Universitat Pompeu Fabra
Departament d'Economia i Empresa

Bayesian Inference in Heterogeneous Dynamic Panel Data Models: Three Essays

Matteo Ciccarelli

Ph.D. Thesis

Supervisor: Fabio Canova

Barcelona, Spain
2001

Acknowledgments

I wish to thank Fabio Canova for his contagious enthusiasm and his continuous encouragement. I have learned a lot from him, although probably not as much as I could. Over the years of my graduate studies he has been a good friend and a generous supervisor. I am indebted to him more than these simple words can express.

I have also a big debt with Alessandro Rebucci for his patience with my stubbornness. I am amazed by his inexhaustible brightness and his serenity in discussing with me many parts of this work. I still don't believe he wishes to combine his clarity with my chaos in our joint projects.

Sabrina Brancato has been my fixed point and my unique source of inspiration for the last eight years. I owe to her my desire to play this game. She is the only person I would like to keep with me on the tower!

Many other people have contributed in various aspects to the accomplishment of this work. I am particularly grateful to my colleagues of the Department of Economics at the Universidad de Alicante for their spirit of collaboration, and to Luigi Guiso and the members of the "Ente Luigi Einaudi" for their simpathy and encouragement. Part of this work was done when I was a fellow at the "Ente", which provided financial support.

This is also an opportunity to acknowledge my debt towards my sister Rosa and my brother Alessandro, although none of us knows what and how much me owe to each other. I am sure that they know what I am talking about and I hope they can forgive me. This work is dedicated to them.

A Sandro e Rosa

Contents

This Ph.D. thesis is based on the following articles:

Ciccarelli M., Rebucci A., “Has the transmission mechanism of European monetary policy changed in the run-up to EMU”. European Economic Review 2006; 50(3): 737-776.

Ciccarelli M., Canova F. “Forecasting and turning point predictions in a Bayesian panel VAR model”. Journal of Econometrics 2004; 120(2): 323-359.

Ciccarelli M. “Testing restrictions in hierarchical normal data models using Gibbs sampling”. Research in Economics 2004; 58(2): 135-157.

Chapter 1

Introduction

The task of this work is to discuss issues concerning the specification, estimation, inference and forecasting in multivariate dynamic heterogeneous panel data models from a Bayesian perspective. Three essays linked by a few common ideas compose the work.

Multivariate dynamic models (mainly VARs) based on micro or macro panel data sets have become increasingly popular in macroeconomics, especially to study the transmission of real and monetary shocks across economies. This great use of the panel VAR approach is largely justified by the fact that it allows the documentation of the dynamic impact of shocks on key macroeconomic variables in a framework that simultaneously considers shocks emanating from the global environment (world interest rate, terms of trade, common monetary shock) and those of domestic origin (supply shocks, fiscal and monetary policy, etc.).

Despite this empirical interest, the theory for panel VAR is somewhat underdeveloped. The aim of the thesis is to shed more light on the possible applications of the Bayesian framework in discussing estimation, inference, and forecasting using multivariate dynamic models where, beside the time series dimension we can also use the information contained in the cross sectional dimension. The Bayesian point of view provides a natural environment for the models discussed in this work, due to its flexibility in combining different sources of information. Moreover, it has been recently shown that Bayes estimates of hierachical dynamic panel data models have a reduced small sample bias, and help in improving the forecasting performance of these models.

In the first essay, *Forecasting and Turning Point Predictions in a Bayesian Panel VAR model*, we provide methods for forecasting variables and predicting turning points in such an environment. After specifying a flexible model which accounts for both interdependencies in the cross section and time variations in the parameters, we first obtain posterior distributions of the parameters for a particular type of diffuse, for Minnesota-type and for hierarchical priors, and then provide formulas for multistep, multiunit point and average forecasts. An application to the problem of forecasting the growth rate of output and of predicting turning points in the G-7 illustrates the approach. The method proposed is then compared with alternative approaches. It is shown that by allowing interdependencies and some degree of information pooling across units in the model specification, i.e., by introducing an additional level of flexibility with respect to the traditional approaches, the forecasting ability of the model improves substantially.

The second essay, *Asymmetries in the Transmission Mechanism of European Monetary Policy*, investigates the transmission mechanism of European monetary policy by means of dynamic heterogeneous multivariate models. Ideally, one would like to apply the empirical framework proposed in the first essay to a small panel SVAR for output, inflation, interest rates, and the nominal exchange rate. However, the identification of VARs estimated with panel data is a tricky matter because of the restrictions on the variance-covariance matrix of the residuals. Therefore, we combine the methodology developed in the first paper, with the two-stage approach followed by Dornbusch, Favero, and Giavazzi (1998) and do not model inflation and the exchange rate explicitly. In the first stage, a measure of monetary policy is extracted from the data by estimating a reaction function for each central bank, allowing for simultaneity and interdependence in short-term interest rates, and parameters' variation across countries and across time periods. In the second stage, the impact of this measure of monetary policy is analyzed by estimating a dynamic equation for a measure of real economic activity, allowing also for parameters' variation both across countries and time periods. Based on pre-EMU evidence from Germany, France, Italy, and Spain in the 1990s, we show that: (i) there are differences in the timing of the effects of monetary policy on economic activity, but their cumulative impact after two years is rather homogeneous across countries; (ii) the transmission mechanism of monetary policy seems to have changed over time in the run up to EMU, but its degree of heterogeneity has not decreased; (iii) the 'European-wide' effects of monetary on economic activity have become faster in the second half of the 1990s, taking about 6-7 months to appear, peaking after 12-16 months, and disappearing within 18-24 months; (iv) Spain is the most different country among those considered. These results are robust to changes in crucial prior assumptions.

In the third essay, *Testing Restrictions in Normal Data Models Using Gibbs Sampling*, we consider the problem of testing a set of restrictions $R(q) = 0$ in a complex hierarchical model of the kind discussed in the two previous papers. We propose a different approach from the standard PO ratio test. This method can be considered as the Bayesian analogous to the classical Wald type test. With respect to the PO ratio, it has the advantage of being easier to implement and, unlike the PO ratio test, it can be computed also when some prior in the hierarchy is diffuse. Several Monte Carlo simulations show that the procedure scores very well both in terms of power and unbiasedness, generally doing as well as the standard

PO ratio approach, or even better in cases where the degree of coefficient heterogeneity is not high. Moreover, because it is close in spirit to the Wald test and given that the study of its properties is undertaken to a large extent using the sampling properties of the estimators involved, this approach could be useful also to classical econometricians.

Chapter 2

Forecasting and Turning Point Predictions in a Bayesian Panel VAR model

with Fabio Canova

– *Queste dovrebbero essere calze che non si smagliano,*– disse.

– *Tutto dovrebbe essere qualcosa, ma non lo è mai. E' la natura dell'esistenza.*

(Don Delillo, *Libra*)

2.1 Introduction

Panel VAR models have become increasingly popular in macroeconomics to study the transmission of shocks across countries (Ballabriga, Sebastian and Valles (1995)), 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) and Holtz–Eakin et al. (1988), who specify panel VAR models for micro data, to the best of our knowledge only Pesaran and Smith (1995), 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

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

where y_{it} is a G–dimensional 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 (see e.g. Holtz–Eakin et al. (1988)) a specification with time varying slope coefficients and a fixed effect is used. 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 of the system 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

$$y_{it} = \alpha_i + \rho_i y_{it-1} + x_{it}'\beta_i + v_t'\delta_i + \varepsilon_{it}$$

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. In some specifications these vectors of coefficients are assumed to have an exchangeable prior. 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.

The task of this paper is to relax these restrictions and study the issues of specification, estimation and forecasting in a macro-panel VAR model with interdependencies. 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

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

where Y_s $(s < t)$ is a vector of GN elements (G variables for each unit $i = 1, \dots N$). Because coefficients vary across units and along time, estimation of the parameters is impossible without imposing restrictions. However, 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 the other which is time specific. We specify a flexible prior on these two components which parsimoniously takes into account possible interdependencies in the cross section and allows for time variations in the evolution of the parameters over time. The prior shares features with those of Lindley 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.

Besides important considerations concerning the specification of the model, 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 we introduce an additional level of flexibility which may improve the forecasting ability of these models.

We analyze several special cases of our specification and compute Bayesian estimators for the individual coefficients and for their mean values over the cross section. In some cases analytical formulas for the posterior mean are available using standard formulas. Whenever the parameters of the prior are unknown, we employ the predictive density of the model to estimate them and plug-in our estimates in the relevant formulas in an empirical Bayes fashion.

In the case of fully hierarchical priors, a Markov Chain Monte Carlo method (the Gibbs sampler) is employed to calculate posterior distributions. Such an approach is particularly useful in our setup since it exploits the recursive features of the posterior distribution. 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 predictive density of future observation is also used to compute turning point probabilities.

To illustrate the forecasting ability of the proposed approach, we apply the methodology to the problem of predicting output growth, of forecasting turning points in output growth and computing the probability of a recession in the G-7 using three variables (output growth, real stock returns and real money growth) for each country in the panel. To evaluate the performance of the model we also provide a forecasting comparison with other specifications suggested in the literature. We show that our panel VAR approach improves over existing univariate and simple BVAR models when we measure the forecasting performance using the Theil-U and the MAD criteria, both at the one step and at the four steps horizons. The improvements are of the order of 5-10% with the Theil-U and about 2-4% with the MAD. The forecasting performance of our specification is also slightly better then the one of a BVAR model which mechanically extends the Litterman prior to the panel case. In terms of turning point predictions, the two versions of our panel approach 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, we show that the proposed method is competitive with the best specifications in predicting the peak in US economic activity occurred in 1990:3 when using the information available in 1988:4, a peak which was missed by many of the commercial and government forecasting procedures. Depending on the specification, our approach finds 20-55% probability of a downward turn at that date.

The rest of the paper is organized as follows. The next section gives the general model specification and the assumptions we make. Section 3 provides the generalities of Bayesian estimation of the model. Section 4 specifies the prior and discusses the computational issues involved. Section 5 describes formulas for multi-step, multi-units forecasting. Section 6 contains the forecasting application to a panel VAR model for the G-7. Section 7 concludes.

2.2 The general specification

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

$$y_{it} = \sum_{j=1}^{N} \sum_{l=1}^{p} b_{it,l}^{j} y_{jt-l} + d_{it} v_t + u_{it} \tag{2.1}$$

where $i = 1, ..., N$; $t = 1, ..., T$; y_{it} is a G–dimensional 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–dimensional vector of random disturbances. Here p is the number of lags, G the number of endogenous variables and q the number of exogenous variables.

The generality of (2.1) 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, since $b_{it,l}^{j} \neq 0$ 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 we set $d_{it} v_t = a_t, \quad b_{it} = b_t \ \forall\, i, \ u_{it} = \psi_t f_i + \xi_{it}$, $b_{it,l}^{j} = 0, \quad j \neq i, \ \forall\, l$, our specification collapses to the one used by Holtz-Eakin et al. (1988).

We rewrite (2.1) in a stacked regression manner

$$Y_t = W_t \gamma_t + U_t \tag{2.2}$$

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–dimensional vector, β_{it}^{g} are k–dimensional vectors, with $k = NGp + q$, containing, stacked, the G rows of the coefficient matrices b_{it} and d_{it}, while Y_t and U_t are $NG \times 1$ matrices containing the endogenous variables and the random disturbances of the model.

If the γ_{it} are different for each cross–sectional unit in different time periods, there is no way to obtain meaningful estimates of them. 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, that is

$$\gamma_{it} = \alpha_i + \lambda_{it} \tag{2.3}$$

2. For each i, the $Gk \times 1$ vector of time invariant components α_i follows a normal distribution

$$\alpha_i \sim N(R_i\bar{\alpha}, \Delta_i) \tag{2.4}$$

where $R_i = I_G \otimes E_i$, $\Delta_i = V \otimes E_i\Omega_1 E_i$, and the $G \times G$ matrix V and the $k \times k$ matrix Ω_1 are symmetric and positive definite. Here 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 also assume that $cov(\alpha_i, \alpha_j) = 0$ for $i \neq j$.

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

$$\bar{\alpha} \sim N(\mu, \Psi) \tag{2.5}$$

4. For each i we write the vector of the time varying components as $\lambda_{it} = R_i\lambda_t$, where λ_t is independent of α_i for any i. The $Gk \times 1$ vector λ_t evolves according to

$$\lambda_t = B\lambda_{t-1} + e_t, \tag{2.6}$$

where $B = \rho * I_{Gk}$ and, conditional on U_t and W_t, $e_t \sim N(0, \Sigma_\varepsilon)$, with $\Sigma_\varepsilon = V \otimes \Omega_2$, and Ω_2 is a positive definite, symmetric matrix. The initial condition is such that $\lambda_0 \sim N\left(\tilde{\lambda}_0, \Omega_0\right)$.

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

$$U_t \sim N(0, \Sigma_u). \tag{2.7}$$

We assume that $\Sigma_u = \Sigma \otimes H$, where Σ is a $N \times N$ matrix and H is a $G \times G$ matrix, both positive definite and symmetric.

Given the previous assumptions, the structure of the model (2.1) can be summarized with the following *a–priori* hierarchical scheme

$$\begin{aligned} Y_t \;&|\; F_t, \alpha, \lambda_t \sim N(W_t\alpha + Z_t\lambda_t,\ \Sigma_u) \\ \alpha \;&|\; F_t \sim N(S_N\bar{\alpha},\ \Delta) \\ \bar{\alpha} \;&|\; F_t \sim N(\mu,\ \Psi) \\ \lambda_t \;&|\; F_t \sim N\left(\hat{\lambda}_{t|\,t-1}, \hat{\Omega}_{t|\,t-1}\right) \end{aligned} \tag{2.8}$$

where F_t is the information set at t (which includes Y_0, the pre-sample information, and W_t); $S_N = e_N \otimes R_i$; $Z_t = W_t S_N$; $\Delta = diag(\Delta_1, .., \Delta_n)$, $\hat{\lambda}_{t|\,t-1} = B\hat{\lambda}_{t-1|\,t-1}$; $\hat{\Omega}_{t|\,t-1} = B\hat{\Omega}_{t-1|t-1}B' + \Sigma_\varepsilon$, e_N is a vector of ones of dimension N and the notation $t|t-1$ indicates values at t predicted with information at $t-1$.

Assumptions 1-4 decompose the parameters vector 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

$$\gamma_{it} = (1-\rho)\,\alpha_i + \rho\gamma_{it-1} + e_{it} \tag{2.9}$$

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

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 the terminology of 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 the context of a VAR without interdependencies, (i.e. $b^j_{it,l} = 0\,, \;\; j \neq i$), 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 unit i. In the context of forecasting, we may be concerned with point prediction using the average coefficient vector $\bar{\gamma}_t$ or in predicting future values of the variables of interest using information available for each unit.

The assumed Kronecker structure for the variance–covariance matrices is convenient 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 $\Omega_2 = 0$, the model reduces to a standard dynamic panel model with no time–variation in the coefficient vector. Finally, when $V = 0$ neither heterogeneity nor time variation are present in the model.

The prior specification is fully symmetric in the sense that it is the same regardless of the variables and of the units we are considering. In some applications where it is interesting to consider some prior asymmetries, this restriction may not be needed. In that case we set $E_i = I_N$ so that $R_i = I_G \otimes I_N$ and (2.3) becomes $\gamma_{it} = \alpha_i + \lambda_t$ where $\alpha_i \sim N(\bar{\alpha}, \Delta)$ and the prior distributions for $\bar{\alpha}$ and λ_t are the same as before.

As compared to standard BVAR models, we allow for some degree of a-priori pooling of cross sectional information via the exchangeable prior on α. This may be important if there are some similarities in the time series characteristics of the vector of variables considered across units since coefficients of other units may contain useful information for estimating the coefficients of the unit under consideration. A single country VAR with fixed coefficients is nested in our specification and can be obtained by setting $b^j_{it,l} = 0, \forall j \neq i, \forall l$ and letting $\Lambda, \Psi, \Sigma_\varepsilon$ go to zero.

2.3 Posterior Estimates

2.3.1 Fixed effects model

Given prior information on γ_t, and assuming that $\tilde{\lambda}_0$, μ and the covariance matrices are known, we can obtain the posterior distribution of the parameter vector by combining the likelihood function conditional on F_t with the prior distribution for γ_t in the usual way. From (2.8) the likelihood 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 time t, is

$$p(\gamma_t \mid F_t) = N\left(\hat{\gamma}_{t-1},\ \hat{H}_{t-1}\right) \tag{2.10}$$

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 $\pi(\gamma_t \mid F_t,\ Y_t)$ is normal with mean γ_t^* and variance H_t^* where:

$$\begin{aligned} \gamma_t^* &= H_t^*\left(W_t'\Sigma_u^{-1}Y_t + \hat{H}_{t-1}^{-1}\hat{\gamma}_{t-1}\right) \\ H_t^* &= \left[\hat{H}_{t-1}^{-1} + W_t'\Sigma_u^{-1}W_t\right]^{-1} \end{aligned} \tag{2.11}$$

Hence γ_t^* is a standard weighted average of prior and sample information. With a known Σ_u and starting from initial conditions $\hat{\gamma}_0$ and $\hat{H}_0$ we can also obtain posterior moments for

γ_t using the following recursive formulas:

$$
\begin{aligned}
\gamma_t^* &= \hat{\gamma}_{t-1} + \hat{H}_{t-1}W_t' \left[W_t\hat{H}_{t-1}W_t' + \Sigma_u\right]^{-1} (Y_t - W_t\hat{\gamma}_{t-1}) \\
H_t^* &= \hat{H}_{t-1} - \hat{H}_{t-1}W_t' \left[W_t\hat{H}_{t-1}W_t' + \Sigma_u\right]^{-1} W_t\hat{H}_{t-1}
\end{aligned} \qquad (2.12)
$$

Here information about γ_t^* and H_t^* is updated in a Kalman filter fashion.

In some cases attention may be centered in obtaining posterior distributions of α and λ_t separately. It is straightforward to show that:

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

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 $V_\alpha^* = \phi_{11} - \phi_{12}\phi_{22}^{-1}\phi_{21}$.

Repeating the same argument we obtain that 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 $\Omega_t^* = \hat{\Omega}_{t|\,t-1} - \hat{\Omega}_{t|\,t-1}Z_t'\phi_{22}^{-1}Z_t\hat{\Omega}_{t|\,t-1}$. As usual, the mean of the posterior distribution is used as a point estimate for the parameter vector while the variance provides a measure of dispersion.

For the formulas to be operational we need at time $t = 1$ a specification for Σ_u and for the prior distributions of α and λ_t, which in turn requires the specification of the matrices B, Σ_ε, Δ, Ψ, Ω_0 and of the vectors μ and $\tilde{\lambda}$. We will return on this issue in the next section.

2.3.2 Random effects model

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

$$
Y_t = Z_t\bar{\gamma}_t + \eta_t \qquad (2.13)
$$

where $\bar{\gamma}_t = \bar{\alpha} + \lambda_t$ and $\eta_t = u_t + W_t v$.

The posterior distributions of $\bar{\alpha}$ and λ_t can be obtained by combining the priors and the respective likelihoods. The sum of the posterior means of $\bar{\alpha}$ and λ_t then gives us a point estimate of the mean coefficient vector at each t.

Standard manipulations give us that the posterior $\pi_3(\bar{\alpha} \mid Y_t,\ F_t) \sim N(\bar{\alpha}^*, \Psi^*)$ and the posterior $\pi_2(\lambda_t \mid Y_t,\ F_t) \sim N(\lambda_t^*, \Omega_t^*)$ where

$$\bar{\alpha}^* = \mu - \Psi Z_t' \left[Z_t \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) Z_t' + \Sigma_u + W_t \Delta W_t' \right]^{-1} \left[Y_t - Z_t \left(\mu + \hat{\lambda}_{t|\,t-1} \right) \right] \tag{2.14}$$

$$\Psi^* = \Psi - \Psi Z_t' \left[Z_t \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) Z_t' + \Sigma_u + W_t \Delta W_t' \right]^{-1} Z_t \Psi \tag{2.15}$$

while the expressions for λ_t^* and Ω_t^* are the same as before. This implies that the posterior $\pi_4(\bar{\gamma}_t \mid Y_t,\ F_t) \sim N(\bar{\gamma}_t^*, H_t^*)$ where

$$\begin{aligned} \bar{\gamma}_t^* &= \left(\mu + \hat{\lambda}_{t|\,t-1} \right) + \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) Z_t' \left[Z_t \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) Z_t' + \Sigma_u + W_t \Delta W_t' \right]^{-1} \\ & \quad \times \left[Y_t - Z_t \left(\mu + \hat{\lambda}_{t|\,t-1} \right) \right] \end{aligned} \tag{2.16}$$

$$\begin{aligned} H_t^* &= \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) - \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) Z_t' \left[Z_t \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) Z_t' + \Sigma_u + W_t \Delta W_t' \right]^{-1} \\ & \quad \times Z_t \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) \end{aligned} \tag{2.17}$$

2.4 Setting up the priors

For the formulas described in the previous section to be operational, we need to specify the vector $\zeta = (\mu, \tilde{\lambda}_o, \Omega_o, \Sigma_u, \Sigma_\varepsilon, B, \Psi, \Delta)$. The results of section 3 were obtained under the assumption that this vector of parameters was known. In practice, this is hardly the case: to get posterior distributions for the parameters we need to make assumptions on the ζ vector and 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 typically contained in ζ.

There are several ways to proceed. One is to assume a diffuse prior on some of the components of the parameter vector, while still assuming that others are known. Another is to specify a Litterman-type prior where the unknown elements of ζ depend on a small vector of hyperparameters to be estimated from the data in Empirical Bayes fashion. The third is to assume explicit prior distributions for the parameter vector and proceed directly to the numerical integration using Markov Chains-Monte Carlo methods. We examine these approaches in turn.

2.4.1 Diffuse Priors

Imposing diffuse priors is interesting in our context as a way to describe the ignorance of a researcher on some aspects of the prior distribution. It is well known (see Zellner (1971)) that a joint diffuse prior for all the elements of ζ leads to posteriors which contain the sample information summarized in a least square fashion. Also, as shown by Kadiyala and Karlsson (1997), such prior produces posterior dependence among the coefficients of different equations, i.e. the joint posterior for the NGk $\times 1$ vector of coefficients does not factor into the product of the posterior for the k coefficients of each of the NG equations. Here we concentrate attention on two special cases of interest: one where there is no information on the location of the mean of the unit specific effect $\left(\Psi^{-1} = 0\right)$ and one where there is no information on the time varying component of the coefficients either at time zero $\left(\Omega_0^{-1} = 0\right)$ or at a particular point in time $\left(\hat{\Omega}_{t|t-1}^{-1} = 0\right)$. All other components of the vector of parameters are assumed to be known.

Case 1: Ignorance about $\bar{\alpha}$

When the prior distribution of the second stage of the hierarchy is proportional to a constant, the posterior distribution changes according to the following proposition:

Proposition 2.4.1 *Given the prior (2.8), if* $\Psi^{-1} = 0$, *conditional on* Y_t *and* F_t,

(i) The posterior distribution $\pi_3(\bar{\alpha} \mid Y_t,\ F_t)$ is normal with mean $\bar{\alpha}^{**}$ and variance Ψ^{**} where

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

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

(ii) The posterior distribution $\pi_1(\alpha \mid Y_t,\ F_t)$ is normal with mean α^{**} and variance V_α^{**} where

$$\alpha^{**} = V_\alpha^{**} W_t' \left(\Sigma_u + Z_t \hat{\Omega}_{t|\,t-1} Z_t' \right)^{-1} \left(Y_t - Z_t \hat{\lambda}_{t|\,t-1} \right),$$

$$V_\alpha^{**-1} = W_t' \left(\Sigma_u + Z_t \hat{\Omega}_{t|\,t-1} Z_t' \right)^{-1} W_t + F \tag{2.18}$$

with $F = \Delta^{-1} - \Delta^{-1} S_N \left(S_N' \Delta^{-1} S_N \right)^{-1} S_N' \Delta^{-1}$.

(iii) The posterior distribution of λ_t is equal to the prior, i.e.,

$$\pi_2 \left(\lambda_t \mid Y_t, X_t \right) = p \left(\lambda_t \mid X_t \right).$$

(The proof of all propositions is in the appendix).

Notice that the diffuse prior on $\bar{\alpha}$ does not allow to update the prior information we have on λ_t. In fact, in this case, the posterior distribution of γ_t does not depend on the prior for λ_t. To see this note that, with $\Psi^{-1} = 0$, we have that $\hat{H}_{t-1} S_N \left(\Psi + \hat{\Omega}_{t|\,t-1} \right) S_N' + \Delta = F^{-1}$ and using the fact that $\hat{H}_{t-1}^{-1} \hat{\gamma}_{t-1} = F S_N \left(\mu + \bar{\lambda}_{t-1} \right) = 0$ we have $\gamma_t^* = \left[F + W_t' \Sigma_u^{-1} W_t \right]^{-1} \left(W_t' \Sigma_u^{-1} Y_t \right)$ and $H_t^* = \left[F + W_t' \Sigma_u^{-1} W_t \right]^{-1}$ where no prior information on λ_t is involved.

Case 2: Ignorance about λ_t

There are two simple ways of attaching a diffuse prior to the time varying component of the coefficient vector. One possibility is to consider lack of information at time zero ($\Omega_0^{-1} = 0$). When the prior distribution for λ_0 is proportional to a constant, given the autoregressive structure for λ_t, and provided $\rho < 1$, the process tends to "forget" the initial condition. In other words, subsequent realizations of λ_t make less and less uncertain our information on the time varying component of the coefficients so that $\Omega_0^{-1} = 0$ does not imply $\hat{\Omega}_{t|t-1}^{-1} = 0$ at all points in time and, for large enough T, the posterior for α and λ_t is the one presented in section 3.

