DISCUSSION PAPER SERIES

No. 2961

FORECASTING AND TURNING-POINT PREDICTIONS IN A BAYESIAN PANEL VAR MODEL

Fabio Canova and Matteo Ciccarelli

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP2961.asp**

ISSN 0265-8003

FORECASTING AND TURNING-POINT PREDICTIONS IN A BAYESIAN PANEL VAR MODEL

Fabio Canova, Universitat Pompeu Fabra, Barcelona and CEPR
Matteo Ciccarelli, Universitat d'Alacant

Discussion Paper No. 2961
September 2001

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 2961

September 2001

ABSTRACT

Forecasting and Turning-Point Predictions in a Bayesian Panel VAR Model*

We provide methods for forecasting variables and predicting turning-points in panel Bayesian VARs. We specify a flexible model that accounts for both interdependencies in the cross-section and time variations in the parameters. Posterior distributions for the parameters are obtained for hierarchical and for Minnesota-type priors. Formulas for multistep, multiunit point and average forecasts are provided. An application to the problem of forecasting the growth rate of output and of predicting turning points in the G-7 illustrates the approach. A comparison with alternative forecasting methods is also provided.

JEL Classification: C11, C15, E32 and E37
Keywords: Bayesian methods, panel VAR, Markov chains, Monte Carlo methods, forecasting and turning points

Fabio Canova
Department of Economics
Universitat Pompeu Fabra
Ramon Trias Fargas, 25-27
08005 Barcelona
SPAIN
Tel: (34 93) 542 2659
Fax: (34 93) 542 1746
Email: canova@upf.es

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=114830

Matteo Ciccarelli
Fundamentos del Análisis Económico
Universitat d'Alacant
Ap. Correos 99
03080 Alicante
SPAIN
Email: matteo@merlin.fae.ua.es

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=153879

* We would like to thank two anonymous referees, Albert Marcet and the participants of seminars at Universitat Pompeu Fabra, Universitat d'Alacant, Universidade do Minho, the European Central Bank, the Bundesbank, the European Meetings of the Econometric Society, the ASSET meetings, the International Symposia in forecasting and the CEPR/BDI conference 'Monitoring the Euro Area Business Cycles' for comments and suggestions. This Paper is produced as part of a CEPR Research Network on New Approaches to the Study of Economic Fluctuations, funded by the European Commission under the Training and Mobility of Researchers Programme (Contract No. ERBFMRX-CT98-0213)

Submitted 13 August 2001

1 Introduction

Panel VAR models have become increasingly popular in macroeconomics to study the transmission of shocks across countries (Ballabriga, Sebastian and Valles (1998)), the propagation effects of monetary policy in the European Union (Gerlach and Smets (1996)) and the average differential response of developed and underdeveloped countries to domestic and external disturbances (Hoffmaister and Roldós (1997), Rebucci (1998)). At the same time, recent developments in computer technology have permitted the estimation of increasingly complex multicountry VAR models in reasonable time, making them potentially usable for a variety of forecasting and policy purposes.

Despite this interest, the theory for panel VAR is somewhat underdeveloped. After the works of Chamberlain (1982, 1984), Holtz–Eakin et al. (1988), and Binder et al. (2000), who specify panel VAR models for micro data, to the best of our knowledge only Pesaran and Smith (1996), Canova and Marcet (1997) and Hsiao et al. (1998) have considered problems connected with the specification and the estimation of (univariate) dynamic macro panels. Garcia Ferrer et al. (1987), Zellner and Hong (1989), Zellner, Hong and Min (1991), on the other hand, have provided Bayesian shrinkage estimators and predictors for similar models. In general, a researcher focuses on the specification

$$y_{it} = A(L)\, y_{it-1} + \varepsilon_{it} \tag{1}$$

where y_{it} is a $G \times 1$ vector, $i = 1, \ldots, N$; $A(L)$ is a matrix in the lag operator; $\varepsilon_{it} = \alpha_i + \delta_t + u_{it}$, where δ_t is a time effect; α_i is a unit specific effect and u_{it} a disturbance term. In some cases a specification with time varying slope coefficients and a fixed effect is used (see e.g. Holtz–Eakin et al. (1988)). Two main restrictions characterize this specification. First, it assumes common slope coefficients. Second, it does not allow for interdependencies across units. With these restrictions, the interest is typically in estimating the *average* dynamics in response to shocks (the matrix $A(L)$).

Garcia Ferrer et al., Canova and Marcet and Pesaran and Smith, instead, use a univariate dynamic model of the form

$$y_{it} = \alpha_i + \rho_i y_{it-1} + x'_{it}\beta_i + v'_t\delta_i + \varepsilon_{it} \tag{2}$$

where y_{it} is a scalar, x_{it} is a set of k exogenous unit specific regressors, v_t is a set of h exogenous regressors common to all units while ρ_i, β_i and δ_i are unit specific vectors of coefficients. Two restrictions are implicit also in this specification. First, no time variation is allowed in the parameters. Second, there are no interdependencies either among different variables within units or among the same variable across units.

In this paper we relax these restrictions and study the issues of specification, estimation and forecasting in a macro-panel VAR model. Our point of view is Bayesian. Such an approach has been widely used in the VAR literature since the works of Doan, Litterman and Sims (1984), Litterman (1986), and Sims and Zha (1998) and provides a convenient framework where one can allow for both interdependencies and meaningful time variations in the coefficients. The specification we consider has the general form

$$y_{it} = A_{it}(L)\, Y_{t-1} + \varepsilon_{it} \tag{3}$$

where Y_s $(s < t)$ is a $GN \times 1$ vector (with G variables for each unit $i = 1, \ldots N$). Because coefficients vary across units and along time, estimation of the parameters is impossible without imposing restrictions. Instead of constraining the coefficients to be the same across units, we assume that they are random and a prior distribution on $A_{it}(L)$ is introduced. We decompose the parameter vector into two components, one which is unit specific and one which is time specific. We specify a flexible prior on these two components which parsimoniously accounts for interdependencies in the cross section and for time variations in the evolution of the parameters. The prior shares features with those of Lindlay and Smith (1972), Doan, Litterman and Sims (1984) and Hsiao et al. (1998) and has a hierarchical structure, which allows for various degrees of ignorance in the researcher's information about the parameters.

Bayesian VARs are known to produce better forecasts than unrestricted VAR and, in many situations, ARIMA or structural models (Canova (1995) for references). By allowing interdependencies and some degree of information pooling across units we introduce an additional level of flexibility which may improve the forecasting ability of these models.

We describe in detail two situations of interest: one with fully hierarchical priors and one Minnesota-type priors. In the former case, a Markov Chain Monte Carlo method (the Gibbs sampler) is employed to calculate posterior distributions. Such an approach is useful in our setup since it exploits the recursive features of the posterior distribution. For the Minnesota-type prior, unknown parameters are estimated using the predictive density and posterior estimates are obtained by plugging in our estimates in the relevant formulas in an empirical Bayes fashion. We provide recursive formulas for multistep, multiunit forecasts, consistent with the information available at each point in time using the posterior of the parameters or the predictive density of future observations. The latter is also used to compute turning point probabilities.

To illustrate the features of proposed approach, we apply the methodology to the problem of predicting output growth and of forecasting turning points in output growth in the G-7 and computing the probability of a recession in the US. To evaluate the forecasting performance we provide an extensive comparison with other specifications suggested in the literature.

Our results indicate that our approach improves the forecasting performance of existing univariate and simple BVAR models, both at the one and at the four steps horizons. The improvements are of the order of 5-15% when the Theil-U is used and about 2-8% when the MAD is used. The forecasting performance of our specification is also preferable to the one of a BVAR model which mechanically extends the Litterman prior to the panel case. In terms of turning point predictions, the three versions of the panel approach we consider are able to recognize about 80% of turning points and they turn out to be the best for this task, along with Zellner's g-prior shrinkage approach. The simple extension of the Litterman's prior to the panel case does poorly along this dimension and it is the second worst among all the procedures employed, Finally, we show that the method is competitive with the best specifications in predicting the downturn in US economic activity occurred in 1990:3 when using the information available in 1988:4, a turning point which was missed by many commercial and government forecasting agencies. Depending on the specification, our approach suggest that downturn at that date occurs with 30-57% probability.

The rest of the paper is organized as follows. The next section gives the general model specification and the assumptions made. Section 3 provides general formulas for the posterior. Section 4 sets up the prior and discusses the computational issues involved. Section 5 describes formulas for multi-step, multi-units forecasts. Section 6 contains the application. Section 7 concludes.

2 The general specification

The statistical (reduced form) model we use is of the form:

$$y_{it} = \sum_{j=1}^{N} \sum_{l=1}^{p} b_{it,l}^{j} y_{jt-l} + d_{it} v_t + u_{it} \tag{4}$$

where $i = 1, ..., N$; $t = 1, ..., T$; y_{it} is a $G \times 1$ vector for each i, $b_{it,l}^{j}$ are $G \times G$ matrices, d_{it} is $G \times q$, v_t is a $q \times 1$ vector of exogenous variables, common to all units, and u_{it} is a $G \times 1$ vector of random disturbances. Here p is the number of lags, G the number of endogenous variables and q the number of exogenous variables (including a constant).

The generality of (4) comes from at least two features. First, the coefficients are allowed to vary both across units and across time. Second, there are interdependencies among units whenever $b_{it,l}^{j} \neq 0$ for $j \neq i$ and for any l. Both features constitute the main difference with the literature

(Holtz-Eakin at al. (1988), Rebucci (1998)) that considers panel VAR models. It is easy to verify that if $d_{it}v_t = a_t, \quad b_{it} = b_t \ \forall\, i, \ u_{it} = \psi_t f_i + \xi_{it} \ b^j_{it,l} = 0, \quad j \neq i, \quad \forall\, l$, our specification collapses to the one used by Holtz-Eakin et al. (1988). We rewrite (4) in a stacked regression manner

$$Y_t = W_t \gamma_t + U_t \tag{5}$$

where $W_t = I_{NG} \otimes X_t'$; $X_t = \left(y_{t-1}',\ y_{t-2}', \cdots y_{t-p}',\ v_t'\right)'$; $\gamma_t = (\gamma_{1t}', \ldots, \gamma_{Nt}')'$ and $\gamma_{it} = (\beta_{it}^{1\prime}, \ldots, \beta_{it}^{G\prime})'$. Here y_s $(s < t)$ is a $NG \times 1$vector, β_{it}^g are $k \times 1$ vectors, $k = NGp + q$, containing, stacked, the g rows of the coefficient matrices b_{it} and d_{it}, while Y_t and U_t are $NG \times 1$ vectors containing the endogenous variables and the random disturbances of the model.

If γ_{it} are different for each cross–sectional unit in different time periods, there is no way to obtain meaningful estimates of them using classical methods. One possibility is to view each coefficient vector as random with a given probability distribution. We make the following assumptions:

1. For each i, the $Gk \times 1$ vector γ_{it} has a time invariant and a time varying component:

$$\gamma_{it} = \alpha_i + \lambda_{it} \tag{6}$$

2. For each i, the $Gk \times 1$ vector α_i is normally distributed

$$\alpha_i \sim N\left(R_i \bar{\alpha}, \Delta_i\right) \tag{7}$$

where $R_i = I_G \otimes E_i$, $\Delta_i = V \otimes E_i \Omega_1 E_i$, the $G \times G$ matrix V and the $k \times k$ matrix Ω_1 are symmetric and positive semidefinite and E_i is a $k \times k$ matrix that commutes the k coefficients of unit i for each of the G equations with those of unit one. We assume $cov\left(\alpha_i, \alpha_j\right) = 0\ \forall i \neq j$.

