JOURNAL OF APPLIED ECONOMETRICS, VOL. 12, 99–132 (1997)

NUMERICAL METHODS FOR ESTIMATION AND INFERENCE IN BAYESIAN VAR-MODELS

K. RAO KADIYALA[a] AND SUNE KARLSSON[b*]

[a] *Krannert Graduate School of Management, Purdue University, W. Lafayette, IN 47907, USA*
[b] *Department of Economic Statistics, Stockholm School of Economics, PO Box 6501, 113 83 Stockholm, Sweden.*
E-mail: stsk@hhs.se

SUMMARY

In Bayesian analysis of vector autoregressive models, and especially in forecasting applications, the Minnesota prior of Litterman is frequently used. In many cases other prior distributions provide better forecasts and are preferable from a theoretical standpoint. Several of these priors require numerical methods in order to evaluate the posterior distribution. Different ways of implementing Monte Carlo integration are considered. It is found that Gibbs sampling performs as well as, or better, then importance sampling and that the Gibbs sampling algorithms are less adversely affected by model size. We also report on the forecasting performance of the different prior distributions. © 1997 by John Wiley & Sons, Ltd. J. appl. econom. 12: 99–132, 1997.

(No. of Figures: 11. No. of Tables: 3. No. of Refs: 28.)

1. INTRODUCTION

Vector Autoregressive (VAR) models are frequently used to model and forecast dynamic economic systems. In these forecasting applications the combination of prior beliefs and family of prior distributions advocated by Litterman (1980) is often utilized. However, Kadiyala and Karlsson (1993) found that families of prior distributions that allow for dependence between the equations give better forecasts than the essentially univariate 'Minnesota prior' of Litterman.

Two of the more general prior distributions considered by Kadiyala and Karlsson (1993) have the disadvantage that no closed forms exist for the posterior moments of the regression parameters. Consequently, these must be evaluated using numerical methods. Even when the posterior moments are known, numerical methods are often required in order to obtain forecasts, impulse responses and other non-linear functions of the regression parameters. This can be quite time consuming — especially for large models. The practicability of these priors (and Bayesian analysis of VAR models in general) can thus be questioned.

This paper attempts to address this issue by studying various ways (importance sampling, Gibbs sampling) of implementing Monte Carlo methods for evaluating the posterior distribution of functions of the regression parameters. We do this in the context of one small and one large VAR model, for a variety of posterior distributions and one function of the parameters, the forecast.

The forecasting performance is, with some exceptions, similar across prior distributions in the applications considered here. One such exception is the forecast of inflation in the model of

* Correspondence to: Sune Karlsson, Stockholm School of Economics, Box 6501, 113 83 Stockholm, Sweden

Contract grant sponsor: Swedish Research Council for Humanities and Social Sciences (HSFR)

CCC 0883–7252/97/020099–34$17·50

Received 13 December 1994
Revised 29 September 1996

Litterman (1986) where the ENC prior outperforms the other priors by a wide margin. For the large (301 parameters) Litterman model the importance sampling procedures perform very poorly whereas the Gibbs sampler performs quite well. The penalty for not being able to sample directly from the posterior is, however, still heavy. The CPU times required for posteriors where Gibbs sampling is used exceeds the CPU times for posteriors where it is possible to sample directly from the posterior by a factor of about 140 for this particular model.

The remainder of the paper is organized as follows. In Section 2 the generic problem of forecasting with Bayesian VAR models is discussed. The prior beliefs embodied in the Minnesota prior and the prior distributions used to parameterize these prior beliefs are introduced in Section 3. In Section 4 the problem of evaluating the posterior distribution by Monte Carlo methods is discussed and some possible solutions proposed. Two forecasting applications are presented in Section 5 and the performance of the various methods of evaluating the posterior distributions is studied. Section 6 concludes.

2. FORECASTING WITH BAYESIAN VAR MODELS

Let y_t be the row vector of m variables of interest observed at time t and x_t a row vector of q exogenous variables influencing y_t. The VAR can then be written as

$$y_t = \sum_{i=1}^{p} y_{t-i}\mathbf{A}_i + x_t\mathbf{C} + u_t \tag{1}$$

where $\mathbf{A}_i$ and $\mathbf{C}$ are parameter matrices, of dimensions $m \times m$ and $q \times m$ respectively.

With a quadratic loss function the optimal forecast is the posterior expectation of y_{t+h}, conditional on the observed data $y_t, y_{t-1}, \ldots$ and future exogenous variables, x_{t+h}. Rewrite equation (1) as a first-order system $\mathbf{y}_t^* = \mathbf{y}_{t-1}^*\mathbf{A} + \mathbf{x}_t\mathbf{D} + \mathbf{u}_t^*$, where $\mathbf{y}_t^* = \{y_t, \ldots, y_{t-p+1}\}$, $\mathbf{D} = \{\mathbf{C}, \mathbf{0}, \ldots, \mathbf{0}\}$, $\mathbf{u}_t^* = \{u_t, 0, \ldots, 0\}$ and

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{I} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{A}_2 & \mathbf{0} & \mathbf{I} & & \vdots \\ \vdots & & & \ddots & \mathbf{I} \\ \mathbf{A}_p & & & \ldots & \mathbf{0} \end{pmatrix}$$

The h-step-ahead forecast is thus obtained by evaluating the integral

$$\mathbf{y}_t(h) = \int F(\mathbf{A}, \mathbf{D}, \mathbf{y}_t^*, \mathbf{X}, h)p(\gamma|\mathbf{y})\mathrm{d}\gamma \tag{2}$$

where $p(\gamma|\mathbf{y})$ denotes the (marginal) posterior density of the regression parameters in $\mathbf{A}$ and $\mathbf{D}$ and $F(\ )$ is the forecast function

$$F(\mathbf{A}, \mathbf{D}, \mathbf{y}_t^*, \mathbf{X}, h) = \mathbf{y}_t^*\mathbf{A}^h + \Sigma_{i=0}^{h-1}\mathbf{x}_{t+h-i}\mathbf{D}\mathbf{A}^i \tag{3}$$

In this paper x_t contain only deterministic variables such as time trends and seasonal dummies. When equation (1) is a partial system and we have stochastic variables in x_t we need the additional

assumption that future u_t is mean independent of future x_t, i.e. $E(u_{t+i}|x_{t+j}) = 0, i, j > 0$, for equation (2) to give the posterior expectation of y_{t+h} conditional on $x_{t+1}, \ldots, x_{t+h}$.

Note that equation (3) is a linear function of the parameters of the VAR for lead time 1 and that the forecast function becomes increasingly non-linear with the lead time. The evaluation of equation (2) thus provides a convenient framework for assessing the practicability of Monte Carlo methods. Note also that the calculation of forecasts is closely related to the calculation of impulse response functions and variance decompositions. See Koop (1992) for a Bayesian treatment of the calculation of impulse response functions in structural VAR models.

For the technical discussion of the prior and posterior distributions, we need the following notation. Write equation (1) as

$$y_t = z_t\Gamma + u_t$$

where $z_t = \{x_t, y_{t-1}, \ldots, y_{t-p}\}$ and the $k = q - pm$ by m matrix Γ is given by $\{\mathbf{C}', \mathbf{A}_1', \ldots, \mathbf{A}_p'\}'$. Performing the conventional stacking of the row vectors y_t, z_t and u_t for $t = 1, \ldots, T$ into Y, $\mathbf{Z}$ and $\mathbf{U}$ we have the multivariate regression model

$$\mathbf{Y} = \mathbf{Z}\Gamma + \mathbf{U}$$

Then, letting the subscript i denote the *ith* column vector, we have the equation for variable i as $y_i = \mathbf{Z}\gamma_i + u_i$. For $\mathbf{y}$, γ and $\mathbf{u}$ the vectors obtained by stacking the columns of $\mathbf{Y}$, Γ and $\mathbf{U}$, the system can be written as $\mathbf{y} = (\mathbf{I} \otimes \mathbf{Z})\gamma + \mathbf{u}$.

In addition, $\tilde{}$ (tilde) denote parameters of the prior distribution, $\bar{}$ (bar) denote the parameters of the posterior distribution, and the OLS estimates of Γ and γ are denoted by $\hat{\Gamma}$ and $\hat{\gamma}$, respectively.

Throughout the paper it is assumed that $\mathbf{u} \sim N(\mathbf{0}, \Psi \otimes \mathbf{I})$, that is, u_t is taken to be i.i.d. $N(\mathbf{0}, \Psi)$. The assumption of normality and homoscedasticity can easily be relaxed along the lines of Geweke (1993) to allow for heteroscedasticity or (equivalently) fat-tailed marginal distributions for u_t for some of the priors below. This line of research will, however, not be pursued here and the likelihood is given by

$$L(\gamma, \Psi) \propto |\Psi|^{-T/2} \exp\{-\mathrm{tr}[(\mathbf{Y} - \mathbf{Z}\Gamma)'\Psi^{-1}(\mathbf{Y} - \mathbf{Z}\Gamma)]/2\}$$

After some manipulation we arrive at

$$\begin{aligned} L(\gamma, \Psi) &\propto |\Psi|^{-T/2} \exp\left\{-\frac{1}{2}(\gamma - \hat{\gamma})'(\Psi^{-1} \otimes \mathbf{Z}'\mathbf{Z})(\gamma - \hat{\gamma}) - \frac{1}{2}\,\mathrm{tr}[\Psi^{-1}(\mathbf{Y} - \mathbf{Z}\hat{\Gamma})'(\mathbf{Y} - \mathbf{Z}\hat{\Gamma})]\right\} \\ &= |\Psi|^{-k/2} \exp\left\{-\frac{1}{2}(\gamma - \hat{\gamma})'(\Psi^{-1} \otimes \mathbf{Z}'\mathbf{Z})(\gamma - \hat{\gamma})\right\} \times |\Psi|^{-(T-k)/2} \\ &\times \exp\left\{-\frac{1}{2}\,\mathrm{tr}[\Psi^{-1}(\mathbf{Y} - \mathbf{Z}\hat{\Gamma})^1(\mathbf{Y} - \mathbf{Z}\hat{\Gamma})]\right\} \\ &\propto N(\gamma|\hat{\gamma}, \Psi \otimes (\mathbf{Z}'\mathbf{Z})^{-1}) \times iW(\Psi|(\mathbf{Y} - \mathbf{Z}\hat{\Gamma})'(\mathbf{Y} - \mathbf{Z}\hat{\Gamma}), T - k - m - 1) \end{aligned} \tag{4}$$

the product of an inverse Wishart density for Ψ and a normal density for γ conditional on Ψ.[1]

[1] See Drèze and Richard (1983) for a definition of the inverse Wishart and matricvariate t densities.

3. PRIOR DISTRIBUTIONS FOR BAYESIAN ANALYSIS OF VARS

3.1. Prior Beliefs

Noting that many economic variables behave as if they have random walk components Litterman (1980) suggested specifying the prior means of the regression parameters as a random walk

$$y_{it} = y_{i,t-1} + u_{it} \tag{5}$$

for each variable. That is, the prior mean for the parameter on the first own lag is set to unity and the prior mean of the remaining parameters in γ_i is set to zero.

It is also reasonable that the importance of the lagged variables decreases with the lag length. Consequently, the prior parameter variances are taken to decrease with the lag length, making the prior tighter around zero. The regression parameters on the exogenous variables have a large prior variance, making the prior relatively uninformative for these parameters. The prior covariances between the parameters in γ are set to zero for simplicity.

To simplify the specification of the prior variances, the relative tightness of the prior for the parameters on own lags, foreign lags and exogenous variables is set by the hyper-parameters π_1, π_2 and π_3 and the variances are scaled to account for differing variability in the variables.

More precisely, the prior variances of the parameters in equation i are specified as

$$\mathrm{Var}(\gamma_i) = \begin{cases} \dfrac{\pi_1}{k} & \text{for parameters on own lags} \\ \dfrac{\pi_2 \sigma_i^2}{k \sigma_j^2} & \text{for parameters on lags of variable } j \neq i \\ \pi_3 \sigma_i^2 & \text{for parameters on exogeneous/deterministic variables} \end{cases} \tag{6}$$

where k denotes the lag length and σ_i is a scale factor accounting for the differing variability of the variables. In this paper σ_i is set to s_i, the residual standard error of a p-lag univariate autoregression for variable i. Common values of the hyper-parameters are of the magnitude 0·05 for π_1, 0·005 for π_2 and 10^5 for π_3. The values used in the applications are in Tables II and III.

See also Litterman (1986) for a discussion and motivation of these prior beliefs. Note, however, that we let the prior variances decrease slower with the lag length than Litterman. Litterman uses a factor $1/k^2$ rather than $1/k$.

Although the prior means indicate that the variables are $I(1)$ but not cointegrated, these prior beliefs do not rule out cointegration. The issue of cointegration is, however, not addressed in this paper. See Bauwens and Lubrano (1994), Dorfman (1995), Kleibergen and van Dijk (1994) and Koop (1991) for a discussion of issues arising in a Bayesian analysis of cointegration.

When parameterizing the prior beliefs, a number of prior distributions can be used. Here we will consider the independent normal or Minnesota prior, the Normal-Wishart and Diffuse priors, the Normal-Diffuse and Extended Natural Conjugate priors. The prior and posterior distributions are summarized in Table I.

3.2. The Minnesota Prior

This is the prior distribution used by Litterman (1980, 1986). The residual variance–covariance matrix, Ψ, is taken to be fixed and diagonal and the likelihood in equation (4) reduces to products of independent normal densities for γ_i. With a normal prior, and the prior moments of the

Table I. Prior and posterior distributions

	Prior	Posterior	Importance function	Gibbs sampler
Minnesota	$\gamma_i \sim N(\tilde{\gamma}_i, \tilde{\Sigma}_i)$, Ψ fix and diagonal	$\gamma_i \mid \mathbf{y} \sim N(\bar{\gamma}_i, \bar{\Sigma}_i)$		
Diffuse	$p(\gamma, \Psi) \propto \lvert\Psi\rvert^{-(m+1)/2}$	$\Gamma \mid \mathbf{y} \sim MT(\mathbf{Z}'\mathbf{Z}, (\mathbf{Y} - \mathbf{Z}\hat{\Gamma})' \times (\mathbf{Y} - \mathbf{Z}\hat{\Gamma}), \hat{\Gamma}, T - k)$		
Normal-Wishart	$\gamma \mid \Psi \sim N(\tilde{\gamma}, \Psi \otimes \tilde{\Omega})$, $\Psi \sim iW(\tilde{\Psi}, \alpha)$	$\Gamma \mid \mathbf{y} \sim MT(\bar{\Omega}^{-1}, \bar{\Psi}, \bar{\Gamma}, T + \alpha)$		
Normal-Diffuse	$\gamma \sim N(\tilde{\gamma}, \tilde{\Sigma})$, $p(\Psi) \propto \lvert\Psi\rvert^{-(m+1)/2}$	$p(\gamma \mid y) \propto \exp\{-(\gamma - \tilde{\gamma})'\tilde{\Sigma}^{-1} \times (\gamma - \tilde{\gamma})/2\} \times \lvert(\mathbf{Y} - \mathbf{Z}\hat{\Gamma})'(\mathbf{Y} - \mathbf{Z}\hat{\Gamma}) + (\Gamma - \hat{\Gamma})'\mathbf{Z}'\mathbf{Z}(\Gamma - \hat{\Gamma})\rvert^{-T/2}$	2-0 poly-t	$\gamma \mid \Psi, \mathbf{y} \sim N(\bar{\gamma}(\Sigma^{-1} + \Psi^{-1} \otimes \mathbf{Z}'\mathbf{Z})^{-1})$ $\Psi^{-1} \mid \gamma, \mathbf{y} \sim W([(\mathbf{Y} - \mathbf{Z}\hat{\Gamma})'(\mathbf{Y} - \mathbf{Z}\hat{\Gamma}) + (\Gamma - \hat{\Gamma})'\mathbf{Z}'\mathbf{Z}(\Gamma - \hat{\Gamma})]^{-1}, T)$
Extended Natural Conjugate	$p(\Delta) \propto \lvert\tilde{\Psi} + (\Delta - \tilde{\Delta})' \times \tilde{\mathbf{M}}(\Delta - \tilde{\Delta})\rvert^{-\alpha/2}$ or independent multivariate t's for each equation, $\Psi \mid \Delta \sim iW(\tilde{\Psi} + (\Delta - \tilde{\Delta})' \times \tilde{\mathbf{M}}(\Delta - \tilde{\Delta}), \alpha)$	$p(\Delta \mid \mathbf{y}) \propto \lvert\bar{\Psi} + (\Delta - \bar{\Delta})' \times \mathbf{M}(\Delta - \bar{\Delta})\rvert^{-(T+\alpha)/2}$ $\Psi \mid \Delta, \mathbf{y} \sim iW$ $(\tilde{\Psi} + (\Delta - \bar{\Delta})'\mathbf{M}(\Delta - \bar{\Delta}), T + \alpha)$	STFC, independent multivariate t's obtained as conditional posterior at the posterior mode	$\gamma_i \mid \gamma_i, \ldots, \gamma_{i-1}, \gamma_{i+1}, \ldots, \gamma_m \sim t(\mathbf{d}_i, (q_{ii})/\{T + \alpha - k\})\mathbf{P}_{ii}^{-1}$, $T + \alpha - k$)

preceding section, we have prior and posterior independence between equations and they can be treated separately.