Another possibility is to set $\Sigma_\varepsilon^{-1} = 0$. To implement this diffuse prior, we assume $\Omega_2^{-1} = 0$. Notice that $\hat{\Omega}_{t|t-1} = B\hat{\Omega}_{t-1|t-1}B' + \Sigma_\varepsilon$. Therefore if $\Sigma_\varepsilon^{-1} = 0$, $\hat{\Omega}_{t|t-1}^{-1} = 0$. In this case it is possible to prove the following result

Proposition 2.4.2 *Given the prior (2.8), if* $\Sigma_\varepsilon^{-1} = 0$, *then*

(i) The posterior distribution of $\bar{\alpha}$ is equal to the prior, i.e.,

$$\pi_3(\bar{\alpha} \mid Y_t, X_t) = p(\bar{\alpha} \mid X_t)$$

(ii) The posterior distribution $\pi_1(\alpha \mid Y_t,\ F_t)$ is normal with mean α^{**} and variance V_α^{**} where

$$\begin{aligned} \alpha^{**} &= V_\alpha^{**}\left[W_t'S\left(Y_t - Z_t\hat{\lambda}_{t|t-1}\right) + \left(S_N\Psi S_N' + \Delta\right)^{-1} S_N\mu\right], \\ V_\alpha^{**-1} &= W_t'TW_t + \left(S_N\Psi S_N' + \Delta\right)^{-1} \end{aligned}$$

and $T = \Sigma_u^{-1} - \Sigma_u^{-1}Z_t\left(Z_t'\Sigma_u^{-1}Z_t\right)^{-1}Z_t'\Sigma_u^{-1}$

(iii) The posterior distribution $\pi_2(\lambda_t \mid Y_t,\ F_t)$ is normal with mean λ_t^{**} and variance Ω_t^{**}

where

$$\lambda_t^{**} = \Omega_t^{**}\left\{Z_t'\left[W_t\left(S_N\Psi S_N' + \Delta\right)W_t' + \Sigma_u\right]^{-1}\left(Y_t - Z_t\mu\right)\right\}$$

$$\Omega_t^{**-1} = Z_t'\left[W_t\left(S_N\Psi S_N' + \Delta\right)W_t' + \Sigma_u\right]^{-1}Z_t$$

The assumption $\Sigma_e^{-1} = 0$ implies that $\hat{\Omega}_{t|t-1}^{-1} = 0$, at all points in time. This implication is unreasonable or, at least, excessively myopic, because it prevents researchers to learn from past realizations of λ_t and to be less uncertain on its mean as times goes by. The assumption $\hat{\Omega}_{t|t-1}^{-1} = 0$ can be more realistic if we attach this infinite uncertainty to the coefficients only at a particular point in time (let's say, $t = t_o$), perhaps to take care of a structural break, after which the process restarts and behaves as it did before the break.

It is worth noting that in both cases 1 and 2, the posterior mean and variance for γ_t are the same as those obtained when only prior information on α is used. This is not surprising if we write (2.8) as a three stage hierarchy

$$
\begin{aligned}
Y_t \quad &| \quad F_t,\ \gamma_t \sim N\left(W_t\gamma_t, \Sigma_u\right) \\
\gamma_t \quad &| \quad F_t,\ \bar{\alpha},\ \lambda_t \sim N\left[E_N\left(\bar{\alpha} + \lambda_t\right), \Delta\right] \\
\left(\bar{\alpha} + \lambda_t\right) \quad &| \quad F_t,\ \mu,\ \hat{\lambda}_{t|t-1} \sim N\left[\left(\mu + \hat{\lambda}_{t|t-1}\right), \Psi + \hat{\Omega}_{t|t-1}\right].
\end{aligned}
$$

Assuming $\Psi^{-1} = 0$ or $\Sigma_\varepsilon^{-1} = 0$ is equivalent to assume a diffuse prior on the third stage of the hierarchy.

2.4.2 Litterman-type prior

Next, we modify the so-called Minnesota prior to account for the presence of multiple units in the VAR. 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 non-stationarity 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 Ω_o, Ψ, Δ, are assumed to have the same a-priori structure. Take, for example, $\Delta = diag\left(\Delta_1, ..., \Delta_n\right)$, where $\Delta_i = V \otimes E_i \Omega_1 E_i$.

The matrix Ω_1 is assumed to be diagonal and its elements have the following structure:

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

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

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

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. Also, assume that $V = H$ (see equation (2.7)). Notice that we don't have prior independence between equations. Hence 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. Because of this V and Ω_1 are common to all units and the prior has a symmetric structure (see Sims and Zha (1998)).

The structures for Ψ and Ω_o are similar with $\theta_{1\alpha}$ being replaced by $\theta_{1\bar{\alpha}}$ and $\theta_{1\lambda}$, respectively.

To complete the specification we need to have a measure of the elements of the matrix H and of the σ's. Following Litterman, these parameters are estimated from the data to tune up the prior to the specific application.

The prior time–varying features of the model are determined by specifying the matrices B, Σ_ε. 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 assume $\Sigma_\varepsilon = \theta_6 \Omega_o$. Here θ_5 controls the evolution of the law of motion of λ_t and θ_6 the heteroscedasticity in the coefficients. Note that a time–invariant model is obtained by setting $\theta_5 = 1$ and $\theta_6 = 0$. Homoscedastic time variations are obtained by setting $\theta_6 = 0$.

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 is a function of a 9–dimensional vector of hyperparameters $\Theta = (\theta_{1\alpha}, \theta_{1\lambda}, \theta_{1\bar{\alpha}}, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7)$. 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 then obtained by plugging-in the resulting estimates for $\mu, \tilde{\lambda}_o, \Omega_o, \Sigma_u, \Sigma_\varepsilon, B, \Psi, \Delta$ in the formulas we have derived in 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 allows for unit specific time variations in the variance of the process ($\theta_6 \neq 0$); it separates the prior information for the time and the individual component (they have one parameter in place of $\theta_{1\alpha}, \theta_{1\lambda}, \theta_{1\bar{\alpha}}$) and introduces a further level of uncertainty by specifying a prior for $\bar{\alpha}$. Furthermore, our prior specification is symmetric and it allows for a-priori pooling of the information present in the cross sectional dimension of the panel. None of these features is present in their specification.

2.4.3 Informative priors

When the prior for the vector of parameters is informative, the posterior distribution for the parameter vector 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 $i = 2$ 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 any 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:

$$
\begin{aligned}
Y_t &= W_t\alpha + Z_t\lambda_t + u_t, \\
\alpha_i &= S_N\bar{\alpha} + \varepsilon_i
\end{aligned}
$$

$$\bar{\alpha} = \mu + v$$
$$\lambda_t = B\lambda_{t-1} + e_t$$

where $u_t \sim N(0, \Sigma \otimes H)$; $\varepsilon_i \sim N(0, V \otimes E_i \Omega_1 E_i)$; $v \sim N(0, \Psi)$; $\lambda_o \sim N(0, V \otimes \Omega_2)$ $e_t \sim N(0, V \otimes \eta\Omega_2)$ and η is the tightness on time variation: if $\eta = 0$ and $B = I$ then λ is time invariant. We assume that the covariance matrices are independent, that V, Ψ, η, 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 $\Omega_2 \sim iW_k(w_2, W_2)$, 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 $\vartheta = (\alpha, \Sigma, H, \bar{\alpha}, \Omega_1, \{\lambda_t\}_{t=0}^T, \Omega_2)$ is given by

$$\pi(\vartheta \mid Y_T,\ F_T) \propto f(Y_T \mid \vartheta_T,\ F_T)\, p(\vartheta \mid F_T) \tag{2.19}$$

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

Given the difficulty to obtain marginal posteriors directly from the integration of (2.19), we iterate on the conditional distributions of the parameters, which can easily be obtained from the conditional posterior (2.19). 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:

$$\begin{aligned}
\Omega_1 &\mid \psi_{-\Omega_1},\ Y_T,\ F_T \sim iW_k\left(w_1 + NG, \hat{W}_1\right) \\
\Omega_2 &\mid \psi_{-\Omega_2},\ Y_T,\ F_T \sim iW_k\left(w_2 + TG, \hat{W}_2\right) \\
\Sigma &\mid \psi_{-\Sigma},\ Y_T,\ F_T \sim iW_N\left(\sigma_o + GT, \hat{M}_o\right) \\
H &\mid \psi_{-H},\ Y_T,\ F_T \sim iW_G\left(h_o + NT, \hat{P}_o\right) \\
\alpha &\mid \psi_{-\alpha},\ Y_T,\ F_T \sim N\left(\hat{\alpha}, \hat{V}_\alpha\right) \\
\bar{\alpha} &\mid \psi_{-\bar{\alpha}},\ Y_T,\ F_T \sim N\left(\alpha^*, \hat{V}^*\right)
\end{aligned} \tag{2.20}$$

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 $\pi(\lambda_o, ..., \lambda_T \mid Y_T,\ F_T, \psi_T)$ where $\psi_t \equiv \vartheta_{-\{\lambda_t\}_t}$. We can re–write such a distribution as

$$\pi(\lambda_T \mid Y_T,\ F_T, \psi_T) \times \pi(\lambda_{T-1} \mid Y_T,\ F_T, \psi_{T-1}, \lambda_T) \times \cdots \times \pi(\lambda_o \mid Y_T,\ F_T,, \psi_0, \lambda_1, ...\lambda_T) \tag{2.21}$$

A draw from the joint distribution can be obtained by drawing $\tilde{\lambda}_T$ from $\pi(\lambda_T \mid Y_T,\ F_T, \psi_T)$; 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 (2.21) is

$$\begin{aligned}
&\pi\left(\lambda_t \mid Y_T,\ F_T, \psi_t, \lambda^{t+1}\right) \\
\propto\ &\pi\left(\lambda_t \mid Y^t,\ F_t,, \psi_t\right) \pi(\lambda_{t+1} \mid Y_t,\ F_t, \psi_{t-1},\ \lambda_t) f(Y^{t+1}, \lambda^{t+1} \mid Y_t,\ F_t,\ \lambda_t,\ \lambda_{t+1}) \\
\propto\ &\pi\left(\lambda_t \mid Y^t,\ F_t, \psi_t\right) \pi(\lambda_{t+1} \mid F_t, \psi_{t-1},\ \lambda_t)
\end{aligned} \tag{2.22}$$

The last row 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 of (2.22) is Gaussian with moments $\rho\lambda_t$ and Σ_ε. 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 $\hat{\lambda}_t = \hat{\lambda}_{t|t} + \rho^{-1} M_t \left(\lambda_{t+1} - \rho \hat{\lambda}_{t|t} \right)$; $\hat{\Omega}_t = \hat{\Omega}_{t|t} - M_t \Omega_{t+1|,t} M_t'$ and $M_t = \rho^2 \hat{\Omega}_{t|t} \hat{\Omega}_{t+1|t}^{-1}$.

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, $\hat{\lambda}_t = \hat{\lambda}_{t|t} + \rho^{-1} M_t \left(\tilde{\lambda}_{t+1} - \rho \hat{\lambda}_{t|t} \right)$ and $\hat{\Omega}_t = \hat{\Omega}_{t|t} - M_t \hat{\Omega}_{t+1|,t} M_t'$.

One special case of the setup described in this subsection deserve some attention. Suppose informative priors on all the parameters except that on H, whose prior is now diffuse, so that the prior for Σ_u is diffuse as well. Then the setup resembles the Normal-Diffuse prior of Kadiyala and Karlsson (1997) and implies that posterior dependence among the coefficients of different equations obtains even when there is prior independence. Hence, the major difference of our prior with the specification used by these authors is that we use a three stage hierarchy, so that both the mean and the variance of γ_t are random variables, while they take the mean and the variance of γ_t to be fixed. Note also that our specification does not restrict Σ_u to be diagonal and therefore permits complicated interactions among variables within and across countries.

Finally, it is worth mentioning that in all the setups we have considered in this section, our prior specification maintains a Kronecker structure for the statistical model. Such a specification is useful since, on one hand, it allows to handle the computations for relatively large systems in a simple fashion and, on the other, imposes symmetry restrictions which appear to be desirable in an unrestricted VAR system of the type examined here. 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.

2.5 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, it is convenient to rewrite (2.1) in a companion VAR(1) form

$$Y_{it} = \sum_{j=1}^{N} B_{it}^{j} Y_{jt-1} + D_{it} z_t + U_{it} \tag{2.23}$$

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

$$Y_t = \left[\prod_{r=0}^{h-1} B_{t-r}\right] Y_{t-h} + \sum_{s=0}^{h-1}\left[\prod_{r=0}^{s-1} B_{t-r}\right] D_{t-s} z_{t-s} + \sum_{s=0}^{h-1}\left[\prod_{r=0}^{s-1} B_{t-r}\right] U_{t-s} \tag{2.24}$$

or

$$y_t = J\left[\prod_{r=0}^{h-1} B_{t-r}\right] Y_{t-h} + \sum_{s=0}^{h-1} \Phi_{st} D_{t-s} z_{t-s} + \sum_{s=0}^{h-1} \Phi_{st} u_{t-s} \tag{2.25}$$

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 (2.25) 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

$$y_t(h) = J\left[\prod_{r=0}^{h-1} B_{t+h-r}\right] Y_t + \sum_{s=0}^{h-1} \Phi_{st+h} D_{t+h-s} z_{t+h-s} \tag{2.26}$$

or, recursively

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

where $\tilde{D}_{t+h}$ is the $NG \times q$ matrix $[d_{1t}\ d_{2t} \ldots . d_{Nt}]'$ and $\tilde{B}_{t+h} = diag(B_{1t}, B_{2t}, \ldots, B_{nt})$ with $B_{it} = \left(B_{it}^{1}, B_{it}^{2}, \ldots, B_{it}^{N}\right)$. 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 $y_{t+h} - \hat{y}_t(h) = \sum_{s=0}^{h-1} \Phi_{st+h} u_{t+h-s} + [y_t(h) - \hat{y}_t(h)]$. To measure the forecasting performance it is useful to compute the Mean Square Error (MSE) or the Mean Absolute Error (MAD) of the estimated forecast which are given by

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

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 $Q_s = M^{-1}\sum_{i=s+1}^{M}\left[Y_{t+h}^{(i)} - \hat{Y}_{t+h}\right]\left[Y_{t+h}^{(i)} - \hat{Y}_{t+h}\right]'$.

Note that 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 (2.25), develop two Bayesian methods for computing probability distributions of conditional forecasts. The last term in (2.25) 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)). Let us define turning points as follows:

Definition 2.5.1 *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 2.5.2 *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 $\mathcal{K}(S^1_{it+h} \mid F_t) = \int \ldots \int f(S^1_{it+h} \cdots S^G_{it+h} \mid F_t) dS^2_{it+h} \ldots dS^G_{it+h}$ be the marginal predictive density for the growth rate of the reference variable, which we order to be the 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

$$P_{Dt} = Pr(S^1_{it+1} < S^1_{it} | S^1_{it-2}, S^1_{it-1} < S^1_{it}, F_t) = \int_{-\infty}^{S^1_{it}} \mathcal{K}\left(S^1_{it+1} \mid S^1_{it-2}, S^1_{it-1}, S^1_{it}, F_t\right) dS^1_{it} \quad (2.27)$$

and the probability of an upturn is

$$P_{Ut} = Pr(S^1_{it+1} > S^1_{it} | S^1_{it-2}, S^1_{it-1} > S^1_{it}, F_t) = \int_{S^1_{it}}^{\infty} \mathcal{K}\left(S^1_{it+1} \mid S^1_{it-2}, S^1_{it-1}, S^1_{it}, F_t\right) dS^1_{it} \quad (2.28)$$

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,

$$P_{Dt+h} = Pr(S^1_{it+h+1} < S^1_{it+h} > S^1_{it+h-2}, S^1_{it+h-1} \mid F_t) \quad =$$

$$\int_{-\infty}^{S^1_{it}} \int_{S^1_{it}}^{\infty} \int_{S^1_{it}}^{\infty} \mathcal{K}\left(S^1_{it+h+1} < S^1_{it+h} > S^1_{it+h-2}, S^1_{it+h-1} \mid F_t\right) dS^1_{it+h} dS^1_{it+h-1} dS^1_{it+h-2} \quad (2.29)$$

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:

$$P^m_{Dt} = Pr(S^1_{it+1} < S^1_{it} \; i = 1, \dots m | S^1_{it-2}, S^1_{it-1} < S^1_{it}, F_t,) \quad =$$

$$\int_{-\infty}^{S^1_{1t}} \cdots \int_{-\infty}^{S^1_{mt}} \tilde{\mathcal{K}}\left(S^1_{t+1} \mid S^1_{t-2}, S^1_{t-1} < S^1_t, F_t\right) dS^1_{1t} \dots dS^1_{mt} \quad (2.30)$$

2.6 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.[1] Hence there are 21 variables in the panel VAR. These variables are chosen after a rough

[1] See Figures 1-3 to have a rough idea about the characteristics of the data.

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 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. Data from 1973,1 to 1988,4 is used to estimate the parameters and data from 1989,1 to 1993,4 to evaluate the forecasting performance 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) model 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 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.

Finally, as a 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 for the entire system 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_{u_0} (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 with 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 selected by maximizing the in-sample predictive density of the model with a simplex algorithm.

We produce forecasts from two versions of our panel VAR model: one with a modified Minnesota-prior (PANEL1), and one with a fully hierarchical specification (PANEL2). In the former, 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 posterior estimates of the coefficients are computed numerically using MCMC methods and forecasts are directly obtained from these estimates.

In setting up the panel VAR models we assume that $H = V$, where V is known. 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 specification 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 the degrees of freedom $\sigma_0 = N + 2 + (T - p) * G$, $\omega_1 = k + 2 + N + g$, $\omega_2 = k + 2 + (t - p) * G$, while the scale matrices M_0, W_1 and W_2 are such that Σ_u, Δ, Σ_ϵ have the same structure as in the PANEL1 specification.

We compare the forecasting ability of various models using both the Theil-U Statistics and the Mean Absolute Deviation (MAD) at 1 and 4 periods ahead. These statistics are reported in 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 non-downturns, and the number of upturns and non-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 that there will be a downward turn in the growth rate of US output in 1989:1-1993:4, 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 nine models for each of the 16 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: the maximum, the median, the mean and the minimum value across countries of the Theil-U at both steps are significantly lower than those obtained with OLS.

Unrestricted VAR models are not very successful in forecasting growth rates of output, given the large number of parameters to be estimated. This is noticeable in particular in the case of Japan, Germany and the UK where the Theil-U 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 terms of the median value the gains are of the order of 5-6% over univariate specifications and of more than 10% over the unrestricted VAR. However, the performance at the four step horizon turns out to be inferior to the one of unrestricted VAR, and comparable to the one of univariate shrinkage procedures. This is to be expected since to improve the performance at short horizons BVAR tend to reduce both the memory and the interdependencies of the system, which we have seen are useful exactly when medium-long run forecasts have to be made.

Adding time variation in the coefficients and interdependencies across countries substantially improves the forecasting performance both at short and at medium horizons. For example, the median Theil-U at one step goes from 0.85 with a simple BVAR to 0.82 with the panel version of this model and for 5 countries the Theil-U is lower by as much as 10%. Similarly, the mean across countries drops by about 3% with the PBVAR specification. The improvement is 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.

Our refinement of the Litterman's prior, which allow for both cross sectional and time series a-priori restrictions, gives a performance which is essentially similar to the one of the PBVAR model both at the one and at the four step horizons. Few features of the optimally estimated parameters are worth discussing. First, while θ_6, the time variation parameter in the variance of λ is different from zero, it does not appear to add much to the performance of the model. Hence, at least with quarterly data, allowing for heteroscedasticity does not help in improving the quality of the forecasts. Second, while in the PBVAR, the coefficient vector evolves with a persistence of 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. Note that this difference is inconsequential for forecasting and 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 randomly evolve. 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 (2.9), 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, biases upward estimates of the persistence parameter and this may explain why the two estimated specifications are so different. Third, 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 our specification fits the data for the in-sample period better. However, this superior in-sample fit appears to be unimportant for forecasting out-of-sample. That is, the (wrong) restrictions that the PBVAR imposes and which biases the persistence parameter of the time varying coefficients do not translate in poor forecasts at the horizons we consider. We conjecture that this may have to do with the peculiarity of the forecasting sample more than with true similarities between the two specifications.

The performance of the PANEL2 specification is also comparable to the one obtained with PBVAR at the one step horizon. However, while the ranking of the Theil-U across countries in PBVAR and PANEL1 were identical, there is some reshuffling with the PANEL2 specification. That is, the model is somewhat better for Japan and France and somewhat worse for the US and Italy. At the four step horizon the performance of the model is significantly worse than any other model. While we have not been able to find a reason for this result, we conjecture that this has to do with the fact that the presence of a large amount of randomness in the specification of the model compounds at long horizons and worsens significantly its performance. In fact, the difference between PANEL1 and PANEL2 specifications, apart from problem of precision of estimates is only in the fact that there is an additional layer of uncertainty in the prior of the model.

The relative performance of the various models with the MAD is somewhat similar to the one obtained with the Theil-U at both horizons. However, four features deserve a comment. First, all univariate shrinkage procedures appear to be 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 now significantly worse. Second, unrestricted and simple BVAR display a somewhat mediocre performance both at one and four steps horizons. In general, 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 our refinement of the Litterman's prior produces the best distribution of MAD at the one step horizon. The improvements are primarily concentrated for those countries which are in the central part of the distribution and this is reflected in the lower median value we obtain. Fourth, the PANEL2 specification is better than any other when we use the mean MAD across country to measure the forecasting specifications at both horizons. That is, PANEL2 produces a distribution of MAD across countries which is centered below the one obtained with other models and more concentrated. Notice also that the significant forecasting differences produced by PANEL2 for the Theil-U and the MAD at the four step horizon probably have to do with the different way the two criteria treat forecasting outliers.

In sum, using interdependencies, adding time variation in the coefficients and using cross sectional restrictions in the prior for the coefficients helps in improving forecasts at short-medium horizon. Nevertheless, it should be pointed out that the distribution of forecasting statistics across countries is very wide, for example, the MAD for Italy is 6 times the one of the US. This differences indicate that the process for the growth rate of GDP in some countries does not share much features with the growth rate of GDP of other G-7 countries and that significant improvements on the results we present can be obtained by restricting attention to the subset of the countries which are more similar. Also notice that the forecasting performance for US and Canada GDP growth is very similar across specifications and jointly improves with the complexity of the model, confirming that there are forecasting externalities which can be obtained by cross-sectionally linking the national models for the two countries.

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 primarily emerge when we try to predict upturns and non-upturns and for this type of turning points, Zellner's-g approach is better than the others. Unrestricted VAR models are very poor in this dimension and recognize about 10% less turning points than Zellner's-g approach. The performance of the BVAR model is comparable to the one of univariate 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. Finally, our two Panel approaches produce 73 and 74 turning point forecasts and recognize the same number of upturns and non-upturns. Comparatively speaking, they substantially improve over PBVAR and are competitive with the best approaches.

Three further 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 as good as other more refined approaches. Third, Panel VAR models of the type we have proposed do a better job than any other procedure when we jointly use linear and nonlinear statistics to measure forecasting performance.

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 downward turn in real activity occurred in the US in 1990:3. This is interesting because alternative approaches, which were forecasting pretty well in the sample 1970-1980, failed to find any relevant signs in the data that would predict that a downturn and a short recession were forthcoming (see e.g. Stock and Watson (1993)). Interestingly enough, and contrary to most forecasting models, all procedures predict that there is a significant probability that a peak in economic activity will occur at 1990:3. For univariate procedures this probability is much larger than the threshold of 0.5 which we use to consider the date of a downward turn. In fact all four univariate approaches predict the existence of a peak 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 second Panel VAR approach improves over single country VAR substantially and a produce probability of a downturn in 1990:3 which is comparable with those of univariate approaches. The performance of the first Panel approach is poor and fails to produce a probability in excess of 0.5 in 1990:3. 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 and never exceed 0.5. The PBVAR model, on the other hand, produces a high probability of a downturn in 1991:3, a date where a downturn materialized. The second panel specification also produces a high probability of a downward turn in 1991:3 while the probabilities at other dates are small. Finally notice that the peak in 1989:2 is missed by all approaches: the ones which give highest probability to this event are the PBVAR (0.42) and the first Panel VAR approach (0.41).

In conclusion, our proposed Bayesian PANEL VAR approach is at least as good as any other approach we have examined and in many cases improves the forecasting performance of existing specification. This is true when we compare procedures using linear and non-linear forecasting statistics and when we look at specific historical episodes.

2.7 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 takes into account possible interdependencies in the cross section and allows for time variations in the evolution of the parameters over time. The prior shares features with those of Lindley 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.

We analyze several special cases of our specification and compute Bayesian estimators for the mean parameter in the cross section and for the individual coefficients. In some cases analytical formulas for the posterior mean are available using standard formulas. Whenever the parameters of the prior are unknown, we employ the predictive density of the model to estimate them and plug-in our estimates in the relevant formulas in an empirical Bayes fashion.

In the case of fully hierarchical priors, a Markov Chain Monte Carlo method (the Gibbs sampler) is employed to calculate posterior distributions. Such an approach is particularly useful in our setup since it exploits the recursive features of the posterior distribution. Recursive formulas for multistep, multiunit forecasts, consistent with the information available at each point in time, are provided using the posterior of the parameters or the predictive density of future observations. The predictive density of future observation is also used to compute turning point probabilities.

To illustrate the performance of the proposed approach, we apply the methodology to the problem of predicting output growth, of forecasting turning points in output growth and computing the probability of a recession in the G-7 using a three variables (output growth, real stock returns and real money growth) for each country in the panel. To evaluate the model we also provide a forecasting comparison with other specifications suggested in the literature. We show that our panel VAR approach improves over existing univariate and simple BVAR models when we measure the forecasting performance using the Theil-U and the MAD criteria both at the one step and at the four steps horizons. The improvements are of the order of 5-10% with the Theil-U and about 2-4% with the MAD. The forecasting performance of our specification is also slightly better then the one of a BVAR model which mechanically extends the Litterman prior to the panel case. In terms of turning point prediction, the two versions of our panel approach 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 high probability of a downturn at 90:3, the date selected by the NBER committee to terminate the long expansion of the 80's. In this instance, our approach is competitive with the best and avoids the false alarms that other approaches produce at other dates.

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 the nature of interdependencies, the similarities in the statistical model across units and the existence of time variation in the coefficients. Extensions of the theory outlined here include the formulation of interesting hypothesis on the nature of the interdependencies, on the similarities across units and on time variations and the development of tools to undertake structural identification in these models. The work of Sims and Zha (1998) is the starting point for extensions in this latter case.

2.8 Appendix

2.8.1 Proof of proposition 1

(i) Notice that (2.14) and (2.15) can be written as

$$\bar{\alpha}^* = \Psi^* \left[Z_t' \left(Z_t \hat{\Omega}_{t|t-1} Z_t' + \Sigma_u + W_t \Delta W_t' \right)^{-1} \left(Y_t - Z_t \hat{\lambda}_{t|t-1} \right) + \Psi^{-1} \mu \right] \tag{2.31}$$

$$\Psi^* = \left[\Psi^{-1} + Z_t' \left(Z_t \hat{\Omega}_{t|t-1} Z_t' + \Sigma_u + W_t \Delta W_t' \right)^{-1} Z_t \right]^{-1} \tag{2.32}$$

Setting $\Psi^{-1} = 0$, the result follows.

(ii) The posterior distribution of α is normal with mean $\alpha^* = S_N \mu + \phi_{12} \phi_{22}^{-1} [Y_t - Z_t(\mu + \hat{\lambda}_{t|t-1})]$ and variance $V_\alpha^* = \phi_{11} - \phi_{12} \phi_{22}^{-1} \phi_{21}$ where $\phi_{11} = (S_N \Psi S_N' + \Delta)$; $\phi_{12} = (S_N \Psi S_N' + \Delta) W_t'$; $\phi_{21} = W_t \phi_{11}$; $\phi_{22} = W_t \phi_{11} W_t' + Z_t \hat{\Omega}_{t|t-1} Z_t' + \Sigma_u$. In a more compact way V_α^* can be written as

$$V_\alpha^* = \left[W_t' \left(\Sigma_u + Z_t \hat{\Omega}_{t|t-1} Z_t' \right)^{-1} W_t + \left(S_N \Psi S_N' + \Delta \right)^{-1} \right]^{-1}. \tag{2.33}$$

Notice that $(S_N \Psi S_N' + \Delta)^{-1} = \Delta^{-1} - \Delta^{-1} S_N \left(S_N' \Delta^{-1} S_N + \Psi^{-1}\right)^{-1} S_N' \Delta^{-1}$. Hence, for $\Psi^{-1} = 0$, this expression is equal to F. Moreover the posterior mean can be written as

$$\alpha^* = V_\alpha^* \left[W_t' \left(\Sigma_u + Z_t \hat{\Omega}_{t|t-1} Z_t'\right)^{-1} \left(Y_t - Z_t \hat{\lambda}_{t|t-1}\right) + \left(S_N \Psi S_N' + \Delta\right)^{-1} S_N \mu \right] \tag{2.34}$$

Using the previous result and the fact that $F S_N = 0$, the result follows.

