  $\mathbf{H}$  are given above. The Gibbs-sampler described above satisfies the sufficient condition (Geweke 1999) for the sampler to have the unique invariant distribution of  $p(\delta, \beta, \mathbf{H}|\mathbf{y})$  and an ergodic transition kernel. This implies, following the result in Tierney (1994), that for all absolutely integrable functions,  $g(\delta, \beta, \mathbf{H})$

$$M^{-1} \sum_{i=1}^M g(\delta^{(i)}, \beta^{(i)}, \mathbf{H}^{(i)}) \longrightarrow E[g(\delta, \beta, \mathbf{H})|\mathbf{y}] \tag{15}$$

provided  $E[g(\delta, \beta, \mathbf{H})|\mathbf{y}]$  exists.

### 3 An Illustration

As an illustration of the algorithm defined in Section 2 above, a VAR(5) consisting of a system of two variables is estimated. The data consist of the log of a seasonally adjusted index of industrial production,  $Y$ , and a short interest rate,  $R$ , for the United States (U.S.). Industrial production enters as a logarithms while the interest rate enters as a percentage, with the ordering being that interest rates contemporaneously affect industrial production. Quarterly observations from 1958Q1 through 1995Q2 are used. The data are depicted in Figure 1.

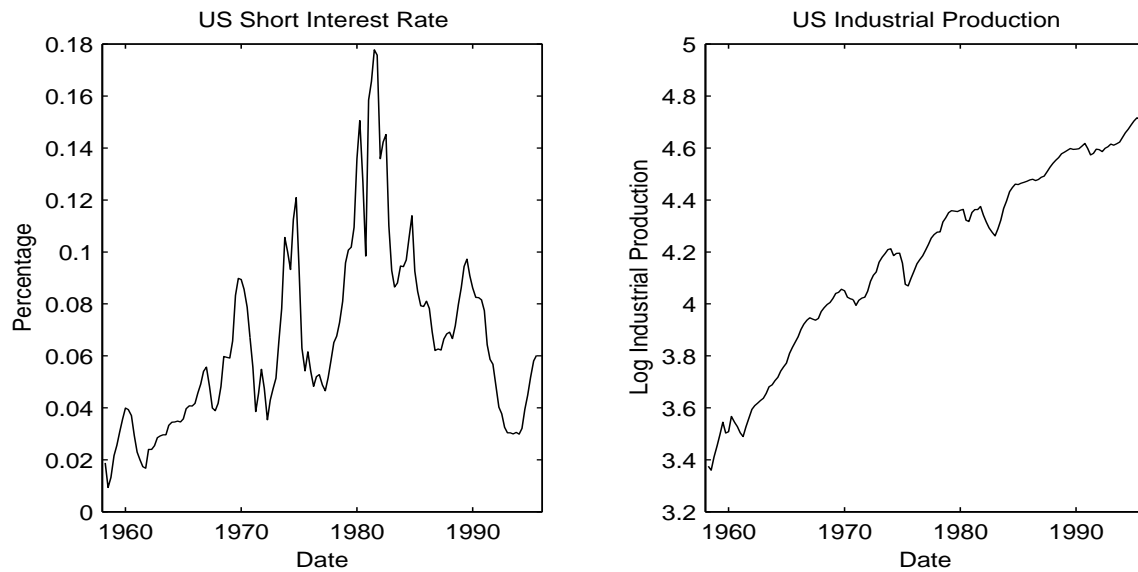


Figure 1: US Interest Rates and Industrial Production

Results from augmented Dickey-Fuller tests, reported in Table 1, indicate that both  $R$  and  $Y$  contain a unit root.

Table 1: Unit Root Tests for  $R$  and  $Y$

	$R$ (constant)	$Y$ (constant + trend)
t-stat	-1.9176	-1.6979
	(-2.88)	(-3.42)

It is clear from Figure 1 that  $R$  and  $Y$  are not co-integrated. Hence, we have a “mixed-drift” VAR.

For this illustration, as is common in the literature, the model that is estimated includes five lags of the dependent variable. This is a common assumption for models that use quarterly data.

