SVERIGES RIKSBANK
WORKING PAPER SERIES
181

SVERIGES RIKSBANK

Inference in Vector Autoregressive Models with an Informative Prior on the Steady State

Mattias Villani

MARCH 2005

WORKING PAPERS ARE OBTAINABLE FROM

Sveriges Riksbank • Information Riksbank • SE-103 37 Stockholm
Fax international: +46 8 787 05 26
Telephone international: +46 8 787 01 00
E-mail: info@riksbank.se

Inference in Vector Autoregressive Models with an Informative Prior on the Steady State*

Mattias Villani†

Sveriges Riksbank Working Paper Series
No. 181
March 2005

Abstract

Vector autoregressions have steadily gained in popularity since their introduction in econometrics 25 years ago. A drawback of the otherwise fairly well developed methodology is the inability to incorporate prior beliefs regarding the system's steady state in a satisfactory way. Such prior information are typically readily available and may be crucial for forecasts at long horizons. This paper develops easily implemented numerical simulation algorithms for analyzing stationary and cointegrated VARs in a parametrization where prior beliefs on the steady state may be adequately incorporated. The analysis is illustrated on macroeconomic data for the Euro area.

Keywords: Cointegration, Bayesian inference, Forecasting, Unconditional mean, VARs.
JEL classification: C11, C32, C53, E50.

*The author thanks Malin Adolfson and Michael Andersson for valuable discussions. 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.

†*Sveriges Riksbank, Research Department, SE-103 37 Stockholm, Sweden*, and *Department of Statistics, Stockholm University*. *E-mail*: mattias.villani@riksbank.se.

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 *Minnesota prior* in Litterman (1986), and subsequent modifications and extensions in Doan, Litterman and Sims (1984), Kadiyala and Karlsson (1997), Robertson and Tallman (1999) for reduced form VARs, Sims and Zha (1998) and Waggoner and Zha (2003) for structural VARs, Kleibergen and van Dijk (1994), Kleibergen and Paap (2002), Strachan (2003), Strachan and Inder (2004), and Villani (2005) 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 on the deterministic component of the model. It is well known that 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. A non-informative prior on the deterministic part of the process therefore has the undesirable consequence that the long run forecasts may converge to values which are in gross conflict with true prior opinions. Thus while the existing Bayesian methods allow the user to incorporate prior information on the dynamics of the economy, they do not allow for prior information regarding its steady state. Such information are 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.

One of the reasons for this missing component of the VAR methodology is probably that applications of Bayesian VARs 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 unconditional mean 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 is illustrated in our empirical example in Section 4.

The recent advancements in dynamic stochastic general equilibrium (DSGE) modelling (see *e.g.* the by now well known model in Smets and Wouters (2003)), have renewed the interest in theoretical models for forecasting and policy analysis. A single model cannot cover all variables of interest and other models, *e.g.* VARs, are used as flexible complements which may be formulated and estimated on short notice to answer the policy maker's question at hand. The complementary models typically have at least a few variables in common with the baseline model and are, almost without exception, estimated independently of the baseline model. This raises the concern that the set of entertained models may be internally inconsistent. A particularly disturbing inconsistency arises when two or more models have widely differing steady states for the same variable. Using a prior on the unconditional mean of the VAR process makes it possible to at least have the same prior opinion on the steady state in the baseline theoretical model and the complementary VAR models.

A related reason for using a prior on the unconditional mean of the process comes from the frequent application of Bayesian VARs (BVAR) as benchmark models in forecasting exercises.

A recent example is Del Negro, Schorfheide, Smets and Wouters (2005), where the out-of-sample forecasting performance of the DSGE in Smets and Wouters (2003) is compared to several BVARs. A fairly informative prior on the steady state of the model is incorporated via the prior on the theoretical model's deep parameters. The comparison of forecasting properties would be more balanced if the same prior had been used on the steady state also in the benchmark BVARs.

In this paper we develop a Bayesian analysis of the VAR model in mean-adjusted form. This parametrization of the VAR is expedient 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(1) process by Schotman and van Dijk (1991a,b) in an influential contribution to the lively unit root debate in the early 90's. Lubrano (1995) extended the Schotman-van Dijk approach to the general AR(k) case, using a convenient approximation. 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. It is also shown that this asymptote disappears when the initial conditions of the process are taken into account. See Bauwens *et al.* (1999) for a clear 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 first consider the stationary, or difference stationary, VAR and subsequently move on to the cointegrated VAR. In the case of cointegrated VAR 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 growth rate.

All results in this paper are derived in the reduced form VAR. Our results are easily seen to apply also to structural or identified VARs. The joint posterior distribution of such models may be sampled by adding an updating step to the Gibbs sampler (see Sections 2 and 3) for the contemporaneous coefficients, as described in Waggoner and Zha (2003) and implemented in the cointegrated case by Villani and Warne (2003).

The paper is organized as follows. The next section develops Bayesian inference for stationary VARs in mean-adjusted form. Section 3 extends the analysis to the cointegrated case. The fourth section illustrates the analysis on a seven-variable model of the Euro area. The final section concludes. The proofs are given in the appendices.

2. Bayesian analysis of the stationary VAR process in mean-adjusted form

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

$$\Pi(L)x_t = \Phi d_t + \varepsilon_t, \tag{2.1}$$

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. $\Pi(L) = I_p - \Pi_1 L - ... - \Pi_k L^k$, 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 $\Pi_1, ..., \Pi_k, \Phi$ and Σ. This is often a daunting task for the user of VAR models and the common approach is to model this joint 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 *e.g.* Litterman, 1986) and this part of the prior is usually taken to be non-informative, either in the form of a uniform distribution or a normal distribution with zero mean and very large variances. 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.

Let $E_\tau(x_t)$ denote the conditional expectation of x_t using information up to time τ. In stationary VARs the usual long horizon forecasts approach the unconditional mean of the process, *i.e.* $E_t(x_{t+h}) \to E_0(x_t)$ as $h \to \infty$. This property makes it clear that the implied prior on $E_0(x_t)$ 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. In the parametrization in (2.1), $E_0(x_t)$ is unfortunately a complex non-linear function of $\Pi_1, ..., \Pi_k$ and Φ. Thus, convenient as this parametrization may be from a computational viewpoint, it is not the preferred parametrization for incorporating prior opinions regarding $E_0(x_t)$.

An alternative parametrization of the model in (2.1) is of the form

$$\Pi(L)(x_t - \Psi d_t) = \varepsilon_t. \tag{2.2}$$

This VAR model is non-linear in its parameters, but the unconditional mean of the process is directly specified by Ψ as $E_0(x_t) = \Psi d_t$. The form of the deterministic component Ψd_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 a VAR model on *mean-adjusted form*.

2.2. **Prior distribution.** Bayesian inference requires a prior distribution on $\Sigma, \Pi_1, .., \Pi_k$, and Ψ. The prior on Σ is here taken as

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

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

$$\operatorname{vec}\Pi \sim N_{kp^2}(\theta_\Pi, \Omega_\Pi).$$

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

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

2.3. **Posterior distribution.** The posterior distribution of the mean-adjusted VAR in (2.2) is intractable. As will be shown below, the posterior distribution of each set of model parameters (*i.e.* one of Σ, Π and Ψ) conditional on the remaining parameters is tractable. The numerical method Gibbs sampling (Smith and Roberts, 1993) exploits this fact and generates a sample of parameter draws from the joint posterior by iteratively sampling from the set of full conditional posterior distributions, always conditioning on the most recent draw of the conditioning parameters. Although the Gibbs draws are dependent, it has been shown that it converges in distribution to the target joint posterior distribution (Tierney, 1994). Note that also for the standard VAR in (2.1) must we resort to numerical methods, but in this case the posterior distribution may be sampled by a two-block Gibbs sampler (one block with Σ and the other with Π and Φ).