Writing the prior for equation i as $\gamma_i \sim N(\tilde{\gamma}_i, \tilde{\Sigma}_i)$ we have the posterior as $\gamma_i \mid \mathbf{y} \sim N(\bar{\gamma}_i, \bar{\Sigma})_i$, with $\bar{\Sigma}_i = (\tilde{\Sigma}_i^{-1} + \psi_{ii}^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$ and $\bar{\gamma}_i = \bar{\Sigma}_i(\tilde{\Sigma}_i^{-1}\tilde{\gamma}_i + \psi_{ii}^{-1}\mathbf{Z}'y_i)$. The diagonal elements of Ψ, ψ_{ii}, are obtained from the data as s_i^2.

The Minnesota prior can be generalized by allowing for a non-diagonal variance–covariance matrix and/or taking Ψ to be unknown. The remaining priors considered here generalize the Minnesota prior in both these directions. An additional possibility is to maintain the assumption of a diagonal Ψ matrix while taking the diagonal elements to be unknown. Independent inverse gamma priors on the diagonal elements then lead to marginal multivariate t priors and posteriors for the parameters of each equation. In this context it is useful to think of the fixed and diagonal Ψ matrix of the Minnesota prior as a restriction on the likelihood or the underlying data generating process rather than being a restriction on the prior distribution.

3.3. The Diffuse and Normal-Wishart Priors

With the diffuse (or Jeffreys') prior distribution (Geisser, 1965; Tiao and Zellner, 1964)

$$p(\gamma, \Psi) \propto |\Psi|^{-(m+1)/2}$$

the posterior distribution is obtained as

$$\gamma|\Psi, \mathbf{y} \sim N(\hat{\gamma}, \Psi \otimes (\mathbf{Z}'\mathbf{Z})^{-1}), \Psi|\mathbf{y} \sim iW((\mathbf{Y} - \mathbf{Z}\hat{\Gamma})'(\mathbf{Y} - \mathbf{Z}\hat{\Gamma}), T - k)$$

Integrating out Ψ of the joint posterior, we have the marginal posterior distribution of the $k \times m$ matrix Γ as

$$p(\Gamma|\mathbf{y}) \propto |(\mathbf{Y} - \mathbf{Z}\hat{\Gamma})'(\mathbf{Y} - \mathbf{Z}\hat{\Gamma}) + (\Gamma - \hat{\Gamma})'\mathbf{Z}'\mathbf{Z}(\Gamma - \hat{\Gamma})|^{-T/2}$$

That is, the marginal posterior distribution of Γ is matricvariate t,

$$\Gamma|\mathbf{y} \sim MT(\mathbf{Z}'\mathbf{Z}, (\mathbf{Y} - \mathbf{Z}\hat{\Gamma})'(\mathbf{Y} - \mathbf{Z}\hat{\Gamma}), \hat{\Gamma}, T - k)$$

When the assumption of a fixed and diagonal residual variance–covariance matrix is relaxed, the natural conjugate prior for normal data is the Normal-Wishart,

$$\gamma|\Psi \sim N(\tilde{\gamma}, \Psi \otimes \tilde{\Omega}), \Psi \sim iW(\tilde{\Psi}, \alpha)$$

with prior mean and variance $E(\gamma) = \tilde{\gamma}$, $\alpha > m$, and $\mathrm{Var}(\gamma) = (\alpha - m - 1)^{-1}\tilde{\Psi} \otimes \tilde{\Omega}$, $\alpha > m + 1$. The posterior distribution is given by

$$\gamma|\Psi, \mathbf{y} \sim N(\bar{\gamma}, \Psi \otimes \bar{\Omega}), \Psi|\mathbf{y} \sim iW(\bar{\Psi}, T + \alpha)$$

with $\bar{\Omega} = (\tilde{\Omega}^{-1} + \mathbf{Z}'\mathbf{Z})^{-1}$, $\bar{\Gamma} = \bar{\Omega}(\tilde{\Omega}^{-1}\tilde{\Gamma} + \mathbf{Z}'\mathbf{Z}\hat{\Gamma})$ and $\bar{\Psi} = \hat{\Gamma}'\mathbf{Z}'\mathbf{Z}\hat{\Gamma} + \tilde{\Gamma}'\tilde{\Omega}^{-1}\tilde{\Gamma} + \tilde{\Psi} + (\mathbf{Y} - \mathbf{Z}\hat{\Gamma})'(\mathbf{Y} - \mathbf{Z}\hat{\Gamma}) - \bar{\Gamma}'(\tilde{\Omega}^{-1} + \mathbf{Z}'\mathbf{Z})\bar{\Gamma}$. The marginal posterior distribution of Γ is again matricvariate t, $\Gamma|\mathbf{y} \sim MT(\bar{\Omega}^{-1}, \bar{\Psi}, \bar{\Gamma}, T + \alpha)$.

The parameters of the Normal-Wishart prior are chosen so the that mean of Ψ coincides with the fixed residual variance–covariance matrix of the Minnesota prior, i.e. the diagonal elements of $\tilde{\Psi}$ are set to

$$\tilde{\psi}_{ii} = (\alpha - m - 1)s_i^2 \tag{7}$$

and $\tilde{\Omega}$ is chosen to match the prior variances of the Minnesota prior with the exception that $\pi_1 = \pi_2$ as discussed below. Finally, choosing the prior degrees of freedom as

$$\alpha = \max\{m + 2, m + 2h - T\} \tag{8}$$

ensures that both the prior variances of the regression parameters and the posterior variances of the forecasts at lead time h exist.

The two main shortcomings of the Minnesota prior, the forced posterior independence between equations and the fixed residual variance–covariance matrix, are not present with these two priors. On the other hand, with the Normal-Wishart, the structure of the variance–covariance matrix of γ forces us to treat all equations symmetrically when specifying the prior. Specifically, the prior variance–covariance matrix of the regression parameters in different equations can only differ by a scale factor. With the specification (6) of the prior variances this requires that the prior beliefs are specified with $\pi_1 = \pi_2$.

Using the prior beliefs outlined above, the Kronecker-structure of the prior variance–covariance matrix is not a serious restriction, but it is also relatively easy to conceive of situations where the prior beliefs are more sophisticated than the simple data-centric specification suggested by Litterman. In these cases the restriction complicates the analysis. Suppose, for example, that our prior beliefs include 'money neutrality', i.e. that the money supply does not Granger-cause real GNP. In a VAR-model this corresponds to zero-restrictions on the coefficients of lagged money supply in the real GNP equation. With an unrestricted variance–covariance matrix these prior beliefs can be approximated by setting the prior means of the relevant parameters to zero and the prior variance to an arbitrarily small number. With the Normal-Wishart prior this leads to problems with the specification of the prior for the other equations. If we want the prior for variables other than lagged money supply to be informative in other equations (relatively small prior variance) we are forced to make the prior for lagged money supply extremely informative in these equations as well. Alternatively, we can make the prior on the lags of money supply informative in the desired way with the prior on the other variables highly uninformative.

The prior can, of course, be amended to take account for this type of prior information. We can, for example, specify the conditional prior of γ as a truncated normal, $\gamma|\Psi \propto \mathrm{I}(\gamma)N(\tilde{\gamma}, \Psi \otimes \tilde{\Omega})$, where $\mathrm{I}(\gamma)$ is an indicator function taking the value 1 when the parameters on lagged money supply in the real GNP equation are sufficiently close to zero and the value zero otherwise. This will, however, rob the Normal-Wishart prior of most of its advantages compared to the other priors discussed here since the conditional posterior of γ will also have the truncated normal form. Posterior moments must be evaluated numerically and the numerical procedures are considerably more complicated and less efficient than the straightforward Monte Carlo procedure possible with the Normal-Wishart prior. One possible algorithm for generating draws from the truncated normal part of the posterior is the acceptance/rejection algorithm discussed in Hajivassiliou and Ruud (1994).

3.4. The Normal-Diffuse Prior

This prior, introduced by Zellner (1971), avoids the Normal-Wishart type restrictions on the variance–covariance matrix of γ and allows for a non-diagonal residual variance–covariance matrix. The multivariate normal prior on the regression parameters of the Minnesota prior is combined with a diffuse prior on the residual variance–covariance matrix. That is, we have prior independence between γ and Ψ with

$$\gamma \sim N(\tilde{\gamma}, \tilde{\Sigma}), p(\Psi) \propto |\Psi|^{-(m+1)/2}$$

Combining this with the likelihood (4) yields the marginal posterior of γ as

$$p(\gamma|\mathbf{y}) \propto \exp\left\{-\frac{1}{2}(\gamma-\tilde{\gamma})'\tilde{\Sigma}^{-1}(\gamma-\tilde{\gamma})\right\}|(\mathbf{Y}-\mathbf{Z}\hat{\Gamma})'(\mathbf{Y}-\mathbf{Z}\hat{\Gamma})+(\Gamma-\hat{\Gamma})'\mathbf{Z}'\mathbf{Z}(\Gamma-\hat{\Gamma})|^{-T/2} \quad (9)$$

the product of the marginal prior distribution and a matricvariate t-distribution. The matricvariate t-factor will give rise to posterior dependence between the equations even when the prior is specified with a diagonal or block diagonal variance–covariance matrix. The matricvariate t-factor is identical to the posterior associated with the Diffuse prior.

The form of the marginal posterior is troublesome in the sense that large differences between the information contained in the prior and the likelihood can cause the posterior to be bimodal, and thus the posterior mean to have low posterior probability. Rather than giving a formal proof of this proposition, we present a simple numerical example for the two-dimensional case. Figure 1 shows the posterior distribution for $\tilde{\Sigma} = \mathbf{I}$, $\hat{\gamma} = (1, 1)'$, $\mathbf{Z}'\mathbf{Z} = 0{\cdot}91$, $(\mathbf{Y}-\mathbf{Z}\hat{\Gamma})'(\mathbf{Y}-\mathbf{Z}\hat{\Gamma}) = 1{\cdot}8 \times \mathbf{I}$, $T = 28$ and for two values of the prior mean $\tilde{\gamma}$, $(8, 9)'$ and $(8, 8{\cdot}5)'$. In the first case the posterior is bimodal but a small decrease in the distance between the centres of the prior and the likelihood makes the posterior unimodal. Note that the distance between the centres of the prior and the likelihood is very large, Mahalanobis distance is close to 100 using the prior variance and over 1000 using the information matrix. Consequently, we do not expect the potential bimodality to be a serious problem in practice.

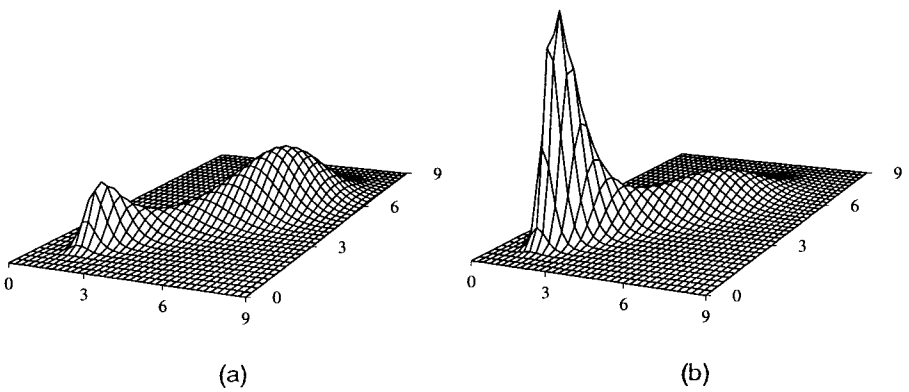

Figure 1. The bivariate Normal-Diffuse posterior: (a) $\tilde{\gamma} = (8, 9)'$; (b) $\tilde{\gamma} = (8, 8{\cdot}5)'$

3.5. The Extended Natural Conjugate Prior

The Extended Natural Conjugate (ENC) prior overcomes the restrictions on Var(γ) of the Normal-Wishart prior by reparameterizing the VAR in equation (1). Let Δ be a $mk \times m$ matrix with the columns γ_i on the diagonal and all other elements zero, that is,

$$\Delta = \begin{pmatrix} \gamma_1 & 0 & \dots & 0 \\ 0 & \gamma_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \gamma_m \end{pmatrix}$$

Also let $\Xi = \iota' \otimes \mathbf{Z}$, where ι is a $m \times 1$ vector of ones. Equation (1) can then be rewritten as $\mathbf{Y} = \Xi\Delta + \mathbf{U}$.

For the prior distribution

$$\begin{aligned} p(\Delta) &\propto |\tilde{\Psi} + (\Delta - \tilde{\Delta})'\tilde{\mathbf{M}}(\Delta - \tilde{\Delta})|^{-\alpha/2} \\ \Psi|\Delta &\sim iW(\tilde{\Psi} + (\Delta - \tilde{\Delta})'\tilde{\mathbf{M}}(\Delta - \tilde{\Delta}), \alpha) \end{aligned}$$

and normal data, the posterior distribution is given by Drèze and Morales (1976) as

$$\begin{aligned} p(\Delta|\mathbf{y}) &\propto |\bar{\Psi} + (\Delta - \bar{\Delta})'\bar{\mathbf{M}}(\Delta - \bar{\Delta})|^{-(T+\alpha)/2} \\ \Psi|\Delta, \mathbf{y} &\sim iW(\tilde{\Psi} + (\Delta - \bar{\Delta})'\mathbf{M}(\Delta - \bar{\Delta}), T + \alpha) \end{aligned} \tag{10}$$

where $\bar{\mathbf{M}} = \tilde{\mathbf{M}} + \Xi'\Xi$, $\bar{\Psi} = \tilde{\Psi} + \tilde{\Delta}'\tilde{\mathbf{M}}\tilde{\Delta} + \mathbf{Y}'\mathbf{Y} - \bar{\Delta}'\bar{\mathbf{M}}\bar{\Delta}$ and $\bar{\Delta}$ is the solution to $\bar{\mathbf{M}}\bar{\Delta} = \tilde{\mathbf{M}}\tilde{\Delta} + \Xi'\mathbf{Y}$. If $\tilde{\mathbf{M}}$ is of full rank, $\bar{\mathbf{M}}$ will be of full rank and $\bar{\Delta}$ is unique. The marginal distribution of Δ has the form of a matricvariate t-density. However, due to the restricted structure of Δ it is *not* matricvariate t.