Two sets of estimates are reported. The first set of estimates are for the unrestricted DVAR

given by

$$\begin{pmatrix} \Delta R_t \\ \Delta Y_t \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \sum_{i=1}^4 \mathbf{B}_i \begin{pmatrix} \Delta R_{t-i} \\ \Delta Y_{t-i} \end{pmatrix} + \epsilon_t \quad (16)$$

and the second set of estimates will be for the modified DVAR

$$\begin{pmatrix} \Delta R_t - \delta_1 \\ \Delta Y_t - \delta_2 \end{pmatrix} = \sum_{i=1}^4 \mathbf{B}_i \begin{pmatrix} \Delta R_{t-i} - \delta_1 \\ \Delta Y_{t-i} - \delta_2 \end{pmatrix} + \epsilon_t, \quad (17)$$

where

$$\mathbf{B}_i = \begin{pmatrix} \beta_{i,11} & \beta_{i,12} \\ \beta_{i,21} & \beta_{i,22} \end{pmatrix}$$

and

$$\epsilon_t \sim N(\mathbf{0}, \mathbf{H}^{-1}).$$

For the first set of estimates, hereafter referred to as the unrestricted estimates, no restrictions are placed on the drift parameters while for the second set of estimates, hereafter referred to as the restricted estimates, the drift parameter for R was restricted to be 0. The parameter  $\beta_u$  refers to the estimates of  $\mathbf{B}_j$  for (16) and the parameter  $\beta_r$  refers to the estimates of  $\mathbf{B}_j$  for (17).

The two sets of estimates will be used to calculate long-run forecasts and orthogonalized impulse response functions for each variable, R and Y. Before analyzing the results a description of the priors used in this illustration is given in the following section.

### 3.1 Prior Specification

Proper, but reasonably diffuse, priors are placed on all the unknown parameters of (16) and (17). For both cases, the prior on the parameter vector  $\beta$  is similar to the standard Litterman prior in that the prior mean for  $\beta$  is  $\mathbf{0}$ . In this case the mean of the prior implies that the two data series are simple random walks, with or without drift. However, as opposed

to the Litterman priors, the prior variances of the  $\mathbf{B}_j$  coefficient matrices are chosen to be relatively large. This is because the Litterman prior places quite stronger priors that  $\mathbf{B}_i = 0$  as  $i$  increases. In terms of forecasting this has little effect as the forecasting performance is dominated by the drift estimates in the long run. However, if we are interested in calculating impulse response functions then we are interested in estimates of the  $\mathbf{B}_i$ 's. In this case, placing restrictive priors that act to shrink the parameter matrices to zero matrices would not be prudent.

In this illustration, the prior precision matrix,  $\mathbf{H}_\beta$ , is set to be  $0.25\mathbf{I}_k$ , where  $k$  is the dimension of  $\beta_r$  or  $\beta_u$  respectively.

The prior for the precision matrix  $\mathbf{H}$ , chosen to be independent of the prior for  $\beta$  and  $\delta$ , is a Wishart distribution with scale parameter equal to 10 and a matrix parameter equal to  $\underline{\mathbf{S}}^{-1}$  where  $\mathbf{S}$  was chosen so that the mean of the prior is equal to the sample variance-covariance matrix of the data. That is the prior for the precision matrix is

$$p(\mathbf{H}) \sim Wish(\underline{\mathbf{S}}^{-1}, 10). \quad (18)$$

One can think of the Wishart distribution as the multivariate version of the  $\chi^2$  distribution with the scale parameter representing the degrees of freedom. The prior for  $\mathbf{H}$  can be thought of as representing information on  $\mathbf{H}$  from a notional sample of size 10. Thus, in this respect one can think of this prior as being reasonably diffuse.

The prior for the drift parameters,  $\delta$ , in (17) was chosen to be the following. As we would not expect interest rates to exhibit a drift in the long run a prior mean of zero was chosen. The prior mean for the drift in the output variable was also chosen to be 0. In practice, one would probably have some strong prior beliefs that the drift in output was non-zero but in order to be consistent with the unrestricted model a prior mean of 0 was chosen. The prior precision for the drift in output was set to be 0.25 (equating to a variance of 4) which implies a quite diffuse prior distribution. The prior variance for  $\delta_1$ , the drift term for short term US interest rates was set to 0 for the constrained case. That is, constraining  $\delta_1$  to be zero can

be thought of as placing a dogmatic prior of mean zero and variance zero on  $\delta_1$ .

That is, for the constrained case the prior for  $\delta$  is

$$p(\delta_1) = \begin{cases} 1 & \text{if } \delta_1 = 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad p(\delta_2) \sim N(0, 4). \quad (19)$$

The prior for  $\beta_u$  has mean 0 and precision  $0.25\mathbf{I}_{18}$ . Note that in this case there is not an explicit prior for the drift, but rather for the constant parameters that appear in the equation. However, using (3), it is easy to see that the mean of implicit prior on the drift parameters is also 0. In order to see how the prior and the posterior distribution for  $\delta$  relate to each other for each model see Figure 2. The prior density for both the unrestricted and restricted estimator are the flat solid lines at the bottom of each graph. It is impossible to distinguish the two prior distributions when compared with the posterior. It is also clear from the Figure 2 that the prior has very little impact on the posterior distribution.