The next result gives the full conditional posterior distribution of Σ, Π and Ψ. We will make use of the following non-standard, but very convenient, notation. Let z_{1t}, z_{2t}, ..., z_{mt}, $t = 1, ..., T$, be m column vectors of possibly differing lengths, l_i, $i = 1, ..., m$. We shall write $Z = \lfloor z_{1t}, z_{2t}, ..., z_{mt} \rfloor_{t=1}^{T}$ to denote that the tth row of the $T \times (\sum_{i=1}^{m} l_i)$ matrix Z equals $(z'_{1t}, z'_{2t}, ..., z'_{mt})$. The symbol $\lfloor \quad \rfloor_{t=1}^{T}$ 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 $Z = \lfloor z_{1t}, z_{2t}, ..., z_{mt} \rfloor$. Furthermore, let $\mathcal{I}_t = \{x_1, ..., x_t, d_1, ..., d_t\}$ denote the available data at time t.

Proposition 2.1.
• *Full conditional posterior of* Σ

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

where $E = \lfloor \Pi(L)(x_t - \Psi d_t) \rfloor$.
• *Full conditional posterior of* Π

$$\operatorname{vec} \Pi | \Sigma, \Psi, \mathcal{I}_T \sim N_{p^2 k}(\bar{\theta}_\Pi, \bar{\Omega}_\Pi),$$

where $\bar{\Omega}_\Pi^{-1} = \Sigma^{-1} \otimes X'_\Psi X_\Psi + \Omega_\Pi^{-1}$, $\bar{\theta}_\Pi = \bar{\Omega}_\Pi [\operatorname{vec}(X'_\Psi Y_\Psi \Sigma^{-1}) + \Omega_\Pi^{-1} \theta_\Pi]$, $Y_\Psi = \lfloor x_t - \Psi d_t \rfloor$ *and* $X_\Psi = \lfloor x_{t-1} - \Psi d_{t-1}, ..., x_{t-k} - \Psi d_{t-k} \rfloor$.
• *Full conditional posterior of* Ψ

$$\operatorname{vec} \Psi | \Sigma, \Pi, \mathcal{I}_T \sim N_{pq}(\bar{\theta}_\Psi, \bar{\Omega}_\Psi),$$

where $\bar{\Omega}_\Psi^{-1} = U'(D'D \otimes \Sigma^{-1})U + \Omega_\Psi^{-1}$, $\bar{\theta}_\Psi = \bar{\Omega}_\Psi [U' \operatorname{vec}(\Sigma^{-1} Y' D) + \Omega_\Psi^{-1} \theta_\Psi]$, $Y = \lfloor \Pi(L) x_t \rfloor$, $U' = (I_{pq}, I_q \otimes \Pi'_1, ..., I_q \otimes \Pi'_k)$ *and* $D = \lfloor d_t, -d_{t-1}, ..., -d_{t-k+1} \rfloor$.

The 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 (Σ, Ψ)-space. The time-consuming Π-step is thus excluded in the reduced Gibbs sampler.

Contrary to the standard VAR in (2.1), the mean-adjusted VAR in (2.2) is locally non-identified (Rothenberg, 1971). 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 $E_0(x_t)$ does not exist and the parameters in Ψ are non-identified. When a flat prior is used on Ψ, this may cause the Gibbs sampler to converge slowly or perhaps not at all in the following way. Once the Gibbs sampler draws a Π in (or very close to) the non-stationary region, the full conditional posterior of Ψ becomes very spread out and the subsequent Ψ-draw may end up far from the underlying posterior distribution. The next draws of Σ and Π will condition on this abnormal Ψ-draw and the Gibbs sampler may enter into a vicious cycle where one bad draw leads to another. This excursion will continue until a draw happens to end up in the stationary region and the Gibbs sampler gets 'back on track' again.

Note however that *both* of the following two conditions must be satisfied for the Gibbs sampler to fail: i) a vague prior (large variance) is used for Ψ, and ii) the posterior of Π has non-negligible probability mass in the non-stationary region. The first condition will typically not be satisfied since the reason for choosing the more complex mean-adjusted VAR model in the first place is that prior information actually are available on Ψ. To explicitly show how prior information on Ψ stabilizes the Gibbs sampler we use straight forward algebra to

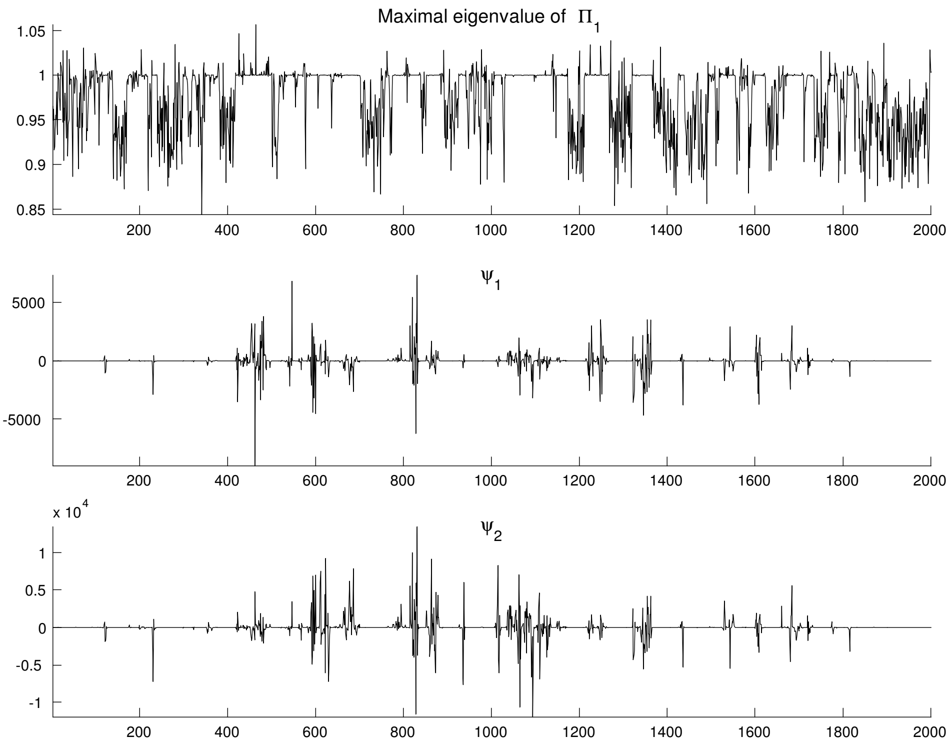

FIGURE 1. Gibbs sampling sequence for the simulated data. Flat prior on the steady state.

rewrite the precision matrix (inverse covariance matrix) in the full conditional posterior of Ψ (assuming for simplicity that $d_t = 1$) as

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

With a flat prior on Ψ it is easily seen that $\bar{\Omega}_{\Psi}$ diverges as we approach the unit root region in Π-space (where $I_p - \sum_{i=1}^{k} \Pi_i$ becomes rank deficient). With an informative prior assigned to Ψ we instead have that $\bar{\Omega}_{\Psi} \to \Omega_{\Psi}$ as the system approaches a unit root.