Unfortunately the ENC is difficult to interpret and there are — except for special cases — no closed forms available for the moments of this distribution. One such special case arises when $\tilde{\Psi}$ is diagonal, $\tilde{\mathbf{M}}$ block diagonal and $\tilde{\Delta}$ has the same structure as Δ. The ENC prior then factors into independent multivariate t-distribution with $\alpha - k$ degrees of freedom for the parameters of each equation (Drèze and Richard, 1983). This makes the specification of the prior straightforward if the prior independence between equations is deemed acceptable. Specifying the ENC prior as independent multivariate t's also gives a nice interpretation of this prior as a generalization of the Minnesota prior not only to an unknown Ψ matrix (marginal multivariate t-priors or normal-inverse gamma priors for each equation) but also to a non-diagonal Ψ matrix.

Using this factorization, the prior means are $\tilde{\gamma}_i$, the elements of $\tilde{\Delta}$ corresponding to γ_i and the prior variance is given by

$$\text{Var}(\gamma_i) = \frac{\tilde{\psi}_{ii}}{\alpha - k - 2}\tilde{\mathbf{M}}_{ii}^{-1} \tag{11}$$

where $\tilde{\mathbf{M}}_{ii}$ is the *ith* diagonal block of $\tilde{\mathbf{M}}$. Once $\tilde{\psi}_{ii}$ is specified, $\tilde{\mathbf{M}}_{ii}$ can be chosen to match the prior variances (6).

$\tilde{\Psi}$ can be specified in different ways. Kadiyala and Karlsson (1993) matched the fixed residual variance matrix of the Minnesota prior with the prior expectation of Ψ, *conditional* on $\Delta = \tilde{\Delta}$.

That is, the diagonal elements of $\tilde{\Psi}$ are set according to equation (7). In this paper we also consider a specification where the *unconditional* prior expectation of Ψ matches the Minnesota residual variance matrix. With diagonal $\tilde{\Psi}$ and $\tilde{\mathbf{M}}$ matrices and $\tilde{\mathbf{M}}_{ii}$ in equation (11) chosen to satisfy equation (6) we have the prior unconditional expectation of Ψ as

$$E(\Psi) = \frac{\alpha - 2}{(\alpha - m - 1)(\alpha - k - 2)}\tilde{\Psi}$$

and setting

$$\tilde{\psi}_{ii} = (\alpha - m - 1)(\alpha - k - 2)s_i^2/(\alpha - 2) \tag{12}$$

gives the desired unconditional expectation.

The existence of prior moments of γ follows from the factorization into independent multivariate t's. The posterior moment of order i exists (Drèze and Richard, 1983) if $T + \alpha - k - m \geqslant i$. As for the Normal-Wishart, we choose the prior degrees of freedom to guarantee that the prior variance of γ and the posterior variance of the forecasts exists.

4. EVALUATING THE POSTERIOR DISTRIBUTION

For the natural conjugate Minnesota and Normal-Wishart priors, as well as the Diffuse prior, we have a closed form solution for the posterior distribution. It is, consequently, straightforward to evaluate the posterior distribution of the parameters for these priors.

The situation is more problematic when it comes to non-linear functions of the parameters, such as the forecasts considered in this paper. In most cases the problems can be solved by estimating the posterior moments by Monte Carlo integration. This is particularly appealing when we can sample directly from the posterior distribution. For the Diffuse and Normal-Wishart posteriors this is possible using, for example, the algorithm of Geweke (1988) to generate pseudo-random numbers from the marginal matricvariate t posterior. Antithetic variates are also easy to implement since the posterior is symmetric.

In forecasting applications it is customary to generate forecasts from the Minnesota posterior using the chain-rule and the posterior means of the parameters, that is equation (3) is evaluated at the posterior means of the parameters. For completeness we also use Monte Carlo integration to obtain the posterior expectation of the forecasts. Monte Carlo integration is straightforward since the posterior distribution of γ is normal. It is, of course, possible to generate forecasts from the posterior means with the other priors as well (with the Diffuse prior this is equivalent to the OLS-based forecasts reported below). For the Normal-Wishart and Diffuse priors this would lead to some savings in CPU-time. With the Normal-Diffuse and ENC priors the posterior means of the parameters must be obtained using numerical methods and the savings in CPU-time are negligible.

For the Normal-Diffuse and ENC priors no closed form solution for the posterior moments exist and Monte Carlo integration is complicated by the lack of algorithms for generating random numbers directly from the posterior distributions. To overcome this problem we use the methods of importance sampling and Gibbs sampling. For readers unfamiliar with these techniques a brief review of the concepts involved is given in the next section. A more complete discussion and the relevant references can be found in Geweke (1995). Sections 4.2 and 4.3

describe the implementations of importance sampling and Gibbs sampling used here and the algorithms are summarized in Table I.

4.1. Monte Carlo Integration

The basic problem is the evaluation of the integral

$$E_p[g(\theta)] = \mu_g = \int g(\theta)p(\theta)\mathrm{d}\theta \tag{13}$$

where $g(\theta)$ is a function of the random vector θ, the forecast say, and $p(\theta)$ is a proper density function. If we can generate pseudo-random numbers from $p(\theta)$, $E_p[g(\theta)]$ can be estimated as the sample mean $\bar{g}_n$ over n draws from $p(\theta)$. Provided that $\mathrm{Var}_p[g(\theta)] = \sigma_g^2$ is finite, $n^{1/2}(\bar{g}_n - \mu_g) \xrightarrow{\mathrm{d}} N(0, \sigma_g^2)$ and probabilistic error bounds are readily available. Estimates of the numerical standard error, $\sigma_g/n^{1/2}$ and equivalent quantities are reported below as measures of the quality of the estimate of $E_p[g(\theta)]$.

For the Diffuse, Normal-Wishart and Minnesota priors we can generate pseudo-random numbers directly from the posterior and equation (13) is operational, for the ENC and Normal-Diffuse priors this is not possible. For the latter priors only the kernel, $p^*(\theta)$ ($p(\theta) = cp^*(\theta)$), of the posterior is known and there are no algorithms for generating pseudo-random numbers directly from the posterior. The integral in equation (13) can, however, still be evaluated using Importance Sampling or Markov chain Monte Carlo methods.

Importance sampling

In importance sampling (see e.g. Kloek and van Dijk, 1978) the problem of evaluating

$$E_p[g(\theta)] = c\int g(\theta)p^*(\theta)\mathrm{d}\theta = \frac{\int g(\theta)p^*(\theta)\mathrm{d}\theta}{\int p^*(\theta)\mathrm{d}\theta} \tag{14}$$

is replaced by an equivalent problem which we can address using the simple Monte Carlo method outlined above. Suppose $I(\theta)$ is a proper density from which we can generate pseudo-random numbers, equation (14) can then be restated as

$$E_p[g(\theta)] = \frac{\int g(\theta)w(\theta)I(\theta)\mathrm{d}\theta}{\int w(\theta)I(\theta)\mathrm{d}\theta} = \frac{E_I[g(\theta)w(\theta)]}{E_I[w(\theta)]}$$

where $w(\theta) = p^*(\theta)/I(\theta)$. Drawing pseudo-random numbers from $I(\theta)$ we can easily estimate $E_p[g(\theta)]$ by

$$\bar{g}_{I,n} = \frac{\Sigma_{i=1}^n g(\theta)w(\theta)/n}{\Sigma_{i=1}^n w(\theta)/n}$$

Geweke (1989, theorem 2) showed that if

$$\text{(i) } w(\theta) \text{ is finite on the support of } p \text{ and, (ii) } \mathrm{Var}_p[g(\theta)] \text{ is finite} \tag{15}$$

then $\bar{g}_{I,n}$ obeys a central limit theorem, $n^{1/2}(\bar{g}_{I,n} - \mu_g) \xrightarrow{\mathrm{d}} N(0, \sigma_{gI}^2)$. The first condition is satisfied if $I(\theta)$ have fatter tails than $p(\theta)$ and the second can often be verified by checking conditions for the existence of moments of the posterior.

In practice, σ_{gI}^2 is often larger than $\mathrm{Var}_p[g(\theta)]$ due to the need to estimate $E_I[w(\theta)]$ and $I(\theta)$ being a poor approximation of $p(\theta)$. The ratio $\mathrm{Var}_p[g(\theta)]/\sigma_{gI}^2$ gives the fraction of draws from the posterior (if this was possible) needed to achieve the same numerical standard error as when drawing from the importance function. Geweke termed this ratio the Relative Numerical Efficiency (*RNE*).

Gibbs sampling

In Markov chain Monte Carlo $p(\theta)$ is approximated by a Markov chain which has $p(\theta)$ as its invariant distribution. Common algorithms for generating the Markov chain are the Metropolis–Hastings algorithm and the Gibbs sampler. In this paper the Gibbs sampler is used.

The Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990) is based on the availability of tractable conditional distributions for a suitable partition of the parameter vector. That is, for $\theta = \{\theta_1, \theta_2, \theta_3\}$, say, we have the conditional distributions

$$p_1(\theta_1|\theta_2, \theta_3), \quad p_2(\theta_2|\theta_1, \theta_3), \quad p_3(\theta_3|\theta_1, \theta_2)$$

from which it is known how to generate pseudo-random numbers. The Gibbs sampler then proceeds by generating the next step in the chain, $\theta^{(i+1)} = \{\theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_3^{(i+1)}\}$ as $\theta_1^{(i+1)}$ from $p_1(\theta_1|\theta_2^{(i)}, \theta_3^{(i)})$, $\theta_2^{(i+1)}$ from $p_2(\theta_2|\theta_1^{(i+1)}, \theta_3^{(i)})$ and $\theta_3^{(i+1)}$ from $p_3(\theta_3|\theta_1^{(i+1)}, \theta_2^{(i+1)})$. If the starting values, $\theta^{(0)}$, somehow, could be obtained from $p(\theta)$, $\theta^{(1)}$ and all the following steps are also distributed as $p(\theta)$. In contrast to simple Monte Carlo integration and importance sampling consecutive draws are, however, not independent.

In practice we do not have a draw from $p(\theta)$ available. The Markov chain is started from some more or less arbitrary point on the support of $p(\theta)$ and the first draws are discarded, allowing the Gibbs sampler to 'burn in'. The convergence of the Markov chain to $p(\theta)$ then becomes a crucial issue, sufficient conditions for convergence can be found in Geweke (1995).

The validity of central limit theorems for the Gibbs sampling-based estimate, $\bar{g}_{\mathrm{G}}$, is hard to verify. In practice, the approximation $\bar{g}_{\mathrm{G}} \sim N(0, \sigma_{g\mathrm{G}}^2/n)$, $\sigma_{g\mathrm{G}}^2 = \Sigma_{-\infty}^{\infty}\lambda_j$ for λ_j the autocovariances of $g(\theta^{(i)})$, seem to work well. We estimate $\sigma_{g\mathrm{G}}^2$ using standard time series techniques as $\hat{\sigma}_{g\mathrm{G}}^2 = \Sigma_{j=-m}^{m}(1 - |j|/m)\hat{\lambda}_j$ with $m = n^{1/3}$. Due to the autocorrelation $\sigma_{g\mathrm{G}}^2$ will in general exceed $\mathrm{Var}_p[g(\theta)] = \lambda_0$. Similar to importance sampling, the numerical efficiency of the Gibbs sampler relative to sampling from the posterior can be assessed by the *RNE* $\mathrm{Var}_p[g(\theta)]/\sigma_{g\mathrm{G}}^2$.

Antithetic variates

Suppose we have available a second stream of random numbers from $p(\theta)$, the antithetic variates $\theta^{*(i)}$, which are negatively correlated with $\theta^{(i)}$. If in addition $g(\theta)$ is monotone (or at least not too non-monotonic) the correlation, ρ, between $g(\theta^{(i)})$ and $g(\theta^{*(i)})$ will be negative as well. It follows that $\bar{g}_{a,n} = [\Sigma_{i=1}^{n} g(\theta^{(i)}) + \Sigma_{i=1}^{n} g(\theta^{*(i)})]/2n$ has a smaller variance than $\bar{g}_{2n} = \Sigma_{i=1}^{2n} g(\theta^{(i)})/2n$. The ratio of the variances $\mathrm{Var}(\bar{g}_{2n})/\mathrm{Var}(\bar{g}_{a,n}) = 1/(1 + \rho)$, is known as the *GAIN* from antithetic variates and gives a rough estimate of the increase in computational time required for the same precision if antithetic variates are *not* used.

For symmetric distributions, antithetic variates with correlation -1 are easily obtained as $\theta^* = 2\mu_\theta - \theta$. With non-symmetric distributions it is slightly more complicated to obtain the antithetic variates and it is in general not possible to obtain a correlation of -1.

With importance sampling antithetic variates are obtained from the importance function. The efficiency gains are generally smaller (although still sizable in many cases) than for simple Monte

Carlo integration because of the increased non-linearity of the problem. When using the Gibbs sampler, antithetic variates can be obtained from each (or some) of the conditional distributions in each step. These variates are negatively correlated with the ordinary output from the Gibbs sampler and will, provided that the Gibbs sampler is convergent, have the required distribution $p(\theta)$.

4.2. The Normal-Diffuse Posterior

Importance sampling

Recall that the Normal-Diffuse posterior is proportional to the product of a multivariate normal density and a matricvariate t-density, which both can be approximated by a multivariate t-density. A natural choice of importance function is thus the 2-0 poly-t-density,[2] which is proportional to the product of two multivariate t-densities. The mean and variance of the two multivariate t-densities are set to the mean and variance of the normal and matricvariate t factors of the posterior (9). The degrees of freedom of the normal factor is set to an arbitrary large number (we use 1000). For the matricvariate t-factor, choosing the degrees of freedom as $T - mk - 1$ ensures that the importance function has heavier tails than the posterior. For VAR-models with many variables and/or a high lag order this might not be possible and the use of the 2-0 poly-t importance function is questionable in these cases.[3]

The algorithm of Bauwens and Richard (1982) is used to generate pseudo-random numbers from the 2-0 poly-t. Antithetic variates are obtained in an intermediate step of the algorithm and the use of antithetic variates is virtually cost less although less efficient than when the importance function is symmetric.[4]

Gibbs sampling

We have, after some manipulation, the conditional posterior distributions

$$\gamma|\Psi, \mathbf{y} \sim N(\bar{\gamma}, (\Sigma^{-1} + \Psi^{-1} \otimes \mathbf{Z}'\mathbf{Z})^{-1}) \tag{16a}$$

and

$$\Psi^{-1}|\gamma, \mathbf{y} \sim W([(\mathbf{Y} - \mathbf{Z}\hat{\Gamma})'(\mathbf{Y} - \mathbf{Z}\hat{\Gamma}) + (\Gamma - \hat{\Gamma})'\mathbf{Z}'\mathbf{Z}(\Gamma - \hat{\Gamma})]^{-1}, T) \tag{16b}$$

where $\bar{\gamma} = (\Sigma^{-1} + \Psi^{-1} \otimes \mathbf{Z}'\mathbf{Z})^{-1}[\Sigma^{-1}\tilde{\gamma} + (\Psi^{-1} \otimes \mathbf{Z}'\mathbf{Z})\hat{\gamma}]$.

The Gibbs sampler is thus easy to implement, switching between equations (16a) and (16b) and using the algorithm given in Geweke (1988) to draw from (16b). One outstanding problem is that equation (16a) requires the factorization of a $mk \times mk$ matrix and inversion of the factor matrix in each step. The speed of the algorithm will thus decrease rapidly as the number of parameters in the VAR-model increases.