3. The mean vector $\bar{\alpha}$ is assumed to have a normal distribution

$$\bar{\alpha} \sim N\left(\mu, \Psi\right) \tag{8}$$

4. For each i, $\lambda_{it} = R_i \lambda_t$, with λ_t independent of $\alpha_i\,\forall i$. The $Gk \times 1$ vector λ_t evolves according to

$$\lambda_t = B\lambda_{t-1} + (I - B)\lambda_0 + e_t \tag{9}$$

where $B = \rho * I_{Gk}$ and, conditional on U_t and W_t, $e_t \sim N(0, V \otimes \Omega_{2t})$, $\Omega_{2t} = \nu_1 \Omega_{2t-1} + \nu_2 \Omega_{20}$ and Ω_{20} is a positive semidefinite, symmetric matrix. The initial conditions are such that $\lambda_0 \sim N\left(\tilde{\lambda}_0, V \otimes \Omega_{20}\right)$.

5. Conditional on W_t, the vector of random disturbances U_t has a normal distribution

$$U_t \sim N(0, \Sigma_u). \tag{10}$$

where $\Sigma_u = \Sigma \otimes H$, Σ is $N \times N$ and H is $G \times G$, both positive definite and symmetric matrices.

Given the previous assumptions, the structure of the model can be summarized with the following *a–priori* hierarchical scheme

$$\begin{aligned} Y_t &\mid F_t, \alpha, \lambda_t \sim N(W_t\alpha + Z_t\lambda_t,\ \Sigma_u) \\ \alpha &\mid F_t \sim N(S_N\bar{\alpha},\ \Delta) \\ \bar{\alpha} &\mid F_t \sim N(\mu,\ \Psi) \\ \lambda_t &\mid F_t \sim N\left(\hat{\lambda}_{t|\,t-1}, \hat{\Omega}_{t|\,t-1}\right) \end{aligned} \tag{11}$$

where F_t is the information set at t (which includes Y_0, the presample information, and W_t); $S_N = diag\{R_i\}$; $Z_t = W_tS_N$; $\Delta = diag\,(\Delta_1, .., \Delta_n)$, $\hat{\lambda}_{t|\,t-1} = B\hat{\lambda}_{t-1|\,t-1} + (I-B)\hat{\lambda}_0$; $\hat{\Omega}_{t|\,t-1} = B\hat{\Omega}_{t-1|\,t-1}B' + (I-B)\,(V \otimes \Omega_{20})\,(I-B)' + V \otimes \Omega_{2,t|t-1}$and the notation $t|t-1$ indicates values at t predicted with information at $t-1$.

Assumptions 1-4 decompose the parameters vector for each i in 2 components: one is unit specific and constant over time; the other is common across units but varies with time. The prior possibility for time–variation increases the flexibility of the specification and provides a general mechanism to account for structural shifts without explicitly modelling the source of the shift. The fact that the time–varying parameter vector is common across units does not prevent unit–specific structural shifts, since γ_{it} can be re-written as

$$\gamma_{it} = (I-B)(\alpha_i + \lambda_{i0}) + (I-B)\gamma_{it-1} + e_{it} \tag{12}$$

where unit specific variations of time occur through the common coefficient vector B.

Assumptions 2 and 3 can be used to recover the vector α or the mean coefficient vector $\bar{\alpha}$. In this sense, we can distinguish between "fixed" and "random" effects, following Lindley and Smith (1972). By fixed effects we mean the estimation of the vector γ_{it}, while the term random effects refers to the estimation of $\bar{\gamma}_t = \bar{\alpha} + \lambda_t$. For example, in a VAR without interdependencies, (i.e. $b^j_{it,l} = 0\,,\quad j \neq i$), we may be more interested in the relationships among the variables of the system for a "typical" unit, in which case interest centers in the estimation of the random effect $\bar{\gamma}_t$. If, instead, we are interested in the relationships across units, for example, wishing to find the effect of a shock in the g variable of unit j on the variables of unit i, we better estimate γ_{it} for each i. In the context of forecasting, we may be concerned with point prediction using the average vector $\bar{\gamma}_t$ or in predicting future values of the variables of interest using information available for each unit. Assumption 2 allows for some degree of a-priori pooling of cross sectional information via an exchangeable prior on α_i. This may be useful in a panel when there are similarities in the characteristics of the vector of variables across units. In this case coefficients of other units may contain useful information for estimating the coefficients of unit i.

The structure underlying assumption 4 is similar to Canova (1993). There it is shown that (9) allows for nonlinearities in the moment structure (of both ARCH-M and Markov switching type) and nonnormalities in the time series under consideration. Note that the common component evolves over time with an heteroskedastic structure (for homoskedastic variations set $\nu_1 = 0$). Besides being useful to directly capture generic volatility clustering which are common across countries, time variations in the variance allow the model to quickly adapt when outliers or regime switches of short mean length are present.

The assumed Kronecker structure for the variance–covariance matrices is computationally convenient and allows us to nest interesting hypothesis. For instance, when $\Omega_1 = 0$, there is no heterogeneity in the cross sectional dimension of the panel. If $B = I_{Gk}$, coefficients evolve over time as a random walk, while when $B = I_{Gk}$ and $\nu_1 = \nu_2 = 0$, the model reduces to a standard dynamic panel model with no time–variation in the coefficient vector. Moreover, when $V = 0$ neither heterogeneity nor time variation are present in the model. Finally, a single country VAR with fixed coefficients can be obtained by setting $b^j_{it,l} = 0, \forall j \neq i, \forall l$ and letting Ω_1, Ψ, Σ_t go to zero. Note that, with the Kronecker structure, the prior is fully symmetric in the sense that it is the same regardless of the variables and of the units we are considering. There are both notational and computation advantages in setting the prior this way. However when the N units display scale and transmission differences, one may want to relax this assumption and let,e.g. $\Delta_i = V \otimes \Omega_{i1}$ while leaving the prior distributions for $\bar{\alpha}$ and λ_t are unchanged.

Stock and Watson (1999), Forni et al. (1999) and others have examined macro panel models where either N or G or both are large. Their approach is to setup the problem so that it can be handled in the context of (dynamic) index models with classical methods. From (11) one can see that also our specification has an index structure, where the two indices we consider are a "common" one and a "time specific" one. Two major features differentiate our approach from their: first the coefficients on our indices are allowed to very over time. Second, their inferential methods require asymptotic approximations, while our approach deliver exact estimates even when N or G are small.

3 Posterior Estimates

3.1 Fixed effects model

From (11) the likelihood function is

$$L(Y_t \mid \gamma_t,\ F_t) = N(W_t\alpha + Z_t\lambda_t,\ \Sigma_u)$$

and the prior, given information at to t, is

$$p(\gamma_t \mid F_t) = N\left(\hat{\gamma}_{t-1},\ \hat{H}_{t-1}\right) \tag{13}$$

where $\hat{\gamma}_{t-1} = S_N\left(\mu + \hat{\lambda}_{t|\,t-1}\right)$ and $\hat{H}_{t-1} = (S_N\Psi S_N' + \Delta) + S_N\hat{\Omega}_{t|\,t-1}S_N'$. Standard calculations give us that the posterior $\pi_0(\gamma_t \mid F_t,\ Y_t)$ is normal with mean γ_t^* and variance H_t^* where:

$$\gamma_t^* = H_t^*\left(W_t'\Sigma_u^{-1}Y_t + \hat{H}_{t-1}^{-1}\hat{\gamma}_{t-1}\right) = \hat{\gamma}_{t-1} + \hat{H}_{t-1}W_t'\left[W_t\hat{H}_{t-1}W_t' + \Sigma_u\right]^{-1}(Y_t - W_t\hat{\gamma}_{t-1}) \tag{14}$$

$$H_t^* = \left[\hat{H}_{t-1}^{-1} + W_t'\Sigma_u^{-1}W_t\right]^{-1} = \hat{H}_{t-1} - \hat{H}_{t-1}W_t'\left[W_t\hat{H}_{t-1}W_t' + \Sigma_u\right]^{-1}W_t\hat{H}_{t-1} \tag{15}$$

where in the second expression, posterior estimates are obtained recursively, given $\hat{\gamma}_0$ and $\hat{H}_0$.

In some cases one may want to obtain posterior distributions of α and λ_t separately. It is straightforward to show that:

$$\begin{pmatrix} \alpha \\ Y_t \end{pmatrix} |F_t \sim N\left[\begin{pmatrix} S_N\mu \\ Z_t\left(\mu + \hat{\lambda}_{t|\,t-1}\right) \end{pmatrix}, \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}\right]$$

where $\phi_{11} = (S_N\Psi S_N' + \Delta)$; $\phi_{12} = \phi_{11}W_t'$; $\phi_{21} = W_t\phi_{11}$; $\phi_{22} = W_t\phi_{11}W_t' + Z_t\hat{\Omega}_{t|\,t-1}Z_t' + \Sigma_u$.

Using the properties of multivariate normal distributions, the conditional marginal $\pi_1(\alpha \mid F_t,\ Y_t)$ is normal with mean $\alpha^* = S_N\mu + \phi_{12}\phi_{22}^{-1}\left[Y_t - Z_t\left(\mu + \hat{\lambda}_{t|\,t-1}\right)\right]$ and variance $V_\alpha^* = \phi_{11} - \phi_{12}\phi_{22}^{-1}\phi_{21}$ and the conditional marginal $\pi_2(\lambda_t \mid Y_t,\ F_t)$ is normal with mean $\lambda_t^* = \hat{\lambda}_{t|\,t-1} + \hat{\Omega}_{t|\,t-1}Z_t'\phi_{22}^{-1}\left[Y_t - Z_t\left(\mu + \hat{\lambda}_{t|\,t-1}\right)\right]$ and variance $\Omega_t^* = \hat{\Omega}_{t|\,t-1} - \hat{\Omega}_{t|\,t-1}Z_t'\phi_{22}^{-1}Z_t\hat{\Omega}_{t|\,t-1}$.

3.2 Random effects model

When interest centers on the estimation of the mean vector $\bar{\gamma}_t$, we rewrite the model as

$$Y_t = Z_t\bar{\gamma}_t + \eta_t \tag{16}$$

where $\bar{\gamma}_t = \bar{\alpha} + \lambda_t$ and $\eta_t = u_t + W_t v$. Standard manipulations give us that the posterior $\pi_3(\bar{\alpha} \mid Y_t,\ F_t) \sim N(\bar{\alpha}^*, \Psi^*)$ where

$$\bar{\alpha}^* = \mu - \Psi Z_t' \left[Z_t \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) Z_t' + \Sigma_u + W_t \Delta W_t' \right]^{-1} \left[Y_t - Z_t \left(\mu + \hat{\lambda}_{t|\,t-1} \right) \right] \tag{17}$$

$$\Psi^* = \Psi - \Psi Z_t' \left[Z_t \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) Z_t' + \Sigma_u + W_t \Delta W_t' \right]^{-1} Z_t \Psi \tag{18}$$

while the posterior $\pi_2(\lambda_t \mid Y_t,\ F_t) \sim N(\lambda_t^*, \Omega_t^*)$ and λ_t^* and Ω_t^* are the same as before. Hence the posterior $\pi_4(\bar{\gamma}_t \mid Y_t,\ F_t) \sim N(\bar{\gamma}_t^*, H_t^*)$ where

$$\begin{aligned} \bar{\gamma}_t^* &= \left(\mu + \hat{\lambda}_{t|\,t-1} \right) + \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) Z_t' \left[Z_t \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) Z_t' + \Sigma_u + W_t \Delta W_t' \right]^{-1} \\ &\quad \times \left[Y_t - Z_t \left(\mu + \hat{\lambda}_{t|\,t-1} \right) \right] \end{aligned} \tag{19}$$

$$\begin{aligned} H_t^* &= \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) - \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) Z_t' \left[Z_t \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) Z_t' + \Sigma_u + W_t \Delta W_t' \right]^{-1} \\ &\quad \times Z_t \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) \end{aligned} \tag{20}$$

4 Setting up the priors

For the formulas derived in the previous section are operational, only if the vector $\zeta = (vec(\mu), vec(\tilde{\lambda}_o), vech(\Sigma_u), vech(\Omega_{20}), vech(B), vech(\Psi), vech(\Delta))$ is known. When this is the case to obtain marginal posteriors we need to integrate nuisance parameters out of the joint posterior density. This integration, in general, is difficult, even with brute force numerical methods, given the large number of parameters contained in ζ. In this section we describe two approaches which make the computation of the posterior feasible.