We will use simulated data to illustrate that a moderately informative prior on Ψ is typically sufficient to prevent the Gibbs sampler from generating 'wild' Ψ's, even if the previous Π-draw happened to end up in the non-stationary region. A bivariate time series of length 100 was generated from the stationary mean-adjusted VAR in (2.2) with $k = 1$, $d_t = 1$ for all t, $\Psi' = (\psi_1, \psi_2) = (1, 4)$, $\Pi_1 = \text{Diag}(0.95, 0.95)$ and $\Sigma = \text{Diag}(0.1, 0.1)$. Note that $E_0(x_t) = (1, 4)'$ and that the data generating process is close to a unit root process. We will display the sequence of Gibbs sampling iterations for three different priors on Ψ: (i) a flat prior, (ii) a mildly informative prior where $\psi_1 \sim N(2.5, 1.25^2)$ independently of $\psi_2 \sim N(5, 2.5^2)$, and (iii) an informative prior where $\psi_1 \sim N(1, 1)$ independently of $\psi_2 \sim N(4, 1)$. A flat prior on Π_1 and the usual non-informative prior $|\Sigma|^{-(p+1)/2}$ is used for Σ in all three cases. Figure 1, 2 and 3 displays 2000 Gibbs iterations for each of the three priors. It is clear from Figure 1 that the Gibbs sampler does not behave well when a flat prior is used for Ψ. The situation improves radically when Ψ is assigned a mildly informative prior (Figure 2), and under the informative prior the mixing of the simulation sequence is excellent (Figure 3).

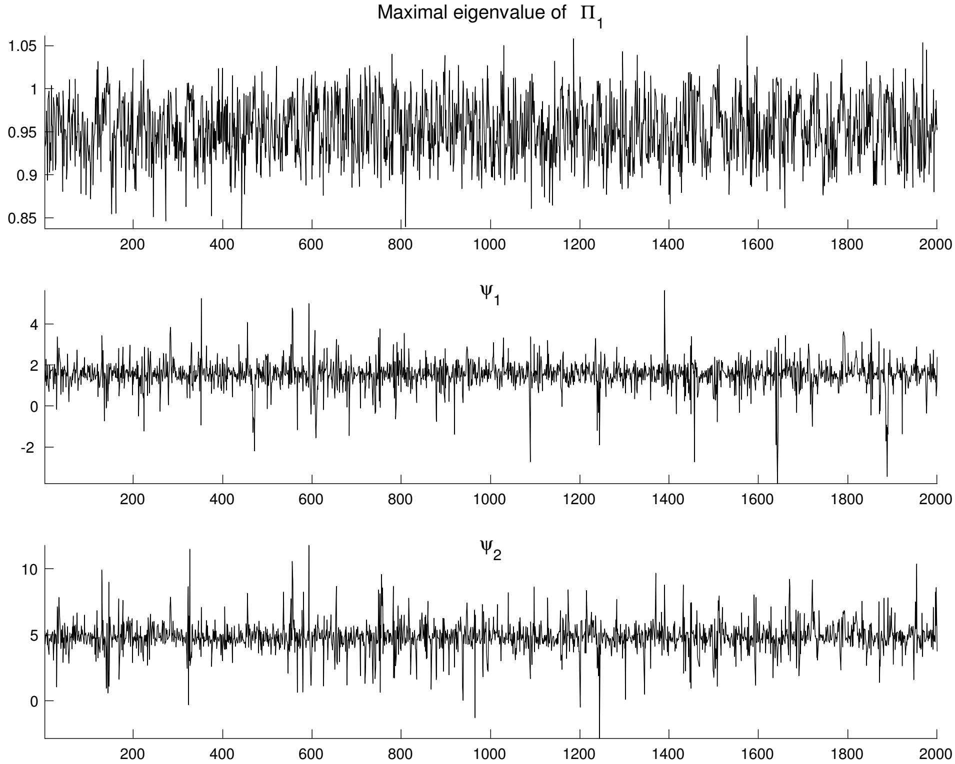

FIGURE 2. Gibbs sampling sequence for the simulated data. Mildly informative prior on the steady state.

It is clear from Proposition 2.1 that there are some additional computing steps in the mean-adjusted VAR compared to the standard form of the model: we need to compute the mean-adjusted data $x_t - \Psi d_t$ following every draw of Ψ, and the U matrix following every draw of $\Pi' = (\Pi_1, ..., \Pi_k)$. One would therefore expect the mean-adjusted VAR to be the more computationally demanding of the two parametrizations. The extra computational steps in the mean-adjusted VAR is only part of the story, however. Regardless of the parametrization, the covariance matrix in the full conditional posterior of Π in Proposition 2.1 consumes the largest fraction of the computing time, especially in larger systems. This covariance matrix is of larger dimension in the standard VAR than in the mean-adjusted model, since the deterministic component in the standard VAR (Φ in (2.1)) is included in Π. In smaller systems with short lag lengths, the additional computational steps in the mean-adjusted VAR dominates the gain in speed from having a smaller dimensional Π. For larger systems with long lags, the opposite holds and the Gibbs sampler for the mean-adjusted form may even be faster than Gibbs sampling in the standard VAR. As an example, the Gibbs sampling for the model analyzed in Section 4 (seven variables and four lags) is slightly faster in mean-adjusted form. It should be noted, however, that the number of iterations of the Gibbs sampler needed to obtain convergence is typically larger in the mean-adjusted model. This is especially true when prior information on the steady state is weak as illustrated in Figures 1-3.

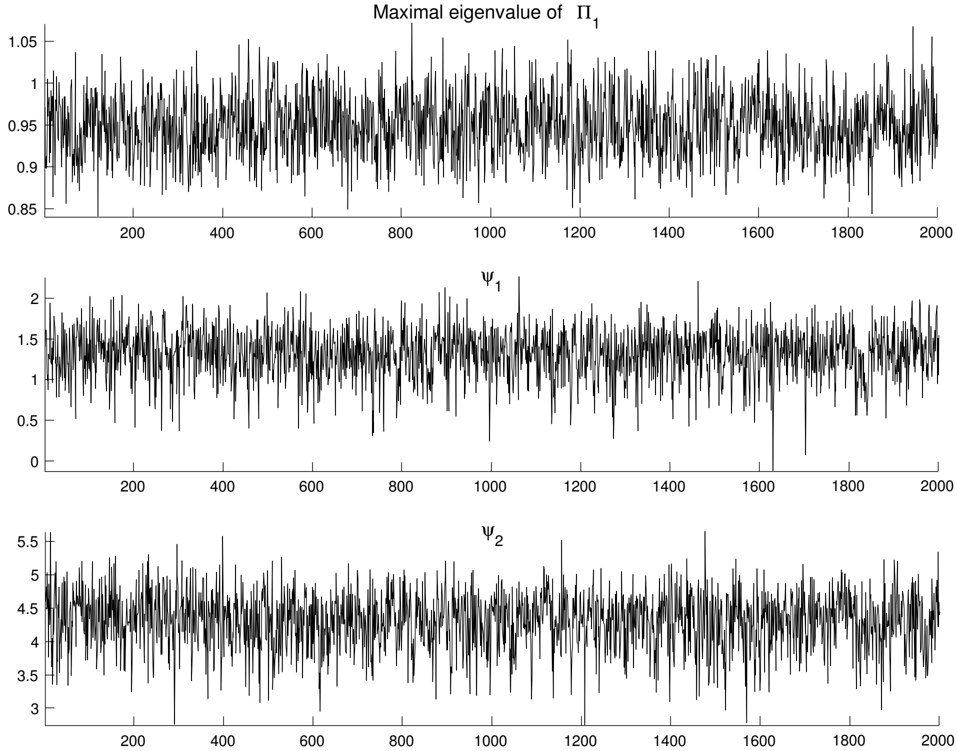

FIGURE 3. Gibbs sampling sequence for the simulated data. Informative prior on the steady state.