(iii) The posterior mean and variance of λ_t can be written as

$$\lambda_t^* = \Omega_t^* \left\{ Z_t' \left[W_t \left(S_N \Psi S_N' + \Delta\right) W_t' + \Sigma_u \right]^{-1} (Y_t - Z_t \mu) + \hat{\Omega}_{t|t-1}^{-1} \hat{\lambda}_{t|t-1} \right\} \tag{2.35}$$

$$\Omega_t^{*-1} = Z_t' \left[W_t \left(S_N \Psi S_N' + \Delta\right) W_t' + \Sigma_u \right]^{-1} Z_t + \hat{\Omega}_{t|t-1}^{-1} \tag{2.36}$$

The matrix $W_t \left(S_N \Psi S_N' + \Delta\right) W_t' + \Sigma_u$ can be written as $Z_t \Psi Z_t' + (W_t \Delta W_t' + \Sigma_u)$ and its inverse is equal to $M^{-1} - M^{-1} Z_t \left(Z_t' M^{-1} Z_t + \Psi^{-1}\right)^{-1} Z_t' M^{-1}$ where $M = (W_t \Delta W_t' + \Sigma_u)$. Setting $\Psi^{-1} = 0$, the last expression reduces to $M^{-1} \left[I - Z_t \left(Z_t' M^{-1} Z_t\right)^{-1} Z_t' M^{-1} \right]$. Premultiplying this matrix by Z_t', we get a zero matrix. Hence, from (2.35) and (2.36) $\Omega_t^{*-1} = \hat{\Omega}_{t|t-1}^{-1}$ $\lambda_t^* = \hat{\lambda}_{t|t-1}$ and the posterior distribution for λ_t is just equal to the prior.

2.8.2 Proof of proposition 2

Recall that $\Sigma_\varepsilon^{-1} = 0$ implies $\hat{\Omega}_{t|t-1}^{-1} = 0$.

(i) Consider the posterior moments (2.31) and (2.32). $[(W_t \Delta W_t' + \Sigma_u) + Z_t \hat{\Omega}_{t|t-1} Z_t']^{-1} = M^{-1} - M^{-1} Z_t \left(Z_t' M^{-1} Z_t + \hat{\Omega}_{t|t-1}^{-1}\right)^{-1} Z_t' M^{-1}$ where M was previously defined. When $\hat{\Omega}_{t|t-1}^{-1} = 0$, this reduces to $M^{-1} - M^{-1} Z_t \left(Z_t' M^{-1} Z_t\right)^{-1} Z_t' M^{-1}$ which gives a zero matrix if premultiplied by Z_t'. Consequently $\Psi^* = \Psi$ $\bar{\alpha}^* = \mu$ and the posterior distribution of $\bar{\alpha}$ is just equal to its prior.

(ii) Consider the posterior moments (2.33) and (2.34). When $\hat{\Omega}^{-1}_{t|t-1} = 0$, $\left(\Sigma_u + Z_t\hat{\Omega}_{t|t-1}Z_t'\right)^{-1} = S = \Sigma_u^{-1} - \Sigma_u^{-1}Z_t\left(Z_t'\Sigma_u^{-1}Z_t\right)^{-1}Z_t'\Sigma_u^{-1}$. Substituting into (2.33) and (2.34), gives the result.

(iii) The proof of this statement comes straight from (2.35) and (2.36).

2.8.3 Definition of the matrices for the Gibbs sampler

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

where $\mathbf{Y}_t$ is $N\times G$, $\mathbf{B}_t$ is $N\times Gk$ and $\mathbf{W}_t = \left(I_G\otimes X_t'\right)$. Model (2.1) is just a raw vectorization of $\mathbf{Y}_t = \mathbf{B}_t\mathbf{W}_t' + \mathbf{U}_t$, where $\mathbf{B}_t = \left[vecr\left(B_{1t}\right), ..., vecr\left(B_{Nt}\right)\right]'$. Here $vecr\left(\right)$ is the row vectorization of a matrix; $B_{it} = A_i + \Lambda_tE_i$ is a $G\times k$ matrix and the parameter vectors α_i and λ_t in (2.4) and (2.6) are the row vectorizations of A_i and Λ_t respectively.

Tables

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
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

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 excheangeable 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.

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

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 excheangeable 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. 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	EXCHANGEABLE	g-PRIOR	PBVAR	PANEL1	PANEL2
89:1	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
89:3	0.020	0.010	0.005	0.010	0.200	0.250	0.010	0.250	0.230
89:4	0.780	0.590	0.625	0.815	0.370	0.280	0.070	0.210	0.470
90:1	0.200	0.375	0.365	0.160	0.070	0.050	0.070	0.230	0.040
90:2	0.000	0.005	0.000	0.000	0.070	0.080	0.040	0.220	0.030
90:3*	0.645	0.660	0.700	0.660	0.320	0.360	0.820	0.300	0.550
90:4	0.005	0.010	0.030	0.015	0.280	0.380	0.040	0.250	0.210
91:1	0.000	0.005	0.000	0.003	0.230	0.050	0.130	0.240	0.020
91:2	0.000	0.000	0.000	0.000	0.170	0.060	0.000	0.250	0.000
91:3*	0.005	0.015	0.000	0.000	0.180	0.490	0.790	0.230	0.630
91:4	0.015	0.005	0.005	0.035	0.250	0.350	0.080	0.240	0.320

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

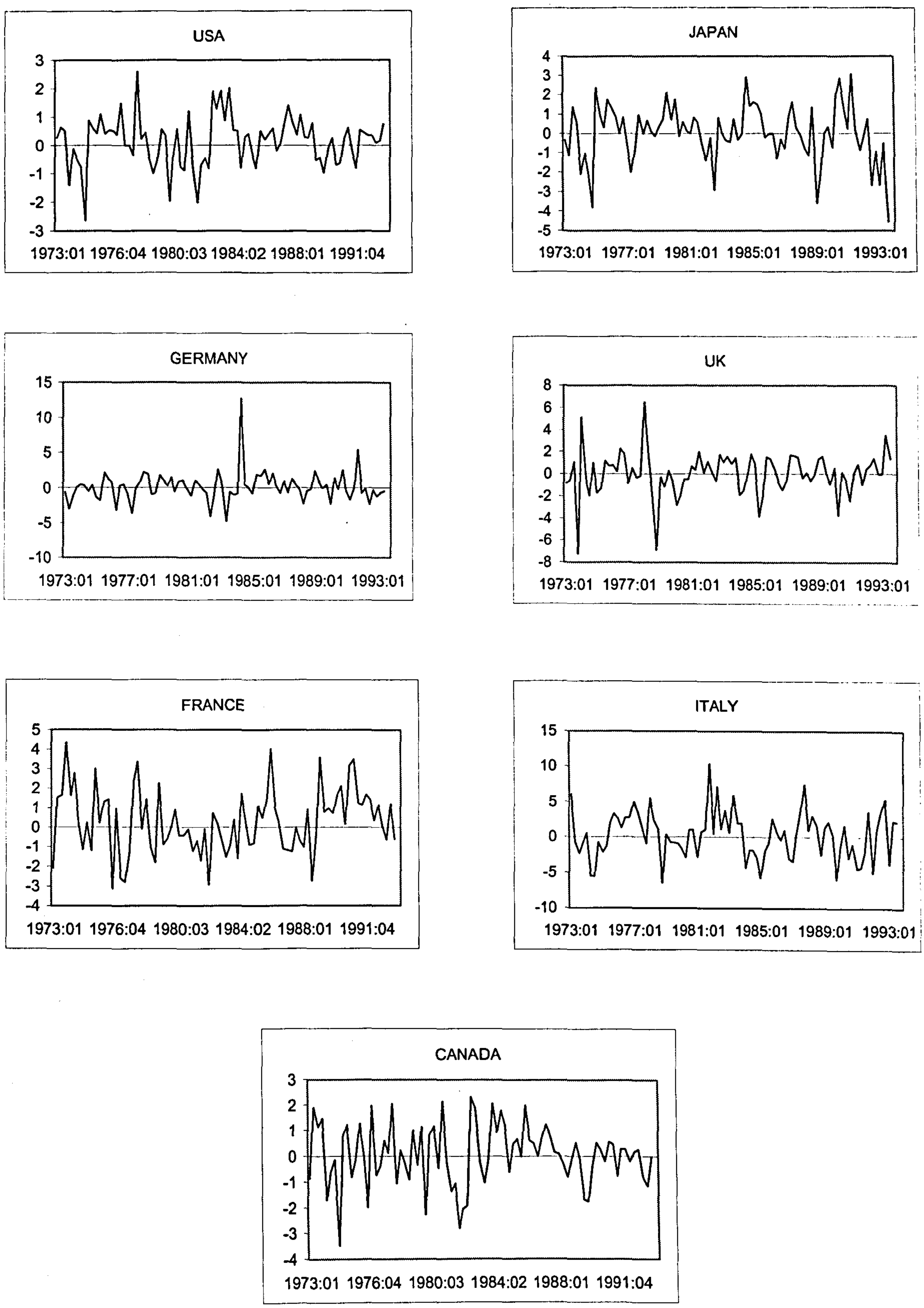


Figure 2: Real stock returns, quarterly data, 1973:I-1993:IV

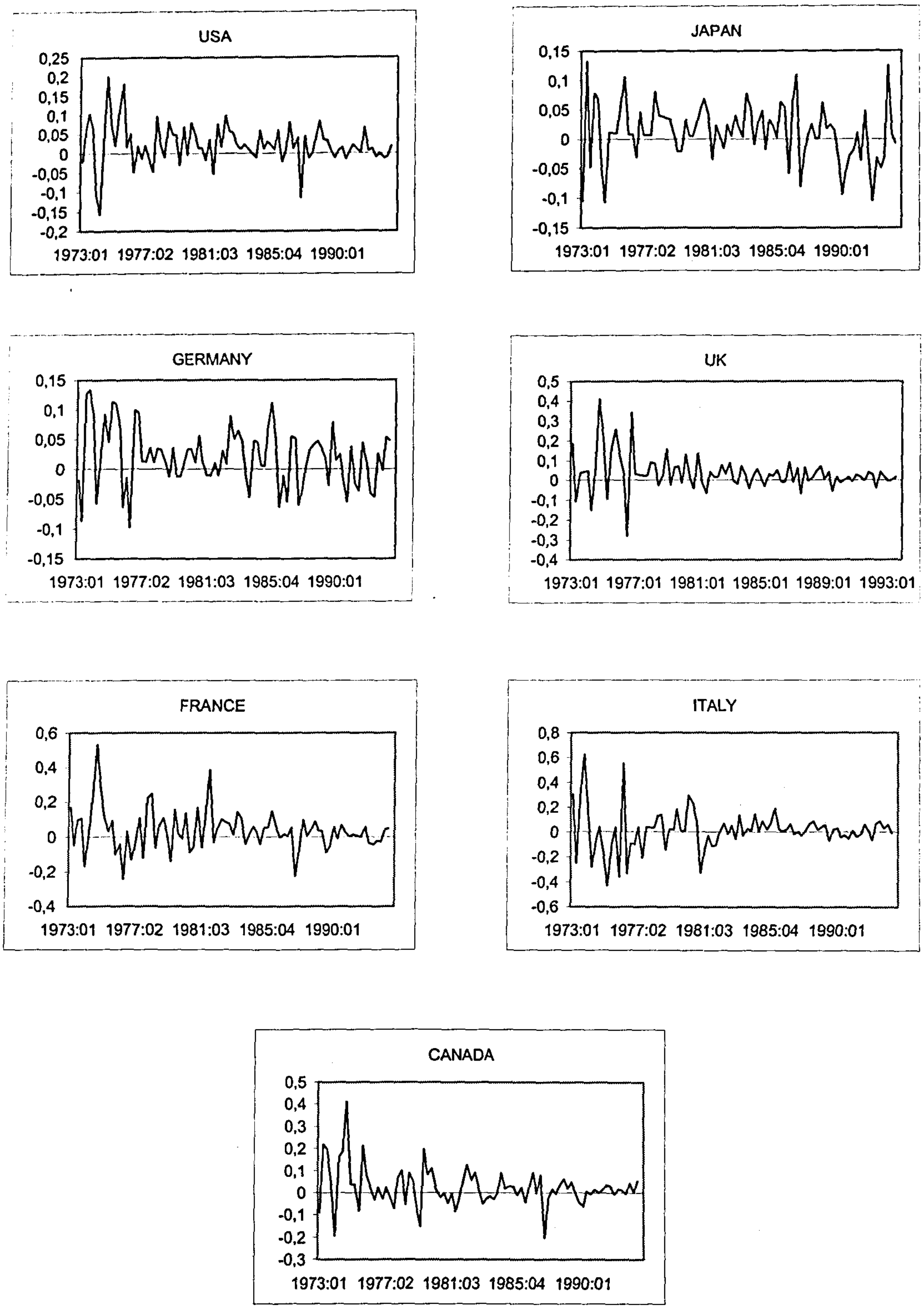
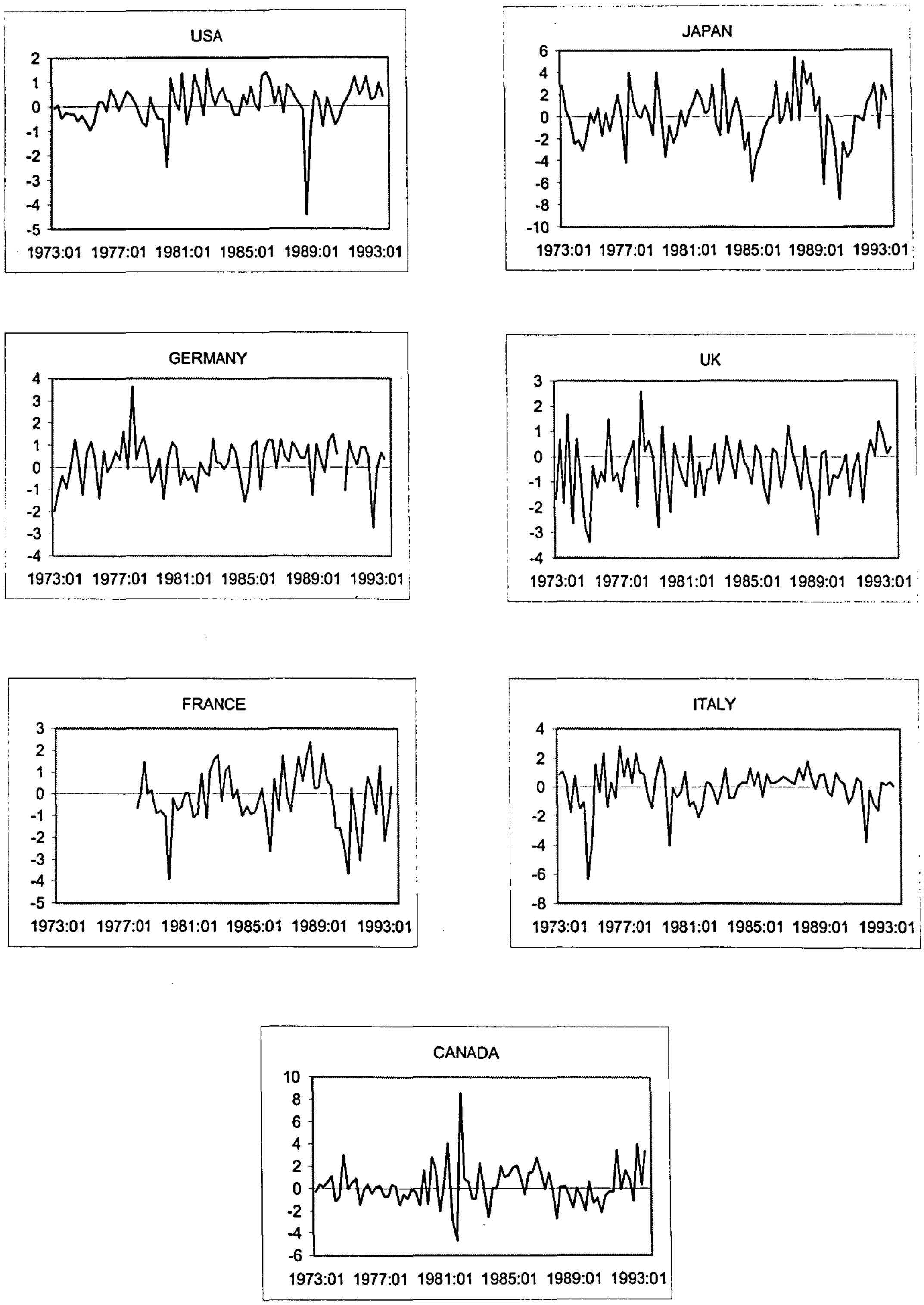


Figure 3: Real money growth, quarterly data, 1973:I-1993:IV



REFERENCES

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

Ballabriga F., M. Sebastián, and J. Vallés (1999), 'European asymmetries', *Journal of International Economics,* 48, 233-253.

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

Canova, F. (1995), 'VAR: Specification, estimation, testing, forecasting', in H. Pesaran and M. Wickens (eds.) *Handbook of Applied Econometrics,* Blackwell Press.

Canova, F. and A. Marcet (1997), 'The poor stay poor: non–convergence across countries and regions', forthcoming, *Review of Economic Studies.*

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

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

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

Chib, S. and E. Greenberg (1995), 'Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models', *Journal of Econometrics,* 68, 409–431.

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

Doan, T., R. Litterman and C. Sims (1984), 'Forecasting and conditional projections using realist prior distributions', *Econometric Review,* 3(1), 1–100.

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.

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.

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.

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

Hoffmaister, W. and J.F.Roldós (1997), 'Are business cycles different in Asia and Latin America?', *IMF Working Paper* no. 97/79.

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

Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu (1997), 'Bayes estimation of short run coefficients in dynamic panel data models', *mimeo*, Cambridge University.

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

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

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

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

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, 1–41.

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

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

Pesaran M.H. and R.Smith (1995), 'Estimating long run relationships from dynamic heterogeneous panels', *Journal of Econometrics*, 68, 79–113.

Rebucci, A (1998), 'External shocks, macroeconomic policy and growth: a panel VAR approach', *Global Economic Institutions*, Working Paper 40.

Sims, C. and T.Zha (1998), 'Bayesian methods for dynamic multivariate models', *International Economic Review*, 39, 949-979.

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

Waggoner, D.F. and T.Zha (1998), 'Conditional Forecasts in Dynamic Multivariate Models', *Federal Reserve Bank of Atlanta*, Working Paper n. 98–22.

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

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

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.

Chapter 3

Asymmetries in the Transmission Mechanism of European Monetary Policy

Ésa es la mejor enseñanza de la presente historia: que no hay contradicción duradera porque el destino no es inteligente. Y que quien crea lo contrario sólo persevera en su debilidad.

(Juan Benet, *En la Penumbra*)

3.1 Introduction

The European Central Bank (ECB) has already moved interest rates several times since it started to operate in January 1999 and yet nobody knows what the magnitude and timing of its actions actually are. What are the effects on prices and output of a change in the common short-term interest rate? How long do these effects take to materialize? Are there differences in the impact across European countries and regions? Are these differences changing over time? Most of these questions have already been asked in the literature. However, the answers provided so far are not entirely satisfactory.

Monticelli and Tristani (1999), for instance, suggest to start considering the European Monetary Union (EMU) as a composite economic system rather than a collection of countries. They analyze the impact of monetary policy on what they call the 'EMU-wide economic system' by estimating a Structural VAR (SVAR) with a GDP-based weighted average of individual time series of the member countries. If the transmission mechanism of monetary policy is similar across European countries, this approach provides a measure of the European-wide effects of monetary policy which is as good as those obtained with alternative estimation methods. But if the transmission mechanism does differ across countries, i.e., if there are cross country differences in the effects of monetary policy, the 'composite' approach is not correct. In this case, as shown by Pesaran and Smith (1995) for standard dynamic panel data models and discussed in Rebucci (2000) for panel VAR specifications, aggregation of individual time series biases the estimates obtained, and the European-wide impact of monetary policy must be measured either by aggregating individual time series estimates or by allowing for explicit variation in the parameters across countries. Before attempting to measure the system-wide effects of a 'synthetic' common monetary policy, therefore, one should try to establish whether or not there are differences across countries in the transmission mechanism of monetary policy. The current consensus view is that, indeed, there are differences across European countries which are likely to decrease over time as real, and especially financial, convergence proceeds.

The existence of some degree of heterogeneity in the transmission mechanism of European monetary policy is supported by a large, albeit sometimes contradicting, body of empirical evidence (see Guiso et al, 2000, among others). Gerlach and Smets (1995), for example, find very different results depending on the type of experiment they run. In their study, the effects on GDP of a one period, one standard deviation shock to short-term interest rates are broadly similar across Germany, France, and Italy. However, when they simulate a 100 basis points increase in interest rates sustained for two-years, they find that German GDP falls almost twice as much as that of France and Italy. On the other hand, Ramaswamy and Sloek (1997) find that the effects on GDP of a one period, one standard deviation shock to short-term interest rates in Germany, the U. K., Finland, the Netherlands, Austria, and Belgium take almost twice as long to occur, but are almost twice as deep as in Denmark, France, Italy, Portugal, Spain and Sweden. Furthermore, Dornbusch, et al. (1998) find evidence suggesting that the long-run effects on output of the anticipated component of monetary policy in Germany, France, Italy, Spain, the U. K., and Sweden are quantitatively sizable and heterogeneous, while the short run effects are quantitatively smaller but relatively homogenous across these countries. Indeed, standard macroeconomic theory predicts that monetary policy is neutral in the long run, and thus its effects should be rather homogenous across countries over this time horizon. As noted by Dornbusch et al. (1998), there is also a difference between the results based on *large econometric models* and those based on *small econometric models,* whereas small (VAR-type) econometric models do not seem to be able to detect statistically significant cross-country differences in the monetary transmission mechanism, contrary to the evidence coming from large country-specific econometric models.

There is also no clear evidence that these differences are decreasing over time. On the contrary, recent work by Cecchetti (1999) shows that they might persist for a long time because they are due to differences in the financial structure, which in turn are rooted in the legal framework of individual countries. If these differences were to persist for sometime, the ECB's life may become quite complicated as pointed out by Dornbusch et al. (1998) and explicitly modelled by Giovannetti and Marimon (1998). Giovannetti and Marimon (1998) develop a dynamic general equilibrium model where economies differ with respect to the relative efficiency of financial intermediaries and show that, if these differences persist, conflicts of interests in pursuing a common monetary policy may indeed arise. Therefore, it would be useful to have some idea not only on the magnitude of these differences but also on their degree of persistence over time.

All this literature, in addition, may be subject to the Lucas' critique.

We propose to overcome some of these difficulties, by rephrasing the questions above in a dynamic heterogenous panel data model, recently proposed by Canova and Ciccarelli (2000). This is a flexible empirical framework where, in addition to interdependencies among individual units, the parameter of the transmission mechanism can be explicitly allowed to change both across times and individual units. Obviously, such a framework cannot be estimated without introducing some kind of restriction, because of the very large number of parameters involved. Canova and Ciccarelli address this issue by taking a Bayesian approach to estimation and specifying the econometric model in terms of few hyperparameters. This framework allows for the maximum degree of heterogeneity, and thus sets the stage for testing alternative homogeneity assumptions, including parameters' stability over time and equality across individuals. In addition, it allows to recover and measure European-wide behavioral relations regardless of the actual degree of heterogeneity present in the data, and is not subject to the Lucas' critique.

In this version of the paper, we consider a small group of core European countries, (Germany, Italy, France, and Spain), using monthly data from 1985 to 1998. The econometric specification is the same for all countries considered. We measure monetary policy by estimating an empirical model of the behavior of these countries' central banks, and then assess the impact of monetary policy on economic activity by estimating a system of dynamic output equations as done by Dornbusch et al. (1998) and Peersman and Smets (1998). We control for both intra-Europe exchange rate movements, and heterogeneity of central banks' preferences along the line pursued by Sala (2000) and Clements and Kontolemis (2001).

Consistently with the consensus view in the literature, we show that there are cross-country differences in the transmission mechanism of European monetary policy, both with regards to country specific and common monetary policy shocks. However, we show also that these are differences of timing rather than magnitude of the impact of monetary policy; the cumulative effect of both country specific and common shocks, in fact, are rather homogenous, especially when parameters' variation across time periods is allowed for. Differently from the consensus view in the literature, and consistently with what suggested by Cecchetti (1999), we provide evidence showing that the transmission mechanism of monetary policy is changing over time in core European countries, but the degree of heterogeneity of the response of these economies to monetary shocks is not decreasing over time. We finally provide preliminary evidence on the European-wide impact of monetary policy, showing that the effects of monetary policy take about 6-7 months to appear, peak after 12 months, and vanish within 24 months.

The paper is organized as follows. In the next section we present and discuss the econometric framework used. In section 3 we present the empirical results. These include: key estimated parameters of the reaction functions; country specific and common monetary policy shocks obtained from the data; the evidence on their effects on economic activity and the degree of homogeneity across countries and stability over time of these effects; and finally, a first set of results on the European-wide impact of monetary policy. Section 4 concludes, while details of the estimation techniques and the data used are given in appendix.

3.2 The econometrics

Ideally, one would like to apply the empirical framework proposed by Canova and Ciccarelli (2000) to a small SVAR for output, inflation, interest rates, and the exchange rate. However, the identification of VARs estimated with panel data is a tricky business because of the restrictions on the variance-covariance matrix of the residuals, and is beyond the scope of this paper. Here, we take the two stage approach followed by Dornbusch, Favero, and Giavazzi (1998) (DFG henceforth) and do not model inflation and the exchange rate explicitly. In the first stage, a measure of monetary policy is extracted from the data by estimating a reaction function for each central bank, allowing for simultaneity and interdependence in short-term interest rates, and parameters' variation across countries and across time periods. In the second stage, the impact of monetary policy is analyzed by estimating a dynamic equation for a standard measure of real economic activity, allowing also for parameters' variation both across countries and time periods. In the following two sub sections, we present the econometric model of the reaction functions and output equations in turn.