## 3.2 Results

The Gibbs-sampling algorithm described in Section 2 was used to make 10,000 draws from the posterior distribution,  $p(\delta, \beta, \mathbf{H} | R_T, Y_T)$ , for both the unrestricted and restricted cases. The posterior means and standard errors for both models are reported in Table 2<sup>6</sup>. In all cases it was clear that the Markov chains had converged to their invariant distribution after 1000 iterations. Hence only the last 9000 iterations were used in the posterior calculations that follow.

Orthogonalized impulse response functions are also computed for each draw from the posterior distribution. In order to reduce the computational burden of simulating impulse response functions for the 9000 iterations, only 1000 of the 9000 iterations were used. The 1000 it-

---

<sup>6</sup>Computations were undertaken, in part, using the Bayesian Analysis, Computation and Communication software (<http://www.econ.umn.edu/~bacc>) described in Geweke (1999)

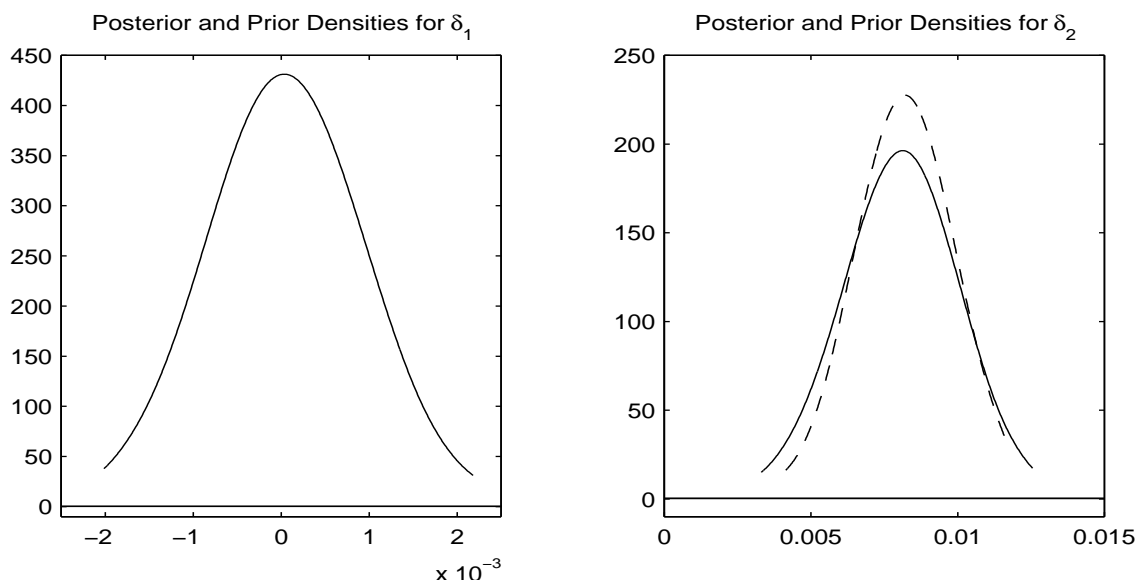


Figure 2: Posterior and Prior Densities for Drift Parameters

Note: The posterior for the unrestricted estimates are depicted by the solid lines while the posterior for the restricted estimates is depicted by the dashed line

erations used for the calculation of the posterior were obtained by taking every 9 iterations starting from iteration 1001 and ending at iteration 10000. In all cases the relative numerical efficiency of the posterior estimates were greater than 0.8.