4.1 Informative priors

When the prior for the parameter vector is informative, the posterior distribution does not have an analytical closed form. Nevertheless, we can implement a hierarchical Bayes analysis using a sampling–based approach, such as the Gibbs sampler, (see e.g. Geman and Geman (1984), Gelfand and Smith (1990), Gelfand and al. (1990) among others).

The basic idea of the approach is to construct a (computable) Markov chain on a general state space such that the limiting distribution of the chain is the joint posterior of interest. Suppose we have a parameter vector ϑ with k components $(\vartheta_1, \vartheta_2, ..., \vartheta_k)$ and that the posterior distributions $\pi(\vartheta_j \mid \vartheta_s,\ s \neq j)$ are available. Then the algorithm works as follows. We start from arbitrary values for $\vartheta_1^{(o)}, \vartheta_2^{(o)}, ..., \vartheta_k^{(o)}$. Setting $i = 1$, we cycle through the conditional distributions sampling $\vartheta_1^{(1)}$ from $\pi\left(\vartheta_1 \mid \vartheta_2^{(o)}, ..., \vartheta_k^{(o)}\right)$ $\vartheta_2^{(i)}$ from $\pi\left(\vartheta_2 \mid \vartheta_1^{(1)}, ..., \vartheta_k^{(o)}\right)$ up to $\vartheta_k^{(i)}$ from $\pi\left(\vartheta_k \mid \vartheta_1^{(1)}, ..., \vartheta_{k-1}^{(1)}\right)$. Next, we set $i = i + 1$ and repeat the cycle. After iterating on this cycle, say, M times, the sample value $\vartheta^{(M)} = \left(\vartheta_1^{(M)}, \vartheta_2^{(M)}, ..., \vartheta_k^{(M)}\right)$ can be regarded as a drawing from the true joint posterior density. Once this simulated sample has been obtained, any posterior moment of interest or marginal density can be estimated, using the ergodic theorem. Convergence to the desired distribution can be checked as suggested in Gelfand and Smith (1990).

In order to apply the Gibbs sampler to our panel VAR model we need to specify prior information so that the conditional posterior distribution for components of the parameter vector can be obtained analytically. Recall that our hierarchical model is given by:

$$
\begin{aligned}
Y_t &= W_t\alpha + Z_t\lambda_t + u_t, \\
\alpha_i &= R_i\bar{\alpha} + \varepsilon_i \\
\bar{\alpha} &= \mu + v \\
\lambda_t &= B\lambda_{t-1} + (1 - B)\lambda_0 + e_t
\end{aligned}
$$

where $u_t \sim N(0, \Sigma \otimes H)$; $\varepsilon_i \sim N(0, V \otimes E_i\Omega_1 E_i)$; $v \sim N(0, \Psi)$; $\lambda_o \sim N(0, V \otimes \Omega_{20})$; $e_t \sim N(0, V \otimes \Omega_{2t})$ and $\Omega_{2t} = \nu_1\Omega_{2t-1} + \nu_2\Omega_{20}$. We assume that the covariance matrices are independent, that V, Ψ, ν_1, ν_2, and μ are known and that $\Sigma \sim iW_N(\sigma_o, M_o)$, $H \sim iW_G(h_o, P_o)$, $\Omega_1 \sim iW_k(w_1, W_1)$, and $\Omega_{20} \sim iW_k(w_2, W_2)$, where the notation $\Phi \sim iW_p(v, Z)$ means that the symmetric positive definite matrix Φ follows a p–dimensional inverted Wishart distribution with v degrees of freedom and scale matrix Z. We also assume that for each of these distributions the degrees of freedom and the scale matrix are known. These assumptions are inconsequential and the analysis goes through, even when consistent estimates are substituted for the true ones.

Given this prior information, the posterior density of the parameter vector $\vartheta_T = (vec(\alpha), vech(\Sigma), vech(H), vec(\bar{\alpha}), vech(\Omega_1), vec(\{\lambda_t\}_{t=0}^T), vech(\Omega_{20}))$ is given by

$$
\pi(\vartheta_T \mid Y_T,\ F_T) \propto f(Y_T \mid \vartheta_T,\ F_T)\, p(\vartheta_T \mid F_T) \tag{21}
$$

where $Y_T = (Y_1, ..., Y_T)$ is the sample data and $p(\vartheta \mid F_T)$ is the prior information available at T.

To obtain marginal posteriors, we iterate on the conditional distributions of the parameters, which can easily be obtained from the conditional posterior (21). To deal with the presence of time varying parameters we adapt the results of Carter and Khon (1994) and Chib and Greenberg (1996). In fact, conditional on $\{\lambda_t\}_{t=0}^T$, the distribution of the remaining parameters can be derived without difficulty. Let ψ_{-x} be the vector ϑ containing all the parameters but x. Then the conditional distributions for parameters other than $\{\lambda_t\}$ are:

$$\begin{aligned} \Omega_1 &\mid \psi_{-\Omega_1},\ Y_T,\ F_T \sim iW_k\left(w_1+NG,\hat{W}_1\right) \\ \Omega_{20} &\mid \psi_{-\Omega_{20}},\ Y_T,\ F_T \sim iW_k\left(w_2+TG,\hat{W}_2\right) \\ \Sigma &\mid \psi_{-\Sigma},\ Y_T,\ F_T \sim iW_N\left(\sigma_o+GT,\hat{M}_o\right) \\ H &\mid \psi_{-H},\ Y_T,\ F_T \sim iW_G\left(h_o+NT,\hat{P}_o\right) \\ \alpha &\mid \psi_{-\alpha},\ Y_T,\ F_T \sim N\left(\hat{\alpha},\hat{V}_\alpha\right) \\ \bar{\alpha} &\mid \psi_{-\bar{\alpha}},\ Y_T,\ F_T \sim N\left(\alpha^*,\hat{V}^*\right) \end{aligned} \tag{22}$$

where the expressions for $\hat{W}_1,\ \hat{W}_2,\ \hat{M}_o,\ \hat{P}_o,\ \hat{\alpha},\ \hat{V}_\alpha,\ \alpha^*,\ \hat{V}^*$ are given in the appendix.

Following Chib (1996) the parameter vector λ_t can be included in the Gibbs sampler via the distribution $\pi\left(\lambda_o,...,\lambda_T \mid Y_T,\ F_T,\psi_T\right)$ where $\psi_t \equiv \vartheta_{-\{\lambda_t\}_t}$. We can re–write such a distribution as

$$\pi\left(\lambda_T \mid Y_T,\ F_T,\psi_T\right)\times\pi\left(\lambda_{T-1} \mid Y_T,\ F_T,\psi_{T-1},\lambda_T\right)\times\cdots\times\pi\left(\lambda_o \mid Y_T,\ F_T,,\psi_0,\lambda_1,...\lambda_T\right) \tag{23}$$

A draw from the joint distribution can be obtained by drawing $\tilde{\lambda}_T$ from $\pi\left(\lambda_T \mid Y_T,\ F_T,\psi_T\right)$; then $\tilde{\lambda}_{T-1}$ from $\pi\left(\lambda_{T-1} \mid Y_T,\ F_T,,\psi_{T-1},\tilde{\lambda}_T\right)$ and so on. Let $\lambda^s=(\lambda_s,...,\lambda_T)$ and $Y^s=(Y_s,...,Y_T)$ for $s\leq T$. The density of the typical term in (23) is

$$\begin{aligned} &\pi\left(\lambda_t \mid Y_T,\ F_T,\psi_t,\lambda^{t+1}\right) \\ \propto\ &\pi\left(\lambda_t \mid Y^t,\ F_t,,\psi_t\right)\pi\left(\lambda_{t+1} \mid Y_t,\ F_t,\psi_{t-1},\ \lambda_t\right)f(Y^{t+1},\lambda^{t+1} \mid Y_t,\ F_t,\ \lambda_t,\ \lambda_{t+1}) \\ \propto\ &\pi\left(\lambda_t \mid Y^t,\ F_t,\psi_t\right)\pi\left(\lambda_{t+1} \mid F_t,\psi_{t-1},\ \lambda_t\right) \end{aligned} \tag{24}$$

The last row of (24) follows from the fact that, conditional on λ_{t+1}, the joint density of (Y^{t+1},λ^{t+1}) is independent of λ_t and, conditional on λ_t, λ_{t+1} is independent of Y_t. The second density in (24) is Gaussian with moments $B\lambda_t+(I-B)\lambda_0$ and $V\otimes\Omega_{2t}$. The first was derived in section 3, and it is Gaussian with mean $\hat{\lambda}_{t|\,t}=\hat{\lambda}_{t|\,t-1}+\hat{\Omega}_{t|\,t-1}Z_t'\phi_{22}^{-1}\left(Y_t-Z_t\mu-Z_t\hat{\lambda}_{t|\,t-1}\right)$ and variance $\hat{\Omega}_{t|\,t}=\hat{\Omega}_{t|\,t-1}-\hat{\Omega}_{t|\,t-1}Z_t'\phi_{22}^{-1}Z_t\hat{\Omega}_{t|\,t-1}$. Hence, $\pi\left(\lambda_t \mid Y_T,\ F_t,\psi_t,\lambda^{t+1}\right)\sim N(\hat{\lambda}_t,\hat{\Omega}_t)$ where $\hat{\lambda}_t=\hat{\lambda}_{t|\,t}+M_t(\lambda_{t+1}-B\hat{\lambda}_{t|\,t}-(I-B)\lambda_0)$; $\hat{\Omega}_t=\hat{\Omega}_{t|\,t}-M_t\Omega^*_{t+1|,t}M_t'$ and $M_t=\rho\hat{\Omega}_{t|\,t}\hat{\Omega}^{*-1}_{t+1|\,t}$, with $\hat{\Omega}^*_{t+1|\,t}=\left[\hat{\Omega}_{t+1|\,t}-(I-B)\left(V\otimes\Omega_{20}\right)(I-B)'\right]$.

To be concrete the following algorithm can be used to sample $\{\lambda_t\}$: first, starting from given initial conditions, we run the Kalman filter to recursively get $\hat{\lambda}_t$ and $\hat{\Omega}_t$; then we simulate $\tilde{\lambda}_T$

from a normal with mean $\hat{\lambda}_{T|T}$ and variance $\hat{\Omega}_{T|T}$; $\tilde{\lambda}_{T-1}$ from $N\left(\hat{\lambda}_{T-1}, \hat{\Omega}_{T-1}\right)$, and so on until $\tilde{\lambda}_o$ is simulated from $N\left(\hat{\lambda}_o, \hat{\Omega}_o\right)$ where, for each t, $\hat{\lambda}_t = \hat{\lambda}_{t|t} + M_t\left(\tilde{\lambda}_{t+1} - \rho\hat{\lambda}_{t|t} - (1-\rho)\,\hat{\lambda}_{\mathbf{0}|\mathbf{0}}\right)$ and $\hat{\Omega}_t = \hat{\Omega}_{t|t} - M_t\hat{\Omega}^*_{t+1|,t}M_t'$.

One special case of this setup deserves some attention. Assume informative priors on all the parameters except on H, whose prior is now diffuse. Then our framework resembles the Normal-Diffuse prior of Kadiyala and Karlsson (1997) where posterior dependence among the coefficients of different equations obtains even when there is prior independence. There are two additional major difference with the specification used by these authors: first, we assume that both the mean and the variance of γ_t are random variables - they take the mean and the variance of γ_t to be fixed. Second, we do not restrict Σ_u to be diagonal and therefore allow complicated interactions among variables within and across countries.

Canova and Ciccarelli (1999) describe in detail two special cases of the general setup of this subsection: (i) no information on the location of the mean of the unit specific effect $\left(\Psi^{-1} = 0\right)$ and (ii) no information on the time varying component of the coefficients at a particular point in time $\left(\hat{\Omega}^{-1}_{t|t-1} = 0\right)$. They show that a diffuse prior on $\bar{\alpha}$ does not allow to update the prior information we have on λ_t and that, in the latter case, the posterior mean and variance for γ_t are the same as those obtained when only prior information on α is used.