We start the Gibbs sampler by generating γ from equation (16a) with Ψ as the least squares estimate and use a burn-in period of 200 draws. Some experimentation showed that the Gibbs sampler is insensitive to the choice of starting value for Ψ.

[2] See Drèze (1977) on poly-t densities.

[3] In addition to the 2-0 poly-t we also experimented with the normal approximation of Zellner (1971) and Geweke's (1989) split-normal modification of the approximation. The results for these importance functions are considerably worse than for the 2-0 poly-t. See Kadiyala and Karlsson (1994) for details of these importance functions.

[4] The algorithm of Bauwens and Richard (1982) is described in Appendix A.

4.3. The Extended Natural Conjugate Posterior

Importance sampling

The tails of the posterior distribution (10) behave like a multivariate or matricvariate t, suggesting that these are suitable importance functions. A mk-dimensional multivariate t importance function can only have thicker tails than the posterior if $T + \alpha > mk$ and might not be suitable for large VAR-models. The degrees of freedom for a matricvariate t importance function can, on the other hand, be chosen to produce thicker tails as long as $T + \alpha > k$. The Kronecker structure of the variance–covariance matrix of the matricvariate t might, however, limit the usefulness of the matricvariate t as an importance function.[5]

A third possibility is to use a product of multivariate t-distributions for the parameters of each equation as importance function. The ENC posterior is in fact quite amenable to this approach. From lemma 6.4 of Drèze and Richard (1983) it follows that the posterior distribution of γ_i condtional on the other regression parameters is a multivariate t with $T + \alpha - k$ degrees of freedom.

Bauwens (1984) used this result to construct a number of importance functions for the ENC posterior in a simultaneous equations setting. Bauwens found that the STFC (student with fixed conditioning) importance function performed best (in terms of CPU-time required to achieve a certain precision). This importance function is constructed as the product of conditional multivariate t-distributions,

$$I(\gamma) = \prod_{i=1}^{m} p(\gamma_i | \gamma_j = \gamma_j^*, j \neq i)$$

where γ_j^* is the parameter vector of equation j at the posterior mode.

The STFC importance function has the disadvantage that the parameters of the different equations are independent. It will consequently be a poor approximation of the posterior if the posterior variance–covariance matrix is not block-diagonal. In addition, conditioning on a fixed set of values tends to give the importance function thinner tails than the posterior.

Gibbs sampling

The Gibbs sampling algorithm for the ENC posterior is also based on lemma 6.4 of Drèze and Richard (1983). That is, the parameters of equation i are distributed as a multivariate t, conditionally on the parameters of the remaining equations,

$$\gamma_i | \gamma_1, \ldots, \gamma_{i-1}, \ldots, \gamma_m \sim t\left(\mathbf{d}_i, \frac{q_{ii}}{T + \alpha - k} \mathbf{P}_{ii}^{-1}, T + \alpha - k\right) \tag{17}$$

The Gibbs sampler is thus implemented by cycling through equation (17) for $i = 1, \ldots, m$. Note that each cycle requires the calculation of q_{ii}, the vectors $\mathbf{d}_i$ and the matrices $\mathbf{P}_{ii}$, which can be time consuming for large models.[6]

[5] We have tried several variations on the matricvariate t importance function similar in the spirit of the split-multivariate t of Geweke (1989), but the results were relatively disappointing. See Kadiyala and Karlsson (1994) for details of these importance functions.

[6] See Drèze and Richard (1983) for explicit expressions for $\mathbf{d}_i$, q_{ii} and $\mathbf{P}_{ii}$.

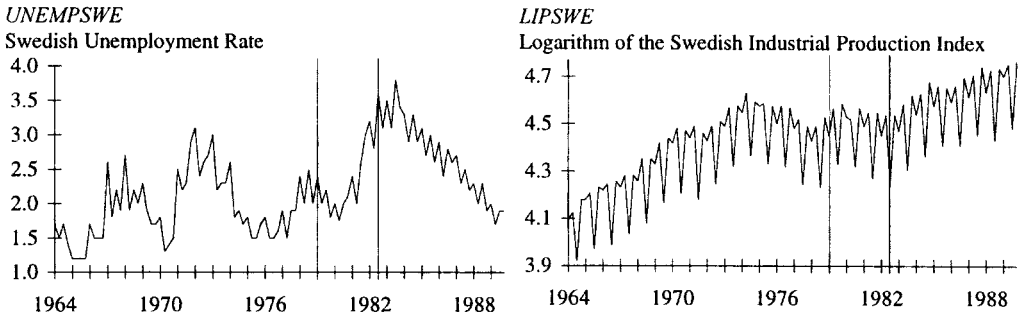

Figure 2. The Swedish data. The vertical lines mark the first and last observation used when determining the hyperparameters

The prior means of γ_i are used as starting values for the Gibbs sampler with a burn-in period of 200 draws. The starting values have virtually no influence on the draws after the burn-in and the Gibbs sampler appears to converge to the true posterior distribution very rapidly.

5. APPLICATIONS

5.1. The Swedish Unemployment Rate

Edlund and Karlsson (1993) compared forecasts of the Swedish unemployment rate from VAR, ARIMA and transfer function models using classical estimation techniques. Here we will consider a modification of the VAR model which produced the best forecasts of the Swedish unemployment rate.[7] This VAR is a simple bivariate model with seasonal dummies, two lags of the unemployment rate (*UNEMPSWE*) and the first difference of the logarithm of the industrial production index (*LIPSWE*).

In the identification of the model a premium was put on short lag lengths. Since Bayesian VAR models handle longer lag lengths better than OLS-estimated ones we increase the lag length to 4. In addition, unit roots do not cause as many problems in a Bayesian analysis and we use the levels of *LIPSWE* rather than the first difference.

The data set in Figure 2 consists of quarterly data for the period 64:1 to 90:4. Edlund and Karlsson used data up to 78:4 to identify the original model. The observations from 79:1 to 82:3 are used to determine the hyperparameters of the prior distributions and the forecast comparison uses the remainder of the data.

The specification of the prior means differs from the random walk (5) for this model. It is unreasonable that the unemployment rate is integrated (although it might be a good local approximation) and the prior mean of the first own lag of *UNEMPSWE* is set to 0·5. The analysis in Edlund and Karlsson indicates that *LIPSWE* is integrated and for this variable we retain the prior mean of unity for the first own lag.

When specifying the prior variances we use the specification (6) and, following Doan *et al.* (1984), the values of the hyperparameters are chosen based on the forecast performance over a calibration period. Four sets of forecasts up to eight periods ahead covering the period 79:1 to 82:3 was generated for different values of the hyperparameters (π_3 was kept constant at $1{\cdot}4 \cdot 10^5$,

[7] Both the ARIMA model and the transfer function model, using *LIPSWE* as the input variable, gave better forecasts.

Table II. Prior hyperparameters: Swedish unemployment

Prior distribution	π_1	π_2	π_3	d.f.
Minnesota	0·0016	0·8	$1{\cdot}4 \cdot 10^5$	—
Normal-Wishart	0·12	0·12	$1{\cdot}4 \cdot 10^5$	9
Normal-Diffuse	0·002	1·3	$1{\cdot}4 \cdot 10^5$	—
ENC, unconditional mean of Ψ	0·0015	0·6	$1{\cdot}4 \cdot 10^5$	16
ENC, conditional mean of Ψ	0·012	3·9	$1{\cdot}4 \cdot 10^5$	16

i.e. the prior is diffuse on the coefficients of the deterministic variables). The sum of the mean square forecast error over the eight lead times for *UNEMPSWE* is plotted against the hyperparameters for the different priors in Figure 3 and the minimizing values are shown in Table II.

The choice of prior degrees of freedom has very little effect on the forecasts for the Normal-Wishart prior and the value that minimizes the sum of MSEs, 9, is close to the minimum degrees of freedom, 4, from (8). In light of this, and the computational cost involved, different values of the prior degrees of freedom were not tried for the ENC prior. The choice $\alpha = 16$ guarantees that the prior variance of the regression parameters exists.

The Minnesota, Normal-Diffuse and ENC priors are all more sensitive to the choice of π_1 than to the choice of π_2. At the minimum the surface is essentially flat in the π_2-direction. They also have in common that the minimizing value of π_2 is considerably larger than the minimizing value of π_1. This gives a large prior variance for foreign lags and little shrinkage towards the prior mean of zero. We interpret this to mean that industrial production is important when determining the unemployment rate and that the unemployment rate is important when determining industrial production. The tighter priors (π_1) on the own lags lend some support for the specification of the prior means for these parameters.

In terms of the minimizing values of π_1 and π_2, the Minnesota prior, Normal-Diffuse prior and the ENC prior using the unconditional prior mean of Ψ are close to each other. For the Normal-Wishart prior $\pi_1 = \pi_2$ by construction and the minimizing choice is a compromise between the tight specification on own lags and diffuse specification on foreign lags of the other priors.

Posterior distributions of the forecasts

With the large number of parameters found in VAR-models it is difficult to summarize the posterior distributions concisely. The parameters are also difficult to interpret by themselves and more insight is often gained by studying suitable functions of the parameters. We have chosen to focus on the posterior distribution of the forecast function (3), in other applications it might be more appropriate to focus on impulse response functions (see Koop, 1992, for an example of this). Note that the posterior distribution of the forecast function differs from the predictive density for y_{t+h}. The latter accounts for the uncertainty due to the disturbances $u_{t+1}, \ldots, u_{t+h}$ as well as the posterior uncertainty about the parameters.

The bivariate distributions of the lead time 4 forecasts of *UNEMPSWE* and *LIPSWE* generated as of 1985:4 (i.e., the forecast for 1986:4) are displayed in Figure 4. The most striking feature of Figure 4 is that the Diffuse prior produces a tighter forecast distribution than the ones obtained from the Normal-Diffuse and ENC priors. This difference is mostly due to the larger variation in the forecast of industrial production with the Normal-Diffuse and ENC priors. As can be seen, the forecasts are negatively correlated, with the Diffuse prior displaying the highest

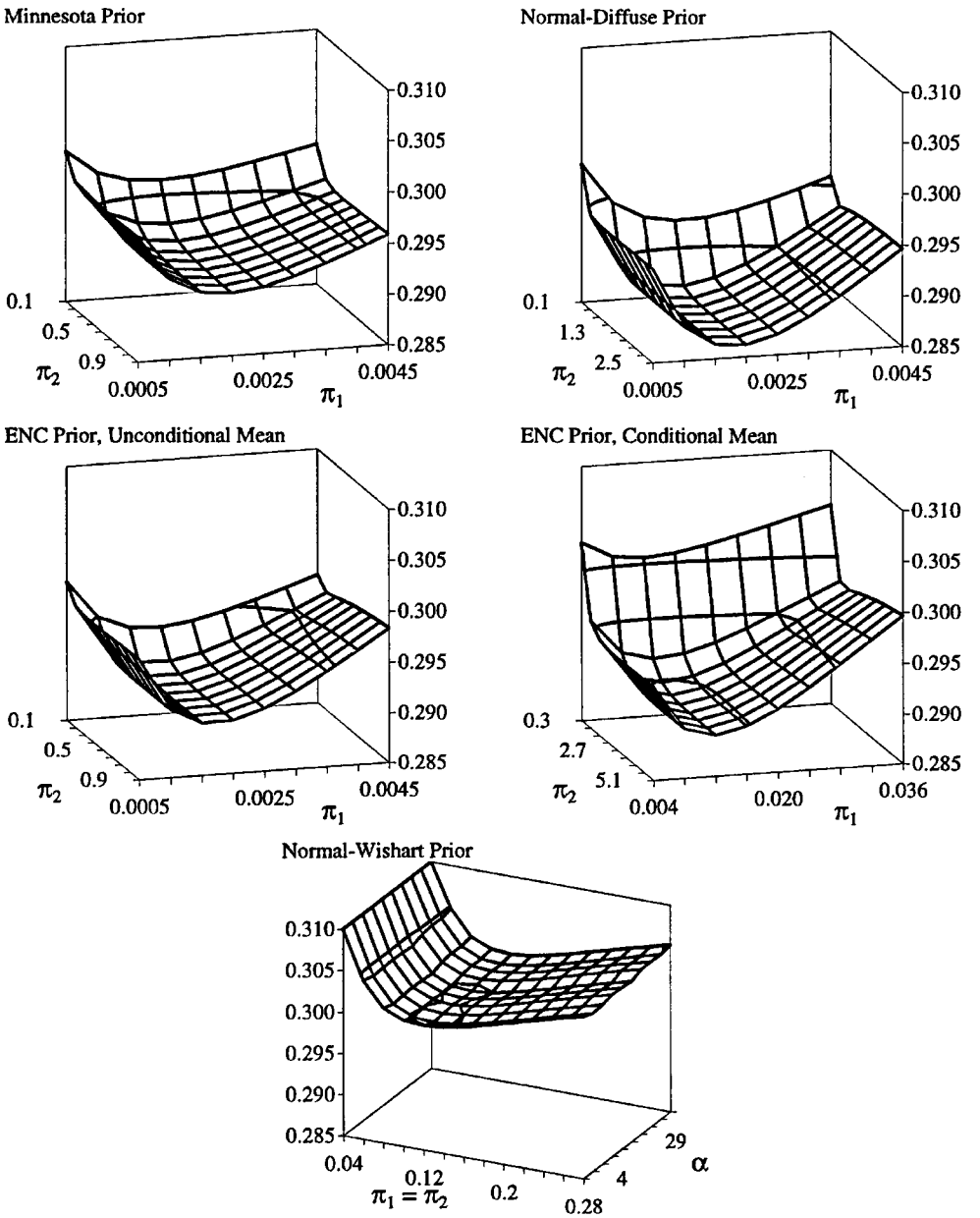

Figure 3. Choice of prior hyperparameters: Swedish unemployment. The figures show the sum of the Mean Square Error for forecasts of unemployment over the lead times 1 to 8 for the period 1979:1 to 1982:3

correlations, −0·56 at lead time 4. With the ENC and Minnesota prior the correlations are low, about −0·25, and the Normal-Diffuse and Normal-Wishart have intermediate correlations of about −0·45 at lead time 4. There are also some differences in location, especially for the forecast