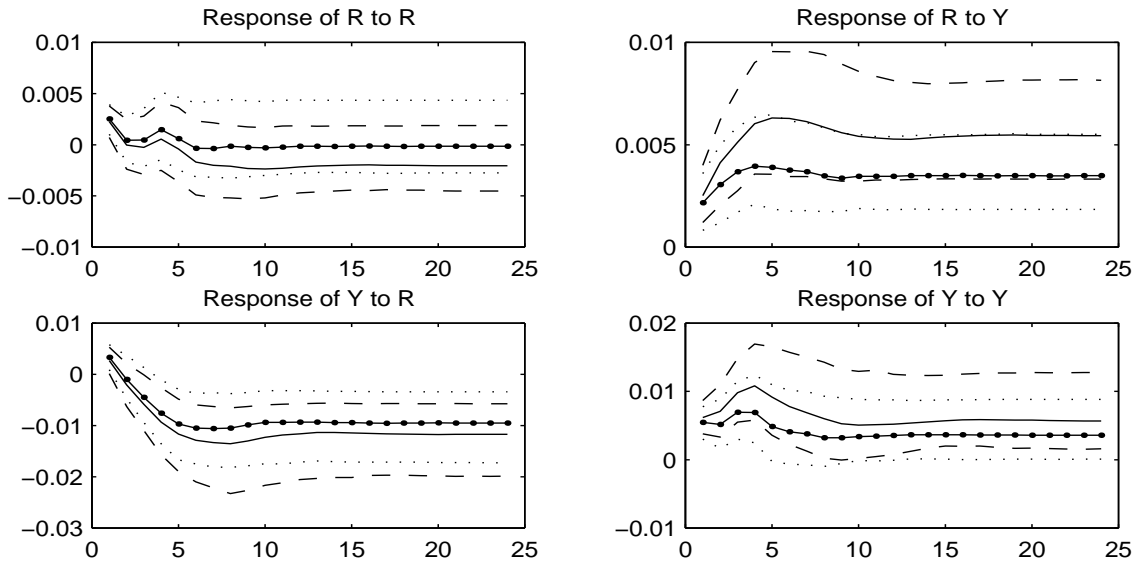
It is clear from the posterior means reported in Table 2 that restricting the drift in the interest rate variable has some impact on the parameter estimates for the coefficient matrices,  $\mathbf{B}_i$ , and the precision matrix  $\mathbf{H}$ . The impact this has on the impulse response functions can be seen in Figure 3. The solid line represents the median impulse response function calculated using the unrestricted estimates while the solid line with dots represents the median impulse response function calculated using the restricted estimates. The dashed line are the 5th and 95th percentiles of the impulse response function using the unrestricted estimates while the dotted lines represent the 5th and 95 percentiles of the impulse response function using the

Table 2: Posterior Estimates for (17)

	Unrestricted Estimates		Restricted Estimates	
$\delta'$	$5.7 \times 10^{-5}$	0.0081	0	0.0082
	(0.0012)	(0.0019)	—	(0.0016)
$\mathbf{B}_1$	0.1172	0.1671	0.1602	0.1414
	(0.0931)	(0.0558)	(0.0890)	(0.0551)
	0.0096	0.4111	0.1028	0.3574
	(0.1495)	(0.0915)	(0.1511)	(0.0911)
$\mathbf{B}_2$	-0.3078	0.0140	-0.2799	-0.0119
	(0.0908)	(0.0597)	(0.0907)	(0.0588)
	-0.5094	-0.1132	-0.4486	-0.1668
	(0.1467)	(0.0960)	(0.1491)	(0.0967)
$\mathbf{B}_3$	0.0915	0.0957	0.1211	0.0743
	(0.0951)	(0.0576)	(0.0932)	(0.0569)
	-0.2222	0.2891	-0.1621	0.2454
	(0.1525)	(0.0922)	(0.1550)	(0.0937)
$\mathbf{B}_4$	0.0388	0.0002	0.0468	-0.0289
	(0.0951)	(0.0554)	(0.0952)	(0.0528)
	-0.3554	-0.0286	-0.3384	-0.0906
	(0.1551)	(0.0904)	(0.1568)	(0.0887)
$\mathbf{H}$	11518.60	-2743.64	11553.05	-2767.15
	(1344.03)	(632.18)	(1355.25)	(628.03)
		4390.90		4345.04
		(505.17)		(514.17)



