STeady State Priors for Vector Autoregressions

Mattias Villani
Research Division, Sveriges Riksbank, Sweden
Department of Statistics, Stockholm University, Sweden

Abstract. Bayesian priors are often used to restrain the otherwise highly over-parametrized vector autoregressive (VAR) models. The currently available Bayesian VAR methodology does not allow the user to specify prior beliefs about the unconditional mean, or steady state, of the system. This is unfortunate as the steady state is something that economists usually claim to know relatively well. This paper develops easily implemented methods for analyzing both stationary and cointegrated VARs, in reduced or structural form, with an informative prior on the steady state. We document that prior information on the steady state leads to substantial gains in forecasting accuracy on Swedish macro data. A second example illustrates the use of informative steady state priors in a cointegration model of the consumption-wealth relationship in the U.S.

Keywords: Cointegration, Bayesian inference, Forecasting, Unconditional mean, VAR, VECM.

JEL Classification: C11, C32, C53, E50.

1. Introduction

Vector autoregressions (VAR) were launched by Sims (1980) as an alternative to the then dominating large scale structural equations models, which he argued imposed incredible identifying restrictions. The flexibility of VARs comes at the cost of greater parameter uncertainty and, as a consequence, erratic model predictions. This was noted by Sims already in his original 1980 paper where he suggested using prior information to increase the precision. Sims’ suggestion spurred the development of prior distributions for VARs, see e.g. the well known Litterman/Minnesota prior documented in Litterman (1986), and subsequent modifications and extensions in Doan, Litterman and Sims (1984), Kadiyala and Karlsson (1997), Robert-son and Tallman (1999) for reduced form VARs, Sims and Zha (1998) and Waggoner and Zha

Correspondence to: Mattias Villani, Research department, Sveriges Riksbank, SE-103 37 Stockholm, Sweden.
Phone: +46 8 7870821. Fax: +46 8 210531. E-mail: mattias.villani@riksbank.se.
(2003b) for structural VARs, Kleibergen and van Dijk (1994), Kleibergen and Paap (2002), Strachan (2003), Strachan and Inder (2004), and Villani (2005a) for cointegrated reduced form VARs, and Villani and Warne (2003) for cointegrated structural VARs.

All available priors for VARs focus on the dynamic coefficients but are largely noninformative about the deterministic component of the model. This is important as long horizon forecasts from stationary VARs converge to the unconditional mean, or steady state, of the process. Similarly, the long run forecasts of growth rates from cointegrated VARs converge to the unconditional mean of the growth rates. Indeed, Clements and Hendry (1998) show that a badly estimated mean of the process is the dominant source of forecast failure at longer forecast horizons. Prior information on the steady state is typically available, and quite often in strong form. The forecasts of inflation undertaken at central banks operating under an explicit inflation target is an apparent example. Decision makers will have a hard time accepting that their prior information may easily be incorporated on the more obscure parts of the model, such as the reduced form dynamic coefficients, but that their strong prior beliefs about the steady state cannot be used for ‘technical reasons’. The purpose of this paper is to remove this straight-jacket from the analyst.

One of the reasons for the lack of prior information on the steady state is probably that applications of Bayesian VAR (BVAR) models for U.S. data have often modelled variables in levels, where the process’ steady state is often non-existing or at least not very relevant (near unit root process). While VARs in levels seem to work well for U.S. data (see e.g. the many papers by Sims and his coauthors), it has not been as successful for other countries, where instead differencing or cointegration have been more widely applied. Theoretical arguments in favor of differencing are given in Clements and Hendry (1995). Another explanation of the rather casual handling of the steady state is perhaps that it is expected to be fairly precisely estimated even in the absence of prior information. This is not always the case, however, as illustrated in Section 4.

BVARs are routinely used in forecasting exercises as benchmark models to more theoretically founded models with a well defined steady state. A recent example is Del Negro, Schorfheide, Smets and Wouters (2007), where the out-of-sample forecasting performance of the DSGE model in Smets and Wouters (2007) is compared to several BVARs. An informative
prior on the steady state of the DSGE model is incorporated via the prior on the structural parameters. The comparison of forecasting properties would be more balanced if the same prior beliefs had also been used on the steady state of the BVARs.

In this paper we develop a Bayesian analysis of the VAR model in so called mean-adjusted form. This parametrization of the VAR is convenient for prior elicitation as the unconditional mean of the process is explicitly modelled. The mean-adjusted form was introduced in Bayesian analysis of the univariate AR process by Schotman and van Dijk (1991a,b, 1993) in an influential contribution to the lively unit root debate in the early 90’s. These papers derive properties of a Bayesian analysis in mean-adjusted form in the case of non-informative priors. It is shown that the marginal posterior distribution of the dynamic coefficients has a non-integrable asymptote when the process has a unit root. See Bauwens et al. (1999) for a survey of the literature in this area. We argue that the reason for using the mean-adjusted form of the VAR in the first place is that prior information actually is available on the steady state and that the use of a proper informative prior alleviates the mentioned difficulties. We will refer to the mean-adjusted form of the VAR as the steady state VAR.

We first consider the stationary, or difference stationary, VAR and subsequently move on to the cointegrated VAR (VECM). In the case of the VECM the mean-adjusted form of Clements and Hendry (1999) is used, where the unconditional mean growth rate of the process and mean of the long run relations are explicitly modelled. Note that cointegration restrictions pin down the long run behavior of variables relative to other variables in the system, but give no control over the long run behavior of the time series in absolute terms. Cointegration restrictions may, for example, be used to force two variables to have the same long run growth, but the common growth rate of the two series is free to take on any value. The methodology presented here give the user the possibility to incorporate prior beliefs about this shared growth rate.

All results in this paper are presented for reduced form models. Appendix C describes the extension of the algorithms to the structural/identified VAR/VECM.

The paper is organized as follows. The next section develops Bayesian inference for the stationary steady state VAR. Section 3 extends the analysis to the cointegrated case. The
fourth section illustrates the analysis on a seven-variable model for the Swedish macro economy, and on a VECM model of the consumption-wealth relationship in the U.S. The final section concludes. The algorithms are described in details in the appendices.

2. Bayesian analysis of the steady state VAR

2.1. The model. The usual parametrization of the VAR model is

\[ \Pi(L)x_t = \Phi d_t + \varepsilon_t, \]

where \( x_t \) is a \( p \)-dimensional vector of time series at time \( t \), \( d_t \) is a \( q \)-dimensional vector of deterministic trends or other exogenous variables. \( \Pi(L) = I_p - \Pi_1 L - \ldots - \Pi_k L^k \), \( L \) is the usual back-shift operator with the property \( Lx_t = x_{t-1} \), and \( \varepsilon_t \sim N_p(0, \Sigma) \) with independence between time periods. We shall initially assume that \( x_t \) is a stationary process (either in its original form or after suitable differencing) and later on treat the extension to the cointegrated case. The model in (2.1) will be referred to as a VAR model on standard form.

A Bayesian analysis requires a joint prior distribution of all model parameters \( \Pi_1, \ldots, \Pi_k, \Phi \) and \( \Sigma \). This daunting task is often simplified by modelling the prior in terms of a small number of hyperparameters which together fully specify the prior. It is often claimed that it is hard to specify prior opinions on \( \Phi \) (see e.g. Litterman, 1986) and this part of the prior is usually taken to be non-informative. This does not mean, however, that prior information on the deterministic component of the model is unavailable, simply that the particular parametrization of the model in (2.1) forces the user to specify her beliefs in an awkward way. In stationary VARs the long horizon forecasts approach the unconditional mean of the process, \( \mu_t = \Pi^{-1}(1)\Phi d_t \). This property makes it clear that the implied prior on \( \mu_t \) is an important aspect of a Bayesian analysis of VARs. Such information is often available and may be very important for the forecasting performance of the VAR, or is at the minimum useful for pedagogical reasons when the forecaster communicates his results. \( \mu_t \) is unfortunately a complicated nonlinear function of \( \Pi_1, \ldots, \Pi_k \) and \( \Phi \). Thus, convenient as parametrization (2.1) may be from a computational viewpoint, it is not the preferred parametrization for incorporating prior opinions regarding \( \mu_t \).
An alternative parametrization of the VAR is

\[ \Pi(L)(x_t - \Psi d_t) = \varepsilon_t. \]

This VAR model is non-linear in its parameters, but the unconditional mean of the process is directly specified by \( \Psi \) as \( \mu_t = \Psi d_t \). The form of the deterministic component \( \Psi d_t \) is flexible, any deterministic function may be used by a suitable definition of \( d_t \), e.g. a constant, a piecewise constant or a linear time trend. The model in (2.2) will be referred to as the \textit{steady state VAR}.

2.2. \textbf{Prior distribution.} Bayesian inference requires a prior distribution on \( \Sigma, \Pi_1, \ldots, \Pi_k, \) and \( \Psi \). The prior on \( \Sigma \) is taken to be

\[ p(\Sigma) \propto |\Sigma|^{-(p+1)/2}. \]

Let \( \Pi = (\Pi_1, \ldots, \Pi_k)' \). The prior for \( \text{vec} \, \Pi \) used here is a general multivariate normal distribution

\[ \text{vec} \, \Pi \sim N(\theta_\Pi, \Omega_\Pi). \]

which includes the well-known Minnesota prior (Litterman, 1986) and variants as special cases. We further assume prior independence between \( \Pi \) and \( \Psi \), and that

\[ \text{vec} \, \Psi \sim N_{pq}(\theta_\Psi, \Omega_\Psi). \]

2.3. \textbf{Posterior distribution.} The posterior distribution of the steady state VAR in (2.2) is intractable. It is shown in Appendix A that the posterior distribution of each set of model parameters (i.e. one of \( \Sigma, \Pi \) and \( \Psi \)) conditional on the other parameters is tractable, and that a simple three-block Gibbs sampler (Smith and Roberts, 1993) can therefore be used draw from the joint posterior. The full conditional posteriors of \( \Sigma \), \( \text{vec} \, \Pi \) and \( \text{vec} \, \Psi \) are inverted Wishart, normal and normal densities, respectively. The extra computing time for the steady state VAR compared to the standard VAR in (2.1) (which is typically analyzed by a two-block Gibbs sampler, one block with \( \Sigma \) and the other with \( \Pi \) and \( \Phi \)) is small, or even negative, see the previous working paper version of this article (Villani, 2005b) for a discussion. Convergence to the target distribution is in general very fast unless the prior on the steady state is very
weakly informative, in which case the steady state VAR should not have been used in the first place, see Appendix A and Villani (2005b).

Zha (1999) gives an in-depth discussion of Bayesian inference under linear restrictions on \( \text{vec}\Pi \). The Gibbs sampler in Appendix A for the steady state VAR can be slightly modified to the case with general linear restrictions on \( \text{vec}\Pi \). To see this, formulate the restrictions as \( \text{vec}\Pi = H\varrho + h \), where \( H \) is a known restriction matrix, \( h \) a known vector and \( \varrho \) the unrestricted parameters in \( \Pi \) (see Johansen (1995) for more details on this formulation). Repeating the derivations in Appendix A with \( \text{vec}\Pi \) replaced by \( H\varrho + h \) (the Jacobian is constant so it can be ignored) shows that the full conditional posterior of \( \varrho \) is a multivariate normal density. The same clearly applies also to linear restrictions on \( \text{vec}\Psi \).

We also note that the generated Gibbs sample may be used to compute the marginal likelihood of a model using \( e.g. \) the methods developed in Chib (1995) and Geweke (1999). Chib’s method is quite efficient in this setting since the additional so called reduced Gibbs sampler (see Chib’s paper for details) can be made to operate solely in \((\Sigma,\Psi)\)-space. The time-consuming \( \Pi \)-step is thus excluded in the reduced Gibbs sampler.

3. Bayesian analysis of the steady state VECM

3.1. The model. So far we have assumed the process to be (trend) stationary. It is of course possible to transform non-stationary \( I(1) \) or \( I(2) \) variables to stationarity by differencing in the usual way and use the steady state VAR in (2.2) directly on these differenced variables. An intermediate case between stationary and difference stationary processes are the cointegrated processes. The steady state VECM is of the form (Clements and Hendry, 1999)

\[
\Gamma(L)(\Delta x_t - \gamma) = \alpha(\beta'x_{t-1} - \mu_0 - \mu_1t) + \varepsilon_t,
\]

where \( \beta \) is a \( p \times r \) matrix with the \( r \) cointegration vectors as columns, \( \alpha \) is a \( p \times r \) matrix of adjustment coefficients determining the speed of adjustment back to equilibrium after a disturbance, \( \Gamma(L) = I_p - \Gamma_1L - \ldots - \Gamma_{k-1}L^{k-1} \), and \( \varepsilon_t \sim N_p(0,\Sigma) \) with independence between time periods. The steady state VECM model is nonlinear in its parameters, but models the
mean of the growth rates explicitly as \( E(\Delta x_t) = \gamma \) and the long-run solution as \( E(\beta'x_{t-1}) = \mu_0 + \mu_1 t \).

The parameters in (3.1) are related by \( \beta' \gamma = \mu_1 \) (Clements and Hendry, 1999, p. 152-153).

Most applications of the VECM model precludes drift in the cointegrating relations, i.e. sets \( \mu_1 \) to the zero vector. This imposes the nonlinear restriction \( \beta' \gamma = 0 \) on the VECM parameters.

When \( \mu_1 = 0 \), we parametrize \( \gamma \) as \( \gamma = \beta_\perp \lambda \), where \( \beta_\perp \) is the \( p \times (p-r) \) matrix orthogonal to \( \beta \) (\( \beta_\perp \beta_\perp = I_p - \beta \)), and \( \lambda \) the \((p-r)\)-dimensional vector of unrestricted parameters.

It is straightforward to extend the model in (3.1) by modelling \( E(\Delta x_t) \) and \( E(\beta'x_t) \) more flexibly (dummies, splines etc). We then need to make sure that the model for the mean growth rate \( E(\Delta x_t) \) and the model for the long run solution \( E(\beta'x_t) \) are related through the equations \( \beta' \Phi_1 d_t = \mu_1 \). If \( \mu_1 = 0 \), then we need to impose the non-linear restriction \( \beta' \Phi_1 = 0 \), i.e. letting \( \Phi_1 = \beta_\perp \lambda \), where \( \lambda \) is now possibly a matrix. It will be clear from Section 3.3 that our posterior sampling algorithm can handle such extensions.

3.2. Prior distribution. We will use the same prior for \( \Sigma \) as in the stationary case. The prior on \( \Gamma = (\alpha, \Gamma_1, ..., \Gamma_{k-1})' \) is of the form

\[
(3.2) \quad \text{vec} \Gamma | \beta \sim N(\text{vec} \theta_\Gamma, \Sigma \otimes \Omega_\Gamma),
\]

where

\[
\Omega_\Gamma = \begin{pmatrix}
\lambda_\alpha (\beta'_S^{-1} \beta)^{-1} & 0 \\
0 & \lambda_\Gamma S_\Gamma
\end{pmatrix},
\]

where \( \lambda_\alpha, \lambda_\Gamma > 0 \) are a shrinkage factors and \( S_\alpha \) are \( S_\Gamma \) are positive definite matrices, usually assumed to be diagonal; see Villani (2005a) for a motivation of the prior on \( \alpha \). The diagonal elements of \( S_\Gamma \) are set as follows. Let \( s_{ij}^{(k)} \) denote the prior standard deviation of the coefficient on \( \Delta x_{j,t-k} \) in the \( i \)th equation, where \( x_{j,t-k} \) is the \( k \)th lag of variable \( j \). Similar to Litterman (1986) we set \( s_{ij}^{(k)} = \lambda_1 / (\hat{\sigma}_j k^{\lambda_3}) \), where \( \lambda_1 \) and \( \lambda_3 \) are hyperparameters and \( \hat{\sigma}_j \) is an estimate of the error standard deviation in a univariate autoregression for variable \( j \) in the system. The \( \hat{\sigma}_j \) adjust the prior to the possibly differing scales of the variables. As in Villani and Warne (2003), we will use the same scale factors for the diagonal of \( S_\alpha \), i.e. we set \( S_\alpha = diag(\hat{\sigma}_1^{-2}, ..., \hat{\sigma}_p^{-2}) \).
It should be noted that the prior covariance matrix of $\Gamma$ is restricted to have a Kronecker structure, which simplifies the posterior sampling algorithm. While a Kronecker structure is probably sufficiently flexible in most applications, it does exclude the possibility to impose the extra cross-equation shrinkage in Litterman (1986), or general linear restrictions on the parameters in $\Gamma$. It is possible to derive (slightly more complicated) algorithms that do not assume a Kronecker structure, but that will not be pursued here.

The prior on $\beta$, $\gamma$ and $\mu_1$ needs to respect that these parameters are functionally dependent through the relation $\beta'\gamma = \mu_1$. We shall here specify priors for the cointegration vectors $\beta$ and the growth rates $\gamma$, and then let these priors imply a prior on $\mu_1$, the drift in the cointegrating relations. Alternatively, we could have specified a prior on subsets of all three of $\beta$, $\gamma$ and $\mu_1$, but this would be more complicated and probably less useful in most applications. We also note that the prior that we propose below for $\mu_0$ is intimately connected to the prior beliefs on $\beta$. Finally, specifying priors directly for $\beta$ and $\gamma$, and therefore only implicitly for $\mu_1$, has the advantage that it makes the prior elicitation similar to the prior we propose for the case with $\mu_1 = 0$, see below.

The prior on $\beta$ is assumed to be of the form

$$\varphi_i \sim N(\mu_{\varphi_i}, \Omega_{\varphi_i}),$$

where $\varphi_i$ are the unrestricted coefficients in the $i$th column of $\beta = (h_1 + H_1\varphi_1, ..., h_r + H_r\varphi_r)$; $h_i$ and $H_i$ are known, see Johansen (1995) for more details on this restriction formulation. The $\varphi_i$ are assumed to be independent apriori.

The interpretation of $\mu_0$ depends crucially on $\beta$, and this must be taken into account when formulating a prior. The prior on $\mu_0$ is of the form

$$(3.3) \quad \mu_0 | \beta \sim N(\theta_{\mu_0}, Q\Omega_{\mu_0}Q'),$$

where $\theta_{\mu_0}$ is the prior mean of $\mu_0$, $\Omega_{\mu_0}$ is a psd matrix (e.g. $\Omega_{\mu_0} = \kappa I_r$). $Q$ is a diagonal matrix with $i$th diagonal element $q_i = 1 + \sqrt{(\varphi_i - \mu_{\varphi_i})^T\Omega_{\varphi_i}^{-1}(\varphi_i - \mu_{\varphi_i})}$. Note that this prior has the attractive property of making the $i$th element of $\mu_0$ more diffuse when the $i$th cointegration vector is far from its prior mean.
The prior on the growth rates $\gamma$ is $\mathcal{N}(\theta_\gamma, \Omega_\gamma)$. As discussed above, the prior for $\mu_1$ is then implicitly given from the relation $\beta' \gamma = \mu_1$ and the priors on $\beta$ and $\gamma$. When $\mu_1 = 0$, $\gamma = \beta_\perp \lambda$ and the prior for $\gamma$ is implicitly given by the prior on $\lambda$. It is in general easier to specify a prior on the growth rates $\gamma$ than it is to specify a prior for $\lambda$. We therefore choose to project the $\mathcal{N}(\theta_\gamma, \Omega_\gamma)$ prior for $\gamma$ down to the subspace $\beta' \gamma = 0$. This way we obtain an implicit prior on $\lambda$ as

$$\lambda|\beta \sim \mathcal{N}(\beta_\perp' \theta_\gamma, \beta_\perp' \Omega_\gamma \beta_\perp'),$$

where $\theta_\gamma$ and $\Omega_\gamma$ are the elicited prior mean and covariance matrix of $\gamma$. Thus, any prior specified on $\gamma$ is automatically adjusted to satisfy the restriction $\beta' \gamma = 0$, in a way that stays as close as possible to the original prior. When $\mu_1 \neq 0$, let $\eta = (\mu_0, \gamma)'$, and when $\mu_1 = 0$, let $\eta = (\mu_0, \lambda)'$. We write the joint prior for $\eta$ as

$$\eta|\beta \sim \mathcal{N}(\theta_\eta, \Omega_\eta).$$

3.3. Posterior distribution. It is in principle possible to set up a Gibbs sampler also in the cointegrated case, but there at least two complications that suggest that this is not the best route to take. First, the full conditional posterior of $\beta$ is non-standard since the prior covariance of $\mu_0$ depends on $\beta$. Moreover, when $\mu_1 = 0$, $\beta$ also enters the model though $\gamma = \beta_\perp \lambda$, which further complicates the situation. Second, we have argued that using informative priors on the steady state VAR solves the Gibbs sampler’s problem of coping with local non-identification. In the case of the VECM, the interpretation of $\mu_0$ is directly dependent on $\beta$ (see the prior in (3.3)), so that being informative on $\mu_0$ requires also an informative prior on $\beta$. This makes it harder to specify priors which are sufficiently informative to stabilize the Gibbs sampler.

We will instead sample from the posterior using the following decomposition

$$p(\varphi, \eta, \Gamma, \Sigma|\mathcal{D}) = p(\varphi, \eta|\mathcal{D})p(\Gamma, \Sigma|\varphi, \eta, \mathcal{D}).$$

Appendix B gives the details of the posterior sampling scheme. Liu, Wong and Kong (1994) prove in a general setting that sampling from decompositions like the one in (3.5) is more
efficient than Gibbs sampling. Note also that the parameters in $\varphi$ and $\eta$ becomes non-identified when $\alpha$ is of reduced rank (Kleibergen and van Dijk, 1994), so that sampling $\varphi$ and $\eta$ directly from their marginal posterior is likely to be cope better with local non-identification than a pure Gibbs sampler (where $\varphi$ and $\eta$ are drawn conditional on $\alpha$).

The marginal likelihood may be computed from the (low-dimensional) posterior draws of $\varphi$ and $\eta$ using the modified harmonic mean estimator in Geweke (1999).

4. Empirical illustrations

4.1. Stationary VAR. We use quarterly data for Sweden over the time period 1980Q1 – 2005Q4 to illustrate the method developed here. The VAR system contains the following seven variables: trade-weighted measures of foreign GDP growth ($\Delta y_f$), CPI inflation ($\pi_f$) and the three-month interest rate ($i_f$), the corresponding domestic variables ($\Delta y, \pi$ and $i$), and the level of the real exchange rate defined as $q = s + p_f - p$, where $p_f$ and $p$ are the foreign and domestic CPI levels (in logs) and $s$ is the (log of the) trade-weighted nominal exchange rate. To incorporate that Sweden is a small economy and therefore not likely to affect the foreign economy, we restrict the upper right submatrix in each $\Pi_i, i = 1, ..., k$, to the zero matrix. We will assume that $x_t = (\Delta y_f, \pi_f, i_f, \Delta y, \pi, i, q)'$ can be modelled as a stationary VAR process.

To model Sweden’s change in monetary policy in the 90’s (move to inflation targeting and flexible exchange rate) we will use the dummy $d_t = (1, d_{MP,t})'$, where

$$d_{MP,t} = \begin{cases} 
1 & \text{if } t \leq 1992Q4 \\
0 & \text{if } t > 1992Q4 
\end{cases} .$$

To formulate a prior on $\Psi$, note that the specification of $d_t$ implies the following parameterization of the steady state

$$\mu_t = \begin{cases} 
\psi_1 + \psi_2 & \text{if } t \leq 1992Q4 \\
\psi_1 & \text{if } t > 1992Q4 
\end{cases} ,$$

where $\psi_i$ is the $i$th column of $\Psi$. The prior on $\psi_1$, which determines the steady state in the latter regime, builds on the implied prior on the steady state in the DSGE model for Sweden in Adolfson, Laseén, Lindé and Villani (in press), and is displayed as the first row of Table I. The prior on $\psi_2$, which determines the difference in steady states between the first and second
regime, is centered on the event that the regime shift is purely nominal, \textit{i.e.} that $\Delta y_f$ and $\Delta y$ have the same steady state throughout the whole sample period. The spread around zero is fairly large, however, making it essentially up to the data to determine if the shift is purely nominal or not.

The prior proposed by Litterman (1986) will be used on the dynamic coefficients in $\Pi$, with the default values on the hyperparameters in the priors advocated by Doan (1992): overall tightness is set to 0.2, cross-equation tightness to 0.5 and a harmonic lag decay with a hyperparameter equal to one. See Litterman (1986) and Doan (1992) for details. Litterman’s prior was designed for data in levels and has the effect of centering the process on the univariate random walk model. We therefore set the prior mean on the first own lag to zero for all the variables in growth rates. The three remaining level variables $r_f$, $r$ and $q$ are assigned a prior which centers on the $AR(1)$ process with a dynamic coefficient equal to 0.9. The usual random walk prior is not used here as it is inconsistent with having a prior on the steady state. Finally, the usual noninformative prior $|\Sigma|^{-(p+1)/2}$ is used for $\Sigma$.

The Gibbs sampler in Appendix A was used to generate a sample of 20,000 draws from the joint posterior distribution of the model parameters. For each parameter draw, a dynamic forecast was generated 20 quarters ahead to form a sample from the predictive distribution. The convergence of the Gibbs sampler was excellent.

Figures 1-3 display the actual outcomes of the three domestic variables along with sequential of forecasts from three different models: a VAR with parameters estimated by maximum likelihood (MLVAR), a standard BVAR with the Litterman prior, and the steady state BVAR. A general observation from Figures 1-3 is that the two BVARs generate smoother forecast paths than the more erratic MLVAR paths. This is due to the added prior information. Figure 4 displays the root mean squared forecast errors (RMSE) for the three models and the no change forecast. The point estimates of the steady state level in the three models are given in Figure 5.

Looking first at the forecasts of GDP growth in Figure 1, it is clear that the standard BVAR overestimates the steady state level of GDP growth and therefore consistently overshoots the actual GDP growth during first part of the evaluation sample (see also Figure 5). It is only towards the end of the sample that the standard BVAR lowers its estimated steady state level
and begins to perform well. The BVAR with the steady state prior seems to have a more realistic steady state level already in the beginning of the evaluation period, and therefore much more accurate forecasts. The gain in forecasting accuracy from using prior information on the steady state is substantial (Figure 4). The black solid line with dots (measured on the right vertical axis) in Figure 5 depicts the posterior uncertainty of the steady state level in the Litterman BVAR (length of a 95% equal-tail probability interval). The posterior uncertainty of the steady state level for GDP growth is large, much larger than what one would expect, especially early in the evaluation period. The prior on the steady state (see Table I) is very informative in comparison with the rather uninformative data, and the posterior estimate of the steady state level is therefore essentially equal to the prior mean over the whole evaluation period.

The same analysis holds also for domestic inflation forecasts in Figure 2. Here the BVAR-Litterman estimate of the steady state level is consistently lower than the Central bank’s explicit inflation target of two percent (see Figure 5), and the forecasts undershoot accordingly (Figure 2). The posterior uncertainty regarding the steady state level of inflation is also very large (Figure 5, right axis), and again the posterior and prior mean of the steady state are therefore very close throughout the evaluation period. The improvement in RMSE for domestic inflation from using a steady state prior are dramatic (Figure 4).

Figure 5 shows that the estimated steady state level of the domestic interest rate is high for BVAR-Litterman model early in the sample, and, that the forecasts overshoot the actual interest rate during the same time period (Figure 3). The steady state estimates from the Litterman prior and the steady state prior converges later in the evaluation sample (Figure 5). Interestingly, the last forecast paths of the interest rate illustrates that a steady state prior may have important effects on the dynamics of the forecasts: the interest rate forecast from the steady state BVAR initial decreases before moving toward the higher steady state, whereas the Litterman prior produces forecasts which immediately increase toward the steady state. Again, the RMSE path in Figure 4 strongly suggests improved forecasting performance from the steady state prior.

The steady states of the foreign variables are more precisely estimated (not shown), and the difference between the two BVAR priors is therefore not as pronounced as for the domestic
variables. The RMSEs in Figure 4 suggests that the steady state VAR is somewhat more accurate than the Litterman BVAR in forecasting foreign GDP growth and inflation, but has slightly inferior forecasting performance for the foreign interest rate.

4.2. Cointegrated VAR. Lettau and Ludvigson (2001, 2004) analyze the U.S. consumption-wealth relationship in the period 1951Q4-2003Q1. They show that a wide range of optimization models imply that the log consumption-wealth ratio is stationary. They further argue that this implies that log consumption \( c_t \), log household net worth \( a_t \) and log labor income \( y_t \) cointegrate. We will here illustrate the prior elicitation and posterior inference for the steady state VECM on an updated data set for the time period 1952Q1-2006Q4\(^1\). We follow Lettau and Ludvigson and present results from the model with \( k = 2 \) and \( r = 1 \) (see Koop, Potter and Strachan, 2005 for a careful investigation of the uncertainty in the model specification for the Lettau-Ludvigson model). Let \( x_t = (c_t, a_t, y_t)' \) and express the cointegration vector as \( \beta = (1, -\varphi_1, -\varphi_2) \). We analyze the model where the log consumption-wealth ratio has a mean, but no trend, i.e. \( \mu_1 = \beta'\gamma = 0 \).

We set \( \lambda_\alpha = \lambda_\Gamma = 10 \), so that the prior on \( \alpha \) and \( \Gamma \) is noninformative. An informative prior for \( \beta \) can be elicited from the observation that \( \varphi_1 \) is in principle equal to the average share of asset holdings in total wealth, and \( \varphi_2 = 1 - \varphi_1 \) the average share of labor income in total wealth. Due to measurement problems this is not quite the case, and Lettau and Ludvigson (2001) estimate that \( \varphi_1 + \varphi_2 \approx 0.9 \). If aggregate production can be described by a Cobb-Douglas technology, then these estimates imply that \( \varphi_1 = \zeta/1.1 \) and \( \varphi_2 = (1 - \zeta)/1.1 \), where \( \zeta \) is the usual share of total payment to capital (Lettau and Ludvigson, 2001). This is important as we may then assign a prior mean for \( \varphi_1 \) and \( \varphi_2 \) based on the estimate \( \hat{\zeta} = 0.353 \) obtained by Solow (1957) on data between 1909-1949, i.e. before our study period. A normal prior that centers \( \varphi_1 \) on 0.353/1.1 with a standard deviation of 0.05 seems to cover all values used to calibrate Real Business Cycle (RBC) models. Similarly, \( \varphi_2 \) is normal a priori with mean \((1 - 0.353)/1.1 \) and standard deviation 0.05. Since we expect to nearly have \( \varphi_2 = 0.9 - \varphi_1 \), we also use a prior correlation coefficient between \( \varphi_1 \) and \( \varphi_2 \) of \(-0.5 \).

Most economic models assume that \( c_t \), \( a_t \) and \( y_t \) grow at the same rate as real GDP per capita \((c_t, a_t \ and \ y_t \ are \ all \ expressed \ in \ real \ per \ capita \ terms)\). The mean growth rate of

\(^1\)The updated data can be downloaded Martin Lettau’s webpage http://pages.stern.nyu.edu/~mlettau/.
real GDP per capita during 1869-1951 (the period before our sample) was 1.98 (Johnston and Williamson, 2005), so it makes sense to center the prior on the annual steady state growth rates of \(c_t, a_t\) and \(y_t\) at 2\%. This is also the value typically used to calibrate RBC/DSGE models. GDP growth was relatively volatile before 1952 so one may expect that prior information about the steady state growth rates in the beginning of 1952 was not as sharp as it is today. We therefore set the standard deviation of the annual growth rates to 0.5, giving a 95\% prior interval approximately ranging between 1 and 3. These growth rates are expected to be positively correlated a priori so we set the prior correlation coefficient to 0.5.

Since \(x_t'\beta = c_t - \varphi_1 a_t - \varphi_2 y_t\) is only (approximately) proportional to the consumption-wealth ratio (Lettau and Ludvigson, 2001), it is hard to elicit an informative prior on \(\mu_0\). We therefore use a non-informative \(N(0,5^2)\) prior for \(\mu_0\). There are other applications where a prior on \(\mu_0\) may be essential. An example is a simple three-variate model for a price index in two countries \((p_t, p_t^*)\) and the nominal exchange rate \((e_t)\) between the two countries. A natural cointegrating relation is then the real exchange rate \(\ln p_t - \ln p_t^* - e_t\), which should be stationary according to the purchasing power parity. \(\mu_0\) would then be the steady state of the real exchange rate, something that economists typically have fairly strong beliefs about.

We generated 50,000 draws from the posterior of the Lettau-Ludvigson model with the independence sampler described in Appendix B \((\kappa = 1, v = 5)\). The posterior sampling, including the numerical maximization, took less than a minute using Matlab 7 on a 2Ghz Pentium M processor. Summaries of the posterior distribution are displayed in Table II (left). The posterior results agree with Lettau and Ludvigson’s estimates. The posterior precision is as expected larger than the prior precision, but the location of the prior and posterior are rather similar. Table II (right) also presents the posterior distribution from a less informative prior with five times larger variances and zero prior correlations. The data appear to be very informative: weakening the information in the prior does not have a large impact on the posterior results. The main difference between the informative and the noninformative prior is in the efficiency of the posterior sampling algorithm. The columns labelled IF in Table II displays the inefficiency factors, i.e. the number of draws needed to obtain the equivalent of a single iid draw. For the informative prior the posterior draws are close to iid, whereas for the weaker prior the sampling is less efficient. The IF’s for \(\Gamma\) are essentially 1 for both priors.
Note that the inefficiency in the sampling of $\varphi$ and $\eta$ under the weaker prior does not spill over to $\alpha$ and $\Gamma$. This is important as $\alpha$ and $\Gamma$ typically contains the majority of the parameters. Finally, we also experimented with the Metropolis version of the sampler. As expected this sampler is less efficient: the IF’s for $\varphi$ and $\eta$ under the informative prior are 3-5 times larger than the IF’s from the independence sampler.

5. Concluding remarks

We have proposed practical algorithms for analyzing both stationary and cointegrated VARs with informative prior beliefs on the steady state of the process. We have argued that this kind of prior information can be very important, especially for long horizon forecasts, and that prior beliefs regarding the steady state are often available in relatively strong form. The methodology was illustrated on a seven-variable model for the Swedish economy, where the improvement in forecasting accuracy from using a steady state prior was substantial, and on a cointegration model of the consumption-wealth relationship in the U.S.

Acknowledgements

Comments from Malin Adolfson, Michael Andersson, Paolo Giordani, Herman van Dijk, Anders Warne, and especially two anonymous referees are gratefully acknowledged. This work was partially supported by a grant from the Swedish Research Council (Vetenskapsrådet, grant no. 412-2002-1007). The views expressed in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank.

Appendix A. Posterior sampling for the steady state VAR

This appendix gives the details on the posterior sampling algorithm for the stationary steady state VAR. We will make use of the following notation. Let $z_{1t}, z_{2t}, \ldots, z_{mt}, t = 1, \ldots, T$, be $m$ column vectors of possibly differing lengths, $l_i, i = 1, \ldots, m$. We shall write $Z = [z_{1t}, z_{2t}, \ldots, z_{mt}]^T_{t=1}$ to denote that the $t$th row of the $T \times (\sum_{i=1}^{m} l_i)$ matrix $Z$ equals $(z'_{1t}, z'_{2t}, \ldots, z'_{mt})$. The symbol $[ ]^T_{t=1}$ is thus simply a symbol for the usual rearrangement of
data vectors into a matrix for the whole sample. Since $T$ will be fixed throughout, we shall merely write $Z = [z_{1t}, z_{2t}, ..., z_{mt}]$. Furthermore, let $D = \{x_1, ..., x_t, d_1, ..., d_t\}$ denote the data.

**Proposition A.1.**
- **Full conditional posterior of** $\Sigma$

$$\Sigma|\Pi, \Psi, D \sim IW(E'E, T),$$

where $E = [\Pi(L)(x_t - \Psi d_t)]$.
- **Full conditional posterior of** $\Pi$

$$\text{vec } \Pi|\Sigma, \Psi, D \sim N(\bar{\theta}_\Pi, \bar{\Omega}_\Pi),$$

where $\bar{\Omega}_\Pi^{-1} = \Sigma^{-1} \otimes X_\Psi X_\Psi + \Omega_\Pi^{-1}$, $\bar{\theta}_\Pi = \bar{\Omega}_\Pi [\text{vec}(X_\Psi Y_\Psi \Sigma^{-1}) + \Omega_\Pi^{-1} \theta_\Pi]$, $Y_\Psi = [x_t - \Psi d_t]$ and $X_\Psi = [x_{t-1} - \Psi d_{t-1}, ..., x_{t-k} - \Psi d_{t-k}]$.
- **Full conditional posterior of** $\Psi$

$$\text{vec } \Psi|\Sigma, \Pi, D \sim N(\bar{\theta}_\Psi, \bar{\Omega}_\Psi),$$

where $\bar{\Omega}_\Psi^{-1} = U'(D'D \otimes \Sigma^{-1}) U + \Omega_\Psi^{-1}$, $\bar{\theta}_\Psi = \bar{\Omega}_\Psi [U' \text{vec}(\Sigma^{-1} Y' D) + \Omega_\Psi^{-1} \theta_\Psi]$, $Y = [\Pi(L)x_t]$, $U' = (I_{pq}, I_q \otimes \Pi_1', ..., I_q \otimes \Pi_k')$ and $D = [d_t, -d_{t-1}, ..., -d_{t-k+1}]$.

**Proof.** Conditional on $\Psi$ the model in (2.2) is a standard VAR model for the time series $x_t - \Psi d_t$. The full conditional posteriors of $\Sigma$ and $\Pi$ therefore follow from standard results, see e.g. Zellner (1971). To derive the full conditional posterior of $\Psi$ we rewrite the model in (2.2) as

$$\Pi(L)x_t = \Pi(L)\Psi d_t + \varepsilon_t = \Psi d_t - \Pi_1 \Psi d_{t-1} - ... - \Pi_k \Psi d_{t-k} + \varepsilon_t.$$ 

Let $Y = [\Pi(L)x_t]$, $D = [d_t, -d_{t-1}, ..., -d_{t-k}]$ and $\Theta' = (\Psi, \Pi_1 \Psi, ..., \Pi_k \Psi)$. The model can now be written $Y = D\Theta + E$. The full conditional posterior of $\Psi$ can therefore be derived
from standard results for multivariate regression by using that $\text{vec } \Theta' = U \text{ vec } \Psi$, where

$$U = \begin{pmatrix} I_{pq} \\ (I_q \otimes \Pi_1) \\ \vdots \\ (I_q \otimes \Pi_k) \end{pmatrix}.$$ 

When at least one of the eigenvalues of the companion matrix of $\Pi$ is equal to or larger than one in modulus the process is non-stationary, the unconditional mean does not exist and the parameters in $\Psi$ are locally non-identified. Similar identification problems exists in the steady state VECM in (3.1); Kleibergen and van Dijk (1994) discuss these problems in a closely related parametrization. This identification problem can lead to convergence problems only if the prior on $\Psi$ is uninformative and the posterior of $\Pi$ has non-negligible probability mass in the non-stationary region. To explicitly show how prior information on $\Psi$ stabilizes the Gibbs sampler we may rewrite the precision matrix in the full conditional posterior of $\Psi$ (assuming for simplicity that $d_t = 1$) as

$$\tilde{\Omega}_\Psi^{-1} = U'(D'D \otimes \Sigma^{-1})U + \Omega_\Psi^{-1} = T(I_p - \sum_{i=1}^k \Pi_i)' \Sigma^{-1} (I_p - \sum_{i=1}^k \Pi_i) + \tilde{\Omega}_\Psi^{-1}.$$ 

With a flat prior on $\Psi$ it is easily seen that $\tilde{\Omega}_\Psi$ diverges as we approach the unit root region (where $I_p - \sum_{i=1}^k \Pi_i$ becomes rank deficient), which may cause the Gibbs sampler to converge slowly or perhaps not at all. With an informative prior assigned to $\Psi$ we instead have that $\tilde{\Omega}_\Psi \rightarrow \Omega_\Psi$ as the system approaches a unit root. Villani (2005b) uses simulated data to illustrate that Gibbs sampling works very well for the Steady State VAR even if the prior on $\Psi$ is only moderately informative. Our experience is that the Gibbs sampler for the Steady State VAR is highly efficient for the type of prior beliefs on the steady state usually held by macroeconomists.

**Appendix B. Posterior sampling for the steady state VECM**

We sample the posterior of the steady state VECM from the decomposition

$$p(\varphi, \eta, \Gamma, \Sigma|\mathcal{D}) = p(\varphi, \eta|\mathcal{D})p(\Gamma, \Sigma|\varphi, \eta, \mathcal{D}).$$
Sampling from \( p(\varphi, \eta) \) is straightforward as the model conditional on \( \varphi \) and \( \eta \) is the standard multivariate regression

\[
\tilde{Y} = \tilde{X}\Gamma + E,
\]

where \( \tilde{Y} = [\Delta x_t - \gamma] \), \( \tilde{X} = [\beta' x_{t-1} - \mu_0 - \mu_1 t, \Delta x_{t-1} - \gamma, ..., \Delta x_{t-k+1} - \gamma] \), \( E = [\varepsilon_t] \) and \( \Gamma = (\alpha, \Gamma_1, ..., \Gamma_{k-1})' \). Straightforward algebra shows that the posterior \( p(\Gamma, \Sigma | \varphi, \eta, D) \) is given by

\[
\Sigma | \varphi, \eta, D \sim IW(\bar{\Sigma}, T)
\]

\[
\text{vec} \Gamma | \Sigma, \varphi, \eta, D \sim N(\text{vec} \bar{\Gamma}, \Sigma \otimes \bar{\Omega}_0)
\]

where \( \bar{\Sigma} = (\tilde{Y} - \tilde{X}\hat{\Gamma})'(\tilde{Y} - \tilde{X}\hat{\Gamma}) + (\hat{\Gamma} - \theta_\Gamma)' \hat{X}' \hat{X} (\hat{X}' \hat{X} + \Omega^{-1}_T)^{-1} \hat{\Omega}_T^{-1} (\hat{\Gamma} - \theta_T) \), \( \hat{\Gamma} = (\hat{X}' \hat{X} + \Omega^{-1}_T)^{-1} (\hat{X}' \hat{X} + \Omega^{-1}_T \theta_T) \), \( \hat{\Omega}_T = (\hat{X}' \hat{X} + \Omega^{-1}_T)^{-1} \). That is, the posterior is of Normal-Inverse Wishart form and can be sampled directly.

To obtain \( p(\varphi, \eta | D) \) we need to integrate \( p(\Gamma, \Sigma, \varphi, \eta | D) \) with respect to \( \Gamma \) and \( \Sigma \). Using properties of the Inverse Wishart to integrate w.r.t. \( \Sigma \) and properties of the matrix \( t \) density to integrate w.r.t. \( \Gamma \) we obtain

\[
(B.1) \quad p(\varphi, \eta | D) \propto |\bar{\Sigma}|^{-T/2} \left| \hat{X}' \hat{X} + \Omega_T^{-1} \right|^{-p/2} p(\varphi, \eta)
\]

The marginal posterior \( p(\varphi, \eta | D) \) in (B.1) is clearly of non-standard form, but is usually not high-dimensional \( \text{Dim}(\varphi, \eta) \leq (p - r) r + p + r \), and we sample from it using the independence Metropolis-Hastings (MH) algorithm. The proposal density is a tailored multivariate \( t \) density with \( v \) degrees of freedom:

\[
(\varphi, \eta) \sim t(\mu, \kappa \Omega, v),
\]

where \( \mu \) is the posterior mode of \( p(\varphi, \eta | D) \), \( \kappa \) is a scaling parameter, and the covariance matrix \( \Omega \) is set to the negative inverse Hessian of \( \ln p(\varphi, \eta | D) \) at \( \mu \). A more robust, but less efficient algorithm, is obtained by replacing \( \mu \) in the proposal density with the most recently accepted draw of the parameters (Metropolis algorithm). The mode and Hessian can be obtained numerically, most easily using a numerical optimizer with sequential updating of the Hessian.\(^2\)

Initial values for \( \varphi \) and \( \eta \) in the optimization can be taken as the maximum likelihood (ML)

\(^2\)We have successfully used Chris Sims’ quasi-Newton optimizer with BFGS update of the Hessian (\textit{csminwel}) which can be freely downloaded for Matlab or R from http://sims.princeton.edu/yftp/optimiz/.
estimates of $\varphi$ in the standard VECM (Johansen, 1995) and $\eta$ can then be estimated from $\hat{\beta}'x_{t-1}$, where $\hat{\beta}$ is the ML estimate of $\beta$.

**APPENDIX C. EXTENSIONS TO STRUCTURAL/IDENTIFIED VARs**

This appendix gives the details for extending the algorithms in the paper to structural or identified VARs. The results presented here were derived in the unpublished paper Villani and Warne (2003). A VAR can be identified by specifying the lag polynomial as $\Pi(L) = \Pi_0 - \Pi_1L - \ldots - \Pi_kL^k$, where $\Pi_0$ is the $p \times p$ matrix of (restricted) contemporaneous coefficients, and setting $\Sigma = I_p$. The same can be done for $\Gamma(L)$ in the VECM. When it comes to the updating of the contemporaneous coefficients we can write both the VAR and the VECM in the form of a simultaneous equation model: $Y\Upsilon = Z + E$, where $\Upsilon = \Pi_0$ and $Z$ contains the linear combinations of lags (and cointegrating relations in the VECM case). The conditional posterior of $\Upsilon$ is therefore of the same form in the VAR and . We assume that the identifying restrictions on $\Upsilon$ is of the form

(C.1) $\omega_i = G_i\phi_i, \ i = 1, \ldots, p,$

where $\omega_i$ is the $i$th column of $\Upsilon$ (the coefficients in the $i$th equation), $G_i$ ($p \times s_i$) determines the restrictions and $\phi_i$ is a $s_i$-dimensional vector of unrestricted coefficients. The equations can be normalized using the method in Waggoner and Zha (2003a).

Formulating an informative prior on $\Upsilon$ is not an easy task. Prior elicitation is easier if the variances of the structural shocks are modelled explicitly (i.e freeing up the diagonal of $\Sigma$), but this leads to other difficulties, such as potentially improper posteriors and less efficient posterior sampling (Sims and Zha, 1994). Here we settle with a uniform prior on $\phi_1, \ldots, \phi_p$, a multivariate normal prior can also be used without changing the algorithm. The full conditional posterior of $\phi_i$ is of the form

(C.2) $p(\phi_i|\cdot) \propto |\Upsilon|^T \text{etr}[(Y\Upsilon - Z)'(Y\Upsilon - Z)] \propto |\Upsilon|^T \exp \left(-\frac{T}{2}(\phi_i - \mu_{\phi_i})'\Omega_{\phi_i}^{-1}(\phi_i - \mu_{\phi_i})\right)$

where $\Omega_{\phi_i}^{-1} = T^{-1}G_i'Y'YG_i$ and $\mu_{\phi_i} = (G_i'Y'YG_i)^{-1}G_i'Y'z_i$, where $z_i$ is the $i$th column of $Z$. This full conditional posterior does not belong to a known family of distributions and
direct sampling is therefore not feasible. Waggoner and Zha (2003b, Theorem 1) devise an
elegant and highly efficient way to sample from \( p(\phi_i|\cdot) \) in the case where \( \mu_{\phi_i} = 0 \). We shall here provide the generalization when \( \mu_{\phi_i} \) is allowed to be non-zero. We need the following preliminary definition.

**Definition C.1.** A random variable \( X \) follows the absolute normal distribution \( \text{AN}(\mu, \rho) \) if it has density function

\[
f_{\text{AN}}(x; \mu, \rho) = c |x|^{\frac{1}{\rho}} \exp \left( -\frac{1}{2\rho} (x - \mu)^2 \right), \quad x \in \mathbb{R},
\]

where \( c \) is a normalizing constant, \( \rho \in \mathbb{R}^+ \) and \( \mu \in \mathbb{R} \).

In the next theorem, let \( B_{-i} \) equal the matrix \( B \) with the \( i \)th column deleted, \( B_\perp \) is the orthogonal complement of \( B \), \( \text{Chol}(B) \) is the Choleski root of \( B \) such that \( B = \text{Chol}(B)\text{Chol}(B)' \), \( \| \cdot \| \) is the usual Euclidean norm and \( \overset{d}{=} \) denotes equality in distribution. With these preliminaries at hand we can prove the following extension of Theorem 1 in Waggoner and Zha (2003b).

**Proposition C.1.** The full conditional posterior of the unrestricted coefficients in the \( i \)th simultaneous relation is

\[
\phi_i|\cdot \overset{d}{=} R_i \sum_{j=1}^{s_i} \xi_j v_j,
\]

where \( R_i = \text{Chol}[T(G_i'Y'YG_i)^{-1}] \), \( \xi_1 \sim \text{AN}(\hat{\xi}_1, T^{-1}) \), \( \xi_j \sim N(\hat{\xi}_j, T^{-1}) \), for \( j = 2, \ldots, s_i \), \( \hat{\xi}_j = \mu_{\phi_i} R_i^{-1} v_j \), \( v_1 = R_i G_i'\Upsilon_{-\perp} / \| R_i G_i'\Upsilon_{-\perp} \| \), \( (v_2, \ldots, v_{s_i}) = v_{1\perp} \) and \( \mu_{\phi_i} = (G_i'Y'YG_i)^{-1}G_i'Y'z_i \), where \( z_i \) is the \( i \)th column of \( Z \).

**Proof.** Following Waggoner and Zha (2003b), decompose \( \phi_i = R_i \sum_{j=1}^{s_i} \xi_j v_j \). Then from (C.2) we have

\[
p(\phi_i|\cdot) \propto |\Upsilon|^T \exp \left( -\frac{T}{2} \sum_{j=1}^{s_i} (\xi_j - \hat{\xi}_j)^2 \right),
\]
where $\hat{\xi}_j = \mu'_i R^i v_j$. As $|Y| = \left| (\omega_1, \ldots, G_i R_i \sum_{j=1}^{s_i} \xi_j v_j, \ldots, \omega_p) \right| \propto |\xi_1|$ (Waggoner and Zha, 2003b), we have

$$p(\phi_i|\cdot) \propto |\xi_1|^T \exp \left[ -\frac{T}{2} (\xi_1 - \hat{\xi}_1)^2 \right] \prod_{j=2}^{s_i} \exp \left[ -\frac{T}{2} (\xi_j - \hat{\xi}_j)^2 \right],$$

which by Definition C.1. proves the result.

In order to use Proposition C.1 we need to be able to efficiently generate variates from the absolute normal distribution. A simple, very accurate and readily sampled approximation can be derived from the fact that the $AN(\mu, \rho)$-distribution is bimodal with modes at $\mu \pm \frac{1}{2} \sqrt{(\mu^2 + 4)}$. Furthermore, the curvature at $x = x_0$ is

$$- \left. \frac{d^2}{dx^2} \ln f_{AN}(x) \right|_{x=x_0} = \frac{x_0^2}{(1 + x_0^2) \rho}.$$

These two facts can be used to build the following mixture of normals approximation to the $AN(\mu, \rho)$ distribution

$$f_{AN}(x; \mu, \rho) \approx w N(x; \mu_1, \sigma_1^2) + (1 - w) N(x; \mu_2, \sigma_2^2),$$

where $N(x; \cdot, \cdot)$ is used as a shorthand for the density of a normal distribution, $\mu_1 = \frac{\mu}{2} - \frac{1}{2} \sqrt{(\mu^2 + 4)}$, $\mu_2 = \frac{\mu}{2} + \frac{1}{2} \sqrt{(\mu^2 + 4)}$, $\sigma_i^2 = \frac{\mu^2}{(1 + \mu^2) \rho}$, $i = 1, 2$, and $w = [1 + \exp(2\mu \rho^{-1})]^{-1}$. The accuracy of this approximation increases inversely with $\rho$ and even in the worst scenario when $\mu = 0$ it is already very accurate for $\rho = 0.1$. In our use of the absolute normal distribution, $\rho = T^{-1}$, where $T$ is the length of the time series, so the approximation can, for all practical purposes, be taken as exact.

References


<table>
<thead>
<tr>
<th>Parameter</th>
<th>Informative Prior</th>
<th></th>
<th></th>
<th>Looser prior</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.5%</td>
<td>50%</td>
<td>97.5%</td>
<td>IF</td>
<td>2.5%</td>
<td>50%</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>-0.349</td>
<td>-0.284</td>
<td>-0.223</td>
<td>3.632</td>
<td>-0.393</td>
<td>-0.251</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>-0.662</td>
<td>-0.594</td>
<td>-0.520</td>
<td>3.492</td>
<td>-0.736</td>
<td>-0.627</td>
</tr>
<tr>
<td>$\gamma_c$</td>
<td>1.625</td>
<td>1.985</td>
<td>2.336</td>
<td>3.892</td>
<td>1.574</td>
<td>2.020</td>
</tr>
<tr>
<td>$\gamma_a$</td>
<td>1.712</td>
<td>2.303</td>
<td>2.878</td>
<td>3.511</td>
<td>1.574</td>
<td>2.464</td>
</tr>
<tr>
<td>$\gamma_y$</td>
<td>1.758</td>
<td>2.238</td>
<td>2.717</td>
<td>4.007</td>
<td>1.614</td>
<td>2.237</td>
</tr>
<tr>
<td>$\alpha_c$</td>
<td>-0.068</td>
<td>-0.029</td>
<td>0.010</td>
<td>2.489</td>
<td>-0.070</td>
<td>-0.028</td>
</tr>
<tr>
<td>$\alpha_a$</td>
<td>0.013</td>
<td>0.198</td>
<td>0.407</td>
<td>2.960</td>
<td>-0.006</td>
<td>0.194</td>
</tr>
<tr>
<td>$\alpha_y$</td>
<td>-0.074</td>
<td>0.005</td>
<td>0.095</td>
<td>1.823</td>
<td>-0.077</td>
<td>0.008</td>
</tr>
<tr>
<td>$\mu_{cay}$</td>
<td>0.500</td>
<td>0.824</td>
<td>1.086</td>
<td>12.637</td>
<td>0.583</td>
<td>0.884</td>
</tr>
</tbody>
</table>

MH acc. prob 65% 55%

Note: The column IF reports the inefficiency factors from the independence algorithm.
Figure 1. Point forecasts of annual GDP growth for Sweden in 1998Q4-2005Q3. The thick line is the actual outcome and the thin lines are the forecasts made at different points in time. The forecasts are plotted every other quarter for clarity in presentation.
Figure 2. Point forecasts of annual CPI inflation for Sweden in 1998Q4-2005Q3. The thick line is the actual outcome and the thin lines are the forecasts made at different points in time. The forecasts are plotted every other quarter for clarity in presentation.
Figure 3. Point forecasts of the 3-month interest rate for Sweden in 1998Q4-2005Q3. The thick line is the actual outcome and the thin lines are the forecasts made at different points in time. The forecasts are plotted every other quarter for clarity in presentation.
Figure 4. Swedish macro data. Out-of-sample evaluation of point forecasts (predictive mean) for 1999Q1-2005Q4. Each subgraph displays the root mean squared forecast errors (RMSE) as a function of the forecast horizon.
Figure 5. Swedish macro data. Sequential posterior median estimates of the steady state over time (measured on the left axis). The black solid line with dots (measured on the right axis) displays the length of the 95% probability interval of the steady state under the Litterman prior.