3.2.1 Measuring monetary policy

Specification

The behavior of the four European central banks considered is modelled empirically by means of the following structural VAR (SVAR):[1]

$$A_t(L)\,R_t = B_t(L)\,W_t + D_t + U_t. \tag{3.1}$$

where $R_t = [r_{1,t}, \cdots, r_{4,t}]'$ is a 4 x 1 vector of instruments of monetary policy, $W_t = [w_{1,t}, \cdots, w_{4,t}]'$ is a 4 x 1 vector of final objectives of monetary policy, $A_t(L)$ and $B_t(L)$ are polynomial matrices in the lag operator L with lag length p, and D_t is a 4 x 1 vector

[1] As pointed out by DFG, this specification can be interpreted as the reduced form of a forward-looking structural model, or as a system of backward-looking reaction functions (see DFG, 1998, p.16, footnote 12). See Clarida and Gali (1997) on the relative performance on these two alternative specifications.

of constants. $U_t = [u_{1,t}, \cdots, u_{4,t}]'$ is a vector of monetary policy shocks assumed to be Gaussian with

$$E\left[U_t U_t' \mid Z_{t-s}, \quad s \geq 0\right] = I, \quad E\left[U_t \mid Z_{t-s}, \quad s \geq 0\right] = 0, \text{ for all } t,$$

where Z_t contains lagged R_t and contemporaneous and lagged W_t, with I denoting the identity matrix. As noted by Clarida et al. (1997), who estimated forward looking reactions functions for the US, Germany, and Japan, and a small group of European central banks with post-1979 data, under the assumption that the central bank's supply of reserves is infinitely elastic and in the absence of exchange rate risk premia, $u_{i,t}$ should be theoretically equivalent to monetary shocks obtained from standard SVAR models. Thus, under these assumptions the estimated residual of equation (3.1), $\widehat{u}_{i,t}$, may be interpreted as the pure random, or unexpected, component of monetary policy. Shocks to money demand not fully accommodated by the central bank or exogenous shocks to the exchange rate premium, however, may invalidate this interpretation.

We use short term interest rates as monetary policy instruments. Each element of the vector of final objectives, $w_{it} = [(\pi_{i,t} - \pi_i^*), (\mathrm{y}_{i,t} - \mathrm{y}_i^*), (e_{it} - e_i^*), \sigma_{i,t}]'$, contains inflation ($\pi$), output (y) and the nominal exchange rate (e) in percent deviation from trend (π^*, y^*, e^*, respectively), and a measure of the intra-month exchange rate volatility (σ) to control for shocks to exchange rate risk premia. The dimension of W_t therefore is 16 x 1.[2]

This specification imposes very few a priori restrictions on the system of reaction functions. First, the model allows for contemporaneous and lagged interdependence among short term interest rates of different countries. Second, given that the degree of each member's commitment to EMS has varied over time, we do not impose that central banks target German variables as done by DGF, but rather leave $B_t(L)$ unrestricted and let the data

[2] See the data appendix for more details on the data and the transformation used, including the definition of π^*, y^*, and e^*.

reveal which objective was actually pursued in a particular time period. Similarly, $A_t(L^P)$ is unrestricted for $p \neq 0$. Third, all parameters except those governing contemporaneous causation among short-term interest rates can vary over time, allowing for the possibility of change in the central banks' behavior over the sample period considered.[3] However, we do impose an arbitrary lag length restriction assuming that $p = 1$; thus, that one lag is enough to obtain white noise residuals.

Identification

The model's identification exploits the Bundesbank's leading role under EMS and the fact that other European countries considered have comparable size. Specifically, we place the German short term interest rate first in the vector R_t, assuming that it affects other European interest rates contemporaneously without being affected by them, and then we assume that the impact of an increase in interest rates in country i on country j is the same as the impact of an increase in country j on country i.

Formally, the leader-follower behavior characterizing EMS is translated into the following block recursive structure for $A(0)$, the coefficient matrix of L^0 in $A_t(L)$:

$$A(0) = \begin{bmatrix} A_{11}(0) & 0' \\ A_{21}(0) & A_{22}(0) \end{bmatrix} \tag{3.2}$$

where $A_{11}(0)$ is 1×1, $A_{21}(0)$ is 3×1,and $A_{22}(0)$ is 3×3. This gives us three restrictions. The remaining three restrictions needed are obtained imposing that $A_{22}(0)$ is symmetric: these six restrictions identify the model exactly regardless of the order of the other interest rates in R_t.[4]

[3] Assuming $A(L^0)$ to be constant over time renders the posterior distributions analytically tractable and is equivalent to assume homoscedasticity of the structural residuals, given that the model is exactly identified.
[4] See Amisano and Giannini (1997, p. 166-67).

The SVAR model (3.1), therefore, can be rewritten as:

$$\begin{bmatrix} A_{11}(0) & 0' \\ A_{21}(0) & A_{22}(0) \end{bmatrix} \begin{pmatrix} R_{1t} \\ R_{2t} \end{pmatrix} = \begin{bmatrix} A_{11}(L) & A_{12}'(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix}_t \begin{pmatrix} R_{1t} \\ R_{2t} \end{pmatrix}$$

$$+ \begin{bmatrix} B_{11}(L) & B_{12}(L)' \\ B_{21}(L) & B_{22}(L) \end{bmatrix}_t \begin{pmatrix} W_{1t} \\ W_{2t} \end{pmatrix} + D_t + \begin{pmatrix} U_{1t} \\ U_{2t} \end{pmatrix} \tag{3.3}$$

where R_{1t}, W_{1t}, and U_{1t} are the German monetary policy instrument, objectives, and shock, respectively; while R_{2t}, W_{2t}, and U_{2t} represent the vectors containing the same variables for other countries.[5]

Estimation

Bayesian estimation of (3.3) exploits its block recursive structure. Following Zha (1999), let k_j be the total number of right-hand-side variables per equation in the jth block, and G_j the number of equations (i.e., countries) in block j. If we normalize (3.3) pre-multiplying it by

$$A_d^{-1}(0) = \begin{bmatrix} A_{11}^{-1}(0) & 0' \\ 0 & A_{22}^{-1}(0) \end{bmatrix}$$

and rearrange terms, the model can be divided into two blocks:

$$R_{jt} = \mathbf{Z}_{jt}\boldsymbol{\delta}_{jt} + v_{jt} \qquad j = 1, 2, \quad \text{for all } t \tag{3.4}$$

---

[5] Note that premultiplying (3.3) by

$$A_d^{-1}(0) = \begin{bmatrix} A_{11}^{-1}(0) & 0' \\ 0 & A_{22}^{-1}(0) \end{bmatrix}$$

and rearranging terms, a standard simultaneous equation model (SEM) is obtained with

$$A(0) = \begin{bmatrix} 1 & 0' \\ \alpha & I \end{bmatrix} \qquad \alpha = A_{22}^{-1}(0)\, A_{21}(0)$$

and

$$var(v_t) = \Sigma_v = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_{22} \end{bmatrix}$$

where $v_t = A_d^{-1}(0)\, U_t$ is the vector of normalized disturbances. This is equivalent to the identification used by DFG.

where R_{1t} and R_{2t} were defined above, $\mathbf{Z}_{jt} = diag\left[Z_{j1}, Z_{j2}, ..., Z_{jG_j}\right]$—with Z_{jg} (of dimension $T \times k_j$) containing lagged endogenous variables, exogenous and deterministic variables of block j, and R_1 (only in block $j = 2$)—$\boldsymbol{\delta}_{jt} = \left[\delta_{j1}, \delta_{j2}, ..., \delta_{jG_j}\right]$—with δ_{jg} (of dimension $k_j \times 1$) containing the model's parameters—and $v_{jt} = A_{jj}^{-1}(0)\, U_{jt}$ are normalized (block orthogonal) disturbances with

$$v_{jt} \sim N\left(0, \Sigma_{jj}\right), \qquad \Sigma_{jj} = A_{jj}^{-1}(0)\, A_{jj}^{-1}(0)'.$$

Normalization is a device to split the model (3.1) in the two blocks of equations (3.4): the reaction function of the Bundesbank ($j = 1$), and the reaction functions of other European central banks ($j = 2$). Given that, in our case, $G_1 = 1$, $G_2 = 3$, $p = 1$ (the number of lags of the endogenous and exogenous variables), $d = 1$ (the constant), and the number of exogenous variables for each country of block is 4, we have a total of

$$k_2 = (G_1 + G_2) \cdot p + (G_1 + G_2) \cdot 4 \cdot (p + 1) + d + 1 = 38$$

and

$$k_1 = (G_1 + G_2) \cdot p + (G_1 + G_2) \cdot 4 \cdot (p + 1) + d = 37$$

parameters for each equation of block j, whereas the larger number of parameters in the second block is due to the fact that the German rate enters contemporaneously in other equations.

Bayesian estimation of (3.4) is then obtained by means of Kalman filtering and Gibbs sampling techniques, modified as suggested by Chib and Greenberg (1995) to take into account the presence of time variation in the model's parameters: a joint prior on $(\boldsymbol{\delta}_{jt},\ \Sigma_{jj})$ is combined with the likelihood of the data to recover the posterior distributions of interest. Given that the matrix $A(0)$ is exactly identified, we can recover the posterior distribution of the structural parameters, and hence the posterior distribution of the structural residuals, for each iteration of the Gibbs sampler. The mean of the empirical distribution of these residuals is then taken as our measure of monetary policy.[6]

The structural residuals of (3.3) can be used to study and compare across countries the transmission mechanism of country specific monetary policies: policies which reflect, or are the result of, each country's individual preferences over their set of possible final objectives. A key feature of EMU, however, is that individual members' preferences and reaction functions have been substituted by, or aggregated into, those of the ECB and its policymaking bodies.[7] Therefore, in order to approximate as closely as possible the conditions prevailing under EMU, one would also like to investigate the response of these economies to a common or coordinated monetary policy: a policy which reflects, or is the result of, the aggregation of countries' preferences over the possible objectives of monetary policy. In our econometric framework, a common monetary policy could be defined either by setting adequate identifying restrictions on the empirical model of central banks' reaction functions (3.3), or constraining the transmission mechanism of country specific monetary policies through restrictions on (3.8) as done by DFG, or by trying to extract common components (i.e., common monetary policy shocks) directly from country specific measures of monetary policy as done by Sala (2000).

Given the difficulties of attempting to identify a common monetary policy in (3.3), and the computational costs of imposing restrictions on the transmission mechanism of country specific shocks in (3.8),[8] we have followed a straightforward principal component analysis approach along the lines pursued by Sala (2000) and Clements and Knotolemis (2001). More specifically, as a measure of a common monetary policy shock, we take the first principal

---

6 See appendix 3.5.1 for more details.
7 See Clements and Knotolemis (2001) for a more rigorous analysis of this point.
8 When the model (3.3) is overidentified a joint prior must be specified on $(\delta_{jt}, A_{jj}(0))$. The posterior distribution of $A_{jj}(0)$ then becomes non-standard and can be obtained only with a second order Taylor expansion around the maximum of the Likelihood, as explained in Zha (1999).

component of the reduced form residuals (i.e., non orthogonalized residuals) of (3.3).[9] Even though this measure might be crude, it should provide at least a term of comparison for our analysis of the effects of country specific monetary policy shocks.

3.2.2 The transmission mechanism of monetary policy

Specification

The impact of monetary policy on economic activity is modelled empirically through a system of output equations in which annual real output growth is regressed on our measure of monetary policy and a set of control variables. Specifically, for each country i, we specify the following equation:

$$y_{it} = X'_{it}\beta_{it} + \varepsilon_{it} \tag{3.5}$$

where y_{it} is the 12-month growth of industrial production of country i at time t, $X'_{it} = [\widehat{u}_{it-l_i}, x_{it}]'$ is a $1 \times k$ vector of regressors with $\widehat{u}_{it-l_i}$ denoting our measure of monetary policy and x_{it} containing a set of control variables—lagged output growth of all countries considered, the exchange rate of country i vis-a-vis the DM and the US dollar, and the inflation rate. In equation (3.5), $\beta_{it} = [\beta^1_{it}, \beta^2_{it}]'$ is the $k \times 1$ parameters' vector with β^1_{it} denoting the coefficients of $\widehat{u}_{it-l_i}$ and β^2_{it} those of x_{it}. The econometric specification is the same for all countries considered and includes a constant, one lag of the endogenous variable and the control variables, and 24 lags of the monetary policy variable for a total of 31 regressors in each equation.

The specification of the system of output equations (3.5) allows the parameters' vector β_{it} to vary across countries and time periods. This is achieved by assuming that β_{it} is drawn from a common distribution across countries for each time t, but changing over time

[9] Principal component analysis is a standard econometric technique to extract common components from series of data. See Theil (1971), for a basic reference. Note that estimation of the reduced form of this model is identical to that of the structural form described in the text, except that it is not done by blocks.

according to a given law of motion. This assumption, which is referred to as 'exchangeability prior' in the Bayesian literature, allows the parameters of interest to be different across countries, though only as different draws from the same distribution and is both mathematically tractable and economically plausible.

Formally, for each country i and time t we assume that:

$$\beta_{it} = \theta_t + \zeta_{it} \qquad \zeta_{it} \sim N(0, b_o) \tag{3.6}$$

$$\theta_t = \theta_{t-1} + \eta_t \qquad \eta_t \sim N(0, B_1). \tag{3.7}$$

where b_o and B_1 denote the variance of the distribution of ζ_{it} and η_t respectively, and they are the same for all individual units. B_1 controls the systematic time variation of the parameters, whereas b_o controls the cross sectional and the erratic time variation.[10] If $B_1 = 0$, the parameters vary randomly across time and individual units and (3.6) becomes $\beta_{it} = \theta + \zeta_{it}\ \forall t$. On the other hand, when $b_o = 0$, no cross sectional heterogeneity is present and the parameter vector β_{it} is pooled towards a common mean changing (systematically) over time. In this case, (3.6) becomes $\beta_{it} = \theta_t\ \forall i$. When both B_1 and b_1 are zero, $\beta_{it} = \theta\ \forall i$ and t. The prior variances of η_t and ζ_{it}, therefore, provide a means to measure the degree of uncertainty on the mean of the parameters of interest across countries and time periods introduced in the model.[11] The assumptions in (3.6), however, are only priors which must

[10]Note that this specification of the law of motion of θ_t implies that the parameters have an unconditional mean equal to zero. An alternative assumption would be:

$$\theta_t = \rho\theta_{t-1} + (1-\rho)\bar{\theta} + \eta_t,$$

where $\bar{\theta}$ is the long run mean of θ_t. However, when we estimated the hyperparameter ρ by maximizing the sample likelihood of the model (3.8) for each country i, we found that the lowest value for ρ was 0.9985 (for Spain), while the highest was 1.0 (for Italy and France). Given this preliminary result, we decided to stick to the computationally simplier specification in (3.6).

[11]This specification is very similar to the one used by Canova and Ciccarelli (2000) who assume that $\beta_{it} = \alpha_i + \lambda_t$ with $\alpha_i = \bar{\alpha} + \phi_i$. Here, we do not identify α_i separately, but rather put it in the idiosyncratic component ζ_{it}, allowing the common component $\bar{\alpha}$ to drift with time (θ_t). Given the assumption of independence between the common time varying component and the country specific effect, the specification above would not allow to test for the persistence over time of any cross country difference. In our specification, time variation in the country specific effects allows the distribution of the parameters to change over time not only due to common shocks but also to idiosyncratic disturbances; as a result, the posterior distributions

be combined with the data to generate posterior distributions of the parameters of interest; and the moments of the posterior distributions of these parameters do not need to be the same as those characterizing the priors, as indeed we shall see in our empirical results.

Estimation

By stacking countries by row, we can write the set of output equations as a standard system of seemingly unrelated regressions (SUR)[12]

$$y_t = X_t\beta_t + \varepsilon_t, \qquad \varepsilon_t \sim N_g(0, \Omega). \tag{3.8}$$

In this equation, $X_t = diag\left[X'_{1t}, ..., X'_{gt}\right]$ is of dimension $G \times h$, where $h = G * k$ with $G = 4$ denoting the number of endogenous variables and $k = 31$ denoting the number of regressors in each equation, and $\beta_t = [\beta_{1t}, ..., \beta_{gt}]'$ is of dimension $h \times 1$. The evolution of the parameters' vector across times is governed by:

$$\beta_t = M_o\theta_t + \zeta_t, \qquad \zeta_t \sim N_h(0, B_o) \tag{3.9}$$

$$\theta_t = \theta_{t-1} + \eta_t, \qquad \eta_t \sim N_m(0, B_1) \tag{3.10}$$

where M_o is a column vector of G identity matrices of order k relating the regression vector β_t to the vector of common shift parameters θ_t of dimension $h \times 1$, and Ω, B_o, and B_1 are unknown variance-covariance matrices of ε_t, ζ_t and η_t, respectively. The latter three random variables are assumed mutually independent, implying that y_t is conditionally independent of θ_t, B_o, and B_1.[13]

Bayesian estimation of the hierarchical model (3.8-3.10) is performed with a procedure similar to the one used by Chib and Greenberg (1995) or Canova and Ciccarelli (2000):

of the parameters of each country may change over time also because of individual specific effects. This permits to check whether or not cross country differences in the transmission mechanism are changing over time.

[12]Note that for the application of the estimation results used, it is key to have no simultaneity in this system.

[13]Although normal distributions have been chosen to model the data and the parameters, the framework presented is flexible enough to accomodate also other distributional forms. See Hsiao et al. (1998) on potential problems deriving from assuming normality.

prior assumptions are set on the hyperparameters of the model (Ω, B_o, B_1) and then are combined with the information contained in the data (in the form of a likelihood function) to obtain posterior distributions. Given that analytical integration is not feasible, the Gibbs sampler is used to obtain numerically the posterior distributions of the parameters of interest.

Testing

Several hypotheses about parameters' homogeneity across countries and time periods can be performed on the posterior distributions of the parameters of interest. We are particularly interested in the overall degree of parameters' stability over time, in the presence of cross country differences in the transmission mechanism of monetary policy, and any tendency of these differences to change over time. More specifically, we want to test the null hypothesis that $B_1 = 0$, i.e., the absence of (systematic) time variation in the parameters, and that the diagonal elements of B_0 are equal to zero, i.e., the null hypothesis that the variances across countries of the posterior distributions of the transmission mechanism's parameters are homogenous, either over the entire sample considered or in each yearly subperiod.

The first hypothesis is tested as follows. B_1 depends on a single hyperparameter, ϕ. If the posterior distribution of ϕ is concentrated around values closer to zero than its prior, then the evidence supporting a time varying specification would be weak. This is checked by following Chib and Greenberg (1995) and calculating, for arbitrarily small values of ξ, the ratio:

$$z = \frac{\Pr(\phi \leq \xi \mid y) \Pr(\phi > \xi \mid y)}{\Pr(\phi \leq \xi) \Pr(\phi > \xi)}, \tag{3.11}$$

where $\Pr(\phi \leq \xi \mid y)$ denotes the conditional *posterior* probability that ϕ is less than ξ, while $\Pr(\phi \leq \xi)$ denotes the corresponding *prior* probability. The numerator is computed from the relative frequencies generated by the Gibbs sampler, while the denominator from the assumed prior distribution.

The presence of differences across countries in the transmission mechanism of monetary policy is performed using a testing procedure proposed by Ciccarelli (2000), which is an empirical-Bayes analogous of the classical Wald-test. Let's write the null hypothesis of homogeneity of the parameters of interest as a set of linear restrictions on the parameter vector β_t:[14]

$$\mathsf{R}\beta_t = \mathsf{r}, \quad \text{for each } t. \tag{3.12}$$

Conditional on the other parameters of the model, and given the specification above, the posterior distribution of β_t is:

$$\beta_t \sim N\left(\hat{\beta}_t, \widehat{\Omega}_t\right).$$

Thus, the conditional posterior distribution of $\mathsf{R}\beta_t$ is:

$$\mathsf{R}\beta_t \sim N\left(\mathsf{R}\hat{\beta}_t, \mathsf{R}\widehat{\Omega}_t\mathsf{R}'\right).$$

The test is based on the comparison of the following two quadratic forms

$$q_t = \left[\mathsf{R}\left(\beta_t - \hat{\beta}_t\right)\right]' \left[\mathsf{R}\widehat{\Omega}_t\mathsf{R}'\right]^{-1} \left[\mathsf{R}\left(\beta_t - \hat{\beta}_t\right)\right] \tag{3.13}$$

and

$$q_{1t} = (\mathsf{R}\beta_t - \mathsf{r})' \left[\mathsf{R}\widehat{\Omega}_t\mathsf{R}'\right]^{-1} (\mathsf{R}\beta_t - \mathsf{r}). \tag{3.14}$$

The former is (conditionally) distributed as a $\chi^2_{(d)}$, with d degrees of freedom, while the latter is just (3.13) with the restrictions.[15] Clearly, if the restrictions are true q_{1t} must have the same distribution as q_t.

---

[14]In our specific case, the restriction matrix $\mathsf{R}=[\mathsf{R}_{i,j}]$ has dimensions $(g-1) \times p_m \times gk$, where g and k have been defined before and p_m is the number of monetary policy coefficients restricted to be the same across countries. In particular, the null hipothesis that all parameters of the transmission mechanisms are equal across countries means $p_m = 24$. R will have 72 rows, whose values are 1 when $i = j$, -1 when $j = i + k$, and zero otherwise. The hypothesis that the impact of monetary policy at specific lags, or the cumulative effect after one and two years, are equal across countries can also be easily accommodated designing R accordingly.

[15]Note that we are not testing exact restrictions in (3.12), but rather whether or not $\mathsf{R}\beta_t$ is distributed around r *a posteriori*.

Given (3.13), the conditional posterior probability of $(\mathsf{R}\beta_t = \mathsf{r})$ for each t is necessarily related to the probability that, at each iteration of the Gibbs sampler, a $\chi^2_{(d)}$ assumes the value (3.14). Hence, the probability that a draw from a $\chi^2_{(d)}$ exceed the magnitude q_1 gives information on the probability that the random variable $\mathsf{R}\beta_t$ is as far from the posterior mean $\mathsf{R}\hat{\beta}_t$ as represented by the point $\mathsf{R}\hat{\beta}_t = \mathsf{r}$. In order to construct a rejection region, therefore, it is enough to compare these two distributions. The larger the distance between q and q_1, the greater is the probability, *a posteriori*, of rejecting the null.[16] The greater is the distance between the two posterior distributions, the more likely the restriction imposed is converting the reference distribution in a *non-central* one, and the more likely the null is false. The distance between these distributions can then be quantified using a standard Kolmogorov-Smirnov statistics.

The intuition is the same as in the classical Wald test, where one compares two distributions: one under the null, which is *asymptotically* $\chi^2_{(d)}$, and the other under the alternative which is a *non-central* $\chi^2_{(d)}$. The greater is the numerical value of the quadratic form in which the set of restrictions has been substituted, the more likely this value belongs to the distribution under the alternative. The empirical posterior distributions of q e q_1 are easily obtained from the Gibbs sampler. The posterior distribution of q may be seen as a reference distribution by construction because, given the normality assumption, each draw of the Gibbs sampler is from an *exact* $\chi^2_{(d)}$. The main difference with the Wald test is that here we know the exact distribution of q, which can be computed numerically and used to make probability assessments in a Bayesian fashion. In our case, the posterior distribution of (3.14) (and not just one value, as in the classical analysis) can also be computed and

[16]The procedure is explained in more detail and evaluated by means of Monte Carlo simulation in Ciccarelli (2000). He shows that the procedure scores very well both in terms of power and size, generally doing as well as a standard posterior odds (PO) ratio approach, or even better in cases where the degree of coefficient heterogeneity is not high. In addition this approach is easier to implement and, unlike the PO ratio test, it can be computed also when some prior in the hierarchy is diffuse.

compared with (3.13).

When the model is specified with time-varying parameters, we can easily compute empirical distributions for q and q_1 for each yearly subperiod considered. Hence, at each time t, we can test the null hypothesis of parameters' homogeneity across countries. If the parameters are changing over time, i.e., the posterior distribution of ϕ is not concentrated around 0, convergence would occur if q and q_1 get closer and closer as time goes by.

3.3 Empirical results

In this section we present estimation results relative to equation (3.3) and (3.8) respectively. We present first the estimated parameters and the residuals–our measure of monetary policy–of the reaction function of each central bank considered, and then parameter estimates and test statistics for the output equations–which capture the impact of monetary policy on economic activity. The estimation sample is 1985:01 through 1998:12. Kalman filter for the estimation of the reaction function is initialized from 1985:01 to 1990:12. Reported estimates then run from 1991:01 for the parameters of the reaction functions and, given the 24 lags of monetary policy shocks, from 1993:01 for output equation.

3.3.1 Central banks' reaction functions and monetary policy shocks

Figure 1 reports selected estimated parameters of the Bundesbank's reaction function. The most striking feature of these estimates is the apparent extent of parameters variation over time at the beginning of the sample period considered. Thus, suggesting that important behavioral changes were already taking place in the run up to EMU. The coefficient of the lagged endogenous variable (i.e., the coefficient of the German own lagged interest rate), for instance, seems to have increased by about 30 percent at the beginning of the 1990s to stabilize around 0.9 after 1992–implying a high persistence in short term interest rates after

1992-93. Other parameters of the German reaction function appear to follow broadly a similar time pattern except for that of the volatility of the DM/dollar rate which stabilizes only in 1995-96. The coefficients of the inflation, output, and nominal exchange rate gaps–the assumed final objective of German monetary policy–have the right signs (i.e., positive those of the inflation and exchange rate gaps and zero that of output), but are quantitatively slightly smaller than what found in previous studies.[17] Interestingly, the parameters of the volatility of the nominal exchange rate vis-a-vis the US dollar and other European countries appear larger than those of the inflation, output, and exchange rate gaps, though a direct comparison cannot be done because of different units of measure. Moreover, their time profile suggests that the Bundesbank's attention has shifted in the run up to EMU from the dollar value of the DM to the external value of the DM vis-a-vis other European currencies: German short-term interest rates start reacting negatively to the volatility of the Italian lira in 1991-92, and that of the French franc and the Spanish peseta thereafter, arguably in response to speculative persistent activity against these currencies.

Selected estimated parameters for the reaction functions of France, Italy, and Spain are also reported in Figure 2, 3, and 4 respectively. Each Figure reports the parameters of the country own monetary policy objectives and the German ones in order to gauge the extent to which 'core' European countries were actually pursuing an independent monetary policy notwithstanding the EMS constraint. For ease of exposition and comparison of the results across countries, the estimated coefficients of exchange rate volatility are grouped together in Figure 5. Three key results emerge from these charts: first, the behavior of other 'core' European central banks became relatively stable later than that of the Bundesbank, judging

[17]The implied weight attached to the inflation gap, a measure of the relative importance of this objective in the central bank's reaction function, is less than 0.5 throughout the period considered. This compares with a point estimate close to 1 found by DGF (see DGF, Table 5.4 and Appendix A1). We think that this could be due to the larger set of objectives allowed for by our econometric specification.

based on the time profile of most estimated parameters; second, there is no discernible common pattern of behavior among these countries except for the fact that all are strongly affected by contemporaneous movements in German interest rates; and third, as in the case of Germany, exchange rate volatility seems to have had a strong impact on the level of short term rates.