Figure 3: Posterior Distribution of Cumulative Impulse Response Functions



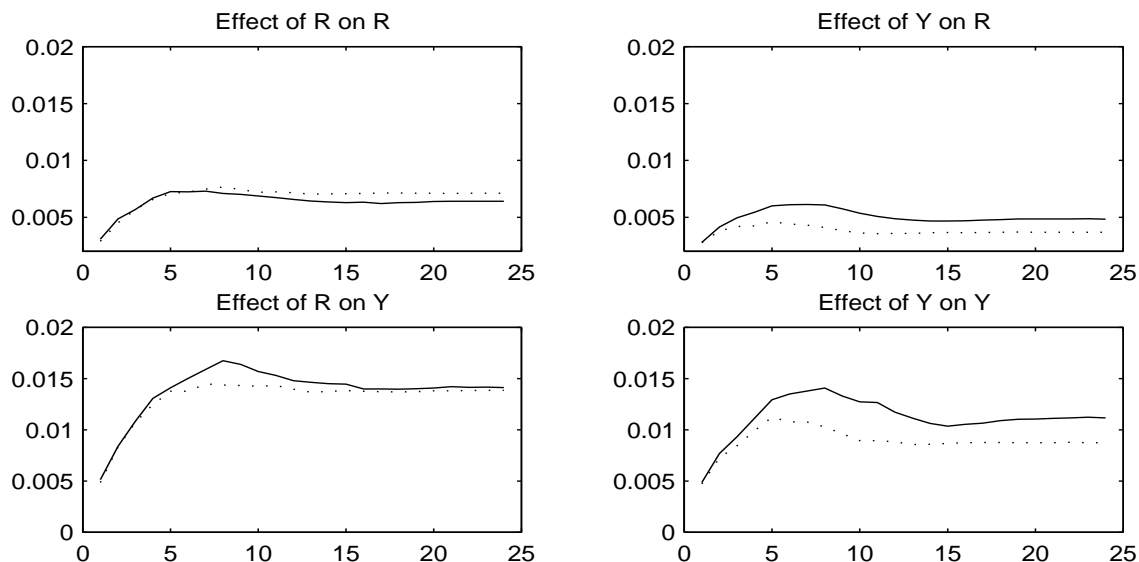
restricted estimates.

There appears that the distribution of the impulse response functions using the restricted estimated is somewhat different from the impulse response functions using the unrestricted estimates. This is apparent by the fact that the median of the distributions do not cross. In all cases the median for the impulse response functions calculated using the restricted estimates are closer to zero than is the median of the impulse response function calculated using the unrestricted estimates. In the case of the impact on interest rates of a one standard deviation shock to output, the median of the distribution of impulse response functions calculated using unrestricted estimates all but lies outside the 90% interquartile range of the restricted distribution. While there is overlap of the distributions, they only overlap in their respective tails. Hence there appears that restricting the drift in the interest rate variable has had a reasonably significant impact on the distribution of the calculated impulse response functions.

Figure 4 contains the 95% interquartile ranges of the distribution for each impulse response function. In three out of the four cases the interquartile range is smaller for the distribution of

the impulse response functions calculated using the restricted estimates than for the impulse response functions calculated using the unrestricted estimates.

Figure 4: Interquantile Range of Cumulative Impulse Response Function Distributions



The orthogonalized impulse response functions reported above are functions of the estimates of the coefficient matrices,  $\mathbf{B}_i$ , and  $\mathbf{H}$ . Another important application of VAR models is in forecasting. The important parameters here are the drift parameters,  $\delta$ , and the coefficient parameters. The drift parameters are important determinants of the long-run forecasting performance of the VAR while the  $\mathbf{B}$  parameters are important for the short-run performance of the model.

The forecast distributions, using the unrestricted and restricted estimates, for the VAR are shown in Figures 5 and 6.

In both Figures, the data is represented by the solid line with solid markers, the median forecast using unrestricted estimates is represented by the solid line, the 5% and 95% quantiles for the unrestricted estimates is represented by the dashed lines, the median forecast using restricted estimates is represented by the dotted-dashed line, and the 5% and 95% quantiles are represented by the dotted lines. Note that the median forecast for the interest