Finally, it is worth reminding that the structure employed imposes symmetry restrictions, which are desirable in an unrestricted VAR system. Clearly these restrictions may be inappropriate for structural or restricted VAR systems and alternative specifications, along the lines of Sims and Zha (1998), should be used.

4.2 Minnesota-type prior

Given the computational complexity involved in calculating posterior Gibbs sampling estimates for large scale problems, one may be interested in knowing whether shortcuts which do not require iterative procedures may be used.

Here, we adapt the so-called Minnesota prior to a panel VAR framework. The Minnesota prior, described in Litterman (1986), Doan, Litterman and Sims (1984), Ingram and Whiteman (1995), Ballabriga, et al. (1998) among others is a way to account for the near nonstationarity of many macroeconomic time series and, at the same time, to weakly reduce the dimensionality of a VAR model. Given that the intertemporal dependence of the variables is believed to be strong, the prior mean of the VAR coefficients on the first own lag is set equal to one and the mean of remaining coefficients is equal to zero. The covariance matrix of the coefficients is diagonal (so we have prior and posterior independence between equations) and the elements are specified in a way that coefficients of higher order lags are likely to be close to zero (the prior variance decreases when the lag length increases). Moreover, since most of the variations in the VAR variables is accounted for by own lags, coefficients of variables other than the dependent one are assigned a smaller relative variance. The prior on the constant term, other deterministic and exogenous variables is diffuse. Finally, the variance-covariance matrix of the error term is assumed to be fixed and known.

For a panel VAR setup we introduce the following modifications. The covariance–matrices $\Omega_{20}, \Psi, \Delta,$ are assumed to have the same structure. Take, for example, $\Delta = diag\,(\Delta_1, ..., \Delta_n)$, where $\Delta_i = V \otimes E_i \Omega_1 E_i$. We assume that $V = H$ (see equation (10)) and that Ω_1 is diagonal with elements :

$$\sigma^2_{g_i j_s} = \left(\frac{\theta_{1\alpha} \theta_3^{\delta(g_i, j_s)}}{l^{\theta_2}} \frac{1}{\sigma_{j_s}} \right)^2 \quad g,\, j = 1, ..., G \quad i, s = 1, ..., N \quad l = 1, \ldots p$$

where $\delta\,(g_i, j_s) = 0$ if $i = s$ and 1 otherwise and

$$\sigma^2_{gm} = (\theta_{1\alpha}\theta_4)^2 \qquad m = 1, ..., q$$

Here, g_i represents equation g of unit i, j_s the endogenous variable j of unit s, l the lag, m exogenous or deterministic variables.

The hyperparameter $\theta_{1\alpha}$ controls the tightness of beliefs for the vector α; θ_2 the rate at which the prior variance decays with the lag; θ_3 the degree of uncertainty for the coefficients of the variables of unit s in the equations of unit i; θ_4 the degree of uncertainty of the coefficients of the exogenous variables and σ_{j_s} are the diagonal elements of the matrix Σ_u used as scale factors to account for differences in units of measurement. Notice that we don't have prior independence between equations: our prior information specifies that, for example, the coefficient on lag 1 of the GNP equation for the US may have some relationship with the same coefficient in the PRICE equation for US. Moreover, we have not specified a hyperparameter which controls the overall tightness of beliefs because the randomness of the coefficients depends on α_i and λ_t and we parametrize the uncertainty in each of them separately. Finally, there is no distinction between own versus other countries variables (see Sims and Zha (1998)). This would not be the case if V_i were country specific. The structures for Ψ and Ω_o are identical with $\theta_{1\alpha}$ being replaced by $\theta_{1\bar{\alpha}}$ and $\theta_{1\lambda}$, respectively.

To complete the specification the elements of the matrix H and the σ's are estimated from the data to tune up the prior to the application.

The prior time–varying features of the model are determined by specifying the matrices B, Ω_{2t}. We assume that B is diagonal and that each of the $k \times k$ diagonal blocks B_g satisfies: $B_g = diag\,(\theta_5)$. Furthermore, we let $\Omega_{2t} = \Omega_{2o}$. Here θ_5 controls the evolution of the law of motion of λ_t and $\theta_6 = \nu_1$ control the time variations in λ_t.

Finally, we assume that the $k \times 1$ vectors μ_g and $\tilde{\lambda}_{og}$ have the following structures:

$$\mu_g = \begin{bmatrix} 0 \\ \vdots \\ \theta_7 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \tilde{\lambda}_{og} = \begin{bmatrix} 0 \\ \vdots \\ 1-\theta_7 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where μ_g and $\tilde{\lambda}_{og}$ are the gth–elements of the mean vectors μ and $\tilde{\lambda}_o$ and θ_7 controls the prior mean on the first own lag coefficient of the dependent variable in equation g for unit i.

Summing up, our prior information can be represented with a 9–dimensional vector of hyper-parameters $\Theta = (\theta_{1\alpha}, \theta_{1\lambda}, \theta_{1\bar{\alpha}}, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7)$. Estimates of Θ can be obtained by maximizing the predictive density of the model as in Doan, Litterman and Sims (1984). Posterior distributions for the parameters are obtained by plugging-in the resulting estimates for $\mu, \tilde{\lambda}_o, \Omega_{2o}, \Sigma_u, B, \Psi, \Delta$ in the formulas of section 3 in an empirical Bayes fashion (see e.g. Berger (1985)).

Compared with Ballabriga et al. (1998), who used a Minnesota prior on a panel VAR model for the Spanish, German and French economies, our specification separates the prior information for the time and the individual component (they have one parameter in place of $\theta_{1\alpha}, \theta_{1\lambda}, \theta_{1\bar{\alpha}}$); it introduces a further level of uncertainty by specifying a prior for $\bar{\alpha}$ and allows for a-priori pooling of the information present in the cross section of the panel. None of these features is present in their specification.

5 Forecasting

Once posterior estimates are obtained, forecasts can be computed. In order to obtain multistep forecasting formulas for a panel VAR and to compute turning points probabilities, we rewrite (4) in a companion VAR(1) form

$$Y_{it} = \sum_{j=1}^{N} B_{it}^{j} Y_{jt-1} + D_{it} z_t + U_{it} \tag{25}$$

where Y_{it} and U_{it} are $Gp \times 1$ vectors, B_{it}^{j} is a $Gp \times Gp$ matrix and D_{it} is a $Gp \times q$ matrix.

Stacking for i, and repeatedly substituting we have

$$Y_t = \left[\prod_{r=0}^{h-1} B_{t-r}\right] Y_{t-h} + \sum_{s=0}^{h-1}\left[\prod_{r=0}^{s-1} B_{t-r}\right] D_{t-s}z_{t-s} + \sum_{s=0}^{h-1}\left[\prod_{r=0}^{s-1} B_{t-r}\right] U_{t-s} \tag{26}$$

or

$$y_t = J\left[\prod_{r=0}^{h-1} B_{t-r}\right] Y_{t-h} + \sum_{s=0}^{h-1} \Phi_{st} D_{t-s}z_{t-s} + \sum_{s=0}^{h-1} \Phi_{st} u_{t-s} \tag{27}$$

where $\Phi_{st} = \prod_{r=0}^{s-1} B_{t-r}$, and $J = I_N \otimes J_1$, $J_1 = [I_G \quad 0]$ and J is a selection matrix such that $JY_t = y_t$, $JU_t = u_t$ and $J'JU_t = U_t$. The expression in (27) can be used to compute the h–steps ahead forecast of the NG–dimensional vector Y_t.

First, we compute a "point" forecast for y_{t+h}. The forecast function is given by

$$y_t\left(h\right) = J\left[\prod_{r=0}^{h-1} B_{t+h-r}\right] Y_t + \sum_{s=0}^{h-1} \Phi_{st+h} D_{t+h-s}z_{t+h-s} \tag{28}$$

or, recursively

$$y_t\left(h\right) = J\tilde{B}_{t+h}Y_t\left(h-1\right) + \tilde{D}_{t+h}z_{t+h}$$

where $\tilde{D}_{t+h}$ is the $NG \times q$ matrix $[d_{1t}\ d_{2t}....d_{Nt}]'$ and $\tilde{B}_{t+h} = diag\left(B_{1t}, B_{2t}, ..., B_{nt}\right)$ with $B_{it} = \left(B_{it}^{1}, B_{it}^{2}, ..., B_{it}^{N}\right)$. One way to obtain a h–step ahead forecasts is to use the posterior mean of $\tilde{B}_{t+h}$ and $\tilde{D}_{t+h}$ and the mean of the predictive density for z_{t+h}, conditional on the information at time t. Estimates for the posterior mean of the coefficients can be obtained from the recursive formulas for λ_t (and, consequently, for γ_t) using expressions like (9) or by drawing from distributions like (20) and (21) in a recursive fashion. Call this estimates $\hat{B}_{t+h|t}$ and $\hat{D}_{t+h|t}$. The forecast error is $y_{t+h} - \hat{y}_t\left(h\right) = \sum_{s=0}^{h-1} \Phi_{st+h}u_{t+h-s} + \left[y_t\left(h\right) - \hat{y}_t\left(h\right)\right]$. To measure the forecasting performance we compute the Mean Square Error (MSE) or the Mean Absolute Error (MAD) of the estimated forecast which are given by

$$\begin{aligned} MSE\left(\hat{y}_t\left(h\right)\right) &= \sum_{s=0}^{h-1} \Phi_{st+h}\Sigma_u\Phi'_{st+h} + MSE\left[y_t\left(h\right) - \hat{y}_t\left(h\right)\right] \\ MAD\left(\hat{y}_t\left(h\right)\right) &= \sum_{s=0}^{h-1} \left|u_{t+h-s}\right| + MAD\left[y_t\left(h\right) - \hat{y}_t\left(h\right)\right] \end{aligned}$$

The first term on the RHS of each equation can be obtained using posterior mean estimates of B_{t+h-r} and of U_t, conditional on the information at time t, while for the second term an approximation can be computed along the lines of Lütkepohl (1991, p.86–89). Clearly, if a researcher is interested in point forecasts using the average value of the parameters, then the previous formulas apply using for $\hat{B}_{t+h|t}$ and $\hat{D}_{t+h|t}$ the posteriors derived in section 3.2.

In many situations, it may be more appealing to compute "average" forecasts h–step ahead using the predictive density $f(Y_{t+h} \mid F_t) = \int f(Y_{t+h} \mid F_t, \vartheta)\, p(\vartheta \mid F_t)$ where $f(Y_{t+h} \mid F_t, \vartheta)$ is the conditional density of the future observation vector given ϑ, and $p(\vartheta \mid F_t)$ is the posterior pdf of ϑ at time t. To compute forecasts for Y_{t+h} we can sample from the predictive density numerically. For each $i = 1, \ldots, M$ we draw $\vartheta^{(i)}$ from the posterior distribution and simulate the vector $Y_{t+h}^{(i)}$ from the density $f\left(Y_{t+h} \mid F_t, \vartheta^{(i)}\right)$. $\left\{Y_{t+h}^{(i)}\right\}_{i=1}^{M}$ constitutes a sample, from which we can compute the necessary moments. The value of the forecast is then the ergodic average $\hat{Y}_{t+h} = M^{-1}\sum_{i=1}^{M} Y_{t+h}^{(i)}$ and its numerical variance can be estimated using $var\left(\hat{Y}_{t+h}\right) = M^{-1}\left[Q_o + \sum_{s=1}^{r}\left(1 - \frac{s}{r+1}\right)(Q_s + Q_s')\right]$ where $Q_s = M^{-1}\sum_{i=s+1}^{M}\left[Y_{t+h}^{(i)} - \hat{Y}_{t+h}\right]\left[Y_{t+h}^{(i)} - \hat{Y}_{t+h}\right]'$.

Since the computation of the impulse response function for orthogonalized shocks is a simple corollary of the calculation of forecasts, the approach we provide here to calculate point and average forecasts can also be used to compute impulse responses. In fact, given the information up to time t, computing impulse response at $t+h$ is equivalent to calculating the difference between the conditional forecasts at $t+h$, given that at $t+1$ there has been a one unit impulse in one of the orthogonal shocks, and the unconditional forecast, i.e. with the value of the vector that would have occurred without shocks (see Koop (1992) for an application to structural VAR models). This idea is exploited in a recent paper by Waggoner and Zha (1998). The authors, using a version of (27), develop two bayesian methods for computing probability distributions of conditional forecasts. The last term in (27) represents the dynamic impact of structural shocks which affect future realizations of variables through the impulse response matrix Φ_{st}. With conditions or constraints imposed on this last term we can produce what they call *conditional forecasts.*

In order to compute structural impulse responses and their error bands we must work with a structural VAR, e.g. impose some restrictions on the contemporaneous coefficient matrix. A prior (flat or informative) can then be assigned to the non-zero elements of this matrix, as suggested by Sims and Zha (1998). The extension of their approach to panel data is however not straightforward and we postpone this issue to future work.

Turning point predictions can also be computed from the predictive density of future observations (see in Zellner, Hong and Min (1991)). We define turning points as follows:

Definition 5.1 *A downward turn for unit i at time $t+h+1$ occurs if S_{it+h} the growth rate of* *the reference variable (typically, GNP) satisfies for all h* $S_{it+h-2},\ S_{it+h-1} < S_{it+h} > S_{it+h+1}$. *An upward turn for unit i at time* $t+h+1$ *occurs if the growth rate of the reference variable satisfies* $S_{it+h-2},\ S_{it+h-1} > S_{it+h} < S_{it+h+1}$.

Similarly, we define a non-downward turn and a non-upward turn:

Definition 5.2 *A non-downward turn for unit i at time* $t+h+1$ *occurs if* S_{it+h} *satisfies for all h* $S_{it+h-2},\ S_{it+h-1} < S_{it+h} \leq S_{it+h+1}$. *A non-upward turn for unit i at time* $t+h+1$ *occurs if the growth rate of the reference variable satisfies* $S_{it+h-2},\ S_{it+h-1} > S_{it+h} \geq S_{it+h+1}$.

Although there are other definitions in the literature (see e.g. Lahiri and Moore (1991)) this is the most used one and it suffices for our purposes. Let $\tilde{f}(Y_{i,t+h} \mid F_t) = \int_{Y_{p,t+h}} f(Y_{t+h} \mid F_t) dY_{p,t+h}$ be the marginal predictive density for the variables of unit i after integrating the remaining p variables and let $\mathcal{K}(S^1_{it+h} \mid F_t) = \int \ldots \int f(S^1_{it+h} \ldots S^G_{it+h} \mid F_t) dS^2_{it+h} \ \ldots dS^G_{it+h}$ be the marginal predictive density for the growth rate of the reference variable, which we order first in the list, in unit i.

Take now the simplest case of $h = 0$. To compute the probability of a turning point we have to calculate S^1_{it+1}. Given the marginal predictive density $\mathcal{K}$, the probability of a downturn in unit i is

$$P_{Dt} = Pr(S^1_{it+1} < S^1_{it} | S^1_{it-2}, S^1_{it-1} < S^1_{it}, F_t) \ = \int_{-\infty}^{S^1_{it}} \mathcal{K}\left(S^1_{it+1} \mid S^1_{it-2}, S^1_{it-1}, S^1_{it}, F_t\right) dS^1_{it} \tag{29}$$

and the probability of an upturn is

$$P_{Ut} = Pr(S^1_{it+1} > S^1_{it} | S^1_{it-2}, S^1_{it-1} > S^1_{it}, F_t) \ = \int_{S^1_{it}}^{\infty} \mathcal{K}\left(S^1_{it+1} \mid S^1_{it-2}, S^1_{it-1}, S^1_{it}, F_t\right) dS^1_{it} \tag{30}$$

Using a numerical sample from the predictive density satisfying $S^1_{it-2},\ S^1_{it-1} < S^1_{it}$, we can approximate these probabilities using the frequencies of realizations which are less then or greater then S_{it}. With a symmetric loss function, minimization of the expected loss leads to predict the occurrence of turning point at $t+1$ if $P_{Dt} > 0.5$ or $P_{Ut} > 0.5$.

For $h \neq 0$ the probability of a turning point can be computed using the joint predictive density for all future observations, i.e. in the case of a downturn,

$$P_{Dt+h} = Pr(S^1_{it+h+1} < S^1_{it+h} > S^1_{it+h-2}, S^1_{it+h-1} \mid F_t) \ = \int_{-\infty}^{S^1_{it}} \int_{S^1_{it}}^{\infty} \int_{S^1_{it}}^{\infty} \mathcal{K}\left(S^1_{it+h+1} < S^1_{it+h} > S^1_{it+h-2}, S^1_{it+h-1} \mid F_t\right) dS^1_{it+h} dS^1_{it+h-1} dS^1_{it+h-2} \tag{31}$$

Given the available panel data structure we may also be interested in computing the probability that a turning point occurs jointly for $m \leq N$ units of panel. For example, we would like to compute the probability that at t+1 there will be a recession in European countries. Let $\tilde{\mathcal{K}}(S^1_{t+h} \mid F_t)$ be the joint predictive density of the reference variable for the m units of interest. Then the probability of a downturn is:

$$
\begin{aligned}
P^m_{Dt} = Pr(S^1_{it+1} < S^1_{it}\ i = 1, \ldots m | S^1_{it-2}, S^1_{it-1} < S^1_{it}, F_t,) \quad = \\
\int_{-\infty}^{S^1_{1t}} \ldots \int_{-\infty}^{S^1_{mt}} \tilde{\mathcal{K}} \left(S^1_{t+1} \mid S^1_{t-2}, S^1_{t-1} < S^1_t, F_t \right) dS^1_{1t} \ldots dS^1_{mt}
\end{aligned}
\tag{32}
$$

6 An application

In this section we apply the methodology to the problem of forecasting growth rates and predicting turning points in the G-7 countries. For each country we consider three national variables (GNP, real stock returns and real money growth) and a world one (the median real stock return in OECD countries) which is assumed to be exogenous in each equation. Figures 1-3 plot the series. Hence there are 21 variables in the panel VAR. These variables are chosen after a rough specification search over about 10 variables because they appear to have the highest in-sample pairwise and multiple correlation with output growth. Among the variables we tried are also the nominal interest rate, the slope of the term structure and inflation. Data is sampled quarterly from 1973,1 to 1993,4 and taken from IMF statistics. The sample 1973,1-1988,4 is used to estimate the parameters and the sample 1989,1-1993,4 to evaluate the forecasting ability and to predict turning points.

We compare the forecasting performance of our panel VAR specifications with those obtained with other models suggested in the literature. As a benchmark we first run two versions of a tri-variable VAR(2) model for each country separately. The first one is an unrestricted (VAR). The second a weakly restricted VAR (BVAR) where we use a standard Litterman-prior with a mean of one on the first lag, a general tightness of 0.15, no decay in the lags and a weight of 0.5 on the lags of other variables. Since these two models do not exploit cross sectional information nor do they allow for time variation, they can be used as a benchmark to measure the improvements obtained by specifications which allow any of these two features in the model.

Also for comparison, we run a single equation AR(3) for GNP growth for each single country, augmented with two lags of real stock returns, 1 lag of real money balances and one lag of the median world real stock return. This is the specification used by Garcia Ferrer et al (1987), Zellner and Hong (1989) and Zellner, Hong and Min (1991) to forecast annual growth rates of output in

18 countries. With the extended sample and the higher frequency of the data we have available, we confirm their results for all of the G-7 countries. This model represents a restricted version of the previous unrestricted VAR where insignificant lags are purged from the specification. The forecasting power of this model is measured when parameters are estimated with OLS (OLS) and with the three shrinkage procedures: a ridge estimator (RIDGE), an estimator obtained assuming an exchangeable prior on the coefficients (as in Garcia Ferrer et al. (1987)) (EXCHANGEABLE) and an estimator obtained using a g-prior (as in Zellner and Hong (1989)) (G-PRIOR). The two latter estimators attempt to improve upon OLS by combining the information coming from each unit with the one from the pooled sample. They differ in the way they combine individual and pooled information. Notice that none of these estimators allows for time variations in the coefficients.

As a final term of comparison, we use a version of the panel VAR specification suggested by Ballabriga et al (1998) (PBVAR). This model specification does not use the information coming from the cross section - every variable is treated in the same way regardless of the country where is from - but allows for time variations in the coefficients of the model. The model has the same structure as Doan, Litterman and Sims (1984) and assumes that the coefficient vector β_t has an AR(1) structure of the form $\beta_t = M\beta_{t-1} + u_t$ where u_t, conditional on the information available, is normal with mean zero and variance Σ_u. The matrices β_0, M, and Σ_u depend on 7 hyperparameters: five parameters controlling the structure of Σ_{u0} (a general tightness (θ_1), a tightness on variables of the same country (θ_3), a tightness on the variables of other countries (θ_4), a geometric lag decay parameter (θ_2), and a tightness on world variables (θ_5)); a parameter describing the structure of M (θ_6); and a parameter controlling the prior mean on the first lag of β_0 (θ_7). Table 1 reports the optimal values obtained by maximizing the in-sample predictive density with a simplex algorithm.

We produce forecasts from three versions of our panel VAR model: one with a modified Minnesota-prior (PANEL1), one with a fully hierarchical homoskedastic specification (PANEL2), one with a fully hierarchical heteroskedastic specification (PANEL3). In the first one, the nine prior parameters are selected to maximize the predictive density using a simplex method. Their optimal values are reported in table 2. For both PBVAR and PANEL1 forecasts are computed using the posterior mean of the coefficients, after we have plugged-in the estimates of the prior parameters in the formula. For PANEL2/PANEL3 posterior estimates of the coefficients are computed numerically using the Gibbs sampler and forecasts are directly obtained from these estimates.

For all panel VAR models we assume that $H = V$. For the PANEL1 specification we compute the scale factors V and the matrix Σ_u as follows. We estimate a trivariate VAR for each country and take the average of the estimated variance–covariance matrix of the residuals across countries as a measure of V. Furthermore, for each of the three variable we estimate a 7–variable VAR (the same variable across countries) and store the variance–covariance matrices of the residuals. An estimate of Σ_u is obtained as:

$$\hat{\Sigma}_u = \sum_{j=1}^{3} \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_7 \end{pmatrix}_j \otimes \begin{pmatrix} 0 & 0 & 0 \\ 0 & v_j & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

where the first matrix contains on the diagonal the estimated standard deviations obtained by running the three 7–variate VARs; while the second matrix contains just one element different from zero, the (j, j) element, which is obtained from the diagonal of the matrix V. For the PANEL2/PANEL3 specifications we need to choose the scale and the degrees of freedom in the various Wishart distribution. We still set $H = V$ with V estimated as before. Following Kadiyala and Karlsson (1997) we set $\sigma_0 = N + 2 + (T - p) * G, \omega_1 = k + 2 + N + g, \omega_2 = k + 2 + (t - p) * G$, while the scale matrices M_0, W_1 and W_2 are such that $\Sigma_u, \Delta, \Sigma_\epsilon$ have the same structure as in the PANEL1 specification. For PANEL3 ν_1 and ν_2 are selected with a rough specification search and Ω_{i1} is drawn from an inverted wishart with degrees of freedom $k + G$ and scale matrix $W_1 + (A_i E_i - \bar{A})' V^{-1} (A_i E_i - \bar{A})$.

We compare the forecasting ability of various models using both the Theil-U Statistics and the Mean Absolute Deviation (MAD), 1 and 4 periods ahead (see table 3). Note that the various specifications we use are in increasing order of complexity and flexibility. Therefore, at each stage we can assess the forecasting improvements obtained adding one extra feature to the model. To examine the performance of various models as business cycle indicators we compute turning points predictions one period ahead. Following Zellner et al. (1991), we compute the total number of turning points, the number of downturns and no-downturns, and the number of upturns and no-upturns in the sample (across all countries) and for each procedure we report the number of correct cases in table 4. Finally, for each model, we compute the probability of a downward turn in the growth rate of US output over the period 1989:1-1991:4, given the information available in 1988:4. According to the official NBER classification the long expansion of the 1980's terminated in 1990:3 and it was followed by a brief and shallow recession. The probabilities for the ten models for each of the 12 periods we consider are presented in table 5.

The forecasting performances of univariate OLS, ridge and exchangeable procedures are very similar. The minimum and maximum values of the Theil-U across countries at one and four steps for the latter two are slightly smaller, but the mean and the median at both steps are practically identical. On the other hand, a univariate model where the parameters are shrunk with a g-prior is somewhat better than OLS in all the dimensions using the Theil-U at both steps.

Unrestricted VAR models are not very successful in forecasting growth rates of output because of the large number of parameters to be estimated. This noticeable in particular for Japan, Germany and the UK where the Theil-Us are significantly worse than those obtained with univariate specifications at the one step horizon. However, unrestricted VAR models outperform all univariate specifications at the four step horizon. Hence, the presence of interdependencies across variables helps in predicting the evolution of the growth rate of output in the medium run. BVAR are significantly better than VAR and univariate approaches at the one step horizon: in the median the gains are of the order of 5-6% over univariate specifications and of more than 10% over the unrestricted VAR. At the four step horizon BVARs turn out to be inferior to unrestricted VARs, and comparable to univariate shrinkage procedures. This is to be expected since to improve the performance at short horizons BVARs reduce both the memory and the interdependencies of the system, which are useful when medium-long run forecasts have to be made.

Adding time variation to the coefficients and interdependencies across countries substantially improves the forecasting performance both at short and at medium horizons. For example, the median (mean) Theil-U at one step goes from 0.85 (0.84) with a simple BVAR to 0.80 (0.80) with the PANEL3 version of the model. There are gains also relative to the PBVAR specification but smaller in magnitude. Relatively speaking, the information contained in the cross section is crucial for the UK (the Theil-U is lower by as much as 15%) and important for Japan, Germany and Italy. For the other three nations country specific models appear to be sufficient to predict output growth. The improvements are noticeable also at longer horizons: the distribution of the Theil-U across countries at the four step horizon is similar to the one obtained with a unrestricted VAR, which is the best among the benchmark models.

The results obtained using the MAD are somewhat similar but four features deserve a comment. First, all univariate shrinkage procedures are better than OLS at the one step horizon. The same is true at four steps horizons except for the case of g-prior, which is significantly worse. Second, unrestricted and simple BVAR display a somewhat mediocre performance at both horizons: the distribution of the MAD across countries is more concentrated but the mean and the median are above those obtained with univariate shrinkage approaches. Third, the improvements obtained with panel VAR approaches are significant and largest for US, Japan, Germany and France and the PANEL3 version of the model produces the best distribution of MAD at the one step horizon. Fourth, at the four step horizon the improvements obtained over univariate shrinkage procedures are small or at times negligible.

It is worth discussing the relative merits of the four models with cross-country interdependences. Our refinement of the Litterman's prior, which employs both cross sectional and time series a-priori restrictions, has, by and large, the same forecasting performance as the PBVAR model both at both horizons. However there are three features which are worth discussing. First, the maximized value of predictive density of the PANEL1 model is significantly higher then the one of the PBVAR model (-36.90 vs. -985.35) suggesting that the former fits the data for the in-sample period better. Second, θ_6, the time variation parameter in the process for λ is significant. Third, while in the PBVAR, the coefficient vector evolves with a persistence equal to 0.95 but with very small variance, in our PANEL1 specification the time varying component of the coefficients is close to be a white noise. This difference can be explained by examining the role of the parameters regulating the cross sectional prior (i.e. the tightness on α and $\bar{\alpha}$). These parameters force a high degree of coherence across countries in the time invariant component and leave the time varying component to evolve randomly. In the PBVAR this distinction is not possible and to produce coefficients which are almost constant over time it is necessary to have close to a random walk dynamics coupled with a small variance. Using equation (12), one can see in fact that coefficients of the PANEL1 model are approximately constant over time and are tightly linked to each other because of the restrictions imposed on α_i. The omission of the fixed effect component, which is precisely what the PBVAR does, is therefore likely to biases upward estimates of the persistence parameter.

The performance of a fully hierarchical model with heteroskedasticity is somewhat better than the two versions of the Minnesota prior we consider. However one should weight this improved performance against the computation costs. While forecasts for the fully hierarchical prior required several hours of computer time on a Pentium III-700, the forecasts produced with a PBVAR or PANEL1 were obtained in a matter of minutes. We are currently experimenting with an alternative specification which drastically reduces the dimensionality of the time component. Preliminary results suggest that the computational costs are reduced by as much as 95% while the quality of the forecasts is only marginally affected.

In sum, using interdependencies, and cross sectional restrictions in the coefficients helps in improving forecasts at short-medium horizon. Nevertheless, because there are still substantial differences in the process for the growth rate of output across countries in the sample, the improvements are not uniform and in some cases considering only domestic variables seems appropriate. Also, the use of time variation the coefficients when coupled with heteroskedasticity appears to be crucial for forecasting probably because the structure of the various economies is changing over time and outliers are present.

How good are various approaches in predicting turning points? Out of 96 total actual turning points in the sample, univariate approaches recognize between 72 and 75. Differences emerge when we try to predict upturns and non-upturns and for this type of turning points, Zellner's-g is best. Unrestricted VARs fare very poorly in this dimension and recognize about 10% less turning points than the best ones. The performance of the BVAR model is comparable to the one of Ridge and Exchangeable approaches but, contrary to them, it predicts upturns and non-upturns better than downturns and non-downturns. The performance of the PBVAR model is surprisingly poor: it is the second worst in recognizing the total number of turning points and is comparable to unrestricted VARs in predicting downturns and non-downturns. The PANEL3 model produces 76 correct turning point forecasts and recognizes the same number of upturns and downturns (38). Comparatively speaking, all PANEL specifications substantially improve over PBVAR and BVAR.

Two other conclusions can be drawn from table 4. First, different models are better in recognizing different types of turning points. If predicting downturns (and non-downturns) is more important than predicting upturns (and non-upturns) our results suggest that VAR, BVAR and PBVAR should not be used. Second, while in terms of linear forecasting statistics there was a clear ranking of procedures, with more complicated ones doing a better job, when we look at nonlinear forecasting statistics, simple univariate approaches, and OLS in particular, are almost as good as other more refined approaches.

Given that our suggested specifications are good in forecasting on average, we would like to know if they are also good in predicting a specific episode of interest, i.e., the downturn in real activity occurred in the US in 1990:3. This is an interesting episode because many commercial and government models, which were forecasting pretty well in the 1980s, failed to find any relevant signs that a downturn and a short recession were forthcoming (see e.g. Stock and Watson (1993)). Interestingly, all procedures predict that there is a nonnegligible probability of a downturn in economic activity at 1990:3. For univariate procedures this probability is much larger than the threshold of 0.5, which we use to consider the date a downturn. In fact all four univariate approaches predict the existence of a downturn with probability above 0.64. Single country VAR, with and without a

Bayesian prior, are worse than univariate procedures (probability 0.32 and 0.36 respectively) but this may be due to the larger number of parameters to be estimated with the information available at 1988:4. The PBVAR specification is overwhelmingly predicting a downward turn in 1990:3 (probability is 0.82) and does not produce any false alarm in the neighborhood of this date. The PANEL2 and PANEL3 approaches improve over single country VAR substantially and the probability of a downturn they produce in 1990:3 is similar to those of univariate approaches. Interestingly, the PANEL3 model gives a strong warning signal also a quarter before the downturn occur, which is indicative of the good anticipatory features of the model. The performance of PANEL1 is poor and the probability of a downturn in 1990:3 is low. Note also that while univariate approaches have the tendency to produce a false alarm in 1989:4, probably due to the stock market crash of the fall of 1989, the probabilities produced by VAR and BVAR at dates other than 1990:3 are small. The latter five models also produce a high probability of a downturn in 1991:3 a date where a downturn materialized. Finally, the downturn in 1989:2 is missed by all approaches but PANEL3 and only the PBVAR (0.42) and the PANEL1 (0.41) provide a non-negligible signal.

7 Conclusions

The task of this paper was to describe the issues of specification, estimation and forecasting in a macro-panel VAR model with interdependencies. The point of view used is Bayesian. Such an approach has been widely used in the VAR literature since the works of Doan, Litterman and Sims (1984), Litterman (1986), and Sims and Zha (1998) and provides a convenient framework where one can allow for both interdependencies and meaningful time variations in the coefficients. We decompose the parameter vector into two components, one which is unit specific and the other which is time specific. We specify a flexible prior on these two components which parsimoniously accounts for possible interdependencies in the cross section and for variations in the evolution of the parameters over time. The prior shares features with those of Lindlay and Smith (1972), Doan, Litterman and Sims (1984) and Hsiao et al. (1998) and it is specified to have a hierarchical structure, which allows for various degrees of ignorance in the researcher's information about the parameters.

Bayesian VARs are known produce better forecasts than unrestricted VAR and, in many situations, ARIMA or structural models (Canova (1995) for references). By allowing interdependencies and some degree of information pooling across units in the model specification we introduce an additional level of flexibility which may improve the forecasting ability of these models.

In the case of fully hierarchical priors, a Markov Chain Monte Carlo method (the Gibbs sampler) is employed to calculate posterior distributions and to construct forecasts. Such an approach is useful in our setup since it exploits the recursive features of the posterior distribution. We also consider a version of the Minnesota prior. In this case we employ the predictive density of the model to estimate unknown parameters and plug-in our estimates in the relevant formulas in an empirical Bayes fashion.

To illustrate the approach, we apply the methodology to the problem of predicting output growth, of forecasting turning points in output growth in the G-7 and computing the probability of a recession in the US. We show that our panel VAR approach is competitive and improves over existing univariate and simple BVAR models using either the Theil-U or the MAD criteria both at the one step and at the four steps horizons. The improvements are of the order of 5-15% with the Theil-U and about 2-8% with the MAD. The forecasting performance of our specification is also better then the one of a BVAR model which mechanically extends the Minnesota prior to the panel case. In terms of turning point prediction, the three Panel VARs are able to recognize about 80% of turning points in the sample and they turn out to be the best for this task, along with Zellner's g-prior shrinkage approach. The simple extension of the Litterman's prior to the panel case does poorly along this dimension and, among all the procedures employed is the second worst. Finally, all the procedures produce a significant probability of a downturn at 90:3, the date selected by the NBER committee to terminate the long expansion of the 80's. However the other two downturn dates in our sample are correctly recognized only by one of our Panel VAR specifications.

We consider the work presented in this paper as the first step in developing a coherent theory for Bayesian Panel VAR models which take into consideration both the nature of interdependencies, the similarities in the model across units and the existence of time variation in the coefficients. Extensions of the theory outlined include the formulation leading indicator index models, the formulation of interesting hypothesis on the nature of the interdependencies, the similarities across units and the variations over time and the development of tools to undertake structural identification in these models. The work of Sims and Zha (1998) is the starting point for developments in this latter case.

Appendix: Definition of the matrices for the Gibbs sampler

$$
\begin{aligned}
\hat{W}_1 &= W_1 + \sum_i \left(A_i E_i - \bar{A}\right)' V^{-1} \left(A_i E_i - \bar{A}\right), \\
\hat{W}_2 &= W_2 + \sum_t \left(\Lambda_t - \rho\Lambda_{t-1} - (1-\rho)\Lambda_o\right)' \left(\delta_t V_1\right)^{-1} \left(\Lambda_t - \rho\Lambda_{t-1} - (1-\rho)\Lambda_o\right) \\
\hat{M}_o &= M_o + \sum_t \left(\mathbf{Y}_t - \mathbf{\Gamma}_t \mathbf{W}_t'\right) H^{-1} \left(\mathbf{Y}_t - \mathbf{\Gamma}_t \mathbf{W}_t'\right)' \\
\hat{P}_o &= P_o + \sum_t^{\prime} \left(\mathbf{Y}_t - \mathbf{\Gamma}_t \mathbf{W}_t'\right)' \Sigma^{-1} \left(\mathbf{Y}_t - \mathbf{\Gamma}_t \mathbf{W}_t'\right) \\
\hat{\alpha} &= \hat{V}_\alpha \left(\sum_t W_t' \left(\Sigma \otimes H\right)^{-1} \left(Y_t - Z_t \lambda_t\right) + \Delta^{-1} S_N \bar{\alpha}\right) \\
\hat{V}_\alpha &= \left(\sum_t W_t' \left(\Sigma \otimes H\right)^{-1} W_t + \Delta^{-1}\right)^{-1} \\
\alpha^* &= \hat{V}^* \left(\left(V \otimes \Omega_1\right)^{-1} \sum_i R_i \alpha_i + \Psi^{-1} \mu\right) \\
\hat{V}^* &= \left(N \left(V \otimes \Omega_1\right)^{-1} + \Psi^{-1}\right)^{-1}
\end{aligned}
$$

where $\delta_t = v_2^t + v_1 \left(1 - v_2^t\right) / \left(1 - v_2\right)$, $\mathbf{Y}_t$ is the $N \times G$ matrix such that $vecr\left(\mathbf{Y}_t\right) = Y_t$, $\mathbf{\Gamma}_t = \left[vecr\left(\Gamma_{1t}\right), ..., vecr\left(\Gamma_{Nt}\right)\right]'$ is $N \times Gk$ and $\mathbf{W}_t = \left(I_G \otimes X_t'\right)$. Here $vecr\left(\right)$ is the row vectorization of a matrix; $\Gamma_{it} = A_i + \Lambda_t E_i$ is a $G \times k$ matrix and $\alpha = vecr(A)$; $\lambda_t = vecr(\Lambda_t)$.

References

[1] Anderson B.D.O., J.B. Moore (1979), *Optimal Filtering*, Prentice–Hall.

[2] Ballabriga F. C., M. Sebastian and J. Vallés (1998) , European Asymmetries, *Journal of International Economics*, 48, 233-53.

[3] Berger J.O. (1985), *Statistical Decision Theory and Bayesian Analysis,* New York: Springer-Verlag, 2nd ed.

[4] Binder M., C. Hsiao and M. H. Pesaran (2000), Estimation and inference in short panel vector autoregressions with unit roots and cointegration, mimeo.

[5] Canova, F. (1993), Modelling and Forecasting exchange rates with a Bayesian time-varying coefficient model, *Journal of Economic Dynamics and Control*, 17, 233-261.

[6] Canova, F. (1995), VAR: Specification, Estimation, Testing, Forecasting in H. Pesaran and M. Wickens (eds.) *Handbook of Applied Econometrics*, Blackwell Press.

[7] Canova, F. and A. Marcet (1997), The poor stay poor: non–convergence across countries and regions, UPF working paper 137.

[8] Canova, F. and Ciccarelli, M. (1999), Forecasting and Turning Point Predictions in a Bayesian Panel VAR model, Universitat Pompeu Fabra, working paper 443, available at www.econ.upf.es.

[9] Carter, C.K and R.Kohn (1994), On Gibbs sampling for state space models, *Biometrika* , 81(3):541–553.

[10] Chamberlain, G. (1982), Multivariate regression model for panel data, *Journal of Econometrics,* 18:5–46.

[11] Chamberlain, G. (1984), Panel data in *Handbook of Econometrics II*, Z. Griliches and M.D.Intriligator (eds.), Amsterdam North Holland, 1247–1318.

[12] Chib, S. and E. Greenberg (1996), Markov chain Monte Carlo simulation methods in econometrics, *Econometric Theory,* 12:409–431.

[13] Doan, T., R. Litterman and C. Sims (1984), Forecasting and conditional projections using realist prior distributions, *Econometric Review,* 3(1), 1–100.

[14] Forni,M., Hallin,M, Lippi, M. and Reichlin, L. (2000), The Generalized Dynamic-Factor Model: identification and estimation, *The Review of Economics and Statistics* , 82(4) 540-54.

[15] Garcia Ferrer, A, Highfield, F., Palm, F. and Zellner, A. (1987), Macroeconomic forecasting using pooled international data, *Journal of Business and Economic Statistics*, 5, 53-67.

[16] Gelfand A.E. and A.F.M. Smith (1990), Sampling–based approaches to calculating marginal densities, *Journal of the American Statistical Association,* 85(410):398–409.

[17] Geman S. and D. Geman (1984), Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 6(6):721–741.

[18] Gerlach, S. and F. Smets (1996), The monetary transmission mechanism: evidence from the G–7 countries, mimeo, Bank of International Settlements, Basle.

[19] Hoffmaister, W. and J.F.Roldós (1997), Are business cycles different in Asia and Latin America?, IMF working paper no. 97/79.

[20] Holtz–Eakin D., W. Newey and H. Rosen (1988), Estimating vector autoregressions with panel data, *Econometrica,* 56(6) pp. 1371–1395.

[21] Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu (1999), Bayes estimation of short run coefficients in dynamic panel data models, in Hsiao et al. (eds.) *Analysis of panels and limited dependent variable models: in honour of G.S. Maddala,* Cambridge, Cambridge University Press.

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

[23] Koop, G (1992), Aggregate shocks and macroeconomic fluctuations: a Bayesian approach, *Journal of Applied Econometrics,* 7:395–411.

[24] Ingram, B. and Whiteman, C. (1994), "Supplanting the Minnesota Prior: Forecasting Macroeconomic Series using Real Business Cycles Model Prior", *Journal of Monetary Economics* , 34, 497-510.

[25] Lahari, K. and Moore, J. (1991) *Leading Economic Indicators: New Approaches and Forecasting Records*, Chicago, IL: University of Chicago Press.

[26] Lindley, D.V. and A.F.M. Smith (1972), Bayes estimates for the linear model (with discussion), *Journal of the Royal Statistical Society, B,* 34, pp. 1–41.

[27] Litterman, R.B. (1986), Forecasting with Bayesian Vector Autoregressions–Five years of experience, *Journal of Business and Economic Statistics,* 4:25–38.

[28] Lütkepohl, H. (1990), *Introduction to Multiple Time Series Analysis,* Berlin: Springer–Verlag.

[29] Pesaran M.H. and R.Smith (1996), Estimating long run relationships from dynamic heterogeneous panels, *Journal of Econometrics*, 68:79–113.

[30] Rebucci, A (1998), External shocks, macroeconomic policy and growth: a panel VAR approach, Global Economic Institutions, working paper 40.

[31] Sims, C. and T.Zha (1998), Bayesian methods for dynamic multivariate models, *International Economic Review* , 39, 949-979.

[32] Stock, J. and Watson, M. (1993), Predicting Recessions in Stock, J. and M. Watson. (eds) *New Research on Business Cycles, Indicators and Forecasting*, Chicago, Il.: Chicago University Press.

[33] Stock, J. and Watson, M. (1999), Diffusion Indices, Harvard University, manuscript.

[34] Waggoner, D.F. and T.Zha (1998), Conditional Forecasts in Dynamic Multivariate Models, Federal Reserve Bank of Atlanta, Working Paper n. 98–22.

[35] Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics,* New York: Wiley.

[36] Zellner, A. and Hong, C (1989), Forecasting International Growth Rates Using Bayesian Shrinkage and other procedures, *Journal of Econometrics*, 40, 183-202.

[37] Zellner, A., Hong, C. and Min, C. (1991), Forecasting turning point in international output growth rates using Bayesian exponentially weighted autoregression, time varying parameter, and, pooling techniques, *Journal of Econometrics*, 49, 275-304.

Table 1: Estimated Hyperparameters: PBVAR

General tightness (θ_1)	0.01
Lag decay (θ_2)	13.96
Own country tightness (θ_3)	3.5-e005
Other countries tightness (θ_4)	7.3-e004
World variable tightness (θ_5)	5.0e-007
AR coefficient (θ_6)	0.95
Prior mean on the first lag (θ_7)	0.11048

Table 2: Estimated Hyperparameters: PANEL1

Tightness for α ($\theta_{1\alpha}$)	0.1207
Tightness for $\lambda(\theta_{1\lambda})$	0.1300
Tightness for $\bar{\alpha}$ ($\theta_{\bar{\alpha}}$)	0.0004
Lag decay (θ_2)	1.9156
Tightness on other countries (θ_3)	0.0046
Tightness on world variables (θ_4)	4.7804
Law of motion of λ (θ_5)	0.1211
Time variation (θ_6)	0.4295
Prior mean on first lag (θ_7)	0.0754

Table 3: Theil-U Statistics

Method	Step	US	Japan	Germany	UK	France	Italy	Canada	Median	Mean
VAR	1	1.06	0.88	0.91	0.94	1.00	0.73	0.95	0.94	0.92
	4	0.73	0.95	0.56	0.81	1.32	0.96	0.72	0.81	0.86
BVAR	1	0.83	0.89	0.69	0.91	0.90	0.80	0.85	0.85	0.84
	4	0.75	0.89	0.65	0.79	1.16	1.00	0.70	0.89	0.85
OLS	1	1.21	0.86	0.88	0.86	0.90	0.79	0.91	0.88	0.90
	4	0.77	0.90	1.07	0.76	0.98	1.03	0.67	0.90	0.88
Ridge	1	1.17	0.83	0.89	0.85	0.89	0.79	0.89	0.89	0.90
	4	0.76	0.88	1.06	0.75	0.99	1.01	0.68	0.88	0.87
Exchangeable	1	1.18	0.84	0.90	0.85	0.89	0.78	0.89	0.89	0.90
	4	0.76	0.90	1.09	0.75	0.99	1.01	0.68	0.90	0.88
g-prior	1	1.06	0.86	0.69	0.78	1.00	0.72	0.92	0.86	0.86
	4	0.83	1.07	0.77	0.75	1.12	1.02	0.70	0.83	0.89
PBVAR	1	0.82	0.85	0.68	0.76	0.98	0.73	0.85	0.82	0.81
	4	0.86	0.91	0.77	0.75	1.08	1.03	0.66	0.86	0.87
Panel 1	1	0.81	0.88	0.67	0.75	1.02	0.70	0.88	0.81	0.81
	4	0.86	0.90	0.76	0.74	1.07	1.03	0.66	0.86	0.86
Panel 2	1	0.93	0.81	0.69	0.78	0.99	0.78	0.85	0.81	0.82
	4	0.83	1.59	1.62	1.55	1.47	1.93	0.90	1.55	1.41
Panel 3	1	0.85	0.80	0.63	0.76	0.95	0.75	0.88	0.80	0.80
	4	0.85	0.87	0.78	0.76	1.07	0.99	0.66	0.82	0.85
MAD Statistics										
VAR	1	0.46	1.71	1.74	1.35	1.26	2.91	0.65	1.35	1.44
	4	0.35	1.55	1.18	1.33	1.66	2.74	0.56	1.33	1.34
BVAR	1	0.46	1.62	1.48	1.32	1.15	3.22	0.58	1.32	1.40
	4	0.40	1.39	1.25	1.28	1.42	2.98	0.51	1.28	1.40
OLS	1	0.56	1.59	1.51	1.37	1.06	3.17	0.57	1.37	1.40
	4	0.34	1.54	1.58	1.28	1.14	3.19	0.54	1.28	1.37
Ridge	1	0.54	1.50	1.68	1.31	1.07	3.14	0.56	1.31	1.40
	4	0.36	1.46	1.72	1.25	1.17	3.09	0.53	1.25	1.37
Exchangeable	1	0.54	1.52	1.68	1.32	1.06	3.14	0.56	1.32	1.40
	4	0.35	1.48	1.73	1.26	1.17	3.09	0.53	1.26	1.37
g-prior	1	0.53	1.63	1.33	1.18	1.26	2.89	0.54	1.26	1.34
	4	0.41	1.60	1.35	1.18	1.34	3.12	0.51	1.34	1.36
PBVAR	1	0.46	1.47	1.29	1.17	1.27	2.85	0.53	1.27	1.29
	4	0.44	1.48	1.27	1.12	1.31	3.14	0.51	1.27	1.32
Panel 1	1	0.46	1.53	1.24	1.08	1.37	2.82	0.54	1.24	1.29
	4	0.44	1.48	1.27	1.11	1.31	3.14	0.50	1.27	1.32
Panel 2	1	0.49	1.45	1.27	1.18	1.32	3.09	0.60	1.27	1.34
	4	0.55	1.40	1.25	1.11	1.43	2.96	0.65	1.25	1.33
Panel 3	1	0.49	1.44	1.23	1.18	1.20	2.93	0.56	1.20	1.25
	4	0.43	1.43	1.33	1.21	1.26	2.99	0.52	1.26	1.31

Notes: VAR is a VAR(2) model for output growth, real stock returns and real money growth, BVAR is the same model with a Minnesota prior. OLS refer to a model where the parameters are estimated with OLS, Ridge to a Ridge correction, Exchangeable to a model with an exchangeable prior and g-prior to Zellner's g-prior specification. PBVAR is a 21 VAR model with a Minnesota prior and time variations, Panel 1 is a panel VAR model with all 7 countries with a modified Minnesota prior, Panel 2 is the same model with a hierarchical prior, Panel 3 is a hierarchical prior with heteroskedasticity

Table 4: Turning points forecasts

Method	Turning Points	DT & NDT	UT & NUT
TRUE	96	47	49
VAR	65	32	33
BVAR	72	34	38
OLS	74	37	37
Ridge	72	37	35
Exchangeable	72	37	35
g-prior	75	37	38
PBVAR	68	32	36
Panel 1	73	36	37
Panel 2	74	37	37
Panel 3	76	38	38

Notes: VAR is a VAR(2) model for output growth, real stock returns and real money growth, BVAR is the same model with a Minnesota prior. OLS refer to a model where the parameters are estimated with OLS, Ridge to a Ridge correction, Exchangeable to a model with an exchangeable prior and g-prior to Zellner's g-prior specification. PBVAR is a 21 VAR model with a Minnesota prior and time variations Panel 1 is a panel VAR model with all 7 countries with a modified Minnesota prior and Panel 2 is the same model with a hierarchical prior, Panel 3 is a hierarchical prior with heteroskedasticity. DT means downturn, NDT means non-downturn, UT means upturn and NUT means a non-upturn.

Table 5: Probabilities of a downturn in US GDP growth

quarter	VAR	BVAR	OLS	RIDGE	EXCH	g-PRIOR	PBVAR	PANEL1	PANEL2	PANEL3
89:1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
89:2*	0.000	0.005	0.005	0.010	0.000	0.270	0.420	0.410	0.160	0.625
89:3	0.020	0.010	0.005	0.010	0.200	0.250	0.010	0.250	0.230	0.000
89:4	0.780	0.590	0.625	0.815	0.370	0.280	0.070	0.210	0.470	0.000
90:1	0.200	0.375	0.365	0.160	0.070	0.050	0.070	0.230	0.040	0.000
90:2	0.000	0.005	0.000	0.000	0.070	0.080	0.040	0.220	0.030	0.517
90:3*	0.645	0.660	0.700	0.660	0.320	0.360	0.820	0.300	0.550	0.572
90:4	0.005	0.010	0.030	0.015	0.280	0.380	0.040	0.250	0.210	0.000
91:1	0.000	0.005	0.000	0.003	0.230	0.050	0.130	0.240	0.020	0.000
91:2	0.000	0.000	0.000	0.000	0.170	0.060	0.000	0.250	0.000	0.572
91:3*	0.005	0.015	0.000	0.000	0.180	0.490	0.790	0.230	0.630	0.633
91:4	0.015	0.005	0.005	0.035	0.250	0.350	0.080	0.240	0.320	0.000

Notes: A * indicates that a downturn occured in output growth at that date.

Figure 1: Growth rates of GNP, Quarterly series, 1973:I-1993:IV

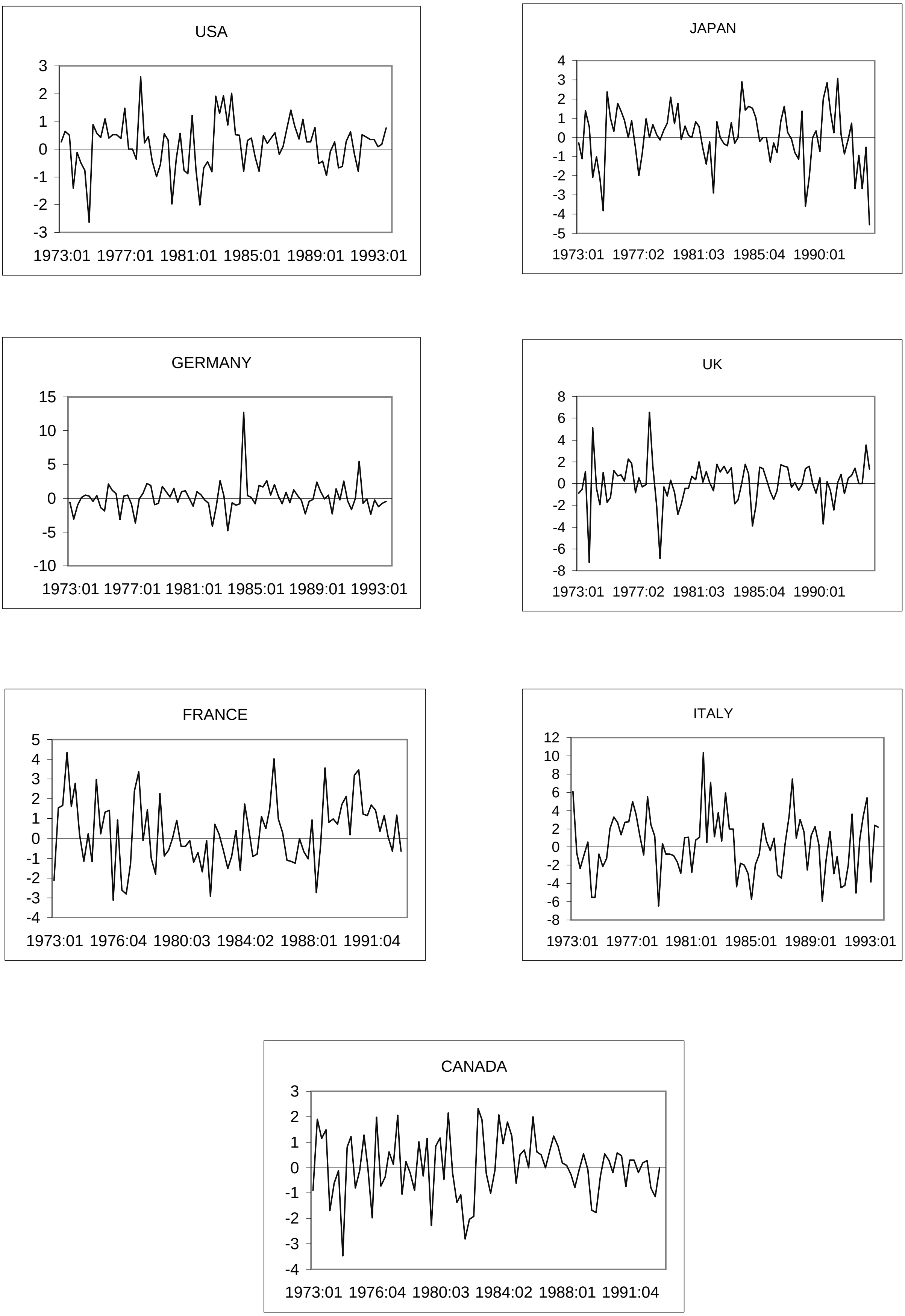

Figure 2: Real stock returns, quarterly data, 1973:I-1993:IV

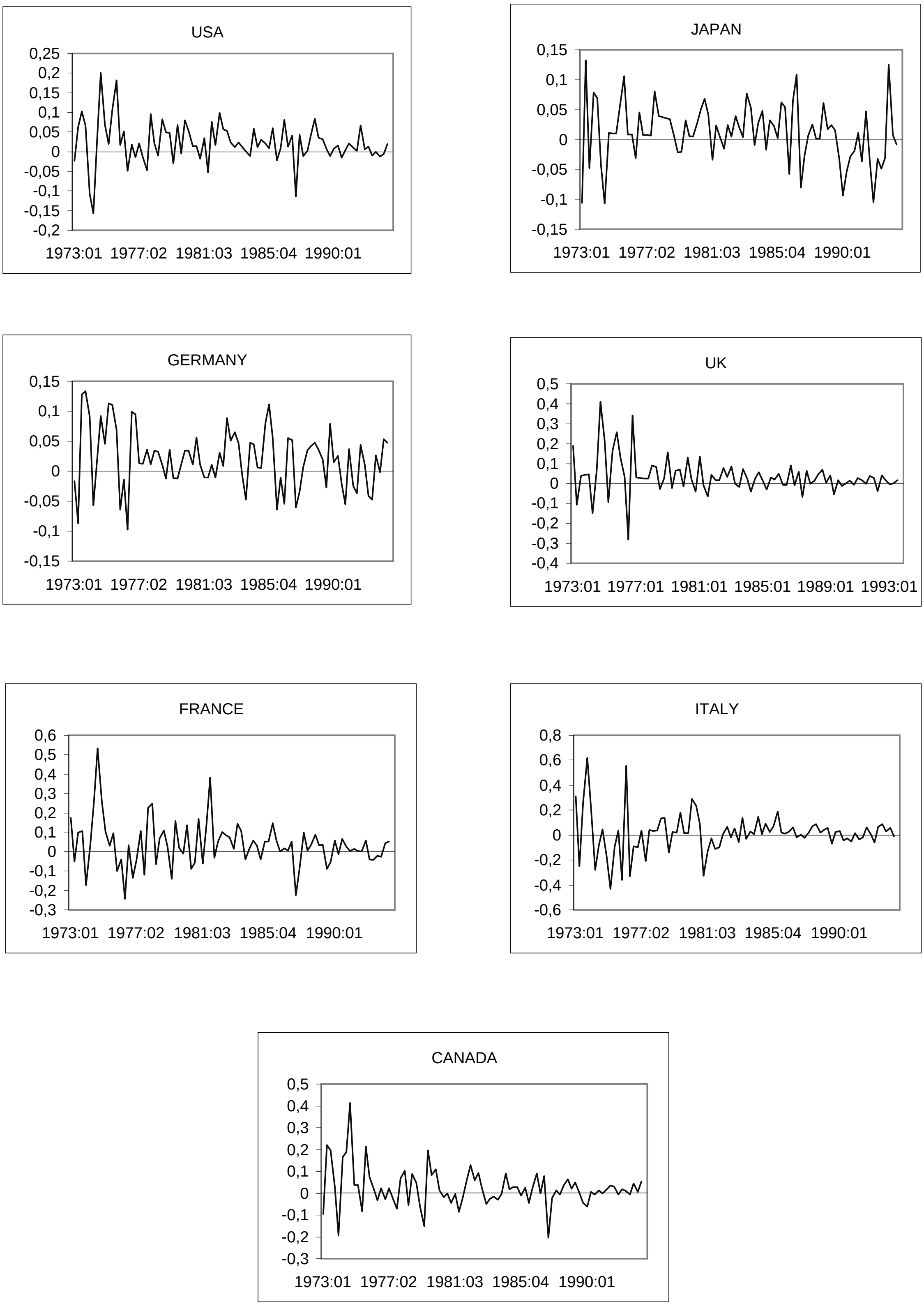

Figure 3: Real money growth, quarterly data, 1973:I-1993:IV