3. Bayesian analysis of cointegrated VAR processes in mean-adjusted form

3.1. **The model.** So far we have assumed the process to be 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 mean-adjusted VAR directly on these differenced variables. An intermediate case between stationary and difference stationary processes are the cointegrated processes introduced by Clive Granger (see *e.g.* Engle and Granger, 1987). A fairly general version of the cointegrated VAR process in mean-adjusted form is (Clements and Hendry, 1999)

$$\Gamma(L)(\Delta x_t - \gamma) = \alpha(\beta' x_{t-1} - \mu_0 - \mu_1 t) + \varepsilon_t, \tag{3.1}$$

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, $\Gamma(L) = I_p - \Gamma_1 L - ... - \Gamma_{k-1} L^{k-1}$, 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. The mean-adjusted form of the cointegrated VAR model is non-linear in its parameters, but models the mean of the growth rates explicitly as $E(\Delta x_t) = \gamma$ and the mean of the cointegration relations as $E(\beta' x_{t-1}) = \mu_0 + \mu_1 t$.

The parameters in (3.1) are subject to the restriction $\beta'\gamma = \mu_1$ (Clements and Hendry, 1999, p. 152-153). An explicit parametrization of γ in terms of its unrestricted elements is

$$\gamma = P_\beta \mu_1 + P_{\beta_\perp} \lambda, \tag{3.2}$$

where $P_\beta = \beta(\beta'\beta)^{-1}$, $P_{\beta_\perp} = \beta_\perp(\beta'_\perp\beta_\perp)^{-1}$, μ_1 is $r \times 1$ and λ is $(p-r) \times 1$. Note that with this parametrization $\beta'\gamma = \mu_1$ and $\beta'_\perp\gamma = \lambda$.

We shall first assume that β is completely known and later in this section discuss the case where it is not. One reason for conditioning the analysis on a fixed β is that since $E(\beta' x_{t-1}) = \mu_0 + \mu_1 t$, the interpretation of μ_0 and μ_1 is dependent on β. Conditioning on β therefore makes it substantially more straight forward to elicit prior opinions on μ_0 and μ_1. When β is sufficiently restricted by over-identifying restrictions to allow the cointegration vectors to be interpreted, it should be possible to specify a prior on μ_0 and μ_1, even if β is partly unknown.

3.2. **Prior distribution.** We need to specify a prior on Σ, α, $\Gamma^* = (\Gamma_1, .., \Gamma_{k-1})'$ and $\eta = (\lambda', \mu_0', \mu_1')'$. We assume prior independence between the parameter blocks Σ, α, Γ and η.

We will use the same prior for Σ as in the stationary case. The prior on α is taken to be of the form

$$\operatorname{vec}\alpha|\beta, \Sigma \sim N_{pr}(\theta_\alpha, \Omega_\alpha). \tag{3.3}$$

Common choices for θ_α and Ω_α are $\theta_\alpha = 0$ and $\Omega_\alpha = (\beta' S \beta)^{-1} \otimes \lambda_\alpha \Sigma$, where $\lambda_\alpha > 0$ is a shrinkage factor and S is a positive definite matrix, usually assumed to be diagonal; see Villani (2004) for a motivation of this prior.

We shall also assume that

$$\operatorname{vec}\Gamma^* \sim N_{p^2(k-1)}(\theta_{\Gamma^*}, \Omega_{\Gamma^*}), \tag{3.4}$$

where θ_{Γ^*} and Ω_{Γ^*} are usually parametrized in terms of a few hyperparameters, see *e.g.* the well-known Minnesota specification (Litterman, 1986). It will be convenient to define $\Gamma = (\alpha, \Gamma^*)$, such that $\operatorname{vec}\Gamma \sim N_{p[r+p(k-1)]}(\theta_\Gamma, \Omega_\Gamma)$.

Finally, it is assumed that

$$\eta \sim N_{p+r}(\theta_\eta, \Omega_\eta),$$

a priori. It follows from (3.2) that $E(\gamma) = P_\beta E(\mu_1) + P_{\beta_\perp} E(\lambda)$, which implies $E(\lambda) = \beta'_\perp E(\gamma)$. If we assume prior independence between λ and μ_1 we also have $V(\gamma) = P_\beta V(\mu_1) P'_\beta + P_{\beta_\perp} V(\lambda) P'_{\beta_\perp}$, so that $V(\lambda) = \beta'_\perp V(\gamma) \beta_\perp$. It is thus sufficient to elicit the mean and covariance matrix of the growth rates, γ, to pin down the prior for λ.

3.3. **Posterior distribution.** The next result shows that the full conditional posteriors of the three parameter blocks Σ, Γ and η are all of standard form. A simple three-block Gibbs sampler may thus be set up also in the cointegrated case.

Proposition 3.1.
- *Full conditional posterior of* Σ

$$\Sigma|\Gamma, \eta, \beta, \mathcal{I}_T \sim IW(E'E, T),$$

where $E = \lfloor \Gamma(L)(\Delta x_t - \gamma) - \alpha(\beta' x_{t-1} - \mu_0 - \mu_1 t) \rfloor$.
- *Full conditional posterior of* $\Gamma = (\alpha, \Gamma_1, ..., \Gamma_{k-1})'$

$$\operatorname{vec}\Gamma|\Sigma, \eta, \beta, \mathcal{I}_T \sim N_{p[r+p(k-1)]}(\bar{\theta}_\Gamma, \bar{\Omega}_\Gamma),$$

where $\bar{\Omega}_\Gamma^{-1} = \Sigma^{-1} \otimes X'_\eta X_\eta + \Omega_\Gamma^{-1}$, $\bar{\mu}_\Gamma = \bar{\Omega}_\Gamma[\operatorname{vec}(X'_\eta Y_\eta \Sigma^{-1}) + \Omega_\Gamma^{-1}\theta_\Gamma]$, $Y_\eta = \lfloor \Delta x_t - \gamma \rfloor$ *and* $X_\eta = \lfloor \beta' x_{t-1} - \mu_0 - \mu_1 t, \Delta x_{t-1} - \gamma, ..., \Delta x_{t-k+1} - \gamma \rfloor$.
- *Full conditional posterior of* η

$$\operatorname{vec}\eta|\Sigma, \Gamma, \beta, \mathcal{I}_T \sim N_{p+r}(\bar{\theta}_\eta, \bar{\Omega}_\eta),$$

where $\bar{\Omega}_\eta^{-1} = Q_\Sigma' Q_\Sigma + \Omega_\eta^{-1}$, $Q_\Sigma' = (Q_1'\Sigma^{-1/2}, ..., Q_T'\Sigma^{-1/2})$, $Q_t = [\Gamma(L)P_{\beta_\perp}, -\alpha, \Gamma(L)P_\beta - t\alpha]$, $\bar{\theta}_\eta = \bar{\Omega}_\eta[Q_\Sigma' \operatorname{vec}(\Sigma^{-1/2}Y_\Gamma') + \Omega_\eta^{-1}\theta_\eta]$ *and* $Y_\Gamma = \lfloor \Gamma(L)\Delta x_t - \alpha\beta' x_{t-1} \rfloor$.

When β cannot realistically be assumed to be known, we can augment the Gibbs sampler in Proposition 3.1 with a Metropolis updating step for the unrestricted elements of β (the full conditional posterior of β is non-standard as a result of the restriction $\beta'\gamma = \mu_1$). One possible class of restrictions is the general linear restrictions $\beta = (h_1 + H_1\psi_1, ..., h_r + H_r\psi_r)$ used by Johansen (1995). A multivariate normal distribution centered over the previous draw may be used as proposal distribution with a covariance matrix equal to a constant times the inverse Hessian at the posterior mode (obtained from standard numerical optimization routines).

4. Empirical illustration

We will use quarterly data for the Euro area, first collected by Fagan, Henry and Mestre (2001), over the time period $1970Q1-2002Q4$ to illustrate the methods developed in this paper. The VAR system consists of the seven variables used in the estimation of the influential DSGE model of Smets and Wouters (2003): the domestic inflation rate π_t, growth rates of real wages Δw_t, consumption Δc_t, investment Δi_t, the short-run interest rate r_t, employment e_t and real GDP growth Δy_t. All series except r_t are in logs. The quarter-to-quarter differences used in the estimation are subsequently aggregated up to annual growth figures in the presented results. The employment series was demeaned and detrended before estimation to remove a large upward trend presumably generated by population growth and increased part-time work during the sample period (data on hours worked are not available for the Euro area).