Figure 5: Interest Rate Forecast Distributions

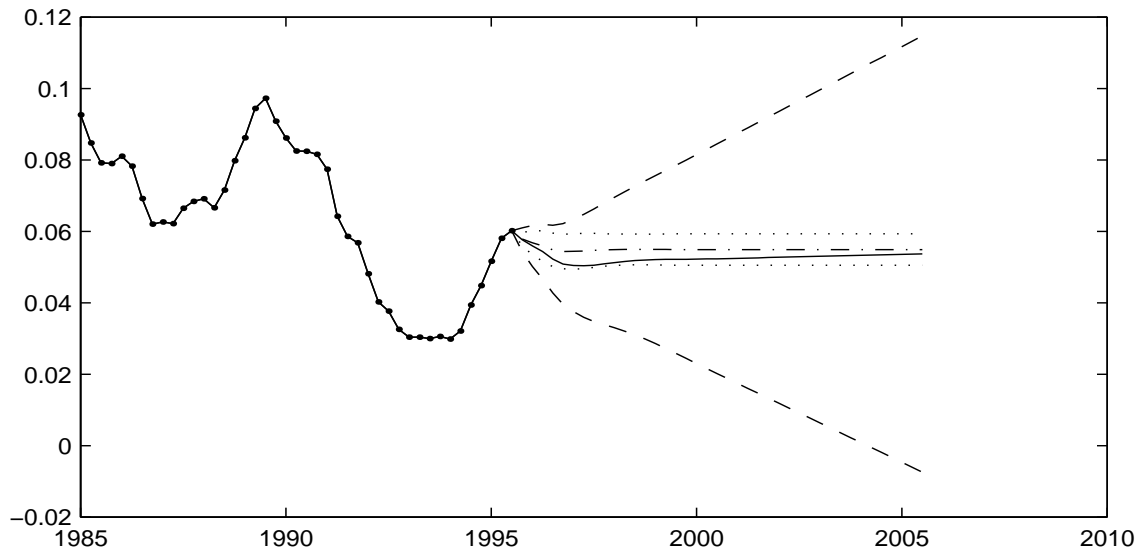
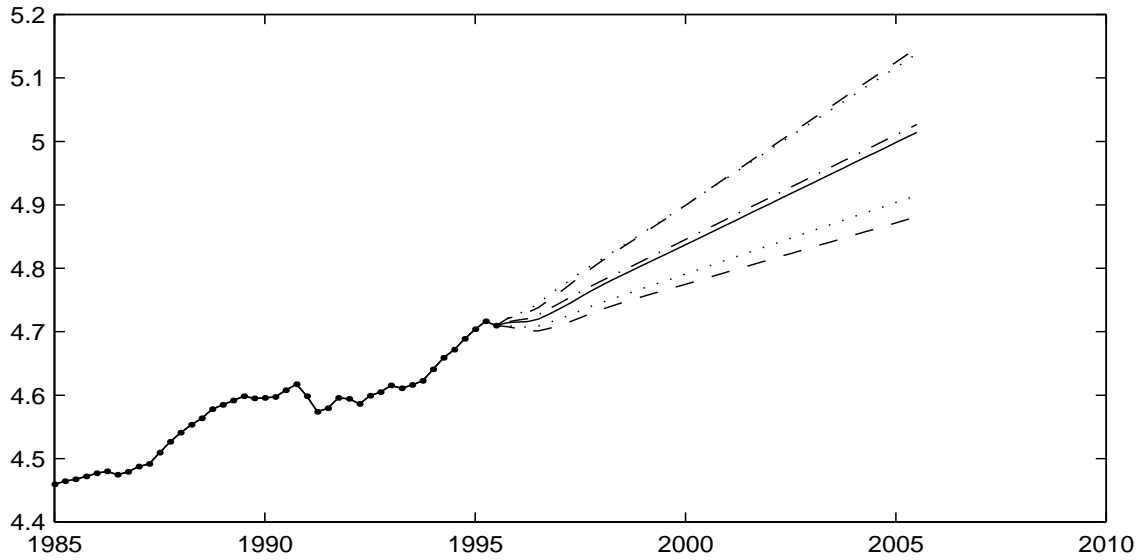


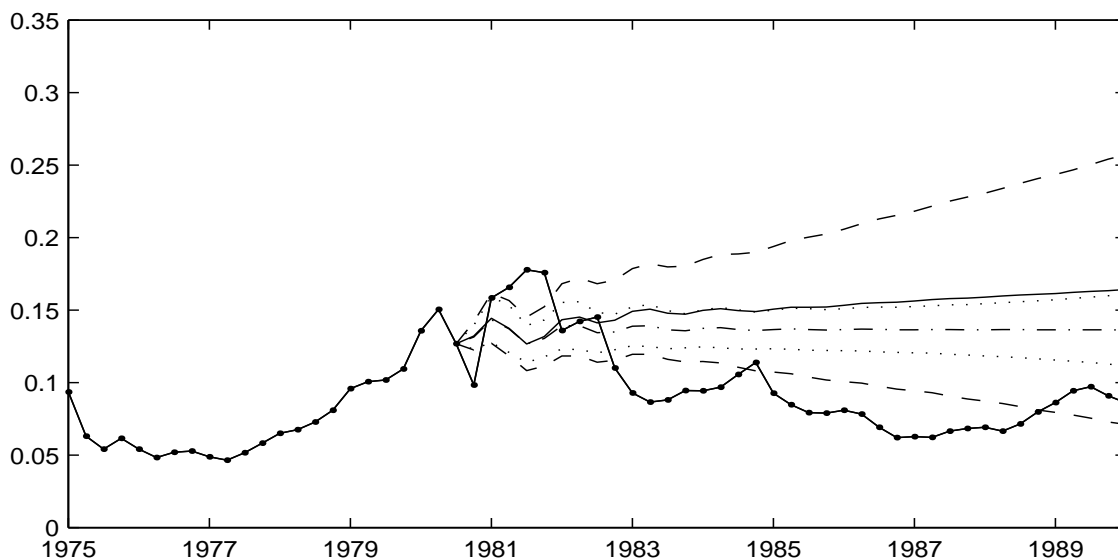
Figure 6: Industrial Production Forecast Distributions



rate using the unrestricted estimates contains a small drift. The interquantile range of the forecast distributions for the interest rate is much smaller using the restricted estimates, clearly due to imposing zero drift exactly. Note however that the interquantile range for industrial output is uniformly smaller using the restricted estimates.