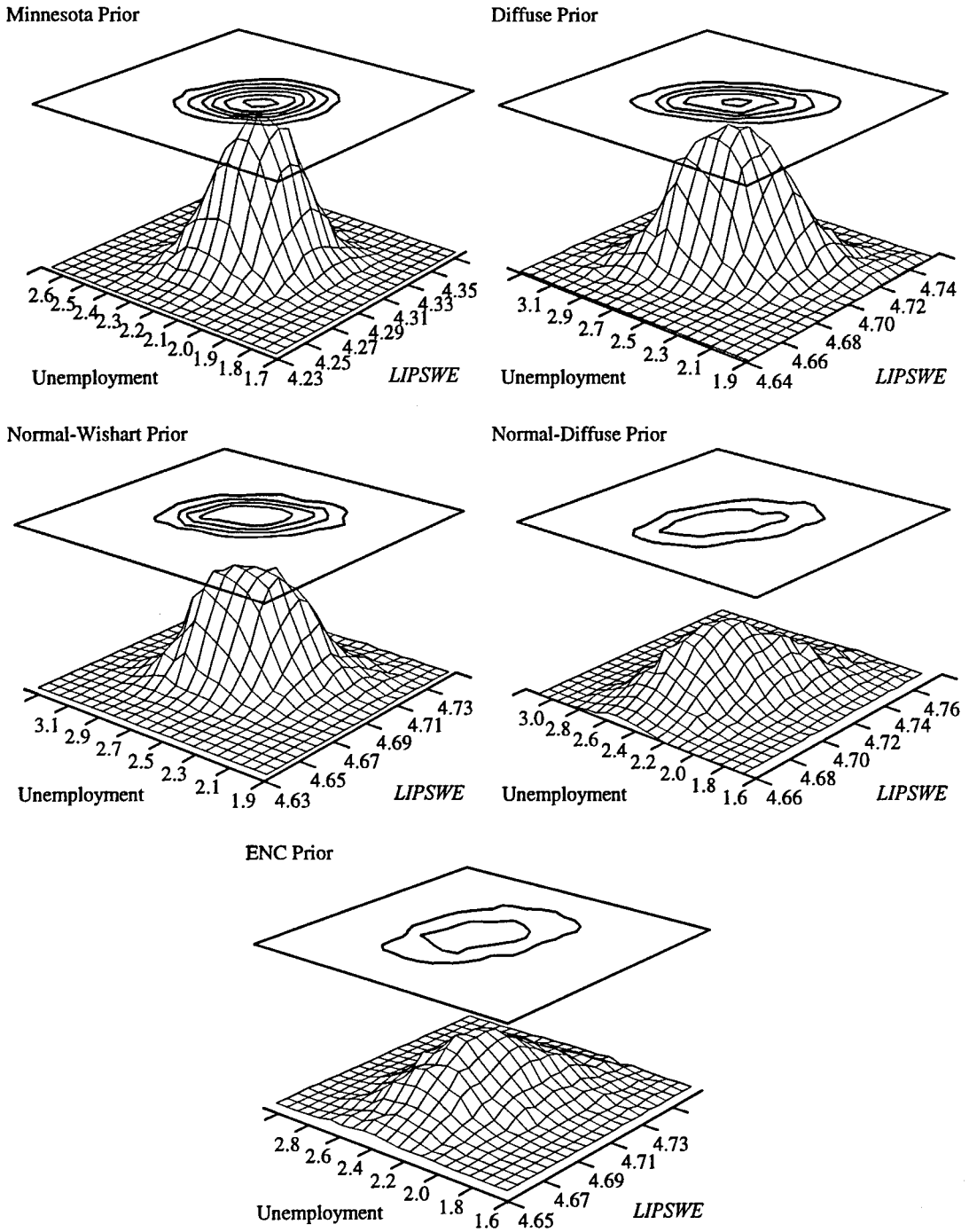

Figure 4. Posterior distributions of forecasts for the Swedish unemployment rate and industrial production at lead time 4. Forecasts generated as of 1985:4. Contours are equidistant level curves

of *UNEMPSWE*. The Minnesota prior gives low forecasts and the Diffuse and Normal-Wishart priors give high forecasts, with the Normal-Diffuse and ENC priors in between but on the low side.

In Figure 5 the lead time 1 to 8 distributions of the forecast function for the unemployment rate is displayed with the actual unemployment rates for the period 1986:1–1987:4. For the Diffuse and Normal-Wishart priors the 90% highest posterior density (HPD) regions contain the unemployment rate for all lead times. For the other priors the lead time 2 and 3 forecasts lie outside the 90% HPD. The priors give rise to similar distributions for the lead time 1 forecasts whereas they behave differently for the higher lead times. The width of the HPDs increases slowly but more or less linearly for the Diffuse and Normal-Wishart prior. As can be expected, the HPDs for the Diffuse prior are wider than for the Normal-Wishart. With the Minnesota, Normal-Diffuse and ENC priors the width of the HPD increases substantially going from lead time 1 to lead time 2. For higher lead times the width of the HPD increases at a decreasing rate for these priors and the width is essentially constant from lead time 5 and up.

Forecasting performance

The root mean square error for the forecasts of the Swedish unemployment rate is shown in Figure 6. The priors divide into two groups with very similar RMSEs within the groups. The Diffuse, Normal-Wishart priors and OLS do well while the ENC, Minnesota and Normal-Diffuse priors produce almost twice as large RMSEs.[8] Note that this is the same grouping as for the location of the forecast distribution for the unemployment rate in Figure 5.

The two specifications of the ENC prior give virtually identical RMSEs and we only report the results when using the conditional expectation of Ψ. There was also very little difference between the RMSEs for the Minnesota prior when using the posterior mean of the forecasts and when the forecasts are generated using the posterior mean of the parameters. We only report the results when using the posterior means.

Numerical performance

The diagnostics on the Monte Carlo integration are displayed in Figure 7. The numerical standard error of the Monte Carlo estimates, the gain from antithetic variates (*GAIN*) and the relative numerical efficiency (*RNE*) are shown for the lead time 1, 4 and 8 forecasts of *UNEMPSWE*. The CPU time used when estimating the lead time 1 to 8 forecasts using 10,000 draws and antithetic variates as well as the CPU time required to achieve a numerical standard error which is 1% of the forecast standard deviation is also shown.[9]

Monte Carlo integration poses no problem with the Normal-Wishart and Diffuse priors and the results are mainly displayed for comparative purposes.

With the Normal-Diffuse we cannot sample directly from the posterior and must resort to importance or Gibbs sampling. Gibbs sampling is slightly more expensive than using the 2-0 poly t importance function in absolute terms. This is, however, offset by the Gibbs samplers higher *RNE* and *GAIN* from antithetic variates so that Gibbs sampling is cheaper when the precision of the estimates is accounted for. The high *RNE* for the Gibbs sampler is a result of the low

[8] The forecasting results for the ENC, Minnesota and Normal-Diffuse priors are sensitive to the specification of the prior means. The forecasting experiment was also run with the prior mean of the first own lag of *UNEMPSWE* set to 1 rather than 0·5. In this case all the priors gave RMSEs very close to the RMSE obtained with OLS.

[9] All computations were carried out on a VAX 6000-410 using a purpose-written FORTRAN program. This machine is slow by today's standards and the CPU times reported should only be taken as relative measures. Those desiring a more up-to-date metric can divide the CPU times by 10 to obtain the approximate timings for a 90 MHz Pentium PC.

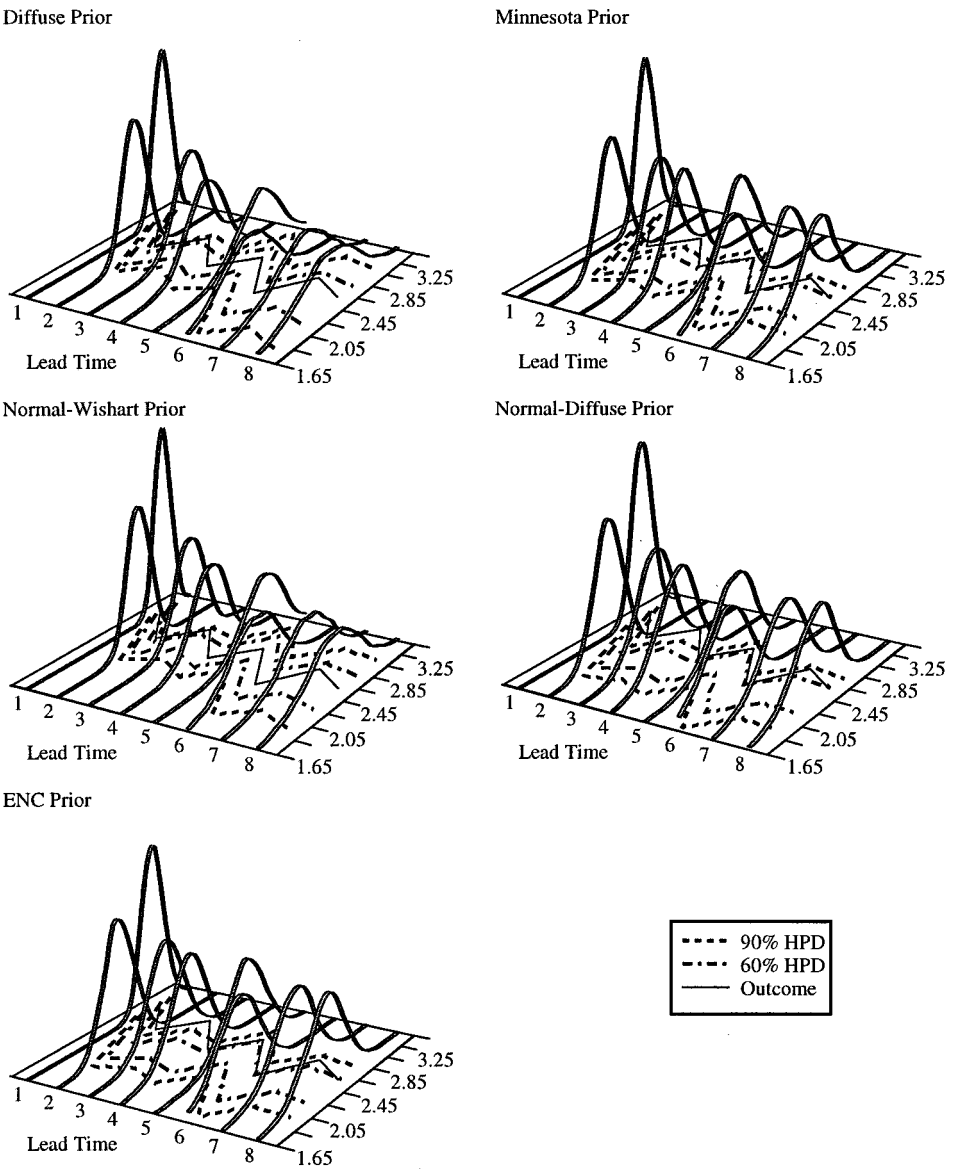

Figure 5. Posterior distributions of forecasts: Swedish unemployment rate. Forecasts generated as of 1985:4

correlation between the draws from the Gibbs sampler. The relatively low *GAIN* from antithetic variates observed for the 2-0 poly *t* is probably due to the way the antithetic variates are obtained with this importance function.[10] Since it is non-symmetric, the antithetic variates are not as strongly negatively correlated with the ordinary variates as for the other sampling schemes.

[10] See Appendix A.

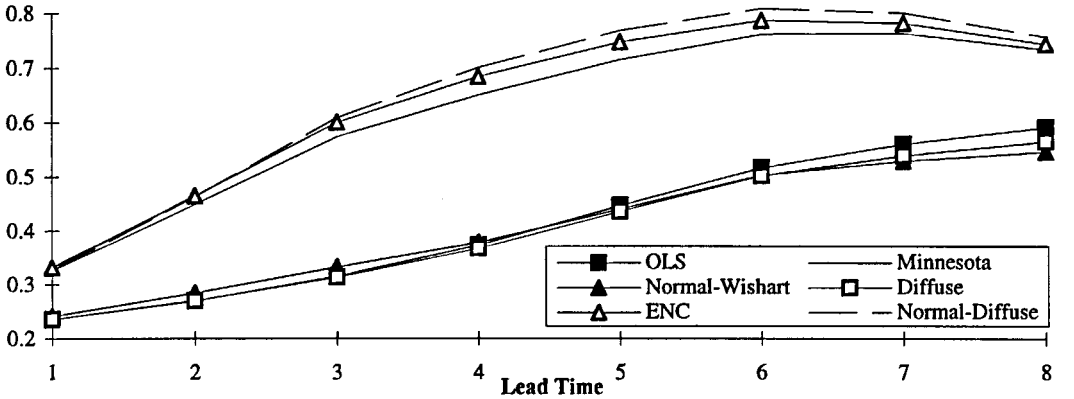

Figure 6. Root Mean Square Error: Swedish unemployment rate. Thirty sets of one- to eight-periods-ahead forecasts covering the period 1982:4 to 1990:4

The ENC posterior also requires importance or Gibbs sampling. With this posterior, both the STFC importance function and the Gibbs sampler do well. The draws from the Gibbs sampler are essentially uncorrelated and the *RNE* is very close to 1, the Gibbs sampler also achieves a higher *GAIN* from antithetic variates than the importance function. As a consequence, the Gibbs sampler also gives the lowest numerical standard errors. The Gibbs sampler is, however, more than twice as expensive as the STFC importance function in absolute terms. If the precision is accounted for, the Gibbs sampler is cheapest for the one-step-ahead forecast and the STFC importance function is cheaper for the longer lead times.

5.2. The Litterman Forecasting Model

The second model we consider is the US forecasting model of Litterman (1986). It consists of seven dependent variables, annual growth rates of real GNP (*RGNPG*), annual inflation rates (*INFLA*), the unemployment rate (*UNEMP*), the logarithm of nominal money supply (*M1*), the logarithm of gross private domestic investment (*INVEST*), the interest rate on four- to six-month commercial paper (*CPRATE*), and the change in business inventories (*CBI*). The right-hand side variables are six lags of each of the seven variables and a constant term, yielding a total of 301 parameters.

The data set consists of quarterly data from 1948:1 and 20 sets of one- to eight-periods-ahead forecasts covering the period 1980:2 to 1986:4 are generated in the forecasting experiment. The variables are displayed in Figure 8 and the data sources are given in Appendix B. The data set we use differs from the one used by Litterman, mainly in that we only include end of quarter data and use the last revisions available at the time of data collection.

While our forecasting results will not be directly comparable to the ones of Litterman (1986), we follow Litterman closely and specify the prior moments in essentially the same way. That is, the prior mean of the regression parameters is the random walk (5) and the prior variances are as in equation (6). The prior hyperparameters in Table III are equivalent to the ones used by Litterman except for π_3. Litterman sets the prior precision for the deterministic variables to zero whereas we set the prior variance to a very large number. Also, recall that the prior variances in

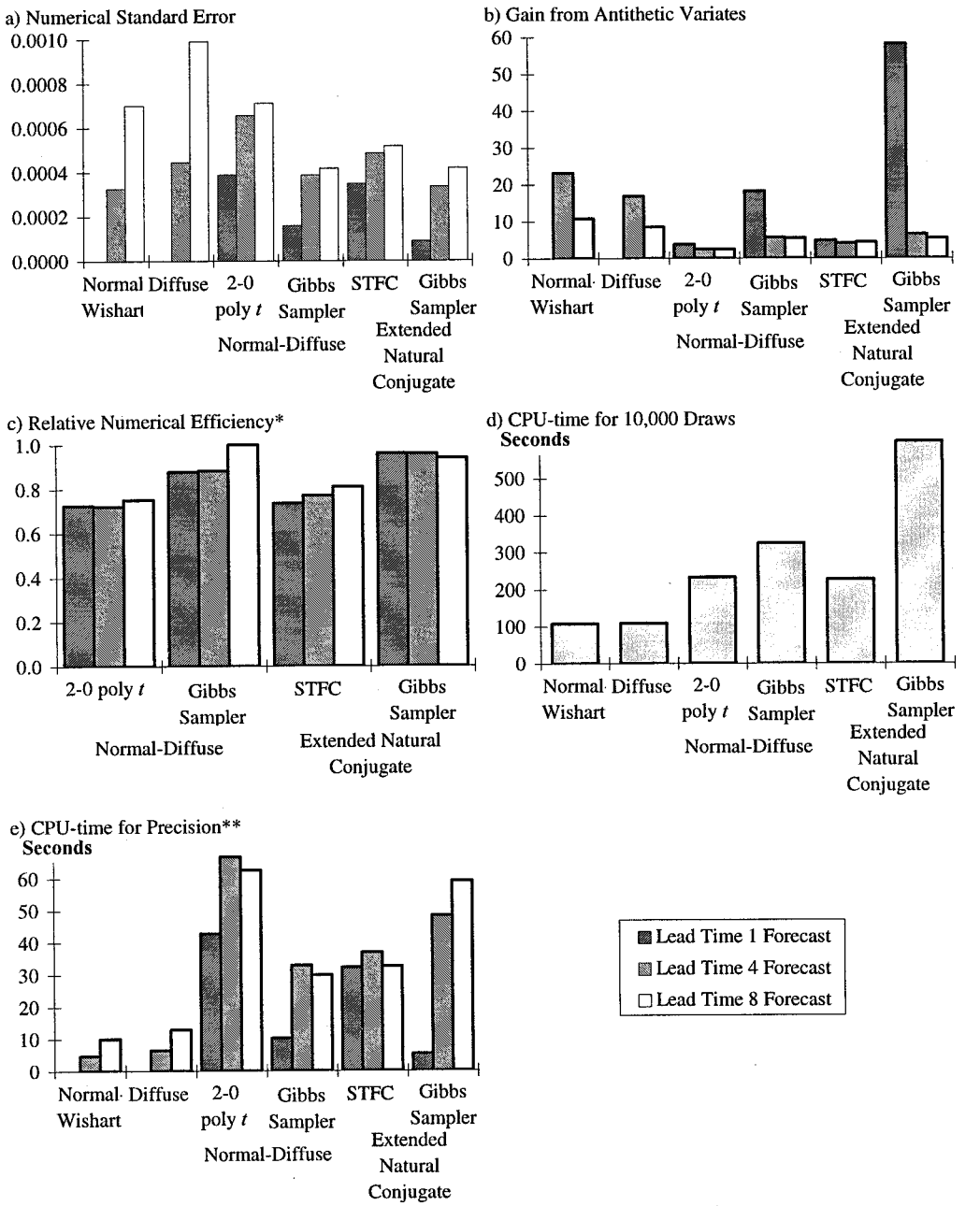

Figure 7. Performance of Monte Carlo methods: Swedish unemployment. All results are based on 10,000 draws from the posterior, importance function or the Gibbs sampler. Statistics are for forecasts of the Swedish unemployment rate (*UNEMPSWE*). The ENC prior is specified using the conditional prior expectation of Ψ. * The *RNE* for the Normal-Wishart and Diffuse priors is 1 by definition. ** The CPU-time required to achieve a numerical standard error which is 1% of the forecast standard deviation