We will assume that the above reported transformation of the original data produces stationary series so that the model in (2.2) can be used. We note that the version of the Smets-Wouter's model used in Del Negro et al. (2005) includes a common trend in the real variables. An alternative would thus have been to analyze the data in levels using the mean-adjusted cointegrated VAR in (3.1) with the six cointegrating relations: $w_t - y_t, c_t - y_t, i_t - y_t, \pi_t, r_t, e_t$.

The DSGE model in Smets and Wouters (2003) features a time varying inflation target to model the changes in monetary policy during the analyzed time period. We will instead use the regime dummy $d_t = (1, d_{MP,t})'$, where $d_{MP,t}$ is a monetary policy dummy

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

The date of the regime shift t^* will be inferred from the data. Figure 4 displays the approximate posterior distribution of the lag length and t^* obtained form Schwarz' first order approximation of the marginal likelihood (Schwarz, 1978). Almost all posterior probability mass is placed on the model with one lag. The Schwarz approximation is well known to favor too small models, however. In addition, the prior used here on the dynamics of the process shrinks longer lags heavily towards zero so the cost of increasing the lag length is much smaller than in a non-Bayesian analysis. This, in combination with the inferior fit of the one lag model (judged visually from the actual-vs-fit graphs and residual autocorrelograms), suggests that a model with more than one lag should be used. For $k > 1$ the posterior probabilities dates the last quarter of the first regime to sometimes in 1992. Given that the data are on a quarterly basis, it seems natural to condition the ensuing analysis on four lags and then use $t^* = 1992Q4$.

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

$$E_0(x_t) = \begin{cases} \psi_1 + \psi_2 & \text{if } t \leq t^* \\ \psi_1 & \text{if } t > t^* \end{cases} ,$$

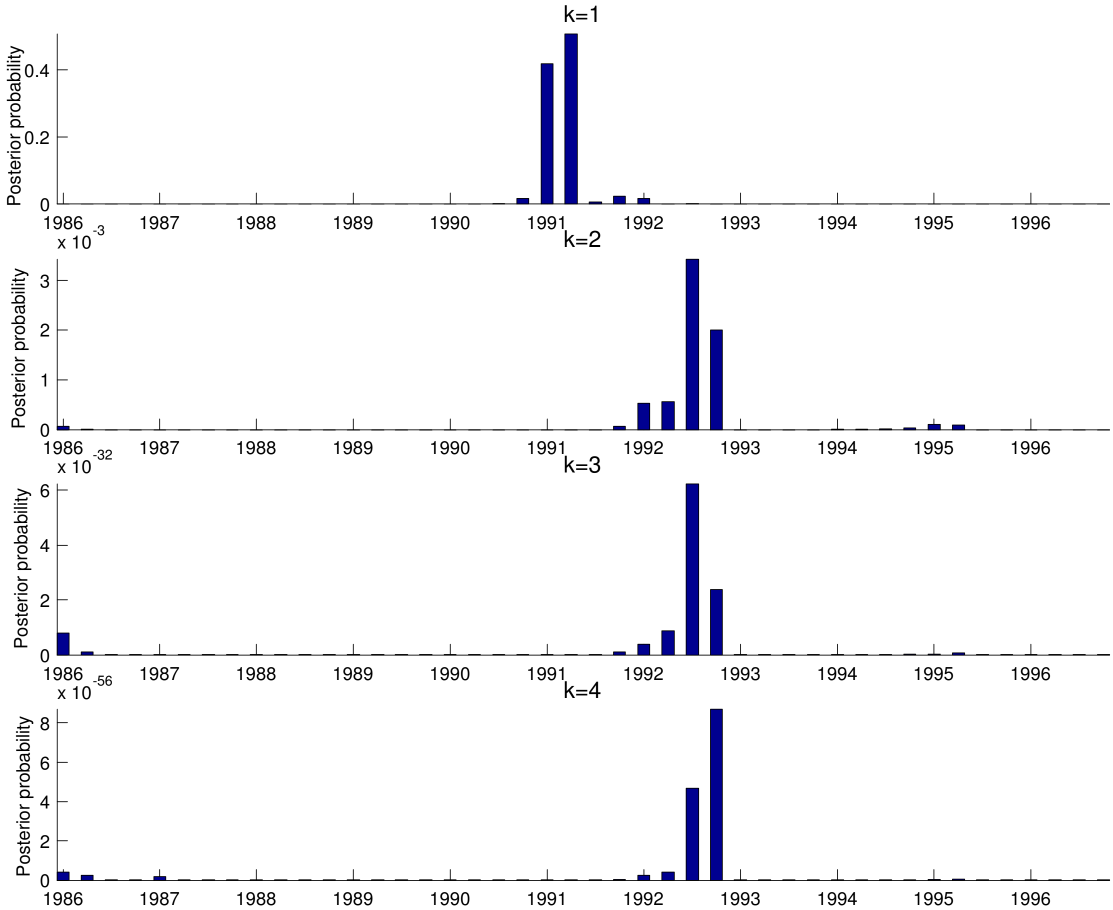

FIGURE 4. Approximate joint posterior distribution (BIC) of the timing of the regime shift (t^*) and the lag length (k) using data from 1980Q1.

	π_t	Δw_t	Δc_t	Δi_t	r_t	e_t	Δy_t
ψ_1	$(1.54, 2.33)$	$(2.02, 2.83)$	$(2.02, 2.83)$	$(2.02, 2.83)$	$(4.93, 6.39)$	$(-10, 10)$	$(2.02, 2.83)$
ψ_2	$(2, 7)$	$(-1, 1)$	$(-1, 1)$	$(-1, 1)$	$(2, 6)$	$(-10, 10)$	$(-1, 1)$

TABLE 1. 95% prior probability intervals of Ψ in the diffuse mean-adjusted model.

where ψ_i is the ith column of Ψ. The prior on ψ_1, which determines the steady state in the latter regime, parallels the implied prior from the deep parameters on the steady state in the DSGE model in Adolfson, Laseén, Lindé and Villani (2005), and is displayed as the first row of Table 1. The prior on ψ_2, which determines the difference in steady states between the first and second regime, is given in the second row of Table 1. Note that the prior on ψ_2 is centered on the event that the regime shift is purely nominal, *i.e.* that the real variables 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. We will also consider a model with the elements in ψ_2 corresponding to the real variables set exactly to zero. We refer to this restricted model as the *mean-adjusted nominal* model and the former model with the prior in Table 1 as the *mean-adjusted diffuse* model. Note also

	π_t	Δw_t	Δc_t	Δi_t	r_t	e_t	Δy_t
Arith. mean	2.179	0.189	1.711	1.641	5.064	0.204	1.967
ML	6.970	2.681	2.435	−1.800	3.923	2.376	2.103
Standard	4.556	2.221	2.380	−0.608	3.374	3.115	2.280
Mean-adj. nominal	2.014	2.160	2.495	2.377	5.455	1.207	2.564
Mean-adj. diffuse	2.012	2.229	2.282	2.399	5.235	1.785	2.516

TABLE 2. Point estimates of the steady state in the subperiod 1993Q1-2002Q4. The model is estimated on data from 1970Q1.