The behavior of the central bank of France conforms well to what one would expect under EMS: in additions to German interest rates, deviations of the nominal exchange rate vis-a-vis the DM and deviations of domestic and German inflation from their targets affect domestic short-term interest rates, while domestic and German output gaps appear to have basically no lasting impact on interest rates. Again, exchange rate volatility vis-a-vis the DM and the US dollar, as well as other 'core' European countries, seems to have a strong impact on local monetary conditions. The behavior of the central bank of Italy is similar to that of the bank of France except for the smaller magnitude of the coefficient of the bilateral rate against the DM and the correspondingly higher value of the coefficient attached to the German output gap. Note also the marked shift in the volatility parameter of the DM/Lt rate during the period in which the lira was floating after the 1992 crisis. The behavior of the central bank of Spain, instead, is quite peculiar: Spain appears to be the country least constrained by EMS, with its own output gap affecting short term interest rates throughout the period considered; secondly, the exchange rate gap vis-a-vis the DM has a persistently negative sign, while the coefficient of the volatility of the bilateral rate against the US dollar is positive throughout the estimation period, even though slightly trending downward.[18]

In summary, this first set of empirical results strongly supports the choice of the general, time-varying specification of the econometric model used to describe central banks' reactions

[18]Spain's peculiar behavior is a feature our results shares with other studies of the transmission of real and monetary shocks in the Euro area, including for example Kim (1998), Ballabriga et al. (1999), and Ortega and Alberola (2000).

functions, and show clearly how difficult it would have been to choose a restricted and yet uniform econometric specification of the econometric model to describe different behaviors over a period of relatively fast structural change.

The relative merit of the econometric approach followed may be appreciated also by looking at the residuals of the estimated reaction functions, which will be used in the rest of the paper as our measure of the unexpected component of country specific and common monetary policy. The estimated structural residuals of equation (3.3)—our measure of a local monetary policy shock—and the first principal component of the reduced form residuals—our measure of a common monetary shock—are plotted in Figure 6 and they look remarkably well behaved: there are very few outliers (most notably a large one for France in April 1993) and there is little evidence of serial autocorrelation and/or heteroscedasticity.[19]

Note also that when we estimate (3.3) without exchange rate volatility and restricting $B(L)$ as done by DFG we find residuals very much like theirs with large outliers at the same dates (DFG, Figure 4), further suggesting that adding exchange rate volatility and letting $B(L)$ unrestricted helps obtaining better residuals, and thereby a better measure of monetary policy shocks.

3.3.2 The impact of monetary policy on economic activity

Even though we have estimated all parameters of the system of output equations (3.8), here we present only the results for the subvector of monetary policy coefficients β^1_{it} and their estimated average or common component θ_t, which we interpret as the European-wide impact of monetary policy. We present four set of estimation and testing results:

[19]The normalized first principal component of the reduced form residuals explains about 50 percent of their total variation, about 25 percent of the residual of the Bundesbank's reaction function, about 10 percent of the Bank of France's reaction function, and about 50 percent of the residuals of the reaction functions of the Bank of Italy and the Bank of Spain. Its simple correlation with the residual of the Bundesbank's reaction function is 0.24.

two sets based on the estimation of (3.8) specified without parameters' time variation to compare these results to those previously found in the literature; and two set based on (3.8) estimated with time-varying parameters. Both the time-varying and the time-invariant specification of the system of output equations are estimated including only $\hat{u}_{it}$ first (the vector of country-specific structural residuals, which we interpret as a local monetary policy shock) and then including only $\hat{u}_t$ (the principal component of the reduced form residuals, which we interpret as a common monetary policy shock).

In order to save computing time and to facilitate the results' interpretation, the time-varying specification actually estimated allows the parameter vector to change only yearly, while in fact we use monthly data (see Appendix). The type of behavioral change we are interested in —presumably induced by anticipation of and preparation to EMU—is likely to have taken place over time rather slowly; in any case, we are not interested in isolating changes at monthly frequency. Hence, some time aggregation in estimating the parameters of the transmission mechanism of monetary policy might be desirable. In addition, when the model was estimated without imposing this restriction for Germany and Spain, we found very similar results, suggesting that the results presented below are robust to this feature of the specification actually used.

Are there differences in the transmission mechanism of monetary policy?

In order to compare our results with those in the literature, in this subsection, we report time-invariant estimates of the system of output equations and we test several homogeneity hypothesis on the transmission mechanism of country specific and common monetary policy shocks. Table 2 reports the mean, the median, the first and the third quartile of the posterior distribution of the coefficients of $\hat{u}_{it}$. For all countries considered, the table reports the coefficients of selected lags and the cumulative impact after one year and two years respectively.

From Table 2, we can see that the effects of country specific (or local) monetary policy shocks become evident within 18-24 months in all countries considered, and that there are some cross country differences in the impact at particular lags, but basically no quantitative differences with respect to their cumulative impact— which is also a measure of their long-term effects on the level of economic activity—as far as Germany, France, and Italy are concerned. The effects of local monetary policy shocks on output growth in Spain, instead, seem to be different from those in other countries both in terms of their timing and cumulative impact, which is lower. These conclusions are borne out clearly by a formal testing of various homogeneity assumptions.

Table 3 reports a set of Kolmogorov-Smirnov statistics (henceforth, KS) for the distance between the posterior distribution of q and q_1 under the corresponding null hypothesis.[20] When we test the null of equality of all the parameters of the transmission mechanism of country specific monetary shocks, either between all countries considered or through pair-wise comparisons (see the column of p-values under 'all lags' in Table 3), we reject the null decisively. This points to the existence of statistically significant difference in the transmission mechanism of European monetary policy across countries. Running the same test for each pair of countries considered on selected lags and the cumulative impacts of monetary policy after 12 and 24 months (see the corresponding columns of p-values in Table 3), however, we find that the overall difference between these four countries is due mainly to Spain, and, perhaps, some other timing difference in the other three countries. Thus, suggesting that the transmission mechanism of country specific monetary shocks in core European countries was already homogenous on the onset of EMU, especially considering

[20]As explained before, a posterior distribution of q_1 far apart from that of q can be interpreted as evidence against the null of equality of the relevant parameters of interest.

its long run impact.

Turning to the analysis of the transmission mechanism of a common monetary policy shock, $\hat{u}_t$, we can see from Table 4 and 5 that the results are broadly similar to those obtained for country specific shocks. Somewhat surprisingly, however, the cumulative impact after two years is now higher in Spain than in other countries. The bilateral differences between Germany, France and Italy look also slightly higher—as measured by lower p-values in Table 5.[21] This latter result suggests that the differences in the transmission mechanism of monetary policy remains significant even after controlling, albeit roughly, for heterogeneity of national central banks' preferences. The fact that the magnitude of the cumulative impact of common monetary policy shocks is smaller than that of country specific shocks, instead, is more difficult to explain, especially in light of its high correlation with Spanish interest rates.

A direct comparison of our results with those obtained in other studies is difficult because of the peculiarities of the empirical framework used in this paper. Nonetheless, Table 6 and 7 attempt to do this, to the extent possible, contrasting our point estimates (i.e., the mean of the posterior distributions of the parameters of interest) with those surveyed by Guiso et al. (2000). On the one hand, none of our estimate appears far away from what previously reported in the literature, giving confidence that our results are not systematically biased by any feature of the empirical framework used. On the other hand, a few sharp differences stand out. First, comparing our results with those obtained with small scale SVAR models (Table 7)—which are based on impulse response function analysis—we can see that our estimated short-term impact of monetary policy is at the lower end of

[21]Ortega and Alberola (2000) find a similar result for Spain. They attribute the different response of Spain to a (temporary) common monetary policy shock to its larger sensitiveness to changes in competitiveness vis-a-vis its European partners. Other core European countries, instead, are found to be more sensitive to the wealth effects of interest rate changes.

those found in the literature. This is not surprising given that our specification includes lagged output growth of all countries considered, thereby providing a better description of the international transmission mechanism of monetary policy. Second, unlike DGF—who analyze only the effects of anticipated changes in monetary policy—we do find some evidence of heterogeneity in the short term impact of monetary policy, but we do not find such evidence with regards to the cumulative or long-term impact. Also, the estimated long-term impact of a common monetary policy shock is much smaller than theirs, possibly due, again, to the richer specification of our econometric model. Finally, our estimated peak effect and the long run impact are very close to those reported in the BIS study.

In summary, and in part consistently with the consensus view in the literature, the evidence presented so far points to some degree of heterogeneity across countries in the transmission mechanism of monetary policy, and especially with regards to the timing of these effects rather than the magnitude of their cumulative impact. In fact, only Spain's response to both local and common monetary policy shocks appears significantly different from that of other core European countries. Differences in the timing of the effects of monetary policy in core European countries, however, are also important from both a methodological and a policy point of view as explained in the introduction. The question of whether or not the degree of heterogeneity of the transmission mechanism of monetary policy has changed over time—and, if this were the case, in which particular direction—remains therefore to be answered.

Are these differences changing over time?

To answer this question, first we reestimate the system of output equations (3.8) allowing for parameter variation over time and test the null hypothesis that the posterior variance of the third stage of the hierarchy (3.8-3.10) is zero, i.e., we test the hypothesis that ϕ_1 , the hyperparameter tightening the time variation of the coefficients describing the transmission mechanism of monetary policy, β_{it}^1, is zero. This is done using the test statistic (3.11) explained in section 2.2.3. As mentioned above, if the posterior distribution of ϕ_1 is less concentrated on values close to zero than the prior distribution, then we can reject the null of overall parameter stability over time; and thus reject a time-invariant specification of (3.8). In fact, the value of z in (3.11), for $\xi = 0.03$, is 0.465 in the case of country specific monetary shocks and 0.012 in the case of a common shock. For $\xi = 0.05$, z takes on a value of 1.838 and 0.054, respectively.[22] Very small values of z for arbitrarily small values of ξ imply that the posterior distribution of ϕ_1 is located more far away from zero than the prior distribution, providing strong evidence in favor of a time-varying specification, and suggesting that the transmission mechanism is indeed changing over time. This feature can also be appreciated from Figures 7-9, where the posterior distributions of some lags of the parameters of interest are plotted in the form of box-plot diagrams, by countries, i.e. β_{it}, (Fig. 7-8), and common across countries, i.e. θ_t, (Fig. 9), over the sample period.[23]

Once established that the transmission mechanism of monetary policy has changed over time, we check whether or not its degree of heterogeneity across countries has also changed in the run up to EMU. This is done by running a battery of KS statistics on the posterior distributions of q and q_1, under the relevant null hypothesis, as in Table 3 and 5, for each yearly subperiod considered. Table 8 reports the results for all countries considered from 1994 to 1998. As we can see from this table, there is some evidence, arguably weak, of decreasing distance between the benchmark distribution and the posterior ones. But

[22]The values of ξ have been chosen arbitrarily small, as in Chib and Greenberg (1995).

[23]A Box plot is a convenient graphical representation of the distribution of a variable which provides descriptive and diagnostic information. The box contains the central 50 percent of the distribution. The line inside the box is the median, while the two top sides represent the first and the third quartile respectively. Consequently, the length of the box measures the dispersion of the distribution and the position of the line inside the box its degree of symmetry. Outliers, i.e., observations falling under the 1 percent tails of the distributions, have been dropped.

the overall picture is one of neither decreasing nor increasing heterogeneity, rather simply persistence. Nonetheless, we now accept the null hypothesis of equality of the cumulative effects of monetary policy after 12 and 24 months between all countries considered, while this was rejected by the data when tested over the entire period 1991-1998 (see Table 3). It is possible, therefore, that some convergence might have taken place in the first half of the 1990s.[24]

An inspection of the posterior distributions of the parameters of interest country-by-country (Table 9), confirms that the short-term effects of idiosyncratic monetary shocks are heterogenous, but their cumulative impact becomes quite similar across countries after about 12 months. Furthermore, the cumulative impact after 12 months is increasing over time in all countries considered, while the impact after two years is decreasing. This could imply that the length of the European-wide transmission mechanisms was becoming shorter in the second half of the 1990s, arguably, as a result of financial development and gradually increasing labor market flexibility at the regional level.

In the case of common monetary policy shocks (Table 10 and 11 and Figure 8) we obtain similar results: the overall degree of heterogeneity of the transmission mechanism does not appear to decrease over time, but the cumulative impact of these shocks is already homogeneous after 12 months. Interestingly, the value of the third quartile of the distribution of the cumulative impact of these shocks after 24 months is always positive, and slightly decreasing over time. This suggests that the posterior distribution of these parameters was becoming progressively less concentrated on negative values, which in turn could be interpreted as evidence of increasing degree of monetary policy neutrality in the long run. At the same time the 12-month impact of common shocks is increasing slightly over time, as we

[24]These tests can be run only starting in 1994 because of the observations missed to initilize the estimation.

found in the case of country specific shocks. Finally, the magnitude of effects of a common shock looks generally smaller than that of country specific shock, as we found estimating the system of output equations without time variation.

In summary, these results show that the hypothesis of overall parameter stability is rejected by the data: the transmission mechanism of European monetary policy seems to have changed in the second half of the 1990s—possibly becoming shorter—but its degree of heterogeneity across countries has neither increased nor decreased during this period. On the other hand, the results presented suggest also that some convergence might have taken place in the first half of the 1990s given that the null hypothesis of equality of the cumulative effects of monetary policy between all countries considered cannot be rejected by the data when the econometric model is estimated allowing for parameters' variation over time. Consistently with these results, Spain's apparently peculiar behavior, found analyzing the effects of idiosyncratic and common shocks over the period 1990-1998 without allowing for time variation, could be explained as a consequence of an econometric specification error.

The European-wide impact of monetary policy

The evidence presented so far supports the view that the effects of monetary policy in core European countries differ in terms of their timing, though not cumulatively. A study of the European-wide effects of monetary policy in the sense of Tristani and Monticelli (1999)—i.e., the study of the effects of monetary policy in the Euro-area—based on averages of country specific time series, or on standard pooled estimators, therefore, may be biased, potentially. Moreover, we have seen that, in the specific case of Spain, a time invariant specification yields very different results from those obtained allowing for the parameters to vary over time. Within the empirical framework used in this study, the European-wide effects of monetary policy are measured by the posterior distribution of θ_t, the cross sectional mean of β_{it}.

Tables 12 and 13 report the mean, the median, the first and the third quartile of the posterior distribution of the elements of θ_t corresponding to selected lags and the cumulative impact of country specific and common monetary policy shocks, respectively. The overall shape of the posterior distributions of the elements of θ_t can be appreciated also from Figure 9, which plots the box-plot diagram of these distributions for each yearly subperiod from 1994 to 1998. Country specific monetary policy shocks appear to have had a system-wide, peak effect between 12 and 18 months in the mid-1990s, while the peak effect seems to occur earlier toward the end of the 1990s, between six and nine months. Similarly, the system-wide effects of common monetary policy shocks in 1997-98 seem to peak earlier than in 1994-95. This evidence is consistent with what shown above and confirms that the European-wide transmission mechanism of monetary policy might have become shorter in the second part of the 1990s. Also, country specific shocks have a sizable negative cumulative effect, while common shocks have a generally smaller effect, possibly not significantly different from zero.

Even though they are not directly comparable with those reported by Tristani e Monticelli (1999, par. 6.3 e Figure 3), our results suggest that the European-wide effects of monetary policy may be less persistent than what suggested by their results. In their exercise, a temporary one standard deviation monetary policy shock becomes statistically insignificant only after 18-20 months, and its effects are quantitatively negligible within two years. We observe a similar pattern when the model is estimated without time-varying coefficients. But when the model is specified with time-varying coefficients this conclusion holds only for the beginning of the 1990s: in the second part of the 1990s, monetary policy seems to affect economic activity sooner.

3.4 Conclusions

In this paper we study empirically the transmission mechanism of monetary policy in four core European countries using dynamic heterogenous models estimated in a Bayesian fashion with pre-EMU data.

Analyzing EMS data to understand what is happening under EMU has been done before, and will continue to be done for quite sometime. The econometric framework used in this paper shares several features with an 'ideal' one to run such an experiment: (i) the model's specification is the same across countries; (ii) no strong a priori restriction is imposed on the behavior of the central banks studied, letting the data reveal which were the relevant objectives in different stages of the run up to EMU; (iii) intra-European exchange rate movements as well as regional (real) interdependencies, through which monetary policy worked in part under EMS, are controlled for in assessing the impact of monetary policy on economic activity; and (iv) the effects of both country specific and common monetary policy shocks are analyzed, thereby controlling for the heterogeneity of central banks' preferences under EMS. Most importantly, however, the parameters of the reaction functions and those describing the transmission mechanism of monetary policy are allowed to change both across countries and time periods in our empirical framework. Therefore, our empirical results should be robust to the Lucas' critique and help understanding how differences in the transmission mechanism of European monetary policy evolved over time. As far as we know, this is the first study of the European transmission mechanism of monetary policy which allows explicitly for parameters' variation over time.

The empirical results presented show that there are differences in the timing of the effects of monetary policy across core European countries, and that the degree of heterogeneity of the transmission mechanism has not decreased over time during the second half of the

1990s, even though the parameters of the transmission mechanism do seem to have changed over time. We have shown also that the European-wide effects of monetary policy take 6-7 months to appear, peak at 12-18, and disappear within 24 months. These results are consistent with what previously found in the literature in that they point to some degree of heterogeneity in the transmission mechanism of monetary policy. Unlike the results found in previous studies, however, they suggest that these cross-country differences are mainly with regards to the short term impact of monetary policy. As standard monetary theory suggests, we have shown that monetary policy is becoming progressively more neutral in all countries considered in the long run.

This work could be extended in several directions. First, it would be desirable to extend the sample of countries analyzed to include all eleven members of EMU, and possibly also other European countries currently outside EMU. Second, it would be interesting to study the effect of monetary policy at regional rather than national level and to compare European countries (and/or regions) with American States. Finally, it would be useful to improve upon our definition of a common monetary policy shock and to attempt at framing the questions asked in this paper in a full blown panel VAR empirical framework.

3.5 Appendix

3.5.1 Estimation

In this appendix we present details of the estimation procedures used in both stages of the empirical analysis. In both stages the estimation is Bayesian. Thus, given the specification of the systems of reaction functions and output equations discussed in the main text, prior distributions and initial conditions must be combined with the information contained in the data in the form of likelihood functions to produce posterior estimates of the parameters of interest. In both stages of the empirical analysis, it is impossible to obtain close-form solutions for the posterior distributions of interest, and hence we must rely on numerical integration. For the latter, we use the Gibbs sampling method. Previous econometric studies with model specifications very similar to the ones used here have shown that the Gibbs sampling approach produces reasonably good results, even in high-dimensional models.[25]

Reaction functions

The probability density function (pdf) of the data for each block j of (3.4), conditional on the exogenous variables in the model and on the initial observations on R_{jt}, is

$$L(\boldsymbol{\delta}_{jt}, \Sigma_{jj}) \propto |\Sigma_{jj}|^{-T_j/2} \exp\left[-\frac{1}{2}\sum_t (R_{jt} - \mathbf{Z}_{jt}\boldsymbol{\delta}_{jt})' \Sigma_{jj}^{-1} (R_{jt} - \mathbf{Z}_{jt}\boldsymbol{\delta}_{jt})\right]. \tag{3.15}$$

The prior assumptions on the model's parameters generalize those introduced by Zellner (1971) to take into account the presence of time-varying coefficients: a time-varying, multivariate normal prior, i.e., a Minnesota-type of prior (Doan et al., 1984), for the regression parameters is combined with a diffuse prior on the variance–covariance matrix of the residuals, Σ_{jj}. Thus, assuming prior independence:

$$p(\boldsymbol{\delta}_{jt}, \Sigma_{jj}) = p(\boldsymbol{\delta}_{jy})\, p(\Sigma_{jj}),$$

with

$$p(\Sigma_{jj}) \propto |\Sigma_{jj}|^{-(G_j+1)/2} \tag{3.16}$$

$$\boldsymbol{\delta}_{jt} = P_j\boldsymbol{\delta}_{jt-1} + (I - P_j)\,\bar{\boldsymbol{\delta}}_j + \eta_{jt} \tag{3.17}$$

$$\eta_{jt} \sim N(0, \Phi_j)$$

where P_j is a $G_jk_j \times G_jk_j$ matrix governing the law of motion of $\boldsymbol{\delta}_{jt}$, $\bar{\boldsymbol{\delta}}$ is the unconditional mean of $\boldsymbol{\delta}_{jt}$, Φ_j governs the time variation of $\boldsymbol{\delta}_{jt}$, and η_{jt} is assumed to be independent from v_{jt}. The assumption of prior independence is needed for analytical tractability.[26] Note also

[25]See, for instance, Chib and Greenberg (1995) and Canova and Ciccarelli (2000).
[26]See Leamer (1978, p.80) for a better justification of prior independence and Kadiyala and Karlsson (1997) on the comparison of alternative prior assumptions in VAR models.

that giving a joint prior on $(\boldsymbol{\delta}_{jt}, \Sigma_{jj})$ is equivalent to considering a prior on $\left(\boldsymbol{\delta}_{jt}, A(0)_{jj}\right)$ as proposed by Sims and Zha (1998) and Zha (1999) if the model is exactly identified, which is the case dealt with here. Therefore, $A(0)_{jj}$ is recovered from Σ_{jj} through the one-to-one mapping between these two matrices.

In order to run the Gibbs sampler, the conditional posterior distributions of Σ_{jj}^{-1} and $\boldsymbol{\delta}_{jt}$ must be obtained. Combining the likelihood (3.15) with (3.16), it is not difficult to see that the conditional posterior distribution of Σ_{jj}^{-1} is a Wishart:

$$\Sigma_{jj}^{-1} \mid \{\boldsymbol{\delta}_{jt}\}_t, R \sim W\left(T, \left[(R_{jt} - \mathbf{Z}_{jt}\boldsymbol{\delta}_{jt})(R_{jt} - \mathbf{Z}_{jt}\boldsymbol{\delta}_{jt})'\right]^{-1}\right). \tag{3.18}$$

The (joint) conditional posterior distribution $\boldsymbol{\delta}_{j0}, \boldsymbol{\delta}_{j1}, ..., \boldsymbol{\delta}_{jT} \mid \Sigma_{jj}$ is obtained in two steps as shown by Chib and Greenberg (1995). First, we initialize $\{\boldsymbol{\delta}_{jt}\}_t$ for each t by Kalman filter and save the output:

$$\begin{aligned}
\hat{\boldsymbol{\delta}}_{jt|t} &= \hat{\boldsymbol{\delta}}_{jt|t-1} + \hat{\Omega}_{jt|t-1}\mathbf{Z}_{jt}'F\left(R_{jt} - \mathbf{Z}_{jt}\hat{\boldsymbol{\delta}}_{jt|t-1}\right) \\
\hat{\Omega}_{jt|t} &= \hat{\Omega}_{jt|t-1} - \hat{\Omega}_{jt|t-1}\mathbf{Z}_{jt}'F\mathbf{Z}_{jt}\hat{\Omega}_{jt|t-1} \\
F &= \left(\mathbf{Z}_{jt}\hat{\Omega}_{jt|t-1}\mathbf{Z}_{jt}' + \Sigma_{jj}\right)^{-1} \\
M_t &= \hat{\Omega}_{jt|t}\hat{\Omega}_{jt+1|t}^{-1}
\end{aligned} \tag{3.19}$$

where $\hat{\boldsymbol{\delta}}_{jt|t-1} = P_j\hat{\boldsymbol{\delta}}_{jt-1|t-1} + (I - P_j)\bar{\boldsymbol{\delta}}_j$ and $\hat{\Omega}_{jt|t-1} = P_j\hat{\Omega}_{jt-1|t-1}P_j + \Phi_j$. Second, the joint conditional posterior distribution $\boldsymbol{\delta}_{j0}, \boldsymbol{\delta}_{j1}, ..., \boldsymbol{\delta}_{jT} \mid \Sigma_{jj}$ is sampled in reverse time order from

$$\begin{aligned}
\boldsymbol{\delta}_{jT} &\sim N\left(\hat{\boldsymbol{\delta}}_{jT|T}, \hat{\Omega}_{jT|T}\right) \\
\boldsymbol{\delta}_{jT-1} &\sim N\left(\hat{\boldsymbol{\delta}}_{jT-1}, \hat{\Omega}_{jT-1}\right) \\
&\vdots \\
\boldsymbol{\delta}_{j0} &\sim N\left(\hat{\boldsymbol{\delta}}_{j0}, \hat{\Omega}_{j0}\right)
\end{aligned} \tag{3.20}$$

where $\hat{\boldsymbol{\delta}}_{jt} = \hat{\boldsymbol{\delta}}_{jt|t} + M_t\left(\boldsymbol{\delta}_{jt+1} - \hat{\boldsymbol{\delta}}_{jt|t}\right)$, and $\hat{\Omega}_{jt} = \hat{\Omega}_{jt|t} - M_t\hat{\Omega}_{jt+1|t}M_t'$.

To make the updating scheme described in (3.18)-(3.20) operational, initial values for P_j, Φ_j, $\hat{\Omega}_{j0}$, and the vector $\hat{\boldsymbol{\delta}}_{j0}$, at time $t = 1$ (the first period of the sample), must be assigned. Following Litterman (1980, 1986), we define the matrices P_j, Φ_j, $\hat{\Omega}_{j0}$, $\hat{\boldsymbol{\delta}}_{j0}$ in terms of a few hyperparameters. These hyperparameters are assumed known and are estimated before starting the Gibbs sampler. More specifically, each $k_j \times 1$ vector δ^0_{jg} is assumed to depend only on one hyperparameter such that $\delta^0_{jg} = (0, ..., 0, \pi_{1,g}, 0, ...0)_j$, where $\pi_{1,g}$ represents the prior mean of the coefficient of the lagged dependent variable in equation g of block j. The individual components of $\hat{\boldsymbol{\delta}}_{j0}$ are assumed to be mutually independent and independent from analogous components in other equations of the block j; thereby, rendering the covariance matrix $\hat{\Omega}_{j0}$ diagonal. The diagonal elements of $\hat{\Omega}_{j0}$ are then defined so that, for each block j, the relative tightness of the prior of the coefficient of the lagged dependent variable, of other lagged endogenous variables, and of deterministic and exogenous variables is controlled by $\pi_{2,g}$, $\pi_{3,g}$, $\pi_{4,g}$, respectively. In practice, the prior variances of the parameters in equation g of block j are specified as follows:

$$Var\left(\delta^0_{jg}\right) = \begin{cases} \frac{\pi_{2,g}}{l} & \text{for lagged dependent variables} \\ \frac{\pi_{2,g}\ \pi_{3,g}}{l}\frac{\sigma_g}{\sigma_i} & \text{for other lagged endogenous variables} \\ \pi_{2,g}\ \pi_4\ \sigma_g & \text{for exogenous and deterministic variables} \end{cases}$$

where l denotes the lag length, and σ_g is a scaling factor which takes into account the range of variation of different variables. Hence, the overall tightness in the system (the overall degree of uncertainty with which prior information is introduced in the model's specification) is controlled by π_2; and if π_2 goes to infinity, the prior becomes diffuse. The tightness of the coefficients of the lagged dependent variable relative to that of other lagged endogenous variables in the equation is controlled by π_3; if $\pi_3 = 0$, the prior defines a set of univariate autoregressive processes of order p. Finally, π_4 controls the degree of uncertainty with respect to the coefficients of exogenous and deterministic variables.