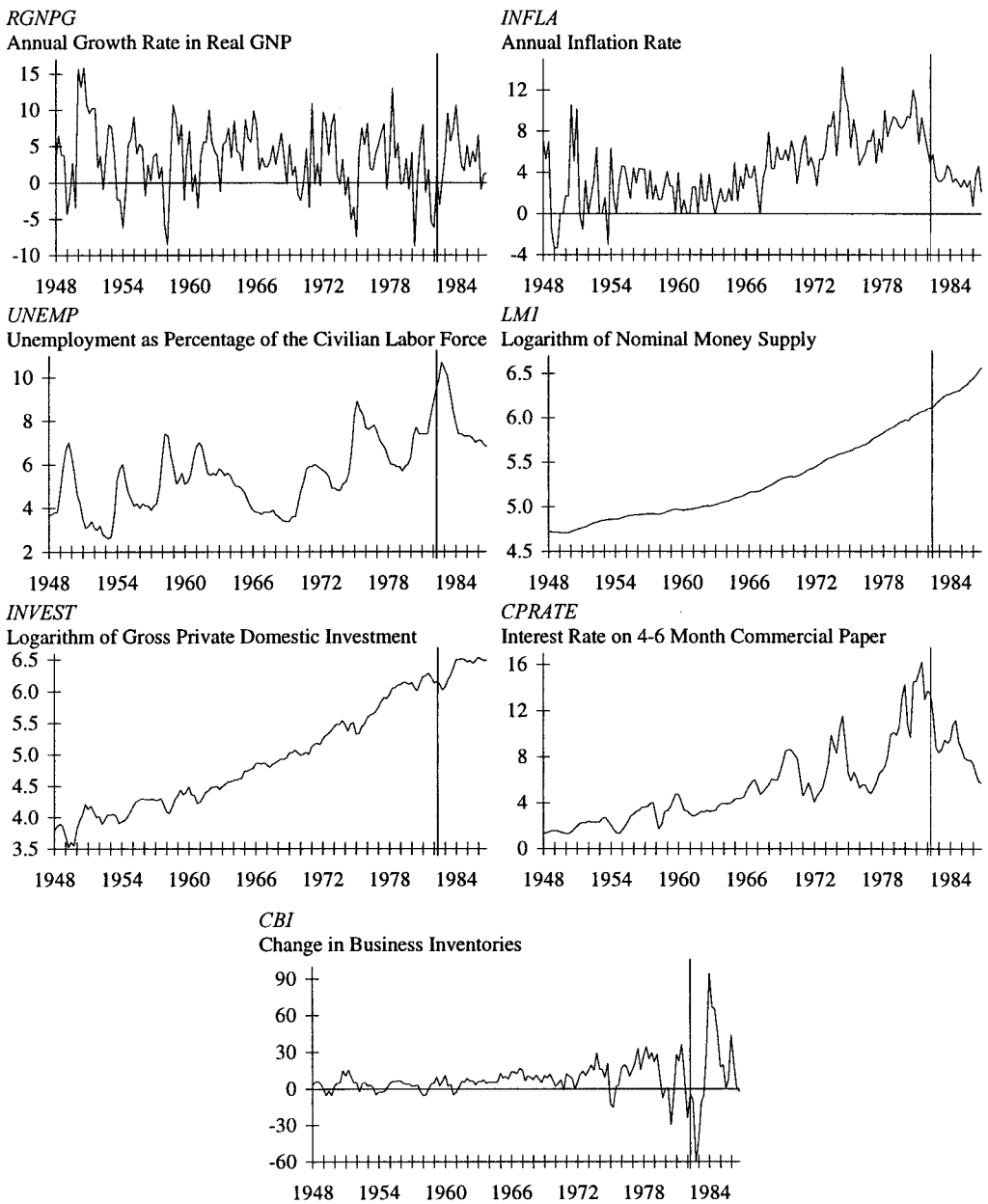

Figure 8. The Litterman data. The vertical line marks the beginning of the forecasting period

equation (6) decay slower with the lag length than in Litterman (1986). The geometric mean of π_1 and π_2 is used with the Normal-Wishart prior.

The ENC prior is specified using the conditional prior expectation of the residual variance matrix Ψ, rather than the unconditional expectation.

Table III. Prior hyperparameters: Litterman forecasting model

Prior distribution	π_1	π_2	π_3	d.f.
Minnesota	0·04	0·0036	$1{\cdot}4 \cdot 10^5$	—
Normal-Wishart	0·012	0·012	$1{\cdot}4 \cdot 10^5$	9
Normal-Diffuse	0·04	0·0036	$1{\cdot}4 \cdot 10^5$	—
Extended Natural Conjugate	0·04	0·0036	$1{\cdot}4 \cdot 10^5$	52

Posterior distributions of the forecasts

The posterior distributions of the forecast function (3) at lead times 1 to 8 (1980:2 to 1982:1) for the variables *RGNPG*, *INFLA* and *UNEMP* are displayed in Figure 9. This is a period with very rapid fluctuations in the growth rate of real GNP and none of the priors do well in forecasting *RGNPG* for this period.

The Diffuse prior gives very wide HPDs for the forecast function compared to the other prior distributions (note that the posterior distributions are drawn to different scales for the Diffuse prior). With the huge number of parameters in the Litterman model, the data speak less clearly and the information embodied in the priors affects the shape of the posterior distributions much more than with the Swedish Unemployment model. The posterior distributions also differ more over the posteriors with informative priors (Minnesota, Normal-Wishart, Normal-Diffuse and ENC) with the Litterman model. The Normal-Diffuse stands out as the posterior with the tightest HPDs for the forecast function, particularly for *RGNPG*.[11] With the ENC prior the mean of the forecast functions for *RGNPG* and *INFLA* is essentially constant over the lead times. The mean of the forecast function for *UNEMP* also increases slowly with the ENC prior, this is in contrast to the other priors which forecast rather dramatic increases in the unemployment rate (note that the scales differ between priors for *UNEMP*). Except for *RGNPG*, the Minnesota and Normal-Diffuse priors give very similar results. The Minnesota and Normal-Wishart also show some resemblance although the Normal-Wishart gives wider HPDs. This is as can be expected when relaxing the assumption of a fixed residual variance–covariance matrix.

It is clear from these differences that the choice of distribution used to embody the prior beliefs is not an innocuous one. Even when we have the same prior mean and variance, the differences in correlations and higher moments between the priors lead to quite different posterior distributions.

Forecasting performance

The RMSEs for three variables, the growth rate of real GNP, the rate of inflation and the unemployment rate, considered by Litterman are given in Figure 10. The results for the Minnesota prior are very close when using the posterior means of the forecasts and forecasts based on the posterior means of the parameters. We only report the results for forecasts using the posterior means of the parameters.

[11] The narrow HPDs for the Normal-Diffuse prior and *RGNPG* could be an indication of bimodality of the posterior distribution and the Gibbs sampler being stuck in one region. To check this the Gibbs sampler was rerun several times with different starting values for Ψ and different random number seeds. In no case did the results from these (shorter) runs differ markedly from the long run used to obtain the posterior distributions.

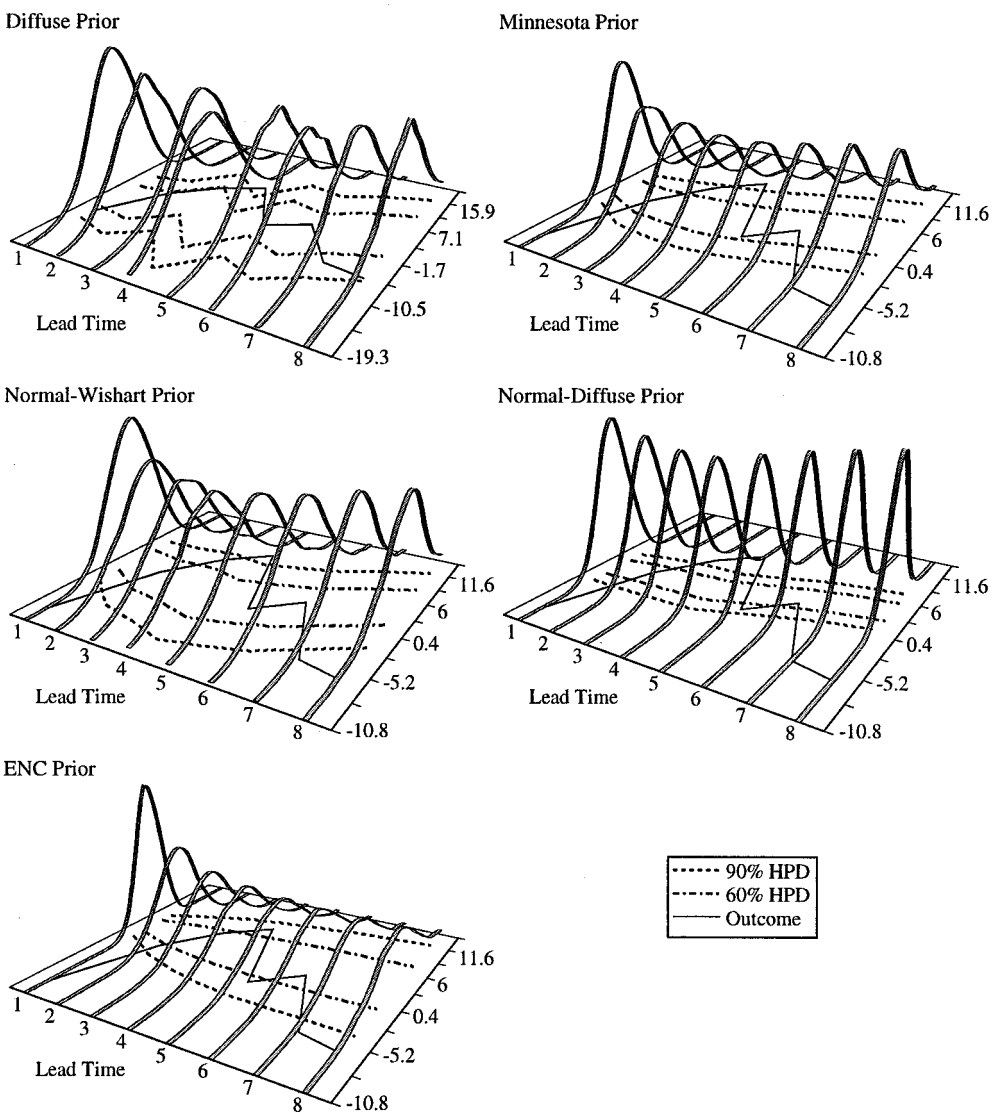

Figure 9a. Posterior distributions of forecasts of *RGNPG*. Forecasts generated as of 1980:1

The Diffuse prior and OLS give essentially identical results for all combinations of lead times and variables. With two exceptions they also produce the worst forecasts. The rankings of the other priors depend on which variable we consider. For *RGNPG* the performance is similar across priors with the Minnesota and Normal-Diffuse priors doing slightly better than the other prior distributions. The ENC prior gives the best forecasts for *INFLA* and the Normal-Diffuse gives slightly better forecasts than the Minnesota and Normal-Wishart priors for longer lead times.

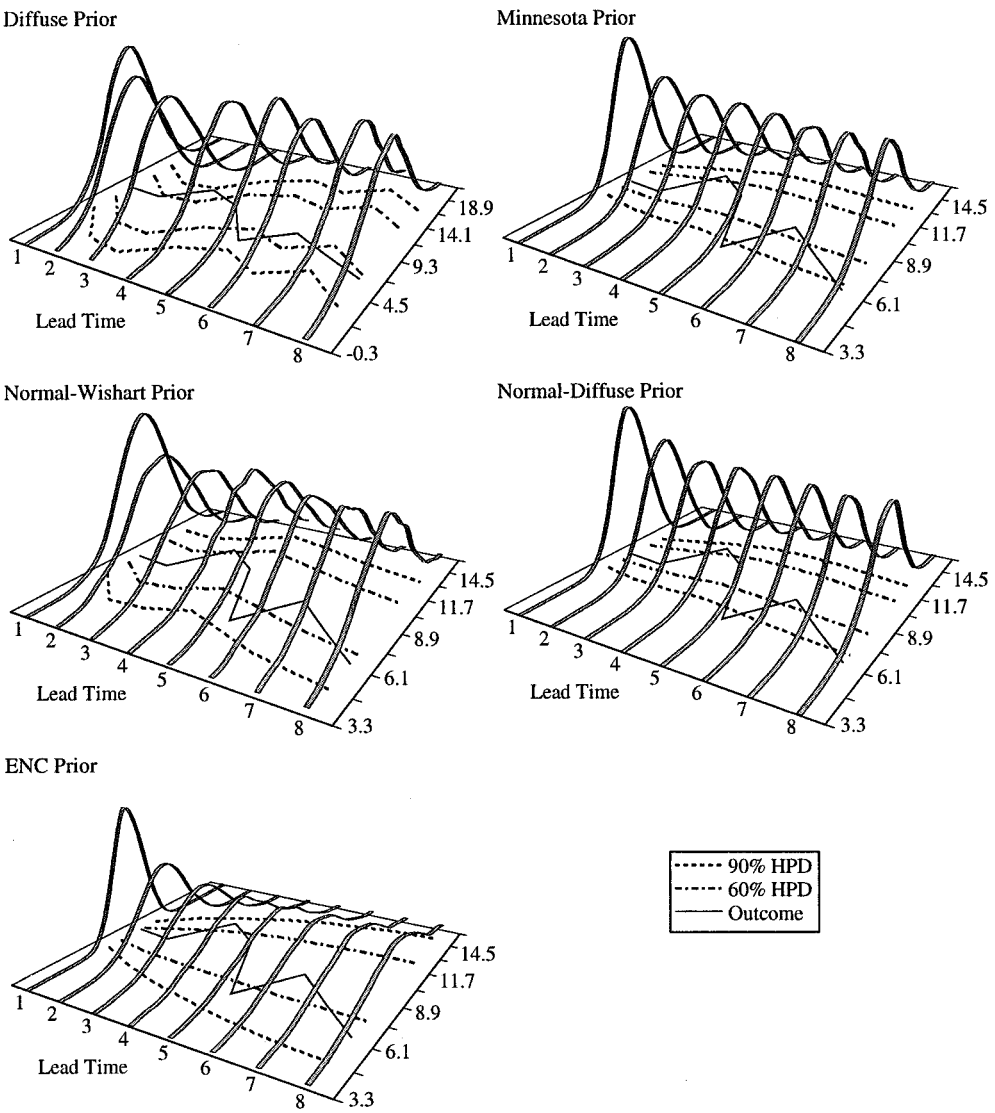

Figure 9b. Posterior distributions of forecasts of *INFLA*. Forecasts generated as of 1980:1

Turning to *UNEMP*, the Minnesota and Normal-Diffuse priors give the best forecasts, closely followed by the Normal-Wishart. The forecast performance of the ENC prior deteriorates faster with the lead time and is worse than that of OLS and the Diffuse prior for lead times 6 and 7.

Overall, the results for the Minnesota, Normal-Wishart and Normal-Diffuse priors are the best with a slight edge for the Normal-Diffuse and Minnesota priors. As mentioned, OLS and the Diffuse prior give the worst forecasts. The performance of the ENC prior is more erratic, it gives the best forecasts for *INFLA*, but it is among the worst when it comes to *UNEMP*.

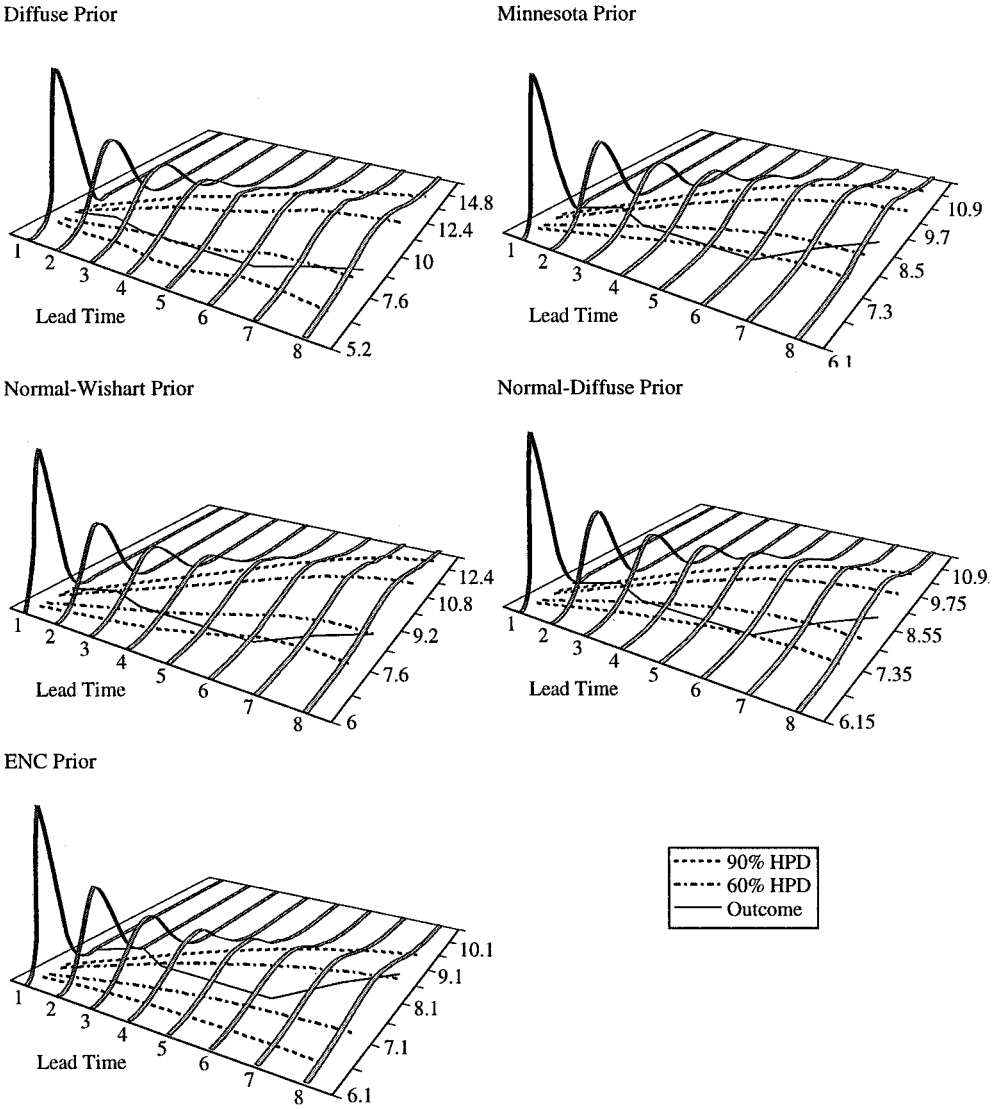

Figure 9c. Posterior distributions of forecasts of *UNEMP*, Litterman model. Forecasts generated as of 1980:1

It should be noted that the forecasting comparison suffers from some drawbacks and is not entirely fair to the Normal-Wishart, Normal-Diffuse and ENC priors in that the same set of hyperparameters, chosen by Litterman for use with the Minnesota prior, is used with all priors.

Numerical performance

Before going into details about the numerical performance of the various methods of evaluating the posterior distributions, it should be noted that the size of the Litterman model

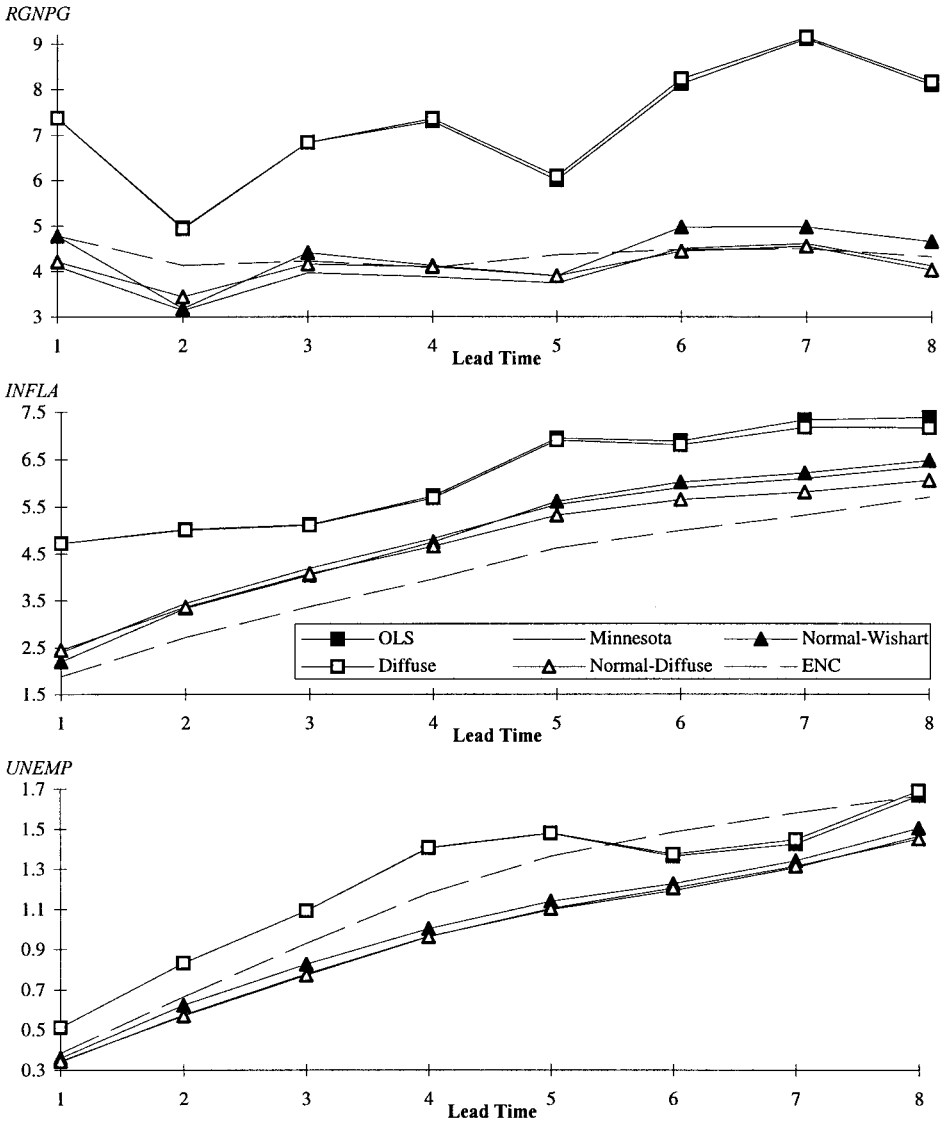

Figure 10. Root Mean Square Error, Litterman forecasting model. Statistics are for 20 sets of one- to eight-periods-ahead forecasts covering the period 1980:2 to 1986:4

introduces some new problems. One obvious effect of the larger size is the increased CPU-time requirements evidenced by Figure 11(d). The Normal-Wishart and Diffuse posteriors — where we sample directly from the posterior — are the least affected with the increase in CPU-time being less than proportional to the increase in the size of the model.

For the STFC and 2-0 poly-t importance functions the increase in CPU-time appears to be proportional to $(mk)^{3/2}$. The more rapid increase for the importance functions is probably due to the more complex calculations involved in evaluating the weights for each draw. The Gibbs

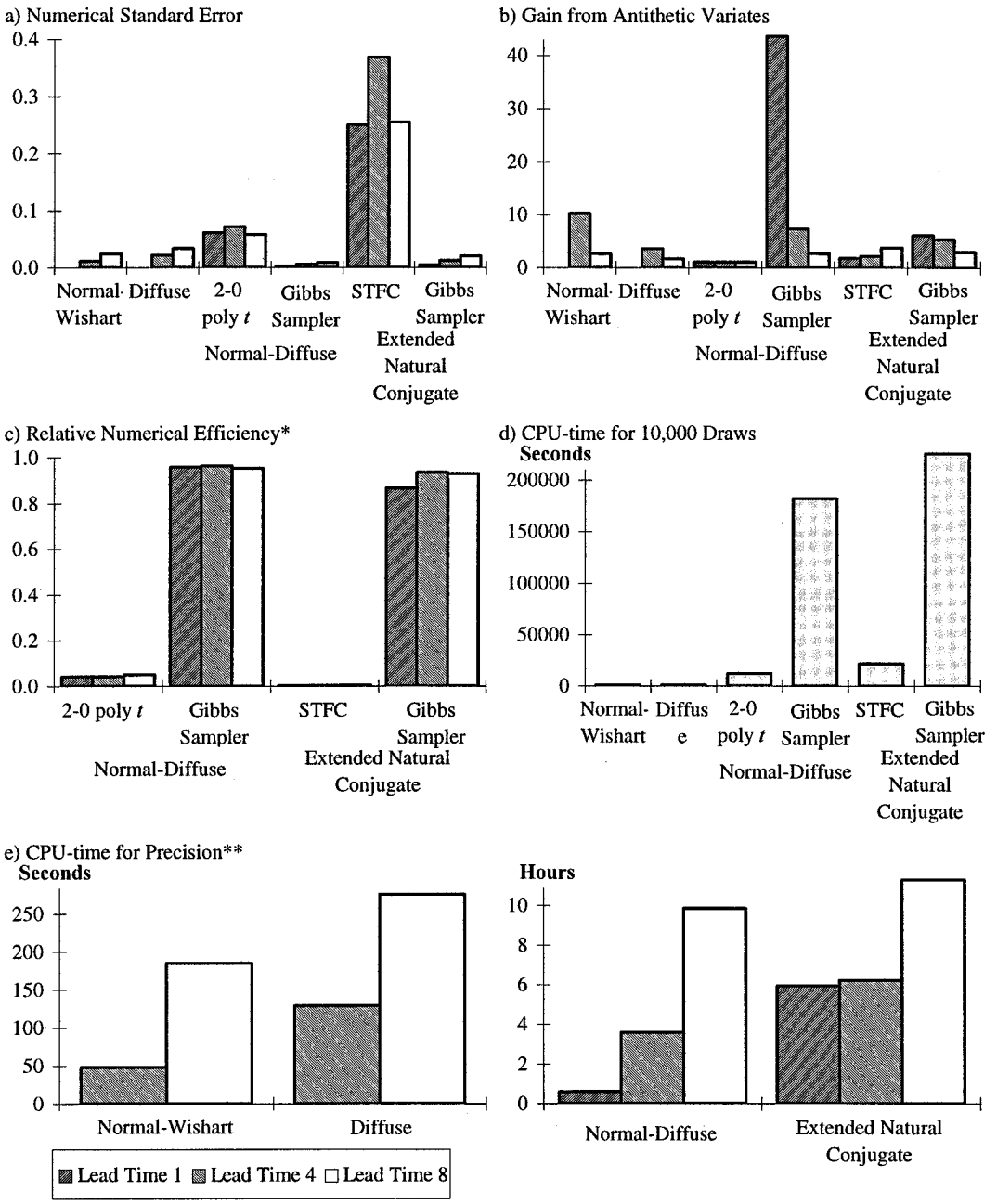

Figure 11. Performance of Monte Carlo methods, Litterman forecasting model. The results are based on 10,000 draws from the posterior, importance function or the Gibbs sampler. Statistics are for forecasts of *RGNPG*. * The *RNE* for the Normal-Wishart and Diffuse priors is 1 by definition. ** CPU-time required to achieve a numerical standard error which is 1% of the forecast standard deviation. **Seconds** for the Normal-Wishart and Diffuse priors, **hours** for the Normal-Diffuse and ENC priors (Gibbs sampler)

sampling algorithms are most severely affected by the larger size and the CPU-time increase appears to be proportional to $(mk)^{5/2}$.[12]

Summary statistics on the evaluation of the posterior mean of the forecasts of *RGNPG* are in Figure 11. The importance functions give imprecise estimates of the posterior mean. This, as evidenced by the very low *RNE*s, is due to the importance functions being poor approximations of the respective posterior distribution. This is in contrast to the smaller model where the importance functions achieved high *RNE*s. It thus appears as if the size of the problem makes it harder to specify a suitable importance function. The Litterman model contains seven equations with posterior dependence between the parameters of the equations and the STFC importance function fails to capture this dependence. The 2-0 poly-t should reflect the posterior dependence reasonably well, but it suffers from having thinner tails than the posterior. Recall that the degrees of freedom of the matricvariate t-factor must be less than $T - mk$ for the importance function to have thicker tails than the Normal-Diffuse posterior. With the Litterman model and the sample sizes used here, $T - mk \leqslant -159$ and the prior degrees of freedom for the matricvariate t-factor are set to 3. Consequently, the condition (15) fails and it is not guaranteed that the Monte Carlo estimates converge in distribution to a normal. The numerical standard errors reported for this importance function must thus be interpreted with some care.

The Gibbs sampling algorithms, on the other hand, perform very well. They give precise estimates and have *RNE*s close to unity. Unfortunately, the algorithms are also very slow, requiring over ten times as much CPU-time as the importance functions. In terms of the CPU-time required to achieve a certain precision, the slowness is more than offset by the high *RNE* and *GAIN* from antithetic variates. The CPU-times required to achieve a numerical standard error which is 1% of the forecast standard deviation are several orders of magnitude higher for the importance functions than for the Gibbs sampling algorithms and are not reported.

Clearly, Gibbs sampling is the only practical method for evaluating the Normal-Diffuse and ENC posteriors with models of this size. Noting that the forecasting experiment required over 200 hours of CPU-time for each of the Normal-Diffuse and ENC posteriors, the practicability of Gibbs sampling can, of course, be questioned as well. This is, however, a problem that will be solved in the near future as computing power becomes cheaper and more readily available.

6. SUMMARY

In Bayesian analysis of VAR-models, and especially in forecasting applications, the Minnesota prior of Litterman is frequently used. In many cases, other prior distributions give better forecasts and/or are preferable from a theoretical standpoint. With these, more general prior distributions, numerical methods are required in order to evaluate the posterior distributions.

It is straightforward to implement Monte Carlo methods with the Normal-Wishart and Diffuse priors. For these prior distributions we can sample directly from the posterior and Monte Carlo methods are entirely practical even for very large VAR models.

The Normal-Diffuse and ENC priors, on the other hand, require importance or Gibbs sampling. The numerical performance of the importance functions considered is unsatisfactory

[12] When Gibbs sampling from the Normal-Diffuse posterior, the $mk \times mk$ matrix $(\Sigma^{-1} + \Psi^{-1} \otimes \mathbf{Z}'\mathbf{Z})$ is factored and inverted for each draw, these are operations of order $(mk)^3$. The implementation of the Gibbs sampling algorithm for the ENC posterior involves computing a $mk \times mk$ matrix for each draw from the m conditional distributions. This adds up to operations of order $m(mk)^2$.

with the large Litterman model. The Gibbs sampling algorithms, while being slow with models this size, proved to be reliable and to give precise estimates.

For smaller models importance sampling is a viable alternative to Gibbs sampling and is sometimes faster when the precision of the estimate is accounted for. The saving in CPU time from importance sampling compared to using Gibbs sampling is, however, modest. Given the greater robustness to the size of the model of the Gibbs sampler we prefer Gibbs sampling over importance sampling.

The evidence on the forecast performance of the different priors is mixed. We get, as one might expect, different result for different models. For the forecasts of the Swedish unemployment rate OLS and the Diffuse and Normal-Wishart priors perform best, but this result is sensitive to the specification of the prior means. For the Litterman model, the Normal-Wishart, Minnesota and Normal-Diffuse priors do best with a slight edge for the Normal-Diffuse and Minnesota priors. The Diffuse prior and the OLS-based forecasts are outperformed by the other priors, indicating that prior information can be quite helpful with a model this size. The performance of the ENC prior is erratic. It gives the best forecasts for the inflation rate but also the worst forecasts of the unemployment rate for some lead times.

The choice of a distribution to embody a set of prior beliefs is not an easy one. All the priors considered here are easy to specify if the prior beliefs are of the type suggested by Litterman. The Normal-Wishart is less suitable if the prior beliefs contain more specific information and the ENC prior require prior independence between equations to be easy to specify. In addition, the Normal-Wishart and ENC priors require that the user specifies the prior degrees of freedom or, equivalently, the existence of prior moments. All prior moments of the parameters exist for the other informative priors.

In terms of analysing the posterior distribution the Minnesota prior has a slight edge over the Diffuse and Normal-Wishart priors in that expressions for all posterior moments are readily available. For more complicated functions of the parameters, Monte Carlo procedures are easy to implement and efficient for these priors. The Normal-Diffuse and ENC prior require numerical evaluation of the posterior distribution and this can — while feasible — be quite expensive for large models.

Finally, or perhaps first, the credibility of the model specification must be considered. The Minnesota prior is at a disadvantage here since it comes with quite severe restrictions on the likelihood in the form of the fixed and diagonal residual variance–covariance matrix. Taking this together our preferred choice is the Normal-Wishart when the prior beliefs are of the Litterman type. For more general prior beliefs or when the computational effort is of minor importance, the Normal-Diffuse and ENC priors are strong alternatives to the Normal-Wishart.

APPENDIX A: DRAWING FROM A 2-0 POLY t DISTRIBUTION

The algorithm for generating pseudo-random numbers from a 2-0 poly t is only available in working paper form (Bauwens and Richard, 1982) and we give a brief outline of the algorithm here.

Define the d-dimensional 2-0 poly t density as proportional to the product of two d–dimensional multivariate t kernels,

$$\kappa_{2-0}(\mathbf{z}|\mu_1, \mathbf{M}_1, \nu_1, \mu_2, \mathbf{M}_2, \nu_2)$$
$$\propto [1 + (\mathbf{z} - \mu_1)'\mathbf{M}_1^{-1}(\mathbf{z} - \mu_1)/\nu_1]^{-(\nu_1+d)/2} \times [1 + (\mathbf{z} - \mu_2)'\mathbf{M}_2^{-1}(\mathbf{z} - \mu_2)/\nu_2]^{-(\nu_2+d)/2}$$

where M_i/v_i are symmetric, positive semidefinite and $v_1 + v_2 + d > 0$. If, in addition, $\mathbf{M}_i$ is positive definite and $v_i > 2$, then kernel i corresponds to a multivariate t density with mean μ_i and variance $v_i\mathbf{M}_i/(v_i - 2)$.

The algorithm follows from the decomposition (Richard and Tompa, 1980)

$$\kappa_{2-0}(\mathbf{z} \mid \mu_1, \mathbf{M}_1, v_1, \mu_2, \mathbf{M}_2, v_2)$$
$$\propto \int_0^1 [1 + (\mathbf{z} - \mu_c)'\mathbf{M}_c^{-1}(\mathbf{z} - \mu_c)/s_c]^{-(v_1+v_2+2d)/2} g(c|\mu_1, \mathbf{M}_1, v_1, \mu_2, \mathbf{M}_2, v_2)dc$$

where

$$\mathbf{M}_c = [c\mathbf{M}_1^{-1} + (1 - c)M_2^{-1}]^{-1}, \mu_c = \mathbf{M}_c[c\mathbf{M}_1^{-1}\mu_1 + (1 - c)\mathbf{M}_2^{-1}\mu_2],$$
$$s_c = c[v_1 + \mu_1'\mathbf{M}_1^{-1}\mu_1] + (1 - c)[v_2 + \mu_2'\mathbf{M}_2^{-1}\mu_2] - \mu_c'\mathbf{M}_c^{-1}\mu_c$$

and

$$g(c|\cdot) = \pi^{d/2}\Gamma((v_1 + v_2 + d)/2)/[\Gamma((v_1 + d)/2)\Gamma((v_2 + d)/2)]$$
$$\times c^{(v_1+d-2)/2}(1 - c)^{(v_2+d-2)/2}|\mathbf{M}_c|^{-1/2}s_c^{-(v_1+v_2+d)/2}$$

Conditional on c, the first factor is a multivariate t density with $v_1 + v_2 + d$ degrees of freedom, mean μ_c and scale matrix $s_c\mathbf{M}_c/(v_1 + v_2 + d)$, i.e. the variance is $s_c\mathbf{M}_c/(v_1 + v_2 + d - 2)$. The distribution of the scalar c can be tabulated by numerical integration and pseudo-random numbers obtained using the inversion algorithm.

The algorithm is thus (1) generate c from $g(c|\cdot)$, (2) generate z from the conditional multivariate t using standard methods. Antithetic variates are obtained in the second step as $\mathbf{z}^* = 2\mu_c - \mathbf{z}$.

The computations are significantly simplified by letting $\mathbf{T}$ be the matrix which simultaneously diagonalizes $\mathbf{M}_1^{-1}$ and $\mathbf{M}_2^{-1}$, $\mathbf{T}'\mathbf{M}_1^{-1}\mathbf{T} = \mathbf{I}, \mathbf{T}'\mathbf{M}_2^{-1}\mathbf{T} = \Lambda$. Let $\Lambda_c = \mathbf{T}'\mathbf{M}_c^{-1}\mathbf{T} = c\mathbf{I} + (1 - c)\Lambda, \phi_i = \mathbf{T}^{-1}\mu_i, \phi_c = \mathbf{T}^{-1}\mu_c = \Lambda_c^{-1}[c\phi_1 + (1 - c)\phi_2]$ and $s_c = c[v_1 + \phi_1'\phi_1] + (1 - c)[v_2 + \phi_2^1\Lambda\phi_2] - \phi_c'\Lambda_c^{-1}\phi_c$. We can then generate $\mathbf{x}$ conditional on c from a multivariate t distribution with mean ϕ_c, scale matrix $s_c\Lambda_c/(v_1 + v_2 + d)$ and $v_1 + v_2 + d$ degrees of freedom, $\mathbf{z}$ is then obtained as $\mathbf{Tx}$. Note that all quantities except ϕ_c, Λ_c and s_c can be precomputed and that the computation of ϕ_c, Λ_c and s_c only involves vectors and diagonal matrices. The calculation of the density $g(c|\cdot)$ is also simplified if, in addition to using the second expression for s_c, we replace the determinant $|\mathbf{M}_c^{-1}|$ by $|T|^{-2}|c\mathbf{I} + (1 - c)\Lambda|$.

APPENDIX B: DATA SOURCES

The Swedish Unemployment Rate

Monthly data on the Swedish unemployment rate and industrial production index for the period 1964:1 to 1990:12 were collected from the OECD Main Economic Indicator database and the quarterly observations were then obtained as three-month averages. The definition of the unemployment rate changed in the first quarter of 1987. It has been estimated that this change of definition decreased the unemployment rate by one half of a percentage point. In order to make the series consistent over time 0·5 was added to each observation, starting at 87:1.

The Litterman Forecasting Model

RGNPG, real GNP, seasonally adjusted, annualized growth rates: *The National Income and Product Accounts* 1929–1982, *Business Statistics* 1986, *Survey of Current Business* July 1988.
INFLA, implicit GNP deflator, seasonally adjusted, annualized growth rates: the same as *RGNPG*.
UNEMP, unemployed (all civilian workers) as percentage of the civilian labour force: *Supplement to the Survey of Current Business* 1979, *Business Statistics* 1986, *Survey of Current Business* July 1988.
M1, seasonally adjusted, quarterly averages of monthly data, billions of dollars: *Banking and Monetary Statistics* 1941–1970, *Annual Statistical Digest* 1970–1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, *Federal Reserve Bulletin* June 1988.
INVEST, Gross Private Domestic Investments, seasonally adjusted, billions of current dollars at annual rates: the same as *RGNPG*.
CPRATE, commercial paper rate, 4–6 months, percentage per annum, averages of daily rates: the same as *M1*.
CBI, change in business inventories, seasonally adjusted, billions of dollars at annual rates: the same as *RGNPG*.

ACKNOWLEDGMENTS

Earlier versions of this paper (Kadiyala and Karlsson, 1994) have been presented at the conference on Econometric Inference Using Simulation Techniques and at the 12th International Symposium on Forecasting. We are grateful for comments from the anonymous referees, an associate editor and participants in these seminars. The second author also wishes to acknowledge the financial support of the Swedish Research Council for Humanities and Social Sciences (HSFR)

REFERENCES

Bauwens, L. (1984), *Bayesian Full Information Analysis of Simultaneous Equation Models Using Integration by Monte Carlo*, Springer-Verlag, Berlin.
Bauwens, L. and J.-F. Richard (1982), 'A poly-*t* random variable generator with applications to Monte Carlo integration', CORE Discussion Paper No. 8124, CORE, Université Catholique de Louvain, Louvain-la-Neuve.
Bauwens, L. and M. Lubrano (1994), 'Identification restrictions and posterior densities in cointegrated Gaussian VAR systems', CORE Discussion Paper 9418, CORE, Université Catholique de Louvain, Louvain-la-Neuve.
Doan, T., R. Litterman and C. Sims (1984), 'Forecasting and conditional projection using realistic prior distributions', with discussion, *Econometric Reviews*, **3**, 1–144.
Dorfman, J. H. (1995), 'A numerical Bayesian test for cointegration of AR processes', *Journal of Econometrics*, **66**, 289–324.
Drèze, J. H. (1977), 'Bayesian regression analysis using poly-t densities', *Journal of Econometrics*, **6**, 329–54.
Drèze, J. H. and J.-A. Morales (1976), 'Bayesian full information analysis of simultaneous equations', *Journal of the American Statistical Association*, **71**, 919–23. Reprinted in A. Zellner (ed.), *Bayesian Analysis in Econometrics and Statistics*, North-Holland, Amsterdam, 1980.
Drèze, J. H. and J.-F. Richard (1983), 'Bayesian analysis of simultaneous equation systems', in Z. Griliches and M. D. Intrilligator (eds), *Handbook of Econometrics, vol I*, North-Holland, Amsterdam.
Edlund, P.-O. and S. Karlsson (1993), 'Forecasting the Swedish unemployment rate. VAR vs. transfer function modelling', *International Journal of Forecasting*, **9**, 61–76.

Geisser, S. (1965), 'Bayesian estimation in multivariate analysis', *Annals of Mathematical Statistics*, **36**, 150–9.
Gelfand, A. E. and A. F. M. Smith (1990), 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association*, **85**, 398–409.
Geman, S. and D. Geman (1984), 'Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–41.
Geweke, J. (1988), 'Antithetic acceleration of Monte Carlo integration in Bayesian inference', *Journal of Econometrics*, **38**, 73–89.
Geweke, J. (1989), 'Bayesian inference in econometric models using Monte Carlo integration', *Econometrica*, **58**, 1317–39.
Geweke, J. (1993), 'Bayesian treatment of the independent Student-*t* linear model', *Journal of Applied Econometrics*, **8**, S19–40.
Geweke, J. (1995), 'Monte Carlo simulation and numerical integration', in H. Amman, D. Kendrick and J. Rust (eds), *Handbook of Computational Economics*, North-Holland, Amsterdam.
Hajivassiliou, H. V. and P. A. Ruud (1994), 'Classical estimation methods for LDV models using simulation', in R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics, Volume IV*, Elsevier Science BV, Amsterdam.
Kadiyala, K. R. and S. Karlsson (1993), 'Forecasting with generalized Bayesian vector autoregressions', *Journal of Forecasting*, **12**, 365–78.
Kadiyala, K. R. and S. Karlsson (1994), 'Numerical aspects of Bayesian VAR-modelling', Working Paper Series in Economics and Finance, No 12, Stockholm School of Economics.
Kleibergen, F. and H. K. Van Dijk (1994), 'On the shape of the likelihood/posterior in cointegration models', *Econometric Theory*, **10**, 514–51.
Kloek, T. and H. van Dijk (1978), 'Bayesian estimates of equation system parameters: an application of integration by Monte Carlo', *Econometrica*, **46**, 1–19. Reprinted in A. Zellner (ed.), *Bayesian Analysis in Econometrics and Statistics*, North-Holland, Amsterdam, 1980.
Koop, G. (1991), 'Cointegration tests in present value relationships: a Bayesian look at the bivariate properties of stock prices and dividends', *Journal of Econometrics*, **49**, 105–39.
Koop, G. (1992), 'Aggregate shocks and macroeconomic fluctuations: a Bayesian approach', *Journal of Applied Econometrics*, **7**, 395–411.
Litterman, R. B. (1980), 'A Bayesian procedure for forecasting with vector autoregressions', mimeo, Massachusetts Institute of Technology.
Litterman, R. B. (1986), 'Forecasting with Bayesian vector autoregressions — five years of experience', *Journal of Business & Economic Statistics*, **4**, 25–38.
Richard, J.-F. and H. Tompa (1980), 'On the evaluation of poly *t* densities', *Journal of Econometrics*, **12**, 335–51.
Tiao, G. C. and A. Zellner (1964), 'On the Bayesian estimation of multivariate regression', *Journal of the Royal Statistical Society*, **B26**, 389–99.
Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, John Wiley, New York.