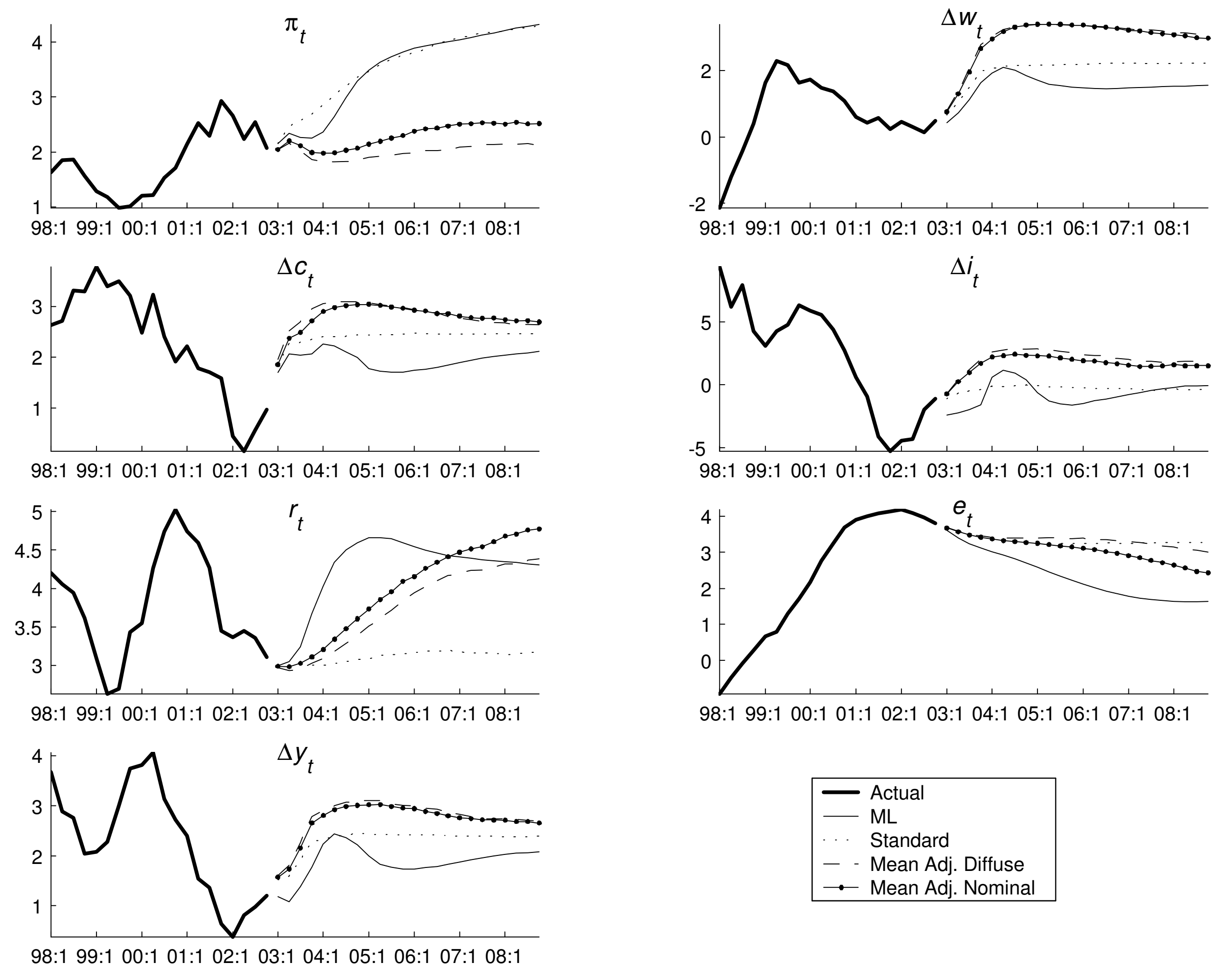

FIGURE 5. Dynamic forecasts using data from 1970Q1. The Bayesian forecasts are the posterior median of the predictive distribution.

that the demeaning and detrending of the employment series makes it problematic to assign a steady state prior to this variable, hence the relatively uninformative prior in Table 1.

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 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

	π_t	Δw_t	Δc_t	Δi_t	r_t	e_t	Δy_t
Arith. mean	2.179	0.189	1.711	1.641	5.064	0.204	1.967
ML	−0.623	0.707	2.227	2.046	3.088	1.848	2.169
Standard	1.387	0.583	1.729	0.373	4.174	2.257	1.831
Mean-adj. nominal	1.873	2.116	2.363	2.409	5.595	1.825	2.409
Mean-adj. diffuse	1.909	2.194	2.282	2.417	5.549	2.137	2.391

TABLE 3. Point estimates of the steady state in the subperiod 1993Q1-2002Q4. The model is estimated on data from 1980Q1.

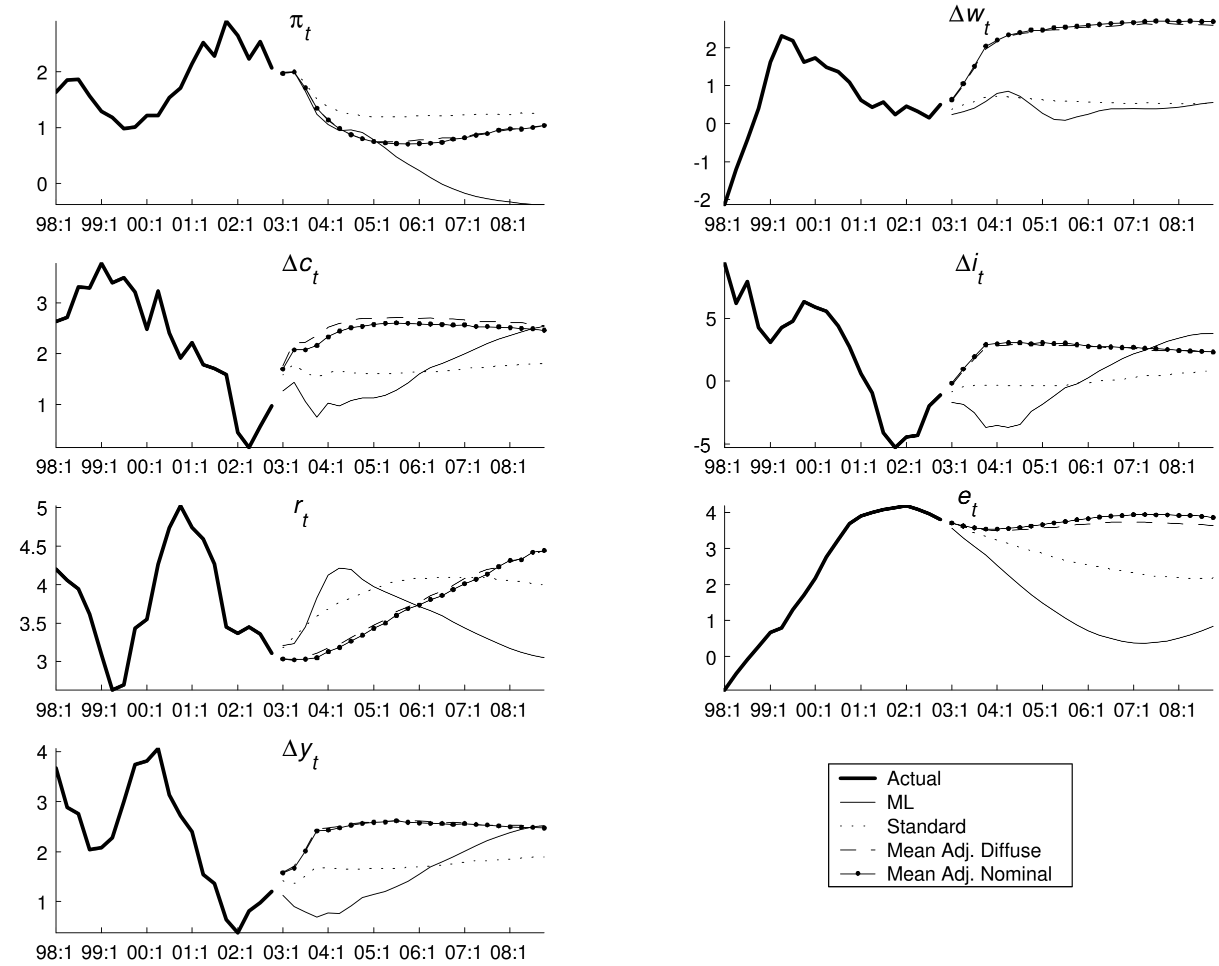

FIGURE 6. Dynamic forecasts using data from 1980Q1. The Bayesian forecasts are the posterior median of the predictive distribution.

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 two remaining level variables r_t and e_t 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 non-informative prior $|\Sigma|^{-(p+1)/2}$ is used for Σ.

The Gibbs sampler in Proposition 2.1 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 24 quarters ahead to form a sample from the predictive distribution. No problems with the convergence of the Gibbs sampler were found. Figure 5 displays the point forecasts from the four different models and Table 2 contains the estimated steady state of the process during the post-break period. We also present results for the standard VAR in (2.1), and a VAR estimated by maximum likelihood (ML) for comparison. Several things are worth noting. First, there are large differences between the ML and standard VAR forecasts, on the one hand, and the two mean-adjusted models on the other. This is especially true for the inflation and interest rate forecasts. The ML and standard VAR estimates of the steady state of inflation in Table 2 are in gross conflict with the current inflation target of the ECB of 'below but close to two percent'. Second, the choice between the nominal and diffuse versions of the prior in the mean-adjusted VAR only matters significantly for the forecasts of inflation and the interest rate. The difference in terms of steady states is small, however. Third, the correspondence between the estimated steady state and the simple arithmetic mean of the observed data is weak for some of the variables (the wage series in particular), even in the Bayesian standard VAR and the ML estimated VAR.

To investigate the robustness of the predictions to changes in the estimation period, we repeated the analysis using only the subsample $1980Q1 - 2002Q4$ for inference. Figure 6 displays the point forecasts from the four different models and Table 3 the estimated post-break steady state. The effect of reducing the estimation sample for the ML and standard VAR predictions is striking, especially for the inflation and interest rate series. The estimated steady state of inflation for the ML and standard VAR in Table 3 is drastically reduced compared to the full sample estimates in Table 2. The estimated steady state inflation from the ML estimates is even negative. The prediction paths from the two mean adjusted models are less affected by changing the sample period. The estimated steady states from these two models are particularly robust to changes in the estimation sample.

5. Concluding remarks

We have developed 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 is often available in relatively strong form. The decision maker will not be pleased to hear that while her prior information may easily be incorporated on the more obscure part of the model, such as the reduced form dynamic coefficients, her 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.

Analysis of regime shifts using VARs with informative priors on the steady state is currently being conducted by the author. Without any special structure on the VAR it is likely, especially in a multivariate framework, that the data point toward a break which may not be of the type we had in mind *a priori*, *e.g.* a change in monetary policy regime with only real variables differing between regimes. Adding this additional structure via an informative prior on the effect of the regime change on the steady state seems to be an attractive option.

Appendix A. Proof of Proposition 2.1

Proof. Conditional on Ψ the model in (2.2) 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 *e.g.* Zellner (1971). To derive the full conditional posterior of Ψ 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 $Y = \lfloor \Pi(L)x_t \rfloor$ and $D = \lfloor d_t, -d_{t-1}, ..., -d_{t-k} \rfloor$ and note that

$$\operatorname{vec}\Theta' \equiv \operatorname{vec}(\Psi, \Pi_1\Psi, ..., \Pi_k\Psi) = \begin{pmatrix} I_{pq} \\ (I_q \otimes \Pi_1) \\ \vdots \\ (I_q \otimes \Pi_k) \end{pmatrix} \operatorname{vec}\Psi = U \operatorname{vec}\Psi.$$

The full conditional likelihood of Ψ can be written

$$\begin{aligned} p(X, D|\Sigma, \Pi, \Psi) &\propto \operatorname{etr}[\Sigma^{-1}(Y - D\Theta)'(Y - D\Theta)] \\ &\propto \operatorname{etr}[(\Theta' - \hat{\Theta}')'\Sigma^{-1}(\Theta' - \hat{\Theta}')D'D], \\ &= \exp\left(-\frac{1}{2}(U\operatorname{vec}\Psi - \operatorname{vec}\hat{\Theta}')'(D'D \otimes \Sigma^{-1})(U\operatorname{vec}\Psi - \operatorname{vec}\hat{\Theta}')\right) \\ &\propto \exp\left(-\frac{1}{2}\left[(\operatorname{vec}\Psi)'A\operatorname{vec}\Psi - (\operatorname{vec}\Psi)'B - B'\operatorname{vec}\Psi\right]\right) \\ &\propto \exp\left(-\frac{1}{2}\left[(\psi - \hat{\psi})'A(\psi - \hat{\psi})\right]\right), \end{aligned}$$

where $\operatorname{etr}(H) = \exp(-\frac{1}{2}\operatorname{tr}H)$ for any quadratic matrix H, $\psi = \operatorname{vec}\Psi$, $A = U'(D'D \otimes \Sigma^{-1})U$, $B = U'(D'D \otimes \Sigma^{-1})\operatorname{vec}\hat{\Theta}'$, $\hat{\Theta} = (D'D)^{-1}D'Y$ and $\hat{\psi} = A^{-1}B = [U'(D'D \otimes \Sigma^{-1})U]^{-1}U'\operatorname{vec}(\Sigma^{-1}Y'D)$. Multiplying the conditional likelihood with the $N_{pq}(\theta_\Psi, \Omega_\Psi)$ prior yields (see e.g. Zellner, 1971)

$$\exp\left(-\frac{1}{2}\left[(\psi - \bar{\theta}_\Psi)'\bar{\Omega}_\Psi^{-1}(\psi - \bar{\theta}_\Psi)\right]\right).$$

where $\bar{\Omega}_\Psi^{-1} = U'(D'D \otimes \Sigma^{-1})U + \Omega_\Psi^{-1}$,

$$\bar{\theta}_\Psi = \bar{\Omega}_\Psi[U'(D'D \otimes \Sigma^{-1})U\hat{\psi} + \Omega_\Psi^{-1}\theta_\Psi] = \bar{\Omega}_\Psi[U'\operatorname{vec}(\Sigma^{-1}Y'D) + \Omega_\Psi^{-1}\theta_\Psi],$$

which proves the result. □

Appendix B. Proof of Proposition 3.1

Proof. Transpose and rearrange (3.1) and stack the resulting row vectors in matrices to obtain the following multivariate regression form of the model

$$Y_\eta = X_\eta\Gamma + E,$$

where $Y_\eta = \lfloor \Delta x_t - \gamma \rfloor$, $X_\eta = \lfloor \beta' x_{t-1} - \mu_0 - \mu_1 t, \Delta x_{t-1} - \gamma, ..., \Delta x_{t-k+1} - \gamma \rfloor$, $E = \lfloor \varepsilon_t \rfloor$ and $\Gamma = (\alpha, \Gamma_1, ..., \Gamma_{k-1})'$. The full conditional posteriors of Γ and Σ now follow from the standard theory of Bayesian multivariate regression, see *e.g.* Zellner (1971).

We will now derive the full conditional posterior of η. Using that $\gamma = P_\beta\mu_1 + P_{\beta_\perp}\lambda$, the model in (3.1) may be rewritten as

$$\begin{aligned} \Gamma(L)\Delta x_t - \alpha\beta' x_{t-1} &= \Gamma(L)(P_\beta\mu_1 + P_{\beta_\perp}\lambda) - \alpha(\mu_0 + \mu_1 t) + \varepsilon_t \\ &= Q_t\eta + \varepsilon_t, \end{aligned} \tag{B.1}$$

where $Q_t = (\Gamma(L)P_{\beta_\perp}, -\alpha, \Gamma(L)P_\beta - t\alpha)$ and $\eta = (\lambda', \mu_0', \mu_1')'$. Transposing (B.1) and stacking the resulting row vectors in matrices yields

$$Y_\Gamma = A + E,$$

where $Y_\Gamma = \lfloor \Gamma(L)\Delta x_t - \alpha\beta' x_{t-1} \rfloor$, $E = \lfloor \varepsilon_t \rfloor$ and $A = \lfloor Q_t \eta \rfloor$. The conditional likelihood function may be written

$$p(X|\Sigma, \Gamma, \eta, \beta, \mathcal{I}_T) \propto \text{etr}[\Sigma^{-1}(Y_\Gamma - A)'(Y_\Gamma - A)] = \text{etr}[(W_\Gamma - C)'(W_\Gamma - C)],$$

where $W_\Gamma = Y_\Gamma \Sigma^{-1/2}$ and $C = A\Sigma^{-1/2} = \lfloor \Sigma^{-1/2} Q_t \eta \rfloor$. Note that

$$\text{vec}(C') = \begin{pmatrix} \Sigma^{-1/2}Q_1\eta \\ \Sigma^{-1/2}Q_2\eta \\ \vdots \\ \Sigma^{-1/2}Q_T\eta \end{pmatrix} = Q_\Sigma \eta, \text{ where } Q_\Sigma = \begin{pmatrix} \Sigma^{-1/2}Q_1 \\ \Sigma^{-1/2}Q_2 \\ \vdots \\ \Sigma^{-1/2}Q_T \end{pmatrix}.$$

.

Thus,

$$\begin{aligned} p(X|\Sigma, \Gamma, \eta, \beta, \mathcal{I}_T) &\propto \text{etr}[(W_\Gamma - C)'(W_\Gamma - C)] \\ &= \exp\left(-\frac{1}{2}[\text{vec}(W_\Gamma') - \text{vec}(C')]'[\text{vec}(W_\Gamma') - \text{vec}(C')]\right) \\ &= \exp\left(-\frac{1}{2}(w - Q_\Sigma\eta)'(w - Q_\Sigma\eta)\right) \\ &\propto \exp\left(-\frac{1}{2}(\eta - \hat{\eta})'Q_\Sigma'Q_\Sigma(\eta - \hat{\eta})\right) \end{aligned}$$

where $w = \text{vec}(W_\Gamma')$ and $\hat{\eta} = (Q_\Sigma' Q_\Sigma)^{-1} Q_\Sigma' w$. Multiplying $p(X|\Sigma, \Gamma, \eta, \beta, \mathcal{I}_T)$ with the $N_{p+r}(\theta_\eta, \Omega_\eta)$ prior yields the full conditional posterior of η (see *e.g.* Zellner, 1971)

$$p(\eta|X, \Sigma, \Gamma, \beta, \mathcal{I}_T) \propto \exp\left(-\frac{1}{2}\left[(\eta - \bar{\theta}_\eta)'\bar{\Omega}_\eta^{-1}(\eta - \bar{\theta}_\eta)\right]\right).$$

where $\bar{\Omega}_\eta^{-1} = Q_\Sigma' Q_\Sigma + \Omega_\eta^{-1}$ and

$$\bar{\theta}_\eta = \bar{\Omega}_\eta[Q_\Sigma' Q_\Sigma \hat{\eta} + \Omega_\eta^{-1}\theta_\eta] = \bar{\Omega}_\eta[Q_\Sigma' w + \Omega_\eta^{-1}\theta_\eta],$$

which proves the result. □

References

Adolfson, M., Laseén, S., Lindé, J. and Villani, M. (2005). The role of sticky prices in an open economy DSGE model: a Bayesian investigation. *Journal of the European Economic Association Paper and Proceedings,* forthcoming.

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

Chib, S. (1995). Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association*, **90**, 1313-21.

Clements, M. P. and Hendry, D. F. (1995). Macro-Economic Forecasting and Modelling, *Economic Journal*, **105**, 1001-1013.

Clements, M. P. and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*, MIT Press, London.

Del Negro, M. Schorfheide, F. Smets, F. and Wouters, R. (2005). On the fit and forecasting performance of new Keynesian models. Manuscript.

Doan, T. A. (1992). RATS User's Manual, Version 4, Evanston, IL.

Doan, T. A., Litterman, R. B. and Sims, C. A. (1984). Forecasting and conditional projection using realistic prior distributions, *Econometrics Reviews*, **5**, 1-100.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation and testing, *Econometrica*, 55, 251–276.

Fagan, G., Henry, J. and Mestre, R. (2001). An Area-Wide Model (AWM) for the Euro Area, Working Paper No. 42. European Central Bank.

Geweke, J. (1999). Using simulation methods for Bayesian econometrics models: inference, development and communication, *Econometric Reviews*, **18** (1), 1-73.

Kadiyala, K. R. and Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR-models, *Journal of Applied Econometrics*, **12**, 99-132.

Kleibergen, F. and Paap, R. (2002). Prior, posteriors and Bayes factors for a Bayesian analysis of cointegration. *Journal of Econometrics* **111**, 223-249.

Kleibergen, F. and van Dijk, H. K. (1994). On the shape of the likelihood/posterior in cointegration models, *Econometric Theory*, **10**, 514-51.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions - Five years of experience, *Journal of Business and Economic Statistics*, **5**, 25-38.

Lubrano, M. (1995). Testing for unit roots in a Bayesian framework, *Journal of Econometrics*, **69**, 81-109.

Robertson, J. C. and Tallman, E. W. (1999). Vector autoregressions: forecasting and reality, Federal Reserve Bank of Atlanta *Economic Review*, **84**, 4-18.

Rothenberg, T. J. (1971). Identification in parametric models, *Econometrica*, **39**, 577-591.

Schotman, P. and van Dijk, H. K. (1991a). A Bayesian analysis of the unit root in real exchange rates, *Journal of Econometrics*, **49**, 195-238.

Schotman, P. and van Dijk, H. K. (1991b). On Bayesian routes to unit roots, *Journal of Applied Econometrics*, **6**, 387-401.

Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**, 461-464.

Sims, C. A. (1980). Macroeconomics and reality, *Econometrica*, **48**, 1–48.

Sims, C. A. and Zha, T. (1998). Bayesian methods for dynamic multivariate models, *International Economic Review*, **39**, 949-968.

Smets, F. and Wouters, R. (2003). An estimated stochastic dynamic general equilibrium model of the Euro area, *Journal of the European Economic Association*, **1**, 1123-1175.

Smith, A. F. M. and Roberts, G. O. (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.

Strachan, R. W. (2003). Valid Bayesian estimation of the cointegrating error correction model. *Journal of Business and Economic Statistics* **21**, 185-195.

Strachan, R. W. and Inder, B. (2004). Bayesian analysis of the error correction model, *Journal of Econometrics*, **123**, 307-325.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion), *Annals of Statistics*, **22**, 1701-1762.

Villani, M. (2005). Bayesian reference analysis of cointegration, *Econometric Theory*, **21**, 326-357.

Villani, M. and Warne, A. (2003). Monetary policy analysis in a small open economy using Bayesian cointegrated structural VARs, ECB Working Paper No. 296.

Waggoner, D. F. and Zha, T. (2003). 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.

Earlier Working Papers:

For a complete list of Working Papers published by Sveriges Riksbank, see www.riksbank.se

Sveriges Riksbank
Visiting address: Brunkebergs torg 11
Mail address: se-103 37 Stockholm

Website: www.riksbank.se
Telephone: +46 8 787 00 00, Fax: +46 8 21 05 31
E-mail: registratorn@riksbank.se

ISSN 1402-9103 Riksbank Printing Office Stockholm 2005