The time variation introduced in the model's parameters a priori is governed by the matrices P_j and Φ_j. These matrices are defined as:

$$P_j = diag\left(P_{j1}, ...P_{jG_j}\right)$$

$$\Phi_j = diag\left(\Phi_{j1}, ..., \Phi_{jG_j}\right)\hat{\Omega}_{j0}$$

where $P_{jg} = diag\left(\pi_{5,g}\right)$ are $k_j \times k_j$ matrices with $\pi_{5,g}$ controlling the coefficients of the law of motion of each δ_{jg}, and $\Phi_{jg} = diag\left(\pi_{6,g}\right)$ are $k_j \times k_j$ matrices with $\pi_{6,g}$ controlling the amount of time variation actually introduced in the model. Thus, a time-invariant model could be obtained by setting $\pi_5 = 1$ and $\pi_6 = 0$.

In sum, we have six hyperparameters for each equation of block j. The hyperparameters are estimated before running the Gibbs sampler by maximizing, equation-by-equation, the sample likelihood of the model written as a function of these hyperparameters themselves, while the model's parameters $(\delta_{jt}, \Sigma_{jj})$ are initialized with a classical SUR estimate of the entire model.[27] Then, the updating scheme (3.19) is run and the Gibbs sampler implemented, switching between (3.18) and (3.20) as if $\pi_1, ..., \pi_6$ were known. The Gibbs sampler runs 5000 times yielding 4000 draws from the posterior distributions after discarding the first 1000 draws.

Output equations

Time variation Let $y_{i,\tau}^s$ denote annual output growth $\left(\ln\left(Y_{i\tau}^s / Y_{i\tau-1}^s\right)\right)$ at the s-th month of the τ-th year for country i. For each country i, $y_{i,\tau}^s$ is modelled as follows:

$$y_{i\tau}^s = X_{i\tau}^{s\prime}\beta_{i\tau} + \varepsilon_{i\tau}^s$$

[27]Note that the first block of the model contains only one equation. In this case (3.18) becomes an inverted gamma and the equation's parameters can be initialized by OLS. All estimated hyperparameters are reported in Table 1.

$$i = 1,...,G; \quad \tau = 1,...,T_1; \qquad s = 1,...,S.$$

In our sample, the number of years (T_1) is 6, the number of countries or endogenous variables (G) is 4, the number of subperiods for each year (S) is 12, and hence the total number of observations for each variable is $T = T_1 * S = 72$.

As noted in the main text, this system can be rewritten as:

$$\begin{aligned} y_\tau^s &= X_\tau^s \beta_\tau + \varepsilon_\tau^s, & \varepsilon_\tau^s \sim N_g(0, \Omega), \\ \beta_\tau &= M_o \theta_\tau + \zeta_\tau, & \zeta_\tau \sim N_s(0, B_o), \\ \theta_\tau &= \theta_{\tau-1} + \eta_\tau, & \eta_\tau \sim N_m(0, B_1). \end{aligned}$$

The likelihood of the data is:

$$\propto |\Omega|^{-T/2} \exp\left\{ -\frac{1}{2} \sum_\tau \sum_s (y_\tau^s - X_\tau^s \beta_\tau)' \Omega^{-1} (y_\tau^s - X_\tau^s \beta_\tau) \right\}.$$

The priors are:

$$\begin{aligned} \Omega^{-1} &\sim W(\omega_o, \Theta), \\ M_o &= e_g \otimes I_k, \\ B_o &= I_g \otimes \Sigma, \quad \Sigma^{-1} \sim W(\sigma_o, \Psi_o), \\ B_1 &= diag(\phi_1 I_{k_1}, \phi_2 I_{k-k_1}), \end{aligned}$$

where e_g is a vector of ones of dimension $g \times 1$, $W(\omega_o, \Theta)$ denotes a Wishart distribution with ω_o degrees of freedom and scale matrix Θ, I_j denotes an identity matrix of dimension j, and k_1 is the number of monetary policy parameters. The time variation of the monetary policy parameters is controlled by ϕ_1, while ϕ_2 tightens the time variation of other parameters.

We set a diffuse prior on ϕ_2 and we assume that the prior distribution of ϕ_1 is an inverted gamma, $\phi_1 \sim IG(\kappa_o/2, \xi_o/2)$. All hyperparameters of the system $(\omega_o, \Theta, \sigma_o, \Psi_o, \kappa_o, \xi_o)$ are assumed known.

The posterior densities of the parameters of interest are obtained by combining the likelihood of the data with the prior distributions above in the form of conditional posterior distributions as before. Letting $Y_T = (y_1, ..., y_T)$ denote the sample data and $\psi = (\{\beta_\tau\}_\tau, \Omega, \{\theta_\tau\}_\tau, \Sigma, \phi_1, \phi_2)$ denote the parameters whose joint distribution needs to be found, we have:

$$\beta_\tau \mid Y_T, \psi_{-\beta_\tau} \sim N\left(\hat{\beta}_\tau, V_\tau\right), \quad \tau \leq T_1;$$

$$\Omega^{-1} \mid Y_T, \psi_{-\Omega} \sim W\left(\omega_o + T, \Theta_T\right);$$

$$\Sigma^{-1} \mid Y_T, \psi_{-\Sigma} \sim W\left(\sigma_o + T_1 g, \Psi_{T_1}\right);$$

$$\phi_1 \mid Y_T, \psi_{-\phi_1} \sim IG\left(\frac{(v_o + T_1 k_1)}{2}, \frac{\zeta_o + \sum_\tau \left(\theta_\tau^1 - \theta_{t-1}^1\right)'\left(\theta_\tau^1 - \theta_{t-1}^1\right)}{2}\right);$$

$$\phi_2 \mid Y_T, \psi_{-\phi_2} \sim IG\left(\frac{T_1(k - k_1)}{2}, \frac{\sum_\tau \left(\theta_\tau^2 - \theta_{t-1}^2\right)'\left(\theta_\tau^2 - \theta_{t-1}^2\right)}{2}\right);$$

where

$$\begin{aligned}
\hat{\beta}_\tau &= V_\tau\left(B_o^{-1} M_o \theta_\tau + \sum_s X_\tau^{s\prime} \Omega^{-1} y_\tau^s\right), \\
V_\tau &= \left(B_o^{-1} + \sum_s X_\tau^{s\prime} \Omega^{-1} X_\tau^s\right)^{-1},
\end{aligned}$$

$$\Theta_T = \left[\Theta^{-1} + \sum_\tau \sum_s \left(y_\tau^s - X_\tau^s \beta_\tau\right)\left(y_\tau^s - X_\tau^s \beta_\tau\right)'\right]^{-1},$$

$$\Psi_{T_1} = \left[\Psi_o^{-1} + \sum_\tau \sum_i \left(\beta_{i\tau} - \theta_\tau\right)\left(\beta_{i\tau} - \theta_\tau\right)'\right]^{-1},$$

with $\psi_{-\gamma}$ denoting ψ without the parameter γ, and θ_τ^1 and θ_τ^2 denoting monetary policy parameters and other parameters, respectively.

The posterior distribution of $\{\theta_\tau\}_{\tau=0}^{T_1}$, conditional on the other parameters, is obtained using an updating scheme as in (3.20) above.

As for the hyperparameters, we set $\omega_o = g+1$, $\sigma_o = k+1$, and $\Psi_o = diag\,(1.0)$, while Θ is initialized with the variance-covariance matrix of a classical SUR estimate of (3.8). The parameters of the gamma distribution of ϕ_1 are $\kappa_o = 6$ and $\xi_o = 1$, implying that the prior mean and the standard deviation of ϕ_1 are 0.25 and 0.25, respectively. To initialize the Gibbs sampler we set also $\phi_1 = \phi_2 = 0.5$, $\Omega = I_g$, and $\Sigma = I_k$, while all β_τ's are initialized with the posterior mean obtained estimating the model without time-variation.

With these starting values the Gibbs sampler begins generating $\{\theta_\tau\}_{\tau=0}^{T_1}$ and then all the other parameters. The Gibbs sampler runs 5000 times yielding 4000 draws from the posterior distribution after discarding the first 1000 draws as before.

Time invariant model The model is also estimated restricting the coefficients to be constant over time. In this case, we used the following hierarchy

$$\begin{aligned} y_t &= X_t\beta + \varepsilon_t, & \varepsilon_t &\sim N_g\,(0,\Omega) \\ \beta &= M_o\theta + \zeta, & \zeta &\sim N_s\,(0,B_o) \\ \theta &= M_1\mu + \eta_t, & \eta &\sim N_m\,(0,B_1) \end{aligned}$$

where now $t = 1,...,T$. The likelihood now become:

$$\propto |\Omega|^{-T/2} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(y_t - X_t\beta)'\,\Omega^{-1}\,(y_t - X_t\beta)\right\}.$$

All the hyperparameters, including μ and B_1, are assumed to be known as before. In particular, we set $B_1^{-1} = 0$, i.e., the third stage of the hierarchy is degenerate.

Using the same notation and priors as before, the conditional posterior distributions now are:

$$\beta \mid Y_T, \psi_{-\beta} \sim N\left(\hat{\beta}, V_T\right);$$

$$\Omega^{-1} \mid Y_T, \psi_{-\Omega} \sim W\left(\omega_o + T, R_T\right);$$

$$\theta \mid Y_T, \psi_{-\theta} \sim N\left(\Delta_1\left(B_1^{-1}M_1\mu + M_o'B_o^{-1}\beta\right), \Delta_1\right);$$

$$\Sigma^{-1} \mid Y_T, \psi_{-\Sigma} \sim W\left(\sigma_o + g, \Psi_g\right);$$

where

$$\hat{\beta} = V_T\left(B_o^{-1}M_o\theta + \sum_t X_t'\Omega^{-1}y_t\right), \qquad V_T = \left(B_o^{-1} + \sum_t X_t'\Omega^{-1}X_t\right)^{-1},$$

$$R_T = \left[R_o^{-1} + \sum_{t=1}^{T}\left(y_t - X_t\beta\right)\left(y_t - X_t\beta\right)'\right]^{-1},$$

$$\Delta_1 = \left(B_1^{-1} + M_o'B_o^{-1}M_o\right)^{-1},$$

$$\Psi_g = \left[\Psi_o^{-1} + \sum_{i=1}^{g}\left(\beta_i - \theta\right)\left(\beta_i - \theta\right)'\right]^{-1}.$$

Finally, the Gibbs sampler is initialized as done in the case of the time-varying model.

3.5.2 Data

All the data used are from the International Financial Statistics (IFS) database of the IMF, except daily exchange rates which were provided by Marcello Pericoli of the Bank of Italy.[28]

The basic dataset is composed of monthly observations from 1985:01 to 1998:12 for the following series:

1. Consumer price index, IFS line 64 (CPI);
2. Industrial production index, IFS line 66 (IP);
3. Nominal exchange rate vis-a-vis the U.S. dollar (period average), IFS line rf (NER);
4. Interest rates (Treasury Bill rate), IFS line 60c (IR);
5. Daily nominal exchange rate, Bank of Italy ($DNER$).[29]

The following transformations of the basic data have been used:

---

[28] A previous version of the paper used the dataset of Dornbusch, Favero, and Giavazzi (1998).

[29] We use the bilateral rate vis-a-vis the DM for France, Italy, and Spain, and vis-a-vis the US dollar for Germany. Bilateral rates vis-a-vis the DM are obtained as cross rates vis-a-vis the U.S. dollar.

6. $\pi_{i,t} = \log(CPI_t/CPI_{t-12})$;

7. $\mathrm{y}_{i,t} = \log(IP_t)$;

8. $R_{i,t} = \log(1 + IR_t/100)$;

9. $y_{i,t} = \log(IP_t/IP_{t-12})$;

10. $\sigma_{i,t} = stdev(\log(DNER_{i,s}/DNER^*)$,

    where *stdev* denotes the intra-month standard deviation and $DNER^*$ the HP trend obtained using a smoothing parameter equal to 130000.

Inflation, output, and exchange rate gaps—denoted $(\pi_{i,t}-\pi_i^*)$, $(\mathrm{y}_{i,t}-\mathrm{y}_i^*)$, and $(e_{it}-e_i^*)$, in the text—were computed as $\log(\pi_{i,t}/\pi_i^*)$, $\log(\mathrm{y}_{t,i}/\mathrm{y}_i^*)$, and $\log(e_{t,i}/e_i^*)$, respectively, where π_i^*, y_i^*, and e_i^* denote the deterministic components of a linear regression of π_t, y_t, and e_t on a constant and a linear trend, a constant and a quadratic trend, and a simple constant, respectively.

Tables

Table 1. Estimated hyperparameters in the reaction functions

	$\pi1$	$\pi2$	$\pi3$	$\pi4$	$\pi5$	$\pi6$	**Likelihood**
GER	0,97582	0,02343	0,18227	13804,4	0,99867	7,7245E-09	1025,122
FRN	0,08534	0,72456	0,02525	1,46355E-05	0,99979	1,12253E-08	953,881
ITL	0,05922	0,05099	0,34916	171320,008	1	9,62304E-09	865,236
SPN	0,01891	0,17227	0,00872	5878,8848	1	1,3967E-08	856,82

Notes:

$\pi1$ = Prior mean on first lag

$\pi2$ = Overall tightness

$\pi3$ = Relative tightness on other variables

$\pi4$ = Relative tightness on the constant

$\pi5$ = Law of motion of the parameter

$\pi6$ = Relative tightness on time variation

Table 2. Estimated impact of idiosyncratic monetary policy shocks. Several lags. All countries

		GER	FRN	ITL	SPN
lag 6	*1st Qu.*	-0,1726	-0,0402	-0,0723	-0,0673
	Mean	**-0,0554**	**0,0638**	**0,0522**	**0,0391**
	Median	**-0,0576**	**0,0623**	**0,0489**	**0,0314**
	3rd Qu.	0,0603	0,1644	0,1666	0,1448
lag 12	*1st Qu.*	0,0627	-0,0427	0,1033	0,0868
	Mean	**0,1659**	**0,0570**	**0,2162**	**0,1831**
	Median	**0,1656**	**0,0571**	**0,2160**	**0,1855**
	3rd Qu.	0,2702	0,1592	0,3279	0,2807
lag 14	*1st Qu.*	-0,1770	-0,1048	-0,1519	-0,0453
	Mean	**-0,0699**	**-0,0084**	**-0,0405**	**0,0558**
	Median	**-0,0687**	**-0,0045**	**-0,0360**	**0,0576**
	3rd Qu.	0,0406	0,0928	0,0741	0,1591
lag 16	*1st Qu.*	-0,2567	-0,1734	-0,2973	-0,2636
	Mean	**-0,1465**	**-0,0667**	**-0,1874**	**-0,1631**
	Median	**-0,1476**	**-0,0705**	**-0,1848**	**-0,1668**
	3rd Qu.	-0,0330	0,0386	-0,0776	-0,0628
lag 18	*1st Qu.*	-0,3311	-0,2140	-0,3798	-0,2961
	Mean	**-0,2203**	**-0,1156**	**-0,2639**	**-0,1963**
	Median	**-0,2225**	**-0,1166**	**-0,2622**	**-0,1977**
	3rd Qu.	-0,1106	-0,0189	-0,1431	-0,0998
lag 24	*1st Qu.*	-0,2455	-0,1391	-0,2892	-0,2235
	Mean	**-0,1391**	**-0,0451**	**-0,1738**	**-0,1269**
	Median	**-0,1369**	**-0,0483**	**-0,1746**	**-0,1273**
	3rd Qu.	-0,0336	0,0471	-0,0597	-0,0324
cumul 12	*1st Qu.*	-0,8818	-0,8361	-0,7235	-0,5365
	Mean	**-0,4093**	**-0,3678**	**-0,2561**	**-0,1521**
	Median	**-0,4156**	**-0,3660**	**-0,2678**	**-0,1537**
	3rd Qu.	0,0764	0,1075	0,2019	0,2450
cumul 24	*1st Qu.*	-2,1080	-2,0268	-2,1834	-1,3751
	Mean	**-1,4115**	**-1,3507**	**-1,5098**	**-0,8963**
	Median	**-1,4005**	**-1,3469**	**-1,5030**	**-0,8947**
	3rd Qu.	-0,7002	-0,6364	-0,8351	-0,3942

Note: For each lag the first quartile, the mean, the median, and the third quartile are reported

Table 3. Testing the null: F(q) = F(q1). Idiosyncratic shocks

	all lags	lag 12	lag 24	cumul 12	cumul 24
joint	0.5020 (0.0000)	0.1843 (0.0000)	0.1385 (0.0000)	0.0788 (0.0001)	0.1530 (0.0000)
Ger vs Frn	0.3370 (0.0000)	0.1223 (0.0000)	0.1008 (0.0000)	0.0152 (0.7280)	0.0138 (0.8327)
Ger vs Itl	0.2528 (0.0000)	0.0342 (0.0175)	0.0285 (0.0745)	0.0370 (0.0079)	0.0172 (0.5781)
Ger vs Spn	0.3058 (0.0000)	0.0207 (0.3452)	0.0095 (0.9920)	0.0600 (0.0000)	0.1045 (0.0000)
Frn vs Itl	0.3223 (0.0000)	0.2068 (0.0000)	0.1658 (0.0000)	0.0192 (0.4372)	0.0198 (0.4051)
Frn vs Spn	0.3162 (0.0000)	0.1675 (0.0000)	0.0807 (0.0000)	0.0767 (0.0000)	0.1135 (0.0000)
Itl vs Spn	0.2388 (0.0000)	0.033 (0.0245)	0.0420 (0.0016)	0.0342 (0.0175)	0.1700 (0.0000)

**Note: numbers represent the Kolmogorov-Smirnov statistics.
P-values in brackets.**

Table 4. Estimated impact of a common monetary Policy shocks. Several lags. All countries

		GER	FRN	ITL	SPN
lag 6	*1st Qu.*	-0,1659	-0,2111	-0,2111	-0,2266
	Mean	**-0,0352**	**-0,0876**	**-0,0627**	**-0,1004**
	Median	**-0,0351**	**-0,0847**	**-0,0621**	**-0,0987**
	3rd Qu.	0,1031	0,0447	0,0893	0,0364
lag 12	*1st Qu.*	-0,0948	-0,1071	-0,0110	-0,0309
	Mean	**0,0240**	**0,0022**	**0,1196**	**0,0879**
	Median	**0,0279**	**0,0027**	**0,1226**	**0,0917**
	3rd Qu.	0,1442	0,1105	0,2484	0,2017
lag 14	*1st Qu.*	-0,1516	-0,1534	-0,1812	-0,0352
	Mean	**-0,0234**	**-0,0369**	**-0,0401**	**0,0860**
	Median	**-0,0233**	**-0,0335**	**-0,0374**	**0,0913**
	3rd Qu.	0,1092	0,0819	0,1013	0,2119
lag 16	*1st Qu.*	-0,3754	-0,2446	-0,2689	-0,4232
	Mean	**-0,2336**	**-0,1230**	**-0,1204**	**-0,2953**
	Median	**-0,2328**	**-0,1263**	**-0,1190**	**-0,2941**
	3rd Qu.	-0,0947	0,0007	0,0259	-0,1644
lag 18	*1st Qu.*	-0,4266	-0,3938	-0,5101	-0,3967
	Mean	**-0,3063**	**-0,2864**	**-0,3803**	**-0,2803**
	Median	**-0,3040**	**-0,2831**	**-0,3768**	**-0,2742**
	3rd Qu.	-0,1803	-0,1739	-0,2481	-0,1633
lag 24	*1st Qu.*	0,0330	-0,0961	-0,1214	-0,1179
	Mean	**0,1530**	**0,0193**	**0,0161**	**0,0011**
	Median	**0,1549**	**0,0176**	**0,0201**	**0,0027**
	3rd Qu.	0,2779	0,1321	0,1584	0,1187
cumul 12	*1st Qu.*	-0,8175	-0,6696	-0,3923	-0,8722
	Mean	**-0,4586**	**-0,3503**	**-0,0138**	**-0,5151**
	Median	**-0,4445**	**-0,3420**	**-0,0148**	**-0,5067**
	3rd Qu.	-0,0920	-0,0160	0,3149	-0,1464
cumul 24	*1st Qu.*	-1,2557	-1,2482	-1,1558	-1,8280
	Mean	**-0,6641**	**-0,6647**	**-0,4765**	**-1,2416**
	Median	**-0,6495**	**-0,6682**	**-0,4726**	**-1,2253**
	3rd Qu.	-0,0471	-0,0924	0,1960	-0,6380

Note: For each lag the first quartile, the mean, the median, and the third quartile are reported

Table 5. Testing the null: F(q) = F(q1). Common shock

	all lags	lag 12	lag 24	cumul 12	cumul 24
joint	0,5640 (0.0000)	0,1275 (0.0000)	0,1853 (0.0000)	0,1822 (0.0000)	0,1935 (0.0000)
Ger vs Frn	0,3615 (0.0000)	0,0222 (0.2668)	0,1795 (0.0000)	0,0268 (0.1100)	0,0268 (0.1100)
Ger vs Itl	0,2965 (0.0000)	0,0833 (0.0000)	0,1648 (0.0000)	0,1728 (0.0000)	0,0370 (0.0079)
Ger vs Spn	0,3095 (0.0000)	0,0707 (0.0005)	0,2150 (0.0000)	0,0212 (0.3175)	0,1442 (0.0000)
Frn vs Itl	0,2658 (0.0000)	0,1430 (0.0000)	0,0148 (0.7644)	0,1328 (0.0000)	0,0308 (0.0436)
Frn vs Spn	0,4812 (0.0000)	0,1060 (0.0000)	0,0185 (0.4881)	0,0435 (0.0010)	0,1570 (0.0000)
Itl vs Spn	0,3405 (0.0000)	0,0305 (0.0464)	0,0197 (0.4951)	0,2290 (0.0000)	0,2188 (0.0000)

Note: numbers represent the Kolmogorov-Smirnov statistics.
P-values in brackets.

Table 6. The short and long-term impact of monetary policy shocks

(Comparison with BIS study and DFG)

	GER			FRA			ITA			SPA		
	BIS	DFG	C	BIS	DFG	C	BIS	DFG	C	BIS	DFG	C
First year	na	-0,64	-0,41	na	-0,18	-0,37	na	-0,80	-0,26	na	0,00	-0,15
	-0.15	na	-0,46	-0.18	na	-0,35	-0.18	na	-0,01	na	na	-0,51
Second year	na	-0,73	-1,41	na	-0,59	-1,35	na	-0,80	-1,50	na	0,00	-0,89
	-0.37	-1,40	-0,66	-0.36	-1,54	-0,67	-0.44	-2,14	-0,48	na	-1,54	-1,24
Peak effect	na	-0,23	-0,22	na	-0,28	-0,12	na	-0,75	-0,26	na	0,00	-0,20
	-0.37	-0,54	-0,31	-0.36	-0,46	-0,29	-0.44	-1,11	-0,38	na	-0,35	-0,30

Source: Guiso et al. (2000), DFG (1997).
Note: the first line refers to idiosyncratic shocks, the second line to a common shock

Table 7. Effect on output one year after idiosyncratic monetary policy shocks
(Comparison with SVAR studies)

	Strength of response	GER	FRA	ITA	SPA
This paper	S<I<F<G	-0,41	-0,37	-0,26	-0,15
Ramaswamy and Sloek (1998)	F<I<G	-0,60	-0,40	-0,50	na
Barron, Coudert, and Mojon (1996)	I<F<G	-0,60	-0,40	-0,30	na
Gerlach and Smets (1995), variant 1	F=I<G	-0,30	-0,20	-0,20	na
Gerlach and Smets (1995), variant 2	G<F=I	-0,10	-0,50	-0,50	na
Ehrmann (1998)	I<F<G	-0,90	-0,50	-0,10	na
Dedola and Lippi (1999)	I<F<G	-2,20	-1,40	-1,10	na

Source: Guiso et al. (2000).