In looking at forecasting performance we should note that the interest variable over the whole sample does not exhibit drift of the same sign. Suppose that we restrict attention to the period 1958:1 to 1980:3. In this period we see significant period of positive drift. Figures 7 and 8 show what the forecast distributions look like if one were to estimate a VAR(5) on this subset of the data.

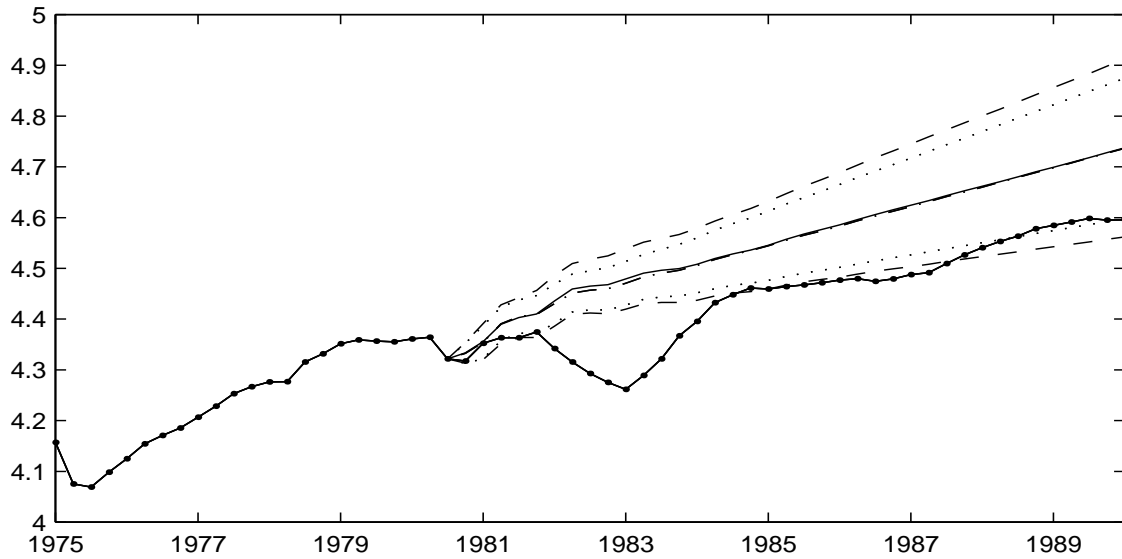
Figure 7: Interest Rate Forecast Distribution: 1958:1-1980:3



For this period we see a much stronger drift estimated for the interest rate variable and this is evidenced by the median forecast (the solid line) in Figure 7. We also see a bigger reduction in forecast uncertainty for industrial production which is evidenced by the smaller interquantile range depicted in Figure 8.

Overall, we see that imposing drift restrictions on a VAR containing solely  $I(1)$  variables impacts the estimates of the parameters and hence the orthogonalized impulse response

Figure 8: Industrial Production Forecast Distribution: 1958:1-1980:3



distributions and the forecast distributions.

## 4 Extension to the Case of Co-integration

The illustration given in Section 3 was for the case where there was no co-integration. In this case the model that was studied was a VAR in difference form. That is (4) was estimated rather than (2). If there are co-integrating relationships then (2) is the appropriate model to estimate.

If the co-integrating relationships were known, that is we knew the value of the  $(r \times m)$  matrix  $\alpha$ , then it would be possible to extend the algorithm as we would be including the same set of variables to each equation. However, if the value of  $\alpha$  is not known then it needs to be estimated. One possible method, which is showing encouraging results is to use Bayesian variable selection methods (Geweke 1996) to decide which of the possible  $(m-1)$  co-integrating vectors to include in the model. This is a matter of ongoing research.

It is also feasible to use variable selection to determine the appropriate lag length in the

VAR. This has also been left for further research.

## 5 Conclusions

It is well known that correctly estimating the drift parameters of a VAR containing solely I(1) variables improves forecasting accuracy. This paper has described a simple Bayesian method that enables the direct estimation of the drift parameters of a VAR. Using the Seemingly Unrelated Regressions formulation of the VAR it is possible to jointly estimate the equations of the VAR, and thus impose cross-equation equality restrictions, without the need to non-linearly transform the regressors in the model. This allows for the restriction of certain drift parameters while still being able to use existing, well-defined, and successful prior distributions on the non-drift parameters of a VAR.

Restricting the drift parameters of variables that are, *a priori*, believed to contain zero drift can impact upon the orthogonalized impulse response function distribution generated by the VAR. In some cases, the difference between the medians of the distributions were quite sizeable. In fact, in the illustration given, the median impulse response functions differed by nearly 45 percentiles of each distribution. Imposing the restriction can improve the accuracy of the calculated impulse response function as in the illustration given in the paper, the 90% interquantile range was uniformly lower in three out of four cases and for the fourth case the interquantile range was smaller over the near horizon but was larger over the far horizon.

Another impact of imposing a zero restriction on one drift parameter was that there was an improvement in the efficiency of the estimates of other drift parameters. This then was shown to long-run improve forecast accuracy in that forecast accuracy was uniformly better for the variable that was not restricted.

Therefore, there is evidence that the forecast and impulse response function distributions are sensitive to the prior on the long-run properties of the variables in a VAR. Imposing

restrictions on the drift of variables in a VAR can alter inferences based on parameter estimates of the VAR. Improvements in accuracy can also be attained while still keeping the basic form of the VAR intact. Finally, the issue of co-integration was only informally discussed in this paper and this topic is left for further research.

## References

- Banerjee, A., J.J. Dolado, J.W. Galbraith, and D.F. Hendrey (1993) *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data* (Great Clarendon Street, Oxford: Oxford University Press)
- Bewley, R. (1979) ‘The direct estimation of the equilibrium response in a linear dynamic model.’ *Economics Letters* 3, 357–361
- (2000) ‘Controlling spurious drift in macroeconomic forecasting models.’ <http://economics.web.unsw.edu.au/people/rbewley/papers/drift.pdf>
- Bewley, R.A., and T.G. Parry (1991) ‘Predicting the monthly and annual current account deficit from provisional data.’ *Economic Record* 67, 317–330
- Clements, M.P., and D.F. Hendry (1998) *Forecasting Economic Time Series* (Cambridge: Cambridge University Press)
- Davidson, J.E.H., D.F. Hendrey, F. Srba, and S. Yeo (1978) ‘Econometric modelling of the aggregate time-series relationship between consumers.’ *Economic Journal* 88, 661–92
- Doan, T., R. Litterman, and C.A. Sims (1984) ‘Forecasting and conditional projection using realistic prior distributions.’ *Econometric Reviews* 3(1), 1–100
- Engle, R.F., and B.S. Yoo (1987) ‘Forecasting and testing in co-integrated systems.’ *Journal of Econometrics* 35, 143–159

- Geweke, J.F. (1996) 'Variable selection and model comparison in regression.' In *Bayesian Statistics 5*, ed. J.O. Berger, J.M. Bernardo, A.P. Dawid, and A.F.M Smith (Oxford: Oxford University Press) pp. 609–620
- (1999) 'Using simulation methods for Bayesian econometric models: Inference, development, and communication.' *Econometric Reviews* 18, 1–126
- Hamilton, J.D. (1994) *Time Series Analysis* (Princeton, N.J.: Princeton University Press)
- LeSage, J.P. (1990) 'A comparison of the forecasting ability of ecm and var models.' *Review of Economics and Statistics* 72, 664–671
- Litterman, R.B. (1986) 'Forecasting with Bayesian Vector Autoregressions - Five years of experience.' *Journal of Business and Economics Statistics* 4, 25–37
- Sargan, J.D. (1964) 'Wages and prices in the United Kingdom: A study in econometric methodology.' In *Econometric Analysis for National Economic Planning*, ed. P.E. Hart, G. Mills, and J.K. Whitiker (London: Butterworth)
- Sims, C.A. (1980) 'Macroeconomics and reality.' *Econometrica* 48, 1–48
- (1992) 'Interpreting the macroeconomic time series facts: The effects of monetary policy.' *European Economic Review* 36, 975–1011
- Sims, C.A., and T. Zha (1998) 'Bayesian methods for dynamic multivariate models.' *International Economic Review* 39(4), 949–967
- Tierney, L. (1994) 'Markov chains for exploring posterior distributions.' *The Annals Of Statistics* 22(4), 1701–1762
- Todd, R.M. (1984) 'Improving economic forecasting with bayesian vector autoregression.' *Federal Reserve bank of Minneapolis Quarterly Review*



Waggoner, D.F., and T. Zha (1999) ‘Conditional forecasts in dynamic multivariate models.’  
*The Review of Economics and Statistics* 81(4), 639–651