JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **24**: 630–650 (2009)
Published online 27 March 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/jae.1065

STEADY-STATE PRIORS FOR VECTOR AUTOREGRESSIONS

MATTIAS VILLANI[ab]*

[a] *Research Department, Sveriges Riksbank, Sweden*
[b] *Department of Statistics, Stockholm University, Sweden*

SUMMARY

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 USA.

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 already noted by Sims in his original 1980 paper, where he suggested using prior information to increase precision. Sims' suggestion spurred the development of prior distributions for VARs; see, for example, the well-known *Litterman/Minnesota prior* documented in Litterman (1986), and subsequent modifications and extensions in Doan *et al.* (1984), Kadiyala and Karlsson (1997), and Robertson and Tallman (1999) for reduced form VARs, Sims and Zha (1998) and Waggoner and Zha (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 non-informative 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

* Correspondence to: Mattias Villani, Research Department, Sveriges Riksbank, SE-103 37 Stockholm, Sweden.
E-mail: mattias.villani@riksbank.se

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 strait-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 US data often have modeled variables in levels, where the process's steady state is typically non-existent or at least not very relevant (near unit root process). While VARs in levels seem to work well for US data (see, for example, the many papers by Sims and his co-authors), it has not been as successful for other countries, where instead differencing or cointegration has 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 *et al.* (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 modeled. 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 1990s. 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 modeled. 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. Section 4 illustrates the analysis on a seven-variable model for the Swedish macro economy, and on a

DOI: 10.1002/jae

VECM model of the consumption–wealth relationship in the USA. Section 5 concludes. The algorithms are described in detail in the appendices.

2. BAYESIAN ANALYSIS OF THE STEADY-STATE VAR

2.1. The Model

The usual parametrization of the VAR model is

$$\Pi(L)x_t = \Phi d_t + \varepsilon_t \tag{1}$$

where x_t is a p-dimensional vector of time series at time t, and 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 (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 Σ. This daunting task is often simplified by modeling 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 Φ (see, for example, 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, but simply that the particular parametrization of the model in (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 μ_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. μ_t is unfortunately a complicated nonlinear function of $\Pi_1, \ldots, \Pi_k$ and Φ. Thus, convenient as parametrization (1) may be from a computational viewpoint, it is not the preferred parametrization for incorporating prior opinions regarding μ_t.

An alternative parametrization of the VAR is

$$\Pi(L)(x_t - \Psi d_t) = \varepsilon_t \tag{2}$$

This VAR model is nonlinear in its parameters, but the unconditional mean of the process is directly specified by Ψ as $\mu_t = \Psi d_t$. The form of the deterministic component Ψd_t is flexible, and 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) will be referred to as the *steady-state VAR*.

2.2. Prior Distribution

Bayesian inference requires a prior distribution on Σ, Π_1, ., Π_k, and Ψ. The prior on Σ is taken to be

$$p(\Sigma) \propto |\Sigma|^{-(p+1)/2}$$

DOI: 10.1002/jae

Let $\Pi = (\Pi_1, ., \Pi_k)'$. The prior for vec Π 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 Π and Ψ, and that

$$\text{vec}\Psi \sim N_{pq}(\theta_\Psi, \Omega_\Psi)$$

2.3. Posterior Distribution

The posterior distribution of the steady-state VAR in (2) is intractable. It is shown in Appendix A that the posterior distribution of each set of model parameters (i.e., one of Σ, Π and Ψ) conditional on the other parameters is tractable, and that a simple three-block Gibbs sampler (Smith and Roberts, 1993) can therefore be used to draw from the joint posterior. The full conditional posteriors of Σ, vec Π and vec Ψ are inverted Wishart, normal and normal densities, respectively. The extra computing time for the steady-state VAR compared to the standard VAR in (1) (which is typically analyzed by a two-block Gibbs sampler, one block with Σ and the other with Π and Φ) 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 vecΠ. The Gibbs sampler in Appendix A for the steady-state VAR can be slightly modified to the case with general linear restrictions on vec Π. To see this, formulate the restrictions as vec $\Pi = H\varrho + h$, where H is a known restriction matrix, h a known vector and ϱ the unrestricted parameters in Π (see Johansen, 1995, for more details on this formulation). Repeating the derivations in Appendix A with vec Π replaced by $H\varrho + h$ (the Jacobian is constant so it can be ignored) shows that the full conditional posterior of ϱ is a multivariate normal density. The same clearly applies also to linear restrictions on vec Ψ.

We also note that the generated Gibbs sample may be used to compute the marginal likelihood of a model using, for example, 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 (Σ, Ψ)-space. The time-consuming Π-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) 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)

$$\Gamma(L)(\Delta x_t - \gamma) = \alpha(\beta x_{t-1} - \mu_0 - \mu_1 t) + \varepsilon_t \tag{3}$$

DOI: 10.1002/jae

where β is a $p \times r$ matrix with the r cointegration vectors as columns, α is a $p \times r$ matrix of adjustment coefficients determining the speed of adjustment back to equilibrium after a disturbance, $\Gamma(L) = I_p - \Gamma_1 L - \ldots - \Gamma_{k-1}L^{k-1}$, 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) are related by $\beta'\gamma = \mu_1$ (Clements and Hendry, 1999, pp. 152–153). Most applications of the VECM model preclude drift in the cointegrating relations, i.e., set μ_1 to the zero vector. This imposes the nonlinear restriction $\beta'\gamma = 0$ on the VECM parameters. When $\mu_1 = 0$, we parametrize γ as $\gamma = \beta_\perp \lambda$, where $\beta_\perp$ is the $p \times (p - r)$ matrix orthogonal to $\beta(\beta'_\perp \beta_\perp = I_{p-r})$, and λ the $(p - r)$-dimensional vector of unrestricted parameters.

It is straightforward to extend the model in (3) by modeling $E(\Delta x_t)$ 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 $\beta' E(\Delta x_t) = \Delta E(\beta' x_t), t = 1, \ldots, T$. Consider for example the model $\Delta x_t = \Phi_1 d_t$, where d_t is a vector with a constant and a set of dummy variables as in Section 2, and $E(\beta' x_t) = \mu_0 + \mu_1 t$. We then have $\beta'\Phi_1 d_t = \mu_1$. If $\mu_1 = 0$, then we need to impose the nonlinear restriction $\beta'\Phi_1 = 0$, i.e., letting $\Phi_1 = \beta_\perp \lambda$, where λ 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 Σ as in the stationary case. The prior on $\Gamma = (\alpha, \Gamma_1, \ldots, \Gamma_{k-1})'$ is of the form

$$\operatorname{vec} \Gamma | \beta \sim N(\operatorname{vec} \theta_\Gamma, \Sigma \otimes \Omega_\Gamma) \tag{4}$$

where

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

where $\lambda_\alpha, \lambda_\Gamma > 0$ are shrinkage factors and S_α and S_Γ are positive definite matrices, usually assumed to be diagonal; see Villani (2005a) for a motivation of the prior on α. The diagonal elements of S_Γ are set as follows. Let $s_{ij}^{(k)}$ denote the prior standard deviation of the coefficient on $\Delta x_{j,t-k}$ in the ith equation, where $x_{j,t-k}$ is the kth lag of variable j. Similar to Litterman (1986) we set $s_{ij}^{(k)} = \lambda_1 / (\hat{\sigma}_j k^{\lambda_3})$, where λ_1 and λ_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_α, i.e., we set $S_\alpha = \operatorname{diag}(\hat{\sigma}_1^{-2}, \ldots, \hat{\sigma}_p^{-2})$. It should be noted that the prior covariance matrix of Γ 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 Γ. 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 β, γ and μ_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 β and the growth rates γ, and then let these priors imply a prior on μ_1, the drift in the cointegrating

DOI: 10.1002/jae

relations. Alternatively, we could have specified a prior on subsets of all three of β, γ and μ_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 μ_0 is intimately connected to the prior beliefs on β. Finally, specifying priors directly for β and γ, and therefore only implicitly for μ_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 β is assumed to be of the form

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

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

The interpretation of μ_0 depends crucially on β, and this must be taken into account when formulating a prior. The prior on μ_0 is of the form

$$\mu_0|\beta \sim N(\theta_{\mu_0}, Q\Omega_{\mu_0}Q') \tag{5}$$

where θ_{μ_0} is the prior mean of μ_0, Ω_{μ_0} is a positive semidefinite (PSD) matrix (e.g., $\Omega_{\mu_0} = \kappa I_r$). Q is a diagonal matrix with ith diagonal element $q_i = 1 + \sqrt{(\varphi_i - \mu_{\varphi_i})'\Omega_{\varphi_i}^{-1}(\varphi_i - \mu_{\varphi_i})}$. Note that this prior has the attractive property of making the ith element of μ_0 more diffuse when the ith cointegration vector is far from its prior mean.

The prior on the growth rates γ is $N(\theta_\gamma, \Omega_\gamma)$. As discussed above, the prior for μ_1 is then implicitly given from the relation $\beta'\gamma = \mu_1$ and the priors on β and γ. When $\mu_1 = 0$, $\gamma = \beta_\perp\lambda$ and the prior for γ is implicitly given by the prior on λ. It is in general easier to specify a prior on the growth rates γ than it is to specify a prior for λ. We therefore choose to project the $N(\theta_\gamma, \Omega_\gamma)$ prior for γ down to the subspace $\beta'\gamma = 0$. This way we obtain an implicit prior on λ as

$$\lambda|\beta \sim N(\beta_\perp'\theta_\gamma, \beta_\perp'\Omega_\gamma\beta_\perp) \tag{6}$$

where θ_γ and Ω_γ are the elicited prior mean and covariance matrix of γ. Thus any prior specified on γ 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 η as

$$\eta|\beta \sim N(\theta_\eta, \Omega_\eta)$$

3.3. Posterior Distribution

It is in principle possible to set up a Gibbs sampler also in the cointegrated case, but there are at least two complications that suggest this is not the best route to take. First, the full conditional posterior of β is non-standard since the prior covariance of μ_0 depends on β. Moreover, when $\mu_1 = 0$, β 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 problem of coping with local non-identification. In the case of the VECM, the interpretation of μ_0 is directly dependent on β (see the prior in (5)), so that being informative on μ_0 requires also an informative prior on β. This makes it harder to specify priors which are sufficiently informative to stabilize the Gibbs sampler.

DOI: 10.1002/jae

We will instead sample from the posterior using the following decomposition:

$$p(\varphi, \eta, \Gamma, \Sigma|\mathcal{D}) = p(\varphi, \eta|\mathcal{D})p(\Gamma, \Sigma|\varphi, \eta, \mathcal{D}) \tag{7}$$

Appendix B gives details of the posterior sampling scheme. Liu *et al.* (1994) prove in a general setting that sampling from decompositions like the one in (7) is more efficient than Gibbs sampling. Note also that the parameters in φ and η becomes non-identified when α is of reduced rank (Kleibergen and van Dijk, 1994), so that sampling φ and η directly from their marginal posterior is likely to cope better with local non-identification than a pure Gibbs sampler (where φ and η are drawn conditional on α).

The marginal likelihood may be computed from the (low-dimensional) posterior draws of φ and η 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 (Δy_f), CPI inflation (π_f) and the 3-month interest rate (i_f), the corresponding domestic variables (Δy, π 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 $\Pi_i, i = 1, \ldots, k$, to the zero matrix. We will assume that $x_t = (\Delta y_f, \pi_f, i_f, \Delta y, \pi, i, q)'$ can be modeled as a stationary VAR process.

To model Sweden's change in monetary policy in the 1990s (move to inflation targeting and flexible exchange rate) we will use the dummy $d_t = (1, d_{MP,t})'$, where

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

To formulate a prior on Ψ, note that the specification of d_t implies the following parametrization of the steady state:

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

where ψ_i is the ith column of Ψ. The prior on ψ_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 *et al.* (2009), and is displayed as the first row of Table I. The prior on ψ_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, i.e., that Δy_f and Δ_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 whether the shift is purely nominal or not.

The prior proposed by Litterman (1986) will be used on the dynamic coefficients in Π, with the default values on the hyperparameters in the priors advocated by Doan (1992): overall tightness

DOI: 10.1002/jae

Table I. Swedish data: 95% prior probability intervals for steady state

	Δy_f	π_f	r_f	Δy	π	r	q
ψ_1	(2,3)	(1.5, 2.5)	(4.5, 5.5)	(2, 2.5)	(1.7, 2.3)	(4, 4.5)	(3.85, 4)
ψ_2	(−1, 1)	(1.5, 2.5)	(1.5, 2.5)	(−1, 1)	(4.3, 5.7)	(3, 5.5)	(−0.5, 0.5)

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 Σ.

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 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 the 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 at 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 explicit inflation target of 2% (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

DOI: 10.1002/jae

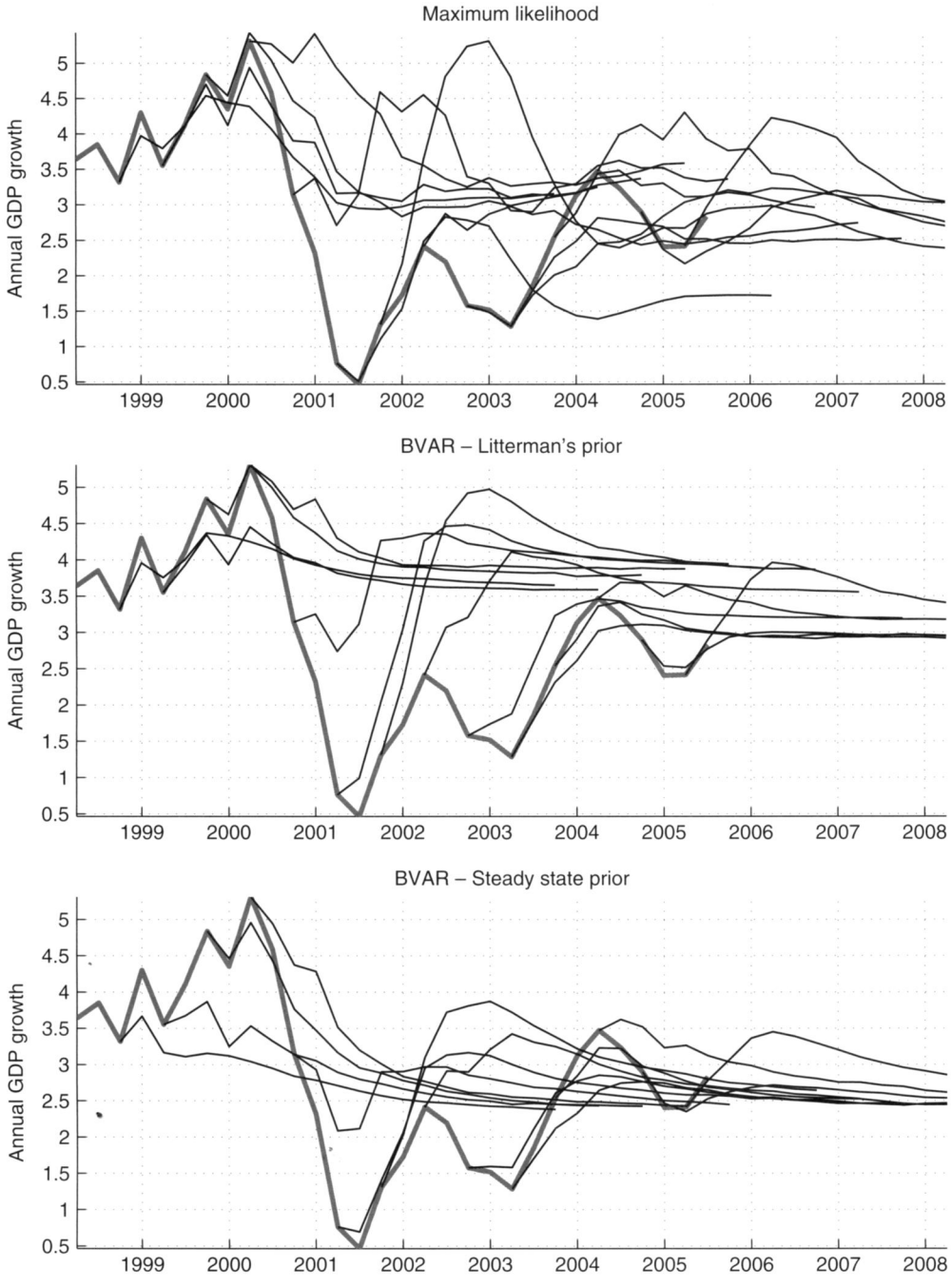

Figure 1. Point forecasts of annual GDP growth for Sweden in 1998Q4–2005Q3. The thick line is the actual outcome and the thin lines are the forecasts made at different points in time. The forecasts are plotted every other quarter for clarity in presentation

J. Appl. Econ. **24**: 630–650 (2009)
DOI: 10.1002/jae

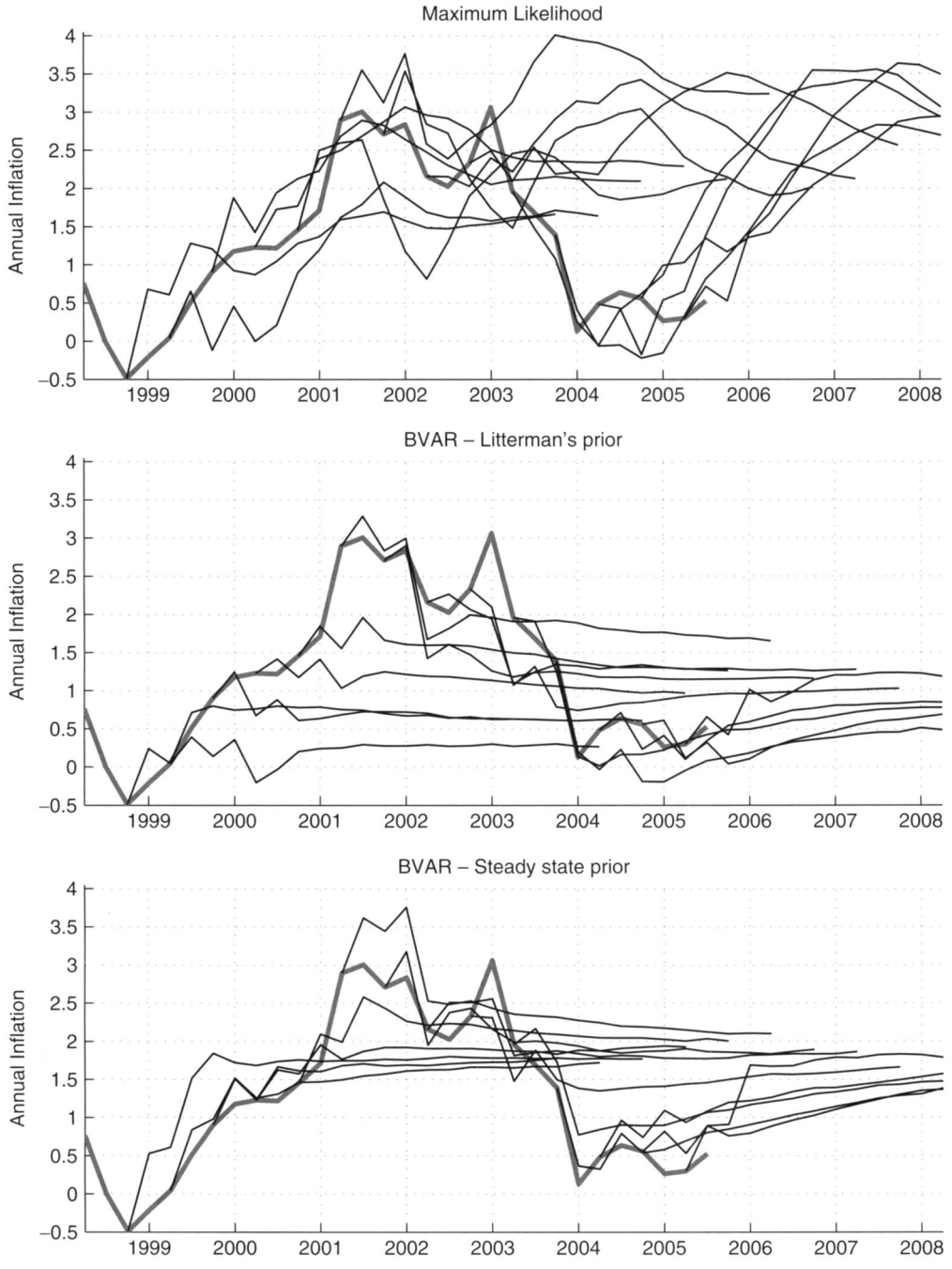

Figure 2. Point forecasts of annual CPI inflation for Sweden in 1998Q4–2005Q3. The thick line is the actual outcome and the thin lines are the forecasts made at different points in time. The forecasts are plotted every other quarter for clarity in presentation

J. Appl. Econ. **24**: 630–650 (2009)
DOI: 10.1002/jae

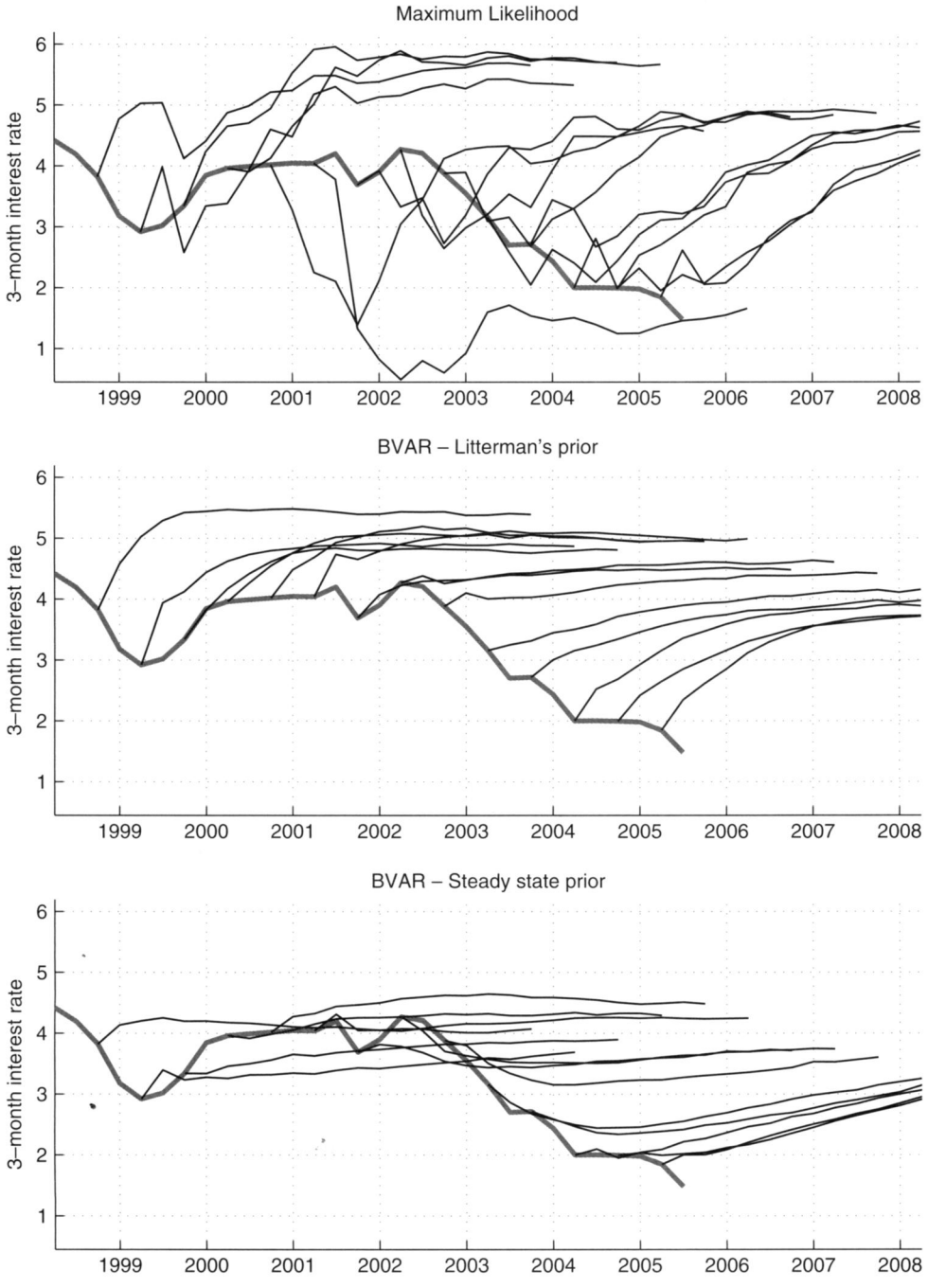

Figure 3. Point forecasts of the 3-month interest rate for Sweden in 1998Q4–2005Q3. The thick line is the actual outcome and the thin lines are the forecasts made at different points in time. The forecasts are plotted every other quarter for clarity in presentation

J. Appl. Econ. **24**: 630–650 (2009)
DOI: 10.1002/jae

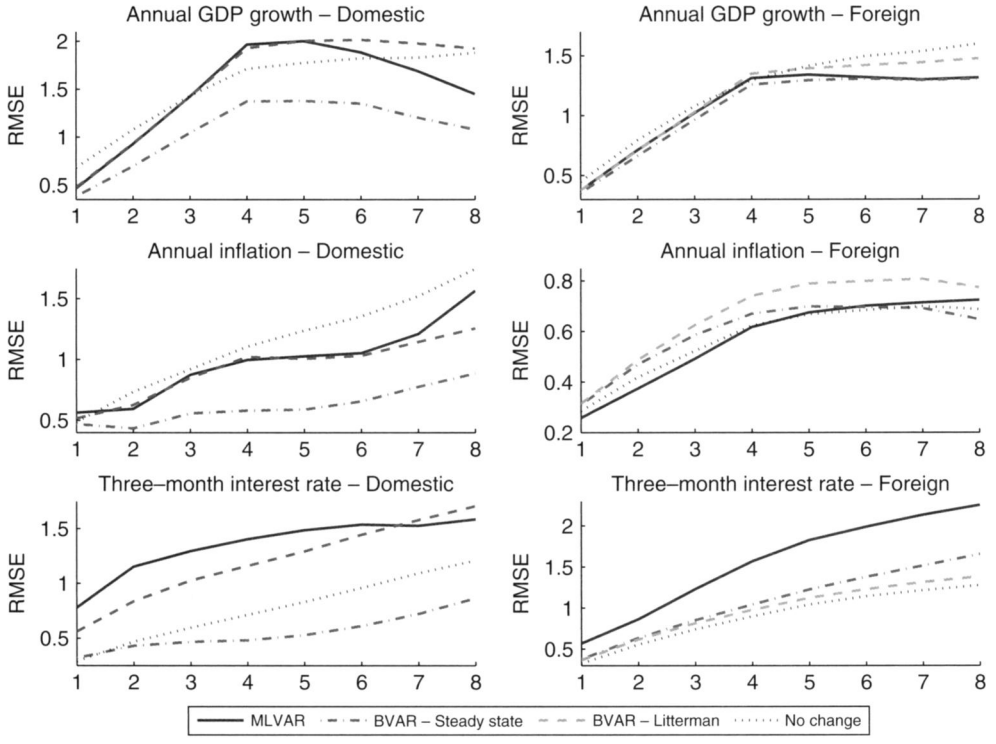

Figure 4. Swedish macro data. Out-of-sample evaluation of point forecasts (predictive mean) for 1999Q1–2005Q4. Each subgraph displays the root mean squared forecast errors (RMSE) as a function of the forecast horizon

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 the 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 converge later in the evaluation sample (Figure 5). Interestingly, the last forecast paths of the interest rate illustrate 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 pronouced as for the domestic variables. The RMSEs in Figure 4 suggest 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.

DOI: 10.1002/jae

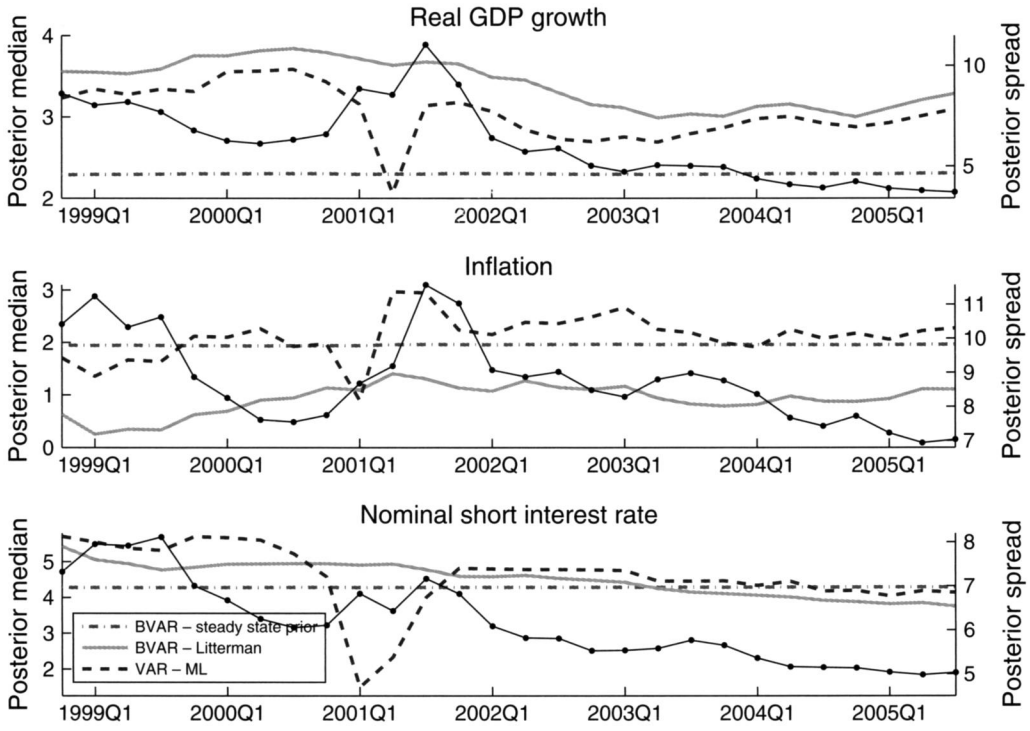

Figure 5. Swedish macro data. Sequential posterior median estimates of the steady state over time (measured on the left axis). The black solid line with dots (measured on the right axis) displays the length of the 95% probability interval of the steady state under the Litterman prior

4.2. Cointegrated VAR

Lettau and Ludvigson (2001, 2004) analyze the US consumption–wealth relationship in the period 1951Q4–2003Q–1. They show that a wide range of optimization models implies 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 illustrate here 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 *et al.*, 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 α and Γ is non-informative. An informative prior for β can be elicitated from the observation that φ_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

[1] The updated data can be downloaded from Martin Lettau's web page:http://faculty.haas.berkeley.edu/lettau/.

DOI: 10.1002/jae

these estimates imply that $\varphi_1 = \zeta/1.1$ and $\varphi_2 = (1-\zeta)/1.1$, where ζ 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 φ_1 and φ_2 based on the estimate $\hat{\zeta} = 0.353$ obtained by Solow (1957) on data between 1909 and 1949, i.e., before our study period. A normal prior that centers φ_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, φ_2 is normal a priori with mean $(1-0.353)/1.1$ and standard deviation 0.05. Since we expect to have nearly $\varphi_2 = 0.9 - \varphi_1$, we also use a prior correlation coefficient between φ_1 and φ_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 real GDP per capita during 1869–1951 (the period before our sample) was 1.98 (Johnston and Williamson, 2008), 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 at 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 ranging approximately 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 μ_0. We therefore use a non-informative $N(0, 5^2)$ prior for μ_0. There are other applications where a prior on μ_0 may be essential. An example is a simple three-variate model for a price index in two countries (p_t and 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. μ_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 ($\kappa = 1, \upsilon = 5$). The posterior sampling, including the numerical maximization, took less than a minute using Matlab 7 on a 2 GHz 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 non-informative prior is in the efficiency of the posterior sampling algorithm. The columns labeled IF in Table II display the inefficiency factors, i.e., the number of draws needed to obtain the equivalent of a single i.i.d. draw. For the informative prior the posterior draws are close to i.i.d. whereas for the weaker prior the sampling is less efficient. The IFs for Γ are essentially 1 for both priors. Note that the inefficiency in the sampling of φ and η under the weaker prior does not spill over to α and Γ. This is important as α and Γ typically contain the majority of the parameters. Finally, we also experimented with the Metropolis version of the sampler. As expected, this sampler is less efficient: the IFs for φ and η under the informative prior are three to five times larger than the IFs from the independence sampler.

DOI: 10.1002/jae

Table II. Posterior summary for the Lettau–Ludvigson model

Parameter	Informative prior				Looser prior			
	2.5%	50%	97.5%	IF	2.5%	50%	97.5%	IF
φ_1	−0.349	−0.284	−0.223	3.632	−0.393	−0.251	−0.156	23.098
φ_2	−0.662	−0.594	−0.520	3.492	−0.736	−0.627	−0.460	44.214
γ_c	1.625	1.985	2.336	3.892	1.574	2.020	2.448	23.929
γ_a	1.712	2.303	2.878	3.511	1.574	2.464	3.364	6.117
γ_y	1.758	2.238	2.717	4.007	1.614	2.237	2.858	17.984
α_c	−0.068	−0.029	0.010	2.489	−0.070	−0.028	0.011	2.673
α_a	0.013	0.198	0.407	2.960	−0.006	0.194	0.420	6.919
α_y	−0.074	0.005	0.095	1.823	−0.077	0.008	0.114	3.697
μ_{cay}	0.500	0.824	1.086	12.637	0.583	0.884	1.287	61.179
MH acc. prob		65%				55%		

Note: The column IF reports the inefficiency factors from the independence algorithm.

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 USA.

ACKNOWLEDGEMENTS

Comments from Malin Adolfson, Michael Andersson, Paolo Giordani, Herman van Dijk, Anders Warne, and especially two anonymous referees are greatfully 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.

REFERENCES

Adolfson M, Laseén S, Lindé J, Villani M. 2009. Evaluating a new-Keynesian model for a small open economy. *Journal of Economic Dynamics and Control* **32**(8): 2690–2721.

Bauwens L, Lubrano M, Richard J-F. 1999. *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press: Oxford.

Chib S. 1995. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* **90**: 1313–1321.

Clements MP, Hendry DF. 1995. Macro-economic forecasting and modelling. *Economic Journal* **105**: 1001–1013.

Clements MP, Hendry DF. 1998. *Forecasting Economic Time Series*. Cambridge University Press: Cambridge, UK.

Clements MP, Hendry DF. 1999. *Forecasting Non-stationary Economic Time Series*. MIT Press: London.

DOI: 10.1002/jae

Del Negro M, Schorfheide F, Smets F, Wouters R. 2007. On the fit and forecasting performance of new Keynesian models (with discussion). *Journal of Business and Economics Statistics* **25**: 123–162.
Doan TA. 1992. *RATS User's Manual*, Version 4. Estima: Evanston, IL.
Doan TA, Litterman RB, Sims CA. 1984. Forecasting and conditional projection using realistic prior distributions. *Econometrics Reviews* **5**: 1–100.
Geweke J. 1999. Using simulation methods for Bayesian econometrics models: inference, development and communication. *Econometric Reviews* **18**: 1–73.
Johansen S. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press: Oxford.
Johnston LD, Williamson SH. 2008. What was the U.S. GDP then? *Measuring Worth*. http://measuring.worth.org/usgdp/.
Kadiyala KR, Karlsson S. 1997. Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics* **12**: 99–132.
Kleibergen F, Paap R. 2002. Prior, posteriors and Bayes factors for a Bayesian analysis of cointegration. *Journal of Econometrics* **111**: 223–249.
Kleibergen F, van Dijk HK. 1994. On the shape of the likelihood/posterior in cointegration models. *Econometric Theory* **10**: 514–551.
Koop G, Potter SM, Strachan RW. 2005. Re-examining the consumption-wealth relationship: the role of model uncertainty. Working paper No. 05/3, University of Leicester, UK.
Lettau M, Ludvigson S. 2001. Consumption, aggregate wealth and expected stock returns. *Journal of Finance* **56**: 815–849.
Lettau M, Ludvigson S. 2004. Understanding trend and cycle in asset values: reevaluating the wealth effect on consumption. *American Economic Review* **94**: 276–299.
Litterman RB. 1986. Forecasting with Bayesian vector autoregressions: five years of experience. *Journal of Business and Economic Statistics* **5**: 25–38.
Liu JS, Wong WH, Kong A. 1994. Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* **81**: 27–40.
Robertson JC, Tallman EW. 1999. Vector autoregressions: forecasting and reality. *Economic Review* **84**: 4–18.
Schotman P, van Dijk HK. 1991a. A Bayesian analysis of the unit root in real exchange rates. *Journal of Econometrics* **49**: 195–238.
Schotman P, van Dijk HK. 1991b. On Bayesian routes to unit roots. *Journal of Applied Econometrics* **6**: 387–401.
Schotman P, van Dijk HK. 1993. Posterior analysis of possibly integrated time series with an application to real GNP. In *New Directions in Time Series Analysis, Part II*, Caines P, Geweke J, Taqqu M (eds). Springer: Heidelberg; 341–361.
Sims CA. 1980. Macroeconomics and reality. *Econometrica* **48**: 1–48.
Sims CA, Zha T. 1994. Error bands for impulse responses. Discussion paper, Cowles Foundation.
Sims CA, Zha T. 1998. Bayesian methods for dynamic multivariate models. *International Economic Review* **39**: 949–968.
Smets F, Wouters R. 2007. Shocks and frictions in US business cycles: a Bayesian DSGE approach. *American Economic Review* **97**: 586–606.
Smith AFM, Roberts GO. 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society B* **55**: 3–24.
Solow RM. 1957. Technical change and the aggregate production function. *Review of Economics and Statistics* **39**: 312–320.
Strachan RW. 2003. Valid Bayesian estimation of the cointegrating error correction model. *Journal of Business and Economic Statistics* **21**: 185–195.
Strachan RW, Inder B. 2004. Bayesian analysis of the error correction model. *Journal of Econometrics* **123**: 307–325.
Villani M. 2005a. Bayesian reference analysis of cointegration. *Econometric Theory* **21**: 326–357.
Villani M. 2005b. Inference in vector autoregressive models with an informative prior on the steady state. Working Paper Series No. 181, Sveriges Riksbank.
Villani M, Warne A. 2003. Monetary policy analysis in a small open economy using Bayesian cointegrated structural VARs. ECB Working Paper No. 296.

DOI: 10.1002/jae

Waggoner DF, Zha T. 2003a. Likelihood preserving normalization in multiple equation models. *Journal of Econometrics* **114**: 329–347.
Waggoner DF, Zha T. 2003b. A Gibbs sampler for structural vector autoregressions. *Journal of Economic Dynamics and Control* **28**: 349–366.
Zellner A. 1971. *An Introduction to Bayesian Inference in Econometrics*. Wiley: New York.
Zha T. 1999. Block recursion and structural vector autoregressions. *Journal of Econometrics* **90**: 291–316.

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 $Z = \lfloor z_{1t}, z_{2t}, \ldots, z_{mt} \rfloor_{t=1}^{T}$ to denote that the tth row of the $T \times \left(\sum_{i=1}^{m} l_i\right)$ matrix Z equals $(z_{1t}', z_{2t}', \ldots, z_{mt}')$. The symbol $\lfloor \ \rfloor_{t=1}^{T}$ 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 $Z = \lfloor z_{1t}, z_{2t}, \ldots, z_{mt} \rfloor$. Furthermore, let $\mathcal{D} = \{x_1, \ldots, x_T, d_1, \ldots, d_T\}$ denote the data.

Proposition A.1

- *Full conditional posterior of* Σ

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

where $E = \lfloor \Pi(L)(x_t - \Psi d_t) \rfloor$.

- *Full conditional posterior of* Π

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

where $\bar{\Omega}_\Pi^{-1} = \Sigma^{-1} \bigotimes X_\Psi' X_\Psi + \Omega_\Pi^{-1}$, $\bar{\theta}_\Pi = \bar{\Omega}_\Pi[\mathrm{vec}(X_\Psi' Y_\Psi \Sigma^{-1}) + \Omega_\Pi^{-1}\theta_\Pi]$, $Y_\Psi = \lfloor x_t - \Psi d_t \rfloor$ *and* $X_\Psi = \lfloor x_{t-1} - \Psi d_{t-1}, \ldots, x_{t-k} - \Psi d_{t-k} \rfloor$.

- *Full conditional posterior of* Ψ

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

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

Proof: *Conditional on* Ψ *the model in (2) is a standard VAR model for the time series* $x_t - \Psi d_t$. *The full conditional posteriors of* Σ *and* Π *therefore follow from standard results (see, for example, Zellner, 1971). To derive the full conditional posterior of* Ψ *we rewrite the model in (2) as*

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

Let $Y = \lfloor \Pi(L)x_t \rfloor$, $D = \lfloor d_t, -d_{t-1}, \ldots, -d_{t-k} \rfloor$ *and* $\Theta' = (\Psi, \Pi_1\Psi, \ldots, \Pi_k\Psi)$. *The model can now be written* $Y = D\Theta + E$. *The full conditional posterior of* Ψ *can therefore be derived from*

DOI: 10.1002/jae

standard results for multivariate regression by using that $vec\Theta' = U\ vec\ \Psi$, *where*

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

□

When at least one of the eigenvalues of the companion matrix of Π is equal to or larger than one in modulus the process is non-stationary, the unconditional mean does not exist, and the parameters in Ψ are locally non-identified. Similar identification problems exist in the steady-state VECM in (3); 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 Ψ is uninformative *and* the posterior of Π has non-negligible probability mass in the non-stationary region. To explicitly show how prior information on Ψ stabilizes the Gibbs sampler we may rewrite the precision matrix in the full conditional posterior of Ψ (assuming for simplicity that $d_t = 1$) as

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

With a flat prior on Ψ it is easily seen that $\overline{\Omega}_{\Psi}$ 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 Ψ we instead have that $\overline{\Omega}_{\Psi} \to \Omega_{\Psi}$ 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 Ψ 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 $p(\Gamma, \Sigma|\varphi, \eta)$ is straightforward as the model conditional on φ and η is the standard multivariate regression

$$\tilde{Y} = \tilde{X}\Gamma + E$$

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

$$\Sigma|\varphi, \eta, \mathcal{D} \sim IW(\overline{\Sigma}, T)$$

$$\text{vec}\ \Gamma|\Sigma, \varphi, \eta, \mathcal{D} \sim N(\text{vec}\overline{\Gamma}, \Sigma \otimes \overline{\Omega}_{\Gamma})$$

DOI: 10.1002/jae

where $\overline{\Sigma} = (\tilde{Y} - \tilde{X}\hat{\Gamma})'(\tilde{Y} - \tilde{X}\hat{\Gamma}) + (\hat{\Gamma} - \theta_\Gamma)'\tilde{X}'\tilde{X}(\tilde{X}'\tilde{X} + \Omega_\Gamma^{-1})^{-1}\Omega_\Gamma^{-1}(\hat{\Gamma} - \theta_\Gamma)$, $\hat{\Gamma} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}$, $\overline{\Gamma} = (\tilde{X}'\tilde{X} + \Omega_\Gamma^{-1})^{-1}(\tilde{X}'\tilde{X}\hat{\Gamma} + \Omega_\Gamma^{-1}\theta_\Gamma)$ and $\overline{\Omega}_\Gamma = (\tilde{X}'\tilde{X} + \Omega_\Gamma^{-1})^{-1}$. That is, the posterior is of normal-inverse Wishart form and can be sampled directly.

To obtain $p(\varphi, \eta|\mathcal{D})$ we need to integrate $p(\Gamma, \Sigma, \varphi, \eta|\mathcal{D})$ with respect to Γ and Σ. Using properties of the inverse Wishart to integrate w.r.t. Σ and properties of the matrix t density to integrate w.r.t. Γ we obtain

$$p(\varphi, \eta|\mathcal{D}) \propto \left|\overline{\Sigma}\right|^{-T/2} \left|\tilde{X}'\tilde{X} + \Omega_\Gamma^{-1}\right|^{-p/2} p(\varphi, \eta) \qquad \text{(B.1)}$$

The marginal posterior $p(\varphi, \eta|\mathcal{D})$ 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 μ is the posterior mode of $p(\varphi, \eta|\mathcal{D})$, κ is a scaling parameter, and the covariance matrix Ω is set to the negative inverse Hessian of $\ln p(\varphi, \eta|\mathcal{D})$ at μ. A more robust, but less efficient algorithm, is obtained by replacing μ 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 φ and η in the optimization can be taken as the maximum likelihood (ML) estimates of φ in the standard VECM (Johansen, 1995) and η can then be estimated from $\hat{\beta}'x_{t-1}$, where $\hat{\beta}$ is the ML estimate of β.

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 Villani and Warne (2003). A VAR can be identified by specifying the lag polynomial as $\Pi(L) = \Pi_0 - \Pi_1 L - \ldots - \Pi_k L^k$, where Π_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 contain the linear combinations of lags (and cointegrating relations in the VECM case). The conditional posterior of Υ is therefore of the same form in the VAR and VECM. We assume that the identifying restrictions on Υ is of the form

$$w_i = G_i\phi_i, \quad i = \ldots, p \qquad \text{(C.1)}$$

where w_i is the ith column of Υ (the coefficients in the ith equation), $G_i(p \times s_i)$ determines the restrictions and ϕ_i is an s_i-dimensional vector of unrestricted coefficients. The equations can be normalized using the method in Waggoner and Zha (2003a).

[2] We have successfully used Chris Sims' quasi-Newton optimizer with BFGS update of the Hessian (*csminwel*) which can be freely downloaded for Matlab or R from http://sims.princeton.edu/yftp/optimize/.

DOI: 10.1002/jae

Formulating an informative prior on Υ is not an easy task. Prior elicitation is easier if the variances of the structural shocks are modeled explicitly (i.e., freeing up the diagonal of Σ), 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 ϕ_i is of the form

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

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 ith 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 μ_{ϕ_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 AN(μ, ρ) if it has density function*

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

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 ith column deleted, $B_\perp$ is the orthogonal complement of B, 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 ith simultaneous relation is*

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

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

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

DOI: 10.1002/jae

$$p(\phi_i|\cdot) \propto |\Upsilon|^T \exp\left(\frac{-T}{2}\left[\sum_{j=1}^{s_i}[(\xi_j - \hat{\xi}_j)^2]\right]\right)$$

where $\hat{\xi}_j = \mu'_{\phi_i} R_i'^{-1} v_j$. As $|\Upsilon| = |(w_1, \ldots, G_i R_i \sum_{j=1}^{s_i} \xi_j v_j, \ldots, w_p| \propto |\xi_1|$ (Waggoner and Zha, 2003b), we have

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

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 $\frac{\mu}{2} \pm \frac{1}{2}\sqrt{(\mu^2 + 4)}$. Furthermore, the curvature at $x = x_0$ is

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

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

$$fAN(x; \mu, \rho) \approx wN(x; \mu_1, \sigma_1^2) + (1 - w)N(x; \mu_2, \sigma_2^2)$$

where $N(x; \cdot, \cdot)$ is used as a shorthand for the density of a normal distribution, $\mu_1 = \frac{\mu}{2} - \frac{1}{2}\sqrt{(\mu^2 + 4)}$, $\mu_2 = \frac{\mu}{2} + \frac{1}{2}\sqrt{(\mu^2 + 4)}$, $\sigma_i^2 = \frac{\mu_i^2}{(1 + \mu_i^2)}\rho$, $i = 1, 2$, and $w = [1 + \exp(2\mu\rho^{-1})]^{-1}$. The accuracy of this approximation increases inversely with ρ 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.

J. Appl. Econ. **24**: 630–650 (2009)
DOI: 10.1002/jae