Table 8. Testing the null: F(q) = F(q1). Time varying model. Idiosyncratic shocks

all countries

	1994	1995	1996	1997	1998
all lags	0.3615 (0.0000)	0.3195 (0.0000)	0.3117 (0.0000)	0.3027 (0.0000)	0.2533 (0.0000)
lag 12	0.0867 (0.0012)	0.0265 (0.1161)	0.0597 (0.0000)	0.0648 (0.0000)	0.039 (0.0043)
lag 24	0.0288 (0.0704)	0.107 (0.0000)	0.0495 (0.0001)	0.0445 (0.0007)	0.072 (0.0000)
cumul 12	0.0153 (0.7280)	0.019 (0.4538)	0.0158 (0.6907)	0.0543 (0.0000)	0.0215 (0.3042)
cumul 24	0.0158 (0.6907)	0.0155 (0.7094)	0.0183 (0.5055)	0.0268 (0.1100)	0.0112 (0.9565)

**Note: numbers represent the Kolmogorov-Smirnov statistics.
P-values in brackets.**

Table 9. Impact of idiosyncratic monetary policy shocks
Several lags. All countries. All years

	lag 6					lag 12				
	1994	1995	1996	1997	1998	1994	1995	1996	1997	1998
	-0,158	-0,170	-0,175	-0,392	-0,222	-0,388	-0,325	-0,347	-0,189	-0,262
GER	**-0,001**	**-0,015**	**-0,022**	**-0,221**	**-0,030**	**-0,212**	**-0,141**	**-0,179**	**-0,009**	**-0,067**
	0,157	0,138	0,131	-0,048	0,163	-0,040	0,019	-0,026	0,162	0,124
	-0,121	-0,218	-0,123	-0,396	-0,196	-0,454	-0,309	-0,356	-0,257	-0,237
FRN	**0,043**	**-0,072**	**0,027**	**-0,229**	**0,003**	**-0,286**	**-0,135**	**-0,206**	**-0,086**	**-0,046**
	0,206	0,073	0,169	-0,061	0,203	-0,128	0,025	-0,067	0,084	0,141
	-0,122	-0,204	-0,175	-0,357	-0,203	-0,412	-0,281	-0,303	-0,234	-0,203
ITL	**0,041**	**-0,050**	**-0,011**	**-0,180**	**-0,005**	**-0,229**	**-0,103**	**-0,134**	**-0,052**	**-0,002**
	0,198	0,108	0,146	-0,003	0,198	-0,058	0,063	0,025	0,122	0,188
	-0,088	-0,181	-0,210	-0,375	-0,196	-0,425	-0,295	-0,313	-0,243	-0,223
SPN	**0,068**	**-0,036**	**-0,052**	**-0,202**	**0,006**	**-0,246**	**-0,113**	**-0,160**	**-0,071**	**-0,028**
	0,224	0,112	0,092	-0,029	0,204	-0,074	0,048	-0,021	0,095	0,164

	lag 18					lag 24				
	1994	1995	1996	1997	1998	1994	1995	1996	1997	1998
	-0,436	-0,380	-0,318	-0,226	-0,269	-0,326	-0,102	-0,094	-0,144	-0,255
GER	**-0,277**	**-0,234**	**-0,151**	**-0,049**	**-0,087**	**-0,160**	**0,047**	**0,063**	**0,012**	**-0,091**
	-0,119	-0,087	0,013	0,125	0,095	0,008	0,188	0,225	0,164	0,082
	-0,396	-0,369	-0,284	-0,122	-0,248	-0,319	-0,041	-0,077	-0,167	-0,144
FRN	**-0,251**	**-0,222**	**-0,127**	**0,042**	**-0,065**	**-0,152**	**0,089**	**0,079**	**-0,031**	**0,015**
	-0,111	-0,080	0,030	0,205	0,118	0,020	0,216	0,242	0,102	0,184
	-0,470	-0,359	-0,284	-0,233	-0,286	-0,296	-0,120	-0,150	-0,183	-0,207
ITL	**-0,316**	**-0,196**	**-0,121**	**-0,057**	**-0,090**	**-0,134**	**0,035**	**0,019**	**-0,029**	**-0,038**
	-0,167	-0,046	0,044	0,120	0,100	0,034	0,184	0,186	0,118	0,135
	-0,465	-0,367	-0,309	-0,181	-0,290	-0,306	-0,161	-0,080	-0,143	-0,190
SPN	**-0,305**	**-0,210**	**-0,153**	**-0,011**	**-0,109**	**-0,143**	**-0,013**	**0,083**	**-0,004**	**-0,029**
	-0,151	-0,058	0,001	0,158	0,076	0,022	0,132	0,248	0,135	0,136

	cumul 12					cumul 24				
	1994	1995	1996	1997	1998	1994	1995	1996	1997	1998
	-2,141	-2,361	-2,302	-2,461	-2,002	-4,681	-4,738	-4,304	-3,880	-2,911
GER	**-1,142**	**-1,459**	**-1,316**	**-1,518**	**-0,958**	**-2,930**	**-2,965**	**-2,586**	**-1,983**	**-0,966**
	-0,283	-0,608	-0,536	-0,712	-0,024	-1,492	-1,570	-1,330	-0,636	0,562
	-2,154	-2,453	-2,358	-2,649	-1,964	-4,832	-4,806	-4,347	-4,021	-2,850
FRN	**-1,161**	**-1,558**	**-1,372**	**-1,716**	**-0,975**	**-3,048**	**-3,023**	**-2,606**	**-2,119**	**-0,976**
	-0,283	-0,727	-0,577	-0,915	-0,069	-1,617	-1,718	-1,380	-0,739	0,527
	-2,160	-2,416	-2,312	-2,414	-1,944	-4,798	-4,753	-4,325	-3,826	-2,781
ITL	**-1,165**	**-1,490**	**-1,324**	**-1,485**	**-0,886**	**-3,045**	**-2,954**	**-2,558**	**-1,951**	**-0,874**
	-0,302	-0,646	-0,555	-0,706	0,029	-1,644	-1,613	-1,311	-0,634	0,664
	-2,117	-2,468	-2,311	-2,361	-1,939	-4,729	-4,834	-4,298	-3,719	-2,760
SPN	**-1,124**	**-1,529**	**-1,318**	**-1,431**	**-0,902**	**-2,983**	**-3,058**	**-2,578**	**-1,790**	**-0,825**
	-0,292	-0,695	-0,540	-0,661	0,014	-1,630	-1,691	-1,347	-0,442	0,694

Note: For each lag the first quartile, the mean, and the third quartile are reported

Table 10. Testing the null: F(q) = F(q1). Time varying model. Common shock

all countries

	1994	1995	1996	1997	1998
all lags	0.1515	0.1535	0.1318	0.0822	0.0702
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
lag 12	0.0115	0.0118	0.075	0.0183	0.0158
	(0.9479)	(0.9385)	(0.0000)	(0.5055)	(0.6907)
lag 24	0.0165	0.0245	0.0195	0.0433	0.0227
	(0.6342)	(0.1749)	(0.421)	(0.0011)	(0.2438)
cumul 12	0.0175	0.0172	0.0068	0.0175	0.0088
	(0.5596)	(0.5781)	(1.0000)	(0.5596)	(0.9973)
cumul 24	0.0062	0.0152	0.0125	0.0063	0.01
	(1.0000)	(0.7280)	(0.9048)	(1.0000)	(0.9857)

Note: numbers represent the Kolmogorov-Smirnov statistics. P-values in brackets.

Table 11. Impact of a common monetary policy shock. Several lags. All countries. All years

	lag 6					lag 12				
	1994	1995	1996	1997	1998	1994	1995	1996	1997	1998
	-0,422	-0,352	-0,368	-0,506	-0,688	-0,221	-0,346	-0,392	-0,308	-0,084
GER	**0,008**	**-0,022**	**-0,050**	**-0,240**	**-0,366**	**0,055**	**-0,073**	**-0,139**	**-0,031**	**0,212**
	0,431	0,324	0,283	0,043	0,019	0,328	0,201	0,128	0,255	0,506
	-0,423	-0,335	-0,390	-0,546	-0,704	-0,196	-0,318	-0,430	-0,327	-0,060
FRN	**0,008**	**-0,020**	**-0,071**	**-0,271**	**-0,372**	**0,075**	**-0,065**	**-0,187**	**-0,056**	**0,232**
	0,433	0,329	0,252	0,005	0,020	0,349	0,210	0,067	0,224	0,525
	-0,427	-0,377	-0,439	-0,518	-0,685	-0,225	-0,330	-0,329	-0,317	-0,087
ITL	**0,010**	**-0,041**	**-0,107**	**-0,251**	**-0,355**	**0,053**	**-0,068**	**-0,078**	**-0,035**	**0,213**
	0,438	0,319	0,236	0,025	0,028	0,324	0,212	0,176	0,260	0,515
	-0,428	-0,407	-0,387	-0,525	-0,687	-0,220	-0,350	-0,386	-0,292	-0,050
SPN	**0,004**	**-0,076**	**-0,071**	**-0,255**	**-0,357**	**0,048**	**-0,085**	**-0,136**	**-0,013**	**0,240**
	0,415	0,276	0,260	0,022	0,029	0,316	0,195	0,119	0,272	0,545

	lag 18					lag 24				
	1994	1995	1996	1997	1998	1994	1995	1996	1997	1998
	-0,716	-0,587	-0,479	-0,627	-0,583	-0,566	-0,505	-0,522	-0,532	-0,511
GER	**-0,456**	**-0,288**	**-0,142**	**-0,277**	**-0,201**	**-0,275**	**-0,281**	**-0,256**	**-0,255**	**-0,206**
	-0,214	-0,025	0,154	-0,046	0,028	-0,008	-0,049	0,001	0,014	0,100
	-0,761	-0,636	-0,488	-0,565	-0,586	-0,556	-0,527	-0,576	-0,622	-0,550
FRN	**-0,497**	**-0,338**	**-0,152**	**-0,212**	**-0,201**	**-0,274**	**-0,311**	**-0,299**	**-0,341**	**-0,246**
	-0,250	-0,078	0,150	0,022	0,034	-0,005	-0,083	-0,041	-0,075	0,052
	-0,775	-0,593	-0,482	-0,583	-0,581	-0,541	-0,494	-0,547	-0,587	-0,550
ITL	**-0,504**	**-0,297**	**-0,149**	**-0,229**	**-0,202**	**-0,250**	**-0,279**	**-0,274**	**-0,311**	**-0,234**
	-0,250	-0,028	0,147	0,012	0,024	0,024	-0,046	-0,017	-0,043	0,080
	-0,698	-0,583	-0,470	-0,573	-0,554	-0,551	-0,543	-0,527	-0,551	-0,507
SPN	**-0,434**	**-0,282**	**-0,138**	**-0,219**	**-0,173**	**-0,262**	**-0,315**	**-0,258**	**-0,277**	**-0,192**
	-0,185	-0,015	0,164	0,021	0,064	0,009	-0,082	0,004	-0,010	0,110

	cumul 12					cumul 24				
	1994	1995	1996	1997	1998	1994	1995	1996	1997	1998
	-0,805	-0,842	-1,332	-1,481	-1,725	-2,343	-2,456	-2,447	-2,244	-1,980
GER	**0,458**	**0,246**	**-0,153**	**-0,506**	**-0,458**	**-0,805**	**-1,151**	**-0,994**	**-0,827**	**-0,204**
	1,664	1,353	1,099	0,491	0,833	0,743	0,222	0,551	0,601	1,507
	-0,781	-0,769	-1,307	-1,560	-1,707	-2,330	-2,450	-2,406	-2,297	-2,000
FRN	**0,439**	**0,300**	**-0,139**	**-0,615**	**-0,468**	**-0,790**	**-1,120**	**-0,971**	**-0,857**	**-0,188**
	1,642	1,418	1,122	0,410	0,794	0,743	0,270	0,577	0,617	1,552
	-0,731	-0,905	-1,402	-1,505	-1,668	-2,330	-2,450	-2,406	-2,297	-2,000
ITL	**0,498**	**0,182**	**-0,203**	**-0,570**	**-0,454**	**-0,790**	**-1,120**	**-0,971**	**-0,857**	**-0,188**
	1,715	1,302	1,039	0,431	0,812	0,743	0,270	0,577	0,617	1,552
	-0,769	-0,923	-1,354	-1,454	-1,660	-2,342	-2,572	-2,468	-2,190	-2,039
SPN	**0,454**	**0,187**	**-0,169**	**-0,491**	**-0,431**	**-0,814**	**-1,256**	**-1,004**	**-0,795**	**-0,209**
	1,667	1,326	1,075	0,498	0,843	0,691	0,151	0,538	0,682	1,557

Note: For each lag the first quartile, the mean, and the third quartile are reported

Table 12. Mean estimated impact of idiosyncratic monetary policy shocks. Several lags. All years

		1994	1995	1996	1997	1998
lag 1	*1st Qu.*	-0,226	-0,175	-0,080	-0,256	-0,510
	Mean	**-0,065**	**-0,044**	**0,078**	**-0,080**	**-0,324**
	Median	**-0,069**	**-0,043**	**0,073**	**-0,091**	**-0,344**
	3rd Qu.	0,096	0,088	0,228	0,086	-0,143
lag 6	*1st Qu.*	-0,114	-0,174	-0,162	-0,347	-0,199
	Mean	**0,034**	**-0,040**	**-0,024**	**-0,189**	**-0,016**
	Median	**0,029**	**-0,036**	**-0,037**	**-0,196**	**-0,020**
	3rd Qu.	0,175	0,089	0,104	-0,036	0,169
lag 9	*1st Qu.*	-0,460	-0,553	-0,409	-0,615	-0,420
	Mean	**-0,295**	**-0,410**	**-0,251**	**-0,457**	**-0,236**
	Median	**-0,306**	**-0,412**	**-0,252**	**-0,447**	**-0,232**
	3rd Qu.	-0,143	-0,264	-0,101	-0,296	-0,052
lag 12	*1st Qu.*	-0,401	-0,293	-0,302	-0,222	-0,210
	Mean	**-0,238**	**-0,131**	**-0,161**	**-0,058**	**-0,037**
	Median	**-0,249**	**-0,139**	**-0,168**	**-0,067**	**-0,038**
	3rd Qu.	-0,086	0,009	-0,030	0,094	0,129
lag 18	*1st Qu.*	-0,423	-0,346	-0,278	-0,184	-0,249
	Mean	**-0,284**	**-0,215**	**-0,136**	**-0,026**	**-0,084**
	Median	**-0,286**	**-0,222**	**-0,136**	**-0,029**	**-0,077**
	3rd Qu.	-0,151	-0,092	0,004	0,128	0,086
lag 24	*1st Qu.*	-0,279	-0,100	-0,084	-0,138	-0,179
	Mean	**-0,138**	**0,031**	**0,056**	**-0,012**	**-0,035**
	Median	**-0,136**	**0,024**	**0,063**	**-0,019**	**-0,027**
	3rd Qu.	0,011	0,152	0,201	0,110	0,120
cumul 12	*1st Qu.*	-2,158	-2,354	-2,278	-2,401	-1,952
	Mean	**-1,167**	**-1,485**	**-1,348**	**-1,499**	**-0,959**
	Median	**-1,245**	**-1,637**	**-1,501**	**-1,574**	**-1,055**
	3rd Qu.	-0,357	-0,672	-0,614	-0,760	-0,094
cumul 24	*1st Qu.*	-3,862	-3,881	-3,588	-3,312	-2,562
	Mean	**-2,222**	**-2,231**	**-2,138**	**-1,700**	**-1,008**
	Median	**-2,703**	**-2,695**	**-2,558**	**-1,995**	**-1,158**
	3rd Qu.	-1,209	-1,172	-1,180	-0,624	-0,324

Note: For each lag the first quartile, the mean, the median, and the third quartile are reported

Table 13. Mean estimated impact of a common monetary policy shock. Several lags. All years

		1994	1995	1996	1997	1998
lag 1	*1st Qu.*	-0,363	-0,337	-0,250	-0,265	-0,270
	Mean	**-0,069**	**-0,058**	**0,028**	**0,047**	**0,072**
	Median	**-0,058**	**-0,081**	**0,021**	**0,074**	**0,126**
	3rd Qu.	0,217	0,205	0,274	0,375	0,467
lag 6	*1st Qu.*	-0,427	-0,359	-0,391	-0,507	-0,673
	Mean	**0,004**	**-0,038**	**-0,078**	**-0,252**	**-0,361**
	Median	**0,060**	**-0,047**	**-0,091**	**-0,257**	**-0,366**
	3rd Qu.	0,420	0,312	0,241	0,017	0,022
lag 9	*1st Qu.*	-0,690	-0,890	-0,982	-1,289	-1,395
	Mean	**-0,340**	**-0,559**	**-0,652**	**-0,959**	**-0,958**
	Median	**-0,366**	**-0,599**	**-0,671**	**-0,922**	**-0,931**
	3rd Qu.	-0,032	-0,219	-0,332	-0,575	-0,550
lag 12	*1st Qu.*	-0,207	-0,313	-0,372	-0,300	-0,071
	Mean	**0,052**	**-0,070**	**-0,130**	**-0,030**	**0,216**
	Median	**0,059**	**-0,059**	**-0,134**	**-0,008**	**0,211**
	3rd Qu.	0,312	0,188	0,119	0,247	0,511
lag 18	*1st Qu.*	-0,718	-0,592	-0,481	-0,567	-0,561
	Mean	**-0,465**	**-0,300**	**-0,154**	**-0,229**	**-0,195**
	Median	**-0,491**	**-0,345**	**-0,242**	**-0,309**	**-0,297**
	3rd Qu.	-0,237	-0,052	0,136	-0,003	0,024
lag 24	*1st Qu.*	-0,545	-0,507	-0,523	-0,553	-0,517
	Mean	**-0,266**	**-0,294**	**-0,272**	**-0,293**	**-0,223**
	Median	**-0,256**	**-0,286**	**-0,274**	**-0,296**	**-0,253**
	3rd Qu.	-0,014	-0,077	-0,030	-0,050	0,071
cumul 12	*1st Qu.*	-0,689	-0,781	-1,354	-1,631	-1,771
	Mean	**0,475**	**0,205**	**-0,252**	**-0,640**	**-0,469**
	Median	**0,465**	**0,283**	**-0,199**	**-0,491**	**-0,412**
	3rd Qu.	1,633	1,249	0,874	0,330	0,773
cumul 24	*1st Qu.*	-2,071	-2,148	-2,192	-1,878	-1,673
	Mean	**-0,625**	**-0,971**	**-0,900**	**-0,701**	**-0,011**
	Median	**-0,528**	**-0,845**	**-0,817**	**-0,599**	**-0,121**
	3rd Qu.	0,808	0,361	0,543	0,562	1,680

Note: For each lag the first quartile, the mean, the median, and the third quartile are reported

Figures

Figure 1. German parameters in own reaction function

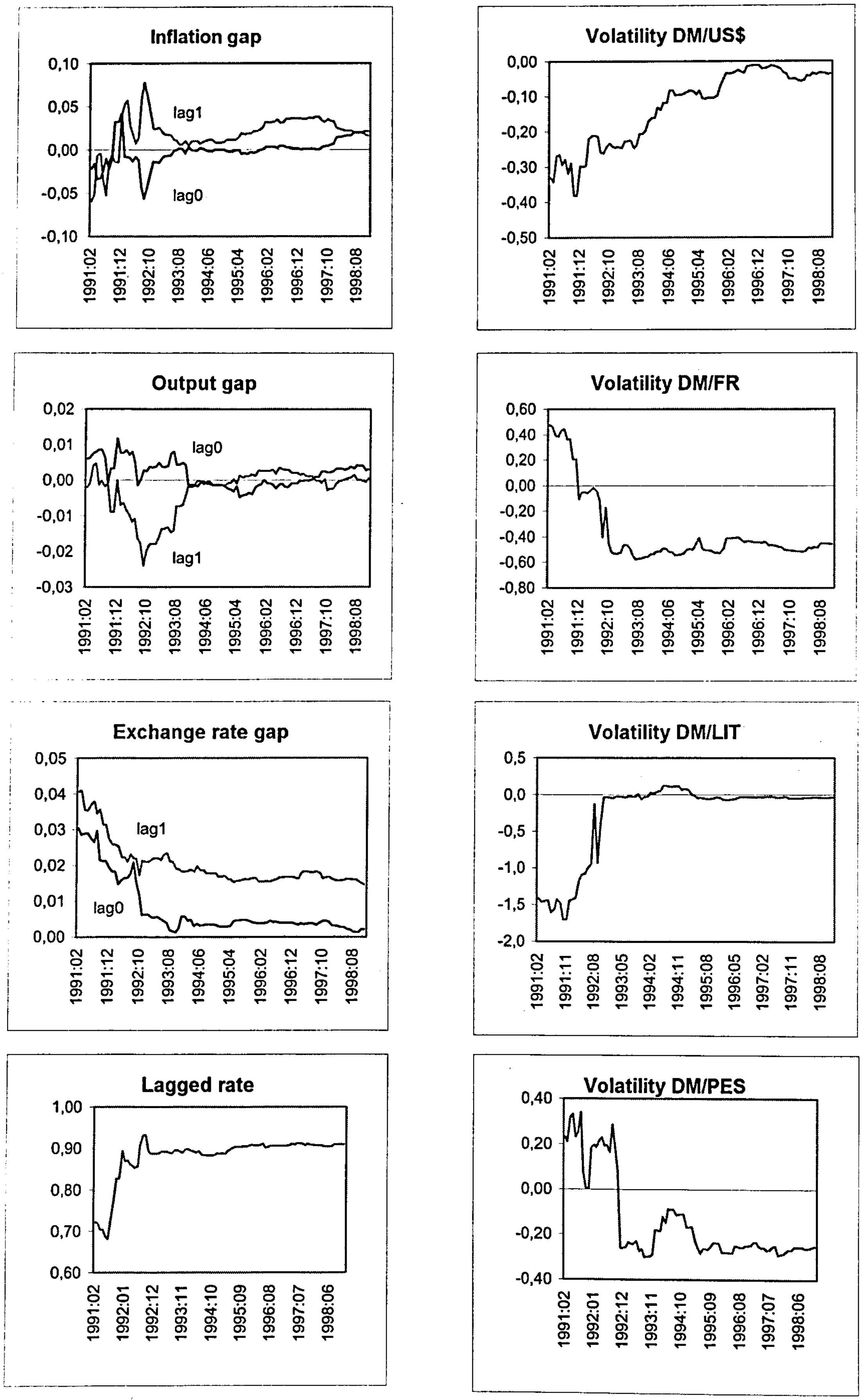


Figure 2. French parameters in own reaction function

Own | **German**

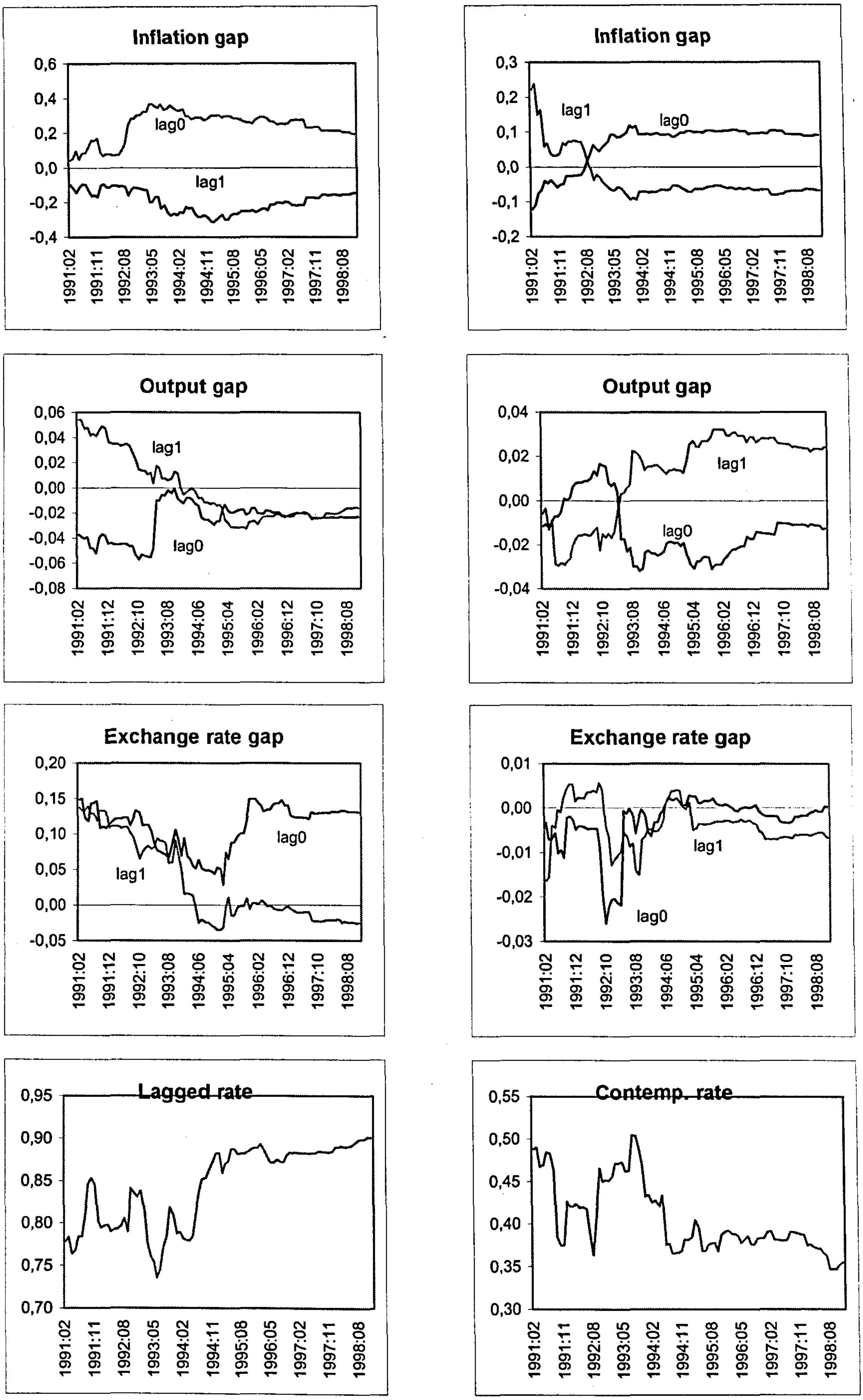


Figure 3. Italian parameters in own reaction function

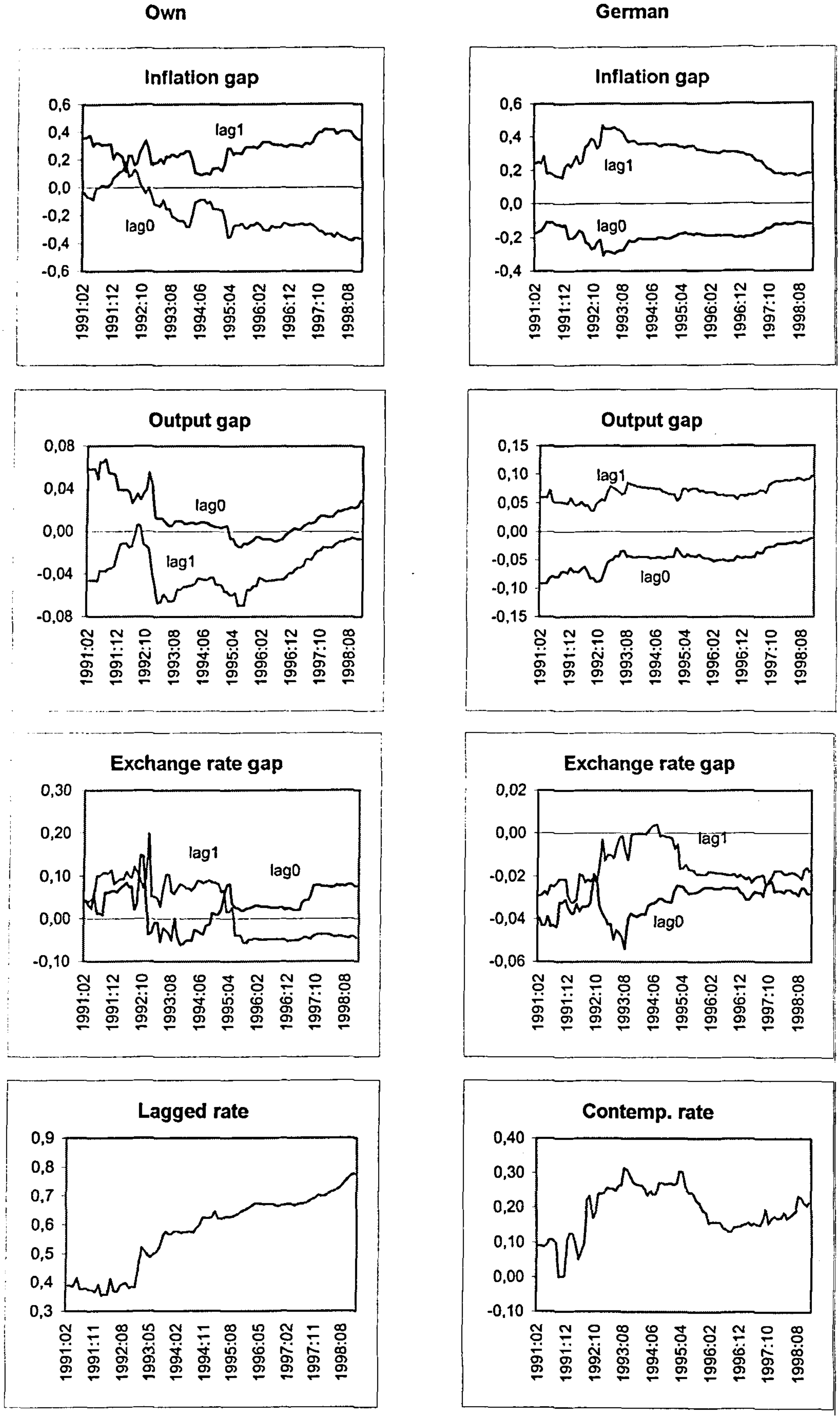


Figure 4. Spanish parameters in own reaction function

Own

German

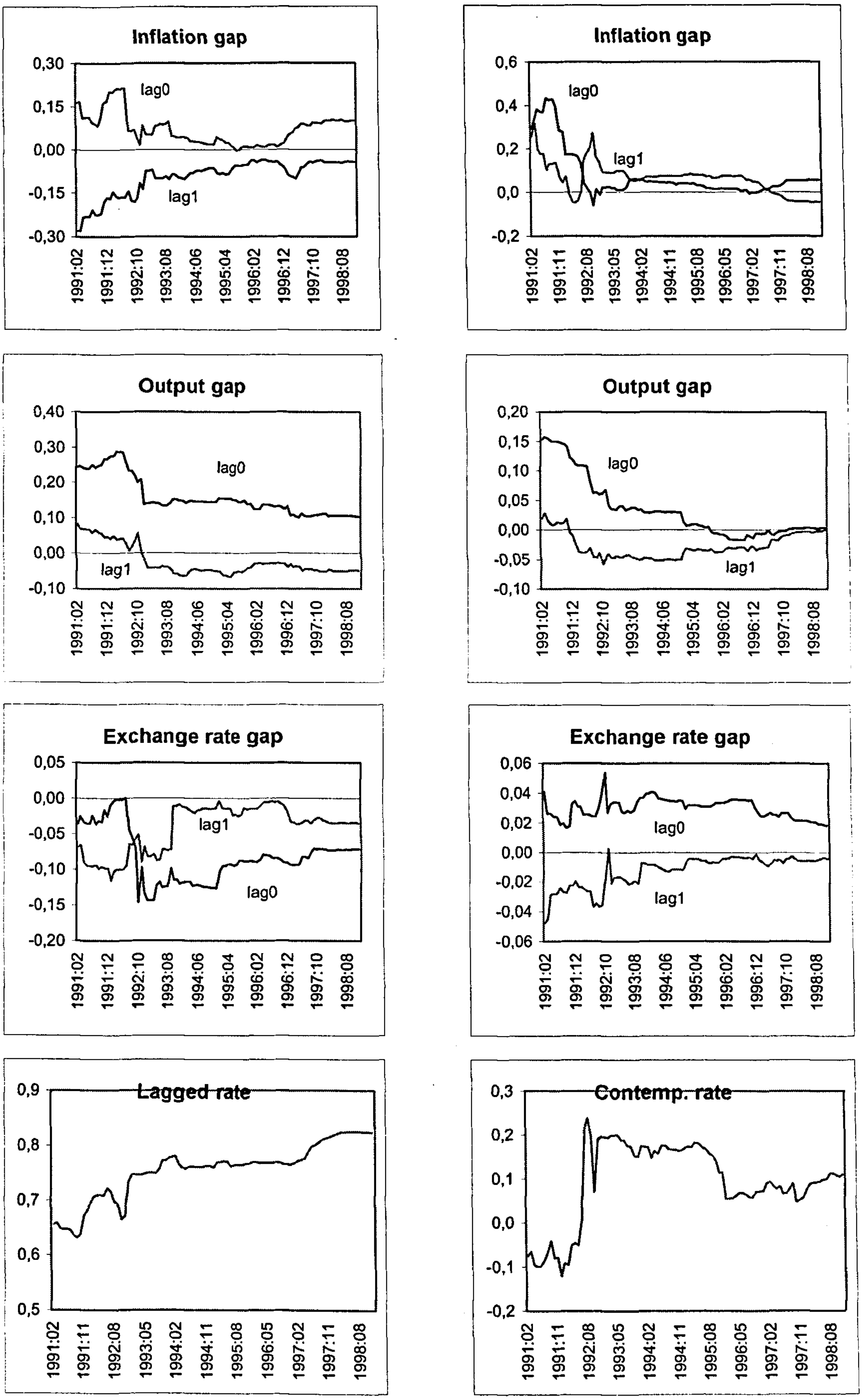


Figure 5. Volatilities in non-German countries

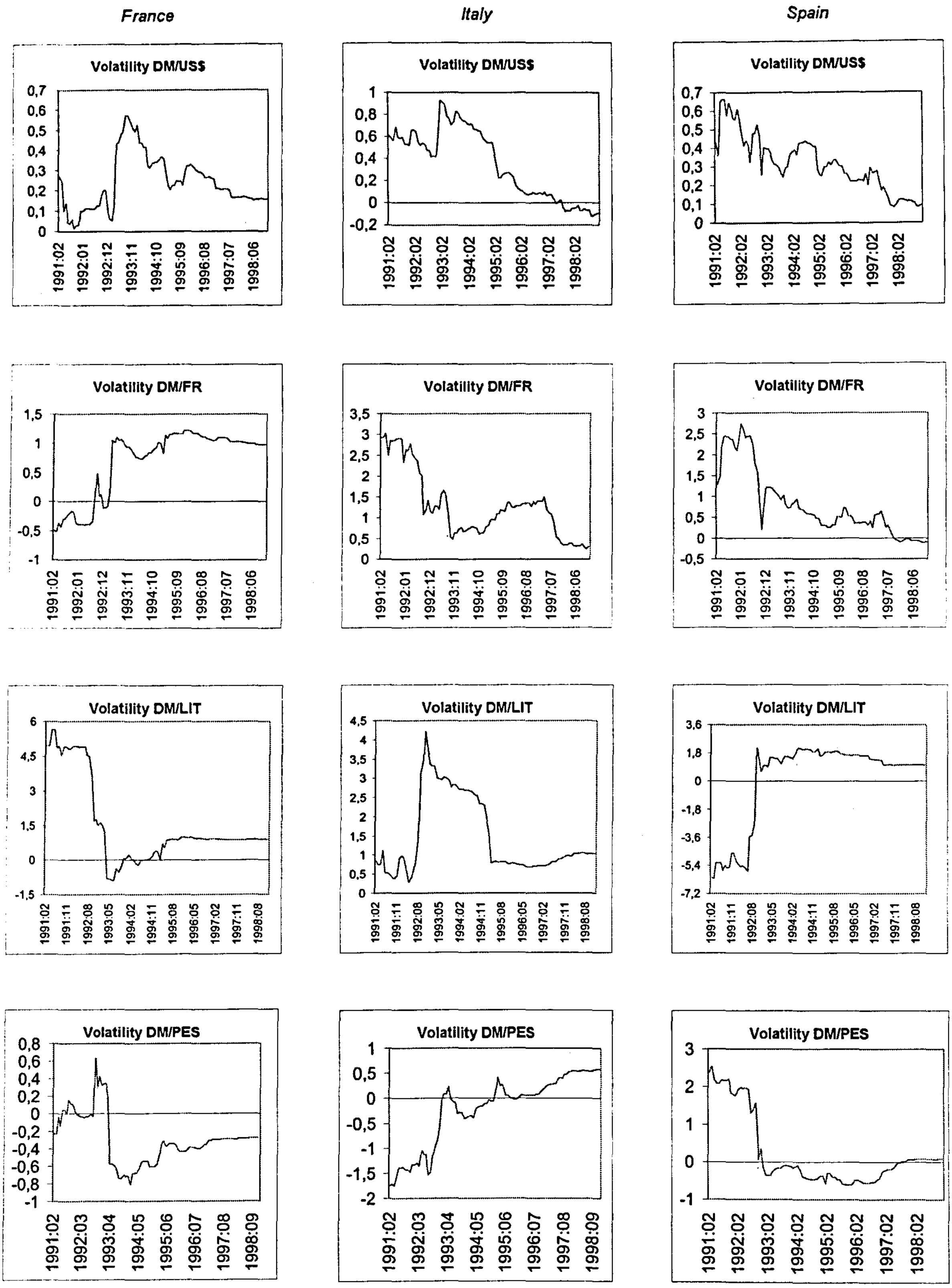


Figure 6. Monetary policy shocks

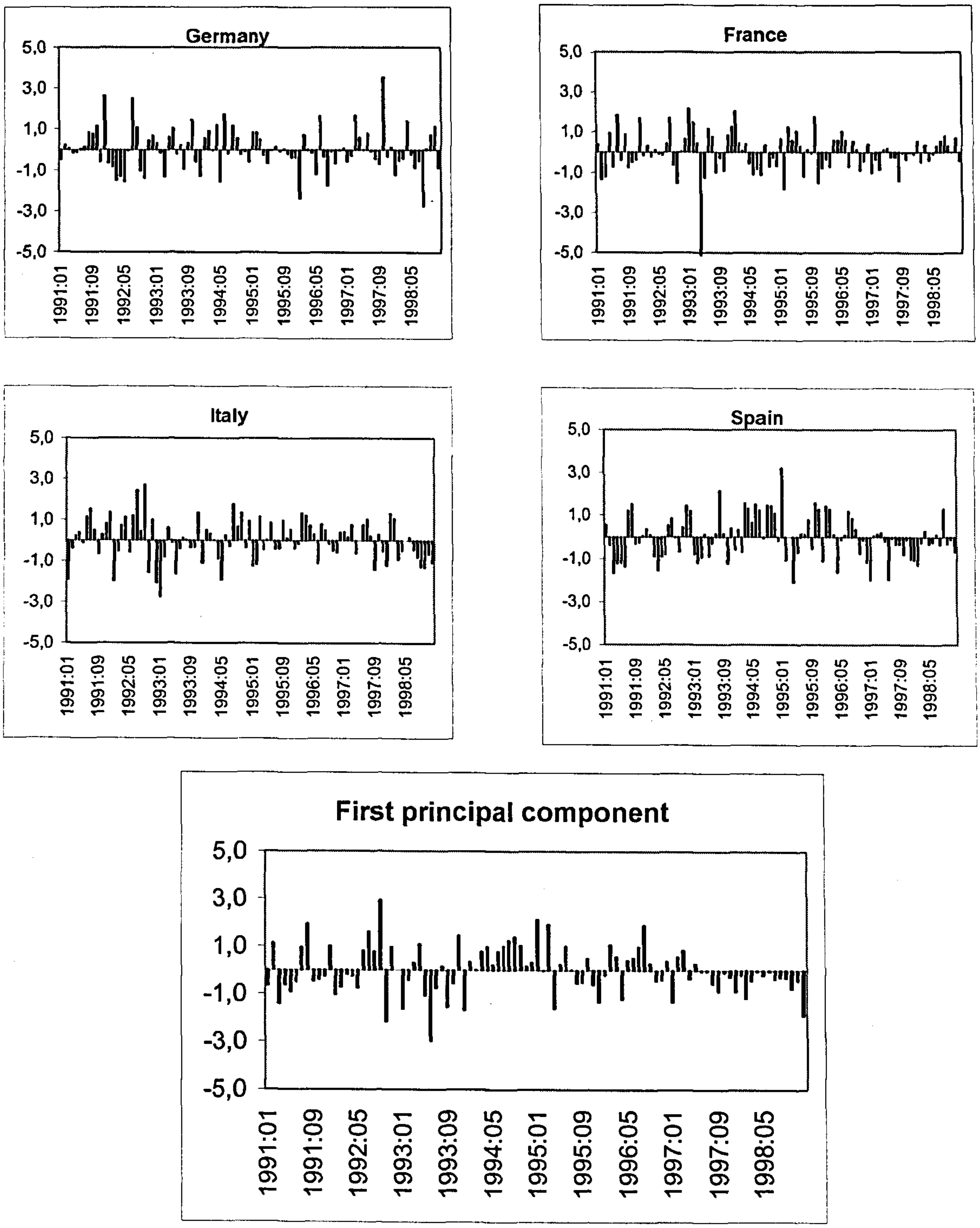


Figure 7. Impact of idiosyncratic shocks. Several lags. All countries. All years

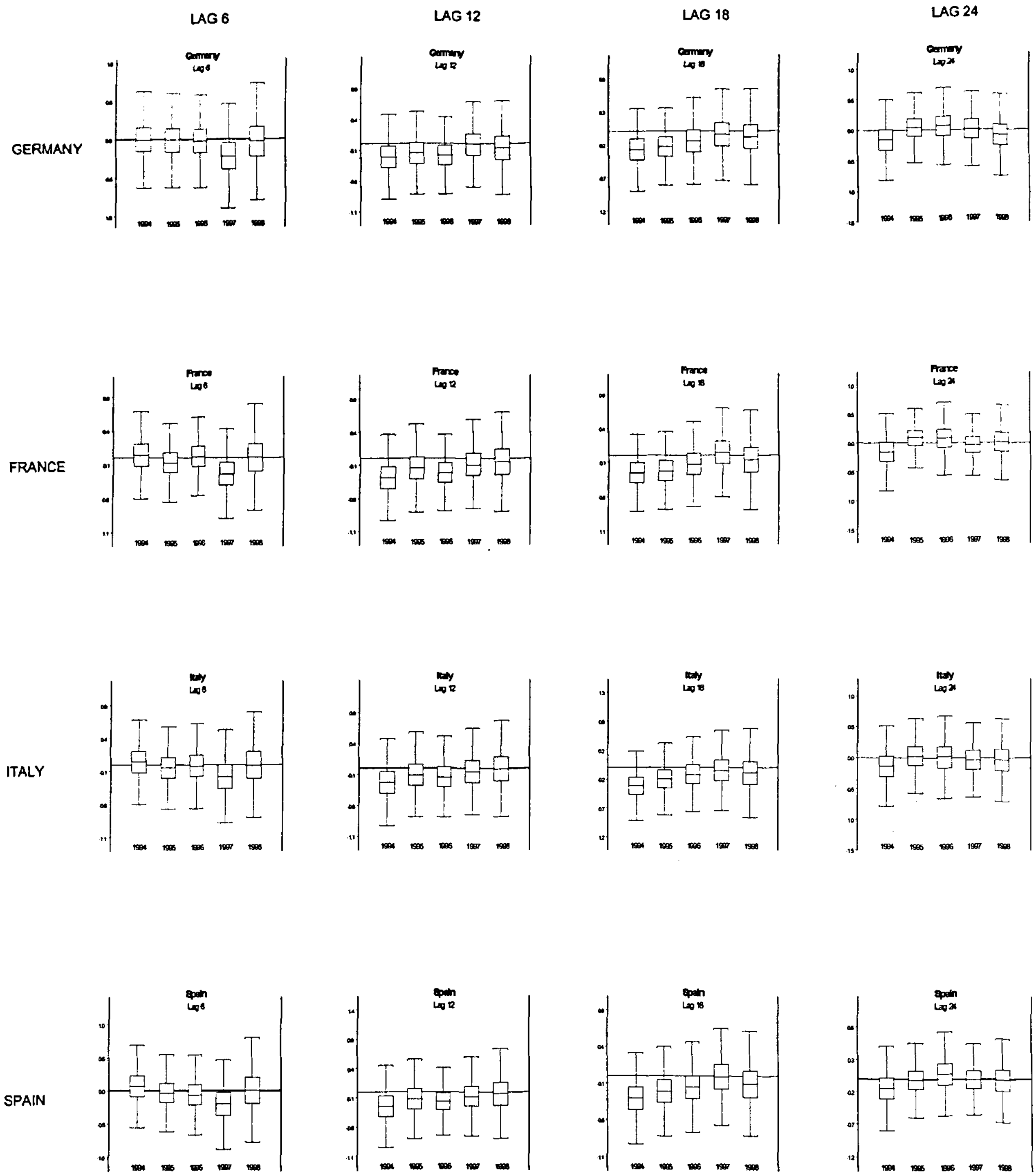


Figure 8. Impact of a common shock. Several lags. All countries. All years

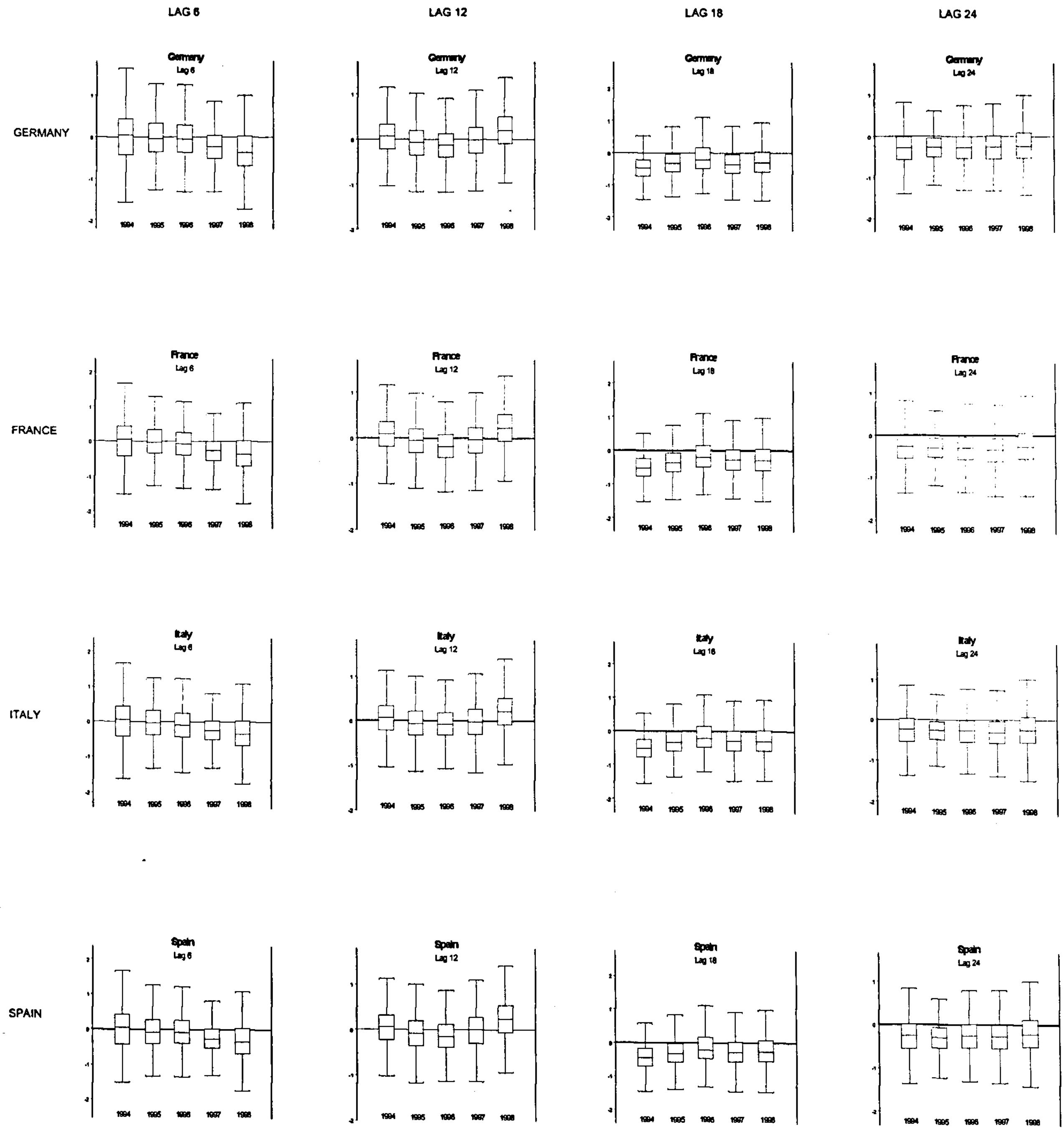
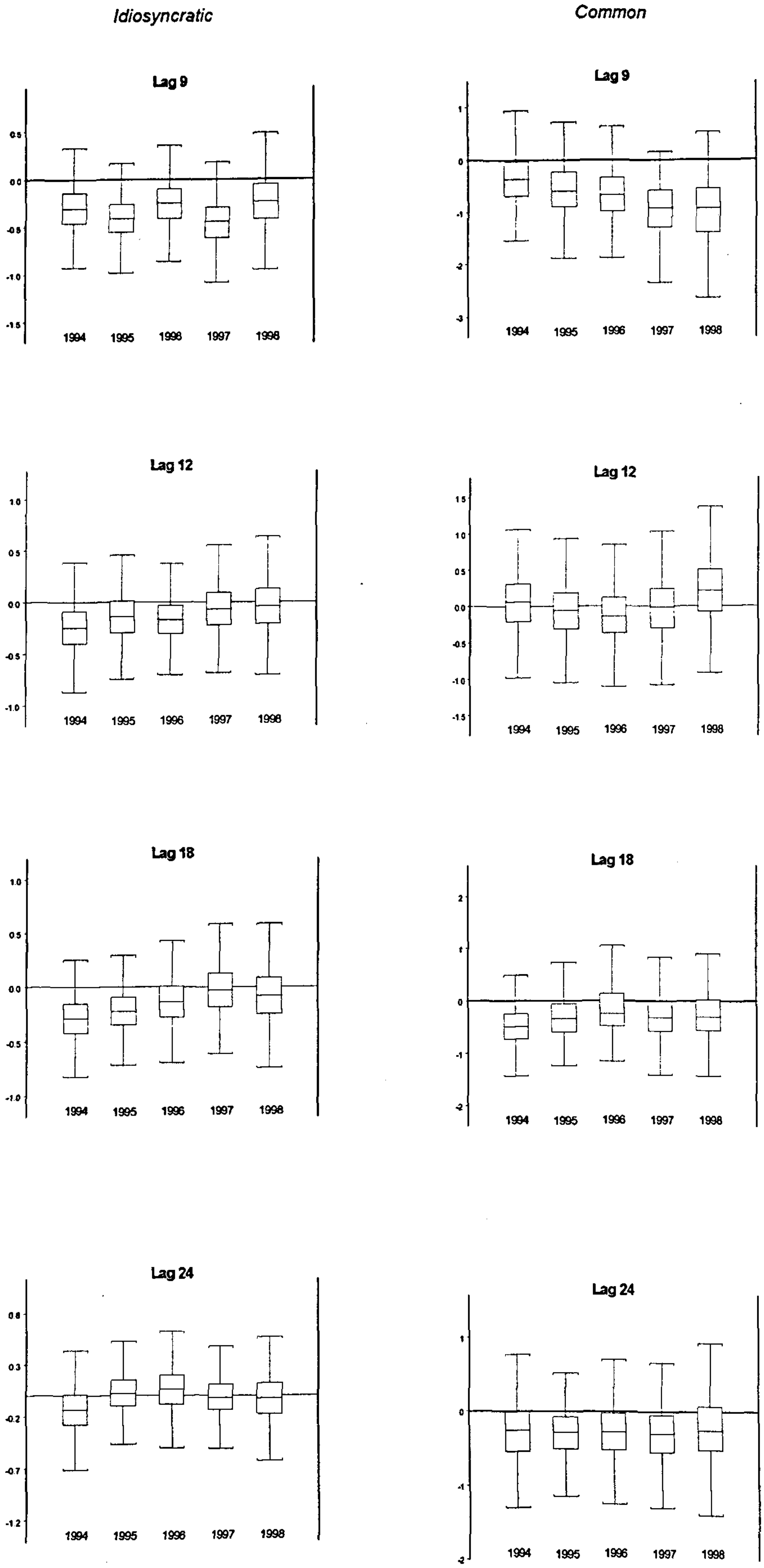


Figure 9. Euro-wide impact of idiosyncratic and common shocks



REFERENCES

Amisano, G. and C. Giannini (1997), *Topics in Structural VAR Econometrics*, 2nd ed., Berlin: Springer.

Ballabriga F., M. Sebastián, and J. Vallés (1999), 'European asymmetries', *Journal of International Economics*, 48, 233-253.

Bank for International Settlements, 1995, 'Financial structure and the monetary policy transmission mechanism', Basel, Switzerland, March, No. C.B. 394.

Barran F., V. Coudert, and B. Mojon (1996), 'The transmission of monetary policy in the European countries', *CEPII, Document de travail No.* 96-03.

Canova F. and M. Ciccarelli (1999), 'Forecasting and turning point prediction in a Bayesian panel VAR model', *mimeo, UPF,* Barcelona, Spain.

Canova F. and A. Marcet (1997), 'The poor stay poor: non–convergence across countries and regions', forthcoming in *Review of Economic Studies.*

Canova F. and G. de Nicoló (1998), 'Did you know that monetary disturbances matter for business cycles fluctuations? Evidence from the G–7 countries', *mimeo* UPF, Barcelona, Spain.

Chib, S. and E. Greenberg (1995), 'Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models', *Journal of Econometrics*, 68, 409–431.

Christiano L., M. Eichenbaum, and C. Evans (1998), 'Monetary policy shocks: what have we learned and to what end?', *NBER Working Paper* No. 6400.

Ciccarelli M. (2000), 'Testing restrictions in normal data models using Gibbs sampling', *mimeo,* Universidad de Alicante.

Clarida R., J. Gali, and M. Gertler (1997), 'Monetary policy rules in practice. Some international evidence', *European Economic Review,* 42, 1033-1067.

Clements B. and Z. Knotolemis (2001), 'Monetary Policy under EMU: Differences in the Transmission Mechanism?', *mimeo,* International Monetary Fund, Washington DC.

Cecchetti S.G. (1999), 'Legal structure, financial structure and the monetary policy', Federal Reserve Bank of New York, *mimeo.*

Dornbusch R., C. Favero, and F. Giavazzi (1998), 'A red letter day', *CEPR Discussion Paper No.* 1804.

Faust J. (1998), 'The robustness of identified VAR conclusions about money", paper prepared for the *Fall 1997 Carnegie-Rochester Conference on Public Policy.*

Favero C. and F. Giavazzi (1999), 'An evaluation of monetary policy transmission in the context of the European Central Bank': a report to the European Parliament.

Gelfand A.E., S. E. Hills, A. Racine-Poon, and A.F.M. Smith (1990), 'Illustration of Bayesian inference in normal data models using Gibbs sampling', *Journal of the American Statistical Association*, 85, 972-985.

Gerlach S. and F. Smets (1995), 'The monetary transmission mechanism: evidence from the G–7 countries', *BIS Discussion Paper.*

Giovannetti, G. and R. Marimón (1998), 'An EMU with different transmission mechanisms?', IUE, *Working Paper 98/33,* Florence.

Guiso L., A.K. Kashyap, F. Panetta, and D. Terlizzese (2000), 'Will a common European monetary policy have asymmetric effects?', *Economic Perspectives,* Federal Reserve Bank of Chicago.

Hsiao, C., M. H. Pesaran and A. K. Tahmiscioglu (1997), 'Bayes estimation of short run coefficients in dynamic panel data models', *mimeo,* Cambridge University.

Kim S. (1998), 'Identifying European monetary policy interactions: French and Spanish system with German variables', Bank of Spain, *Working Paper No. 9811.*

Leamer E. E. (1978), *Specification Searches. Ad Hoc Inference with Nonexperimental Data,* New York: Wiley.

Litterman, R. (1980), 'Techniques for forecasting with vector autoregressions', University of Minnesota, Ph.D. Thesis.

Litterman, R. (1986), 'Forecasting with Bayesian vector autoregressions: five years of experience', *Journal of Business and Economic Statistics,* 4, 25-38.

Monticelli C. and O. Tristani (1999), 'What does the single monetary policy do? A SVAR benchmark for the European Central Bank', *ECB Working Paper, No.* 2.

Ortega E. and E. Alberola (2000), 'Transmission of shocks and monetary policy in the Euro area. An exercise with NIGEM', Bank of Spain, *Working Paper No. 0010.*

Pesaran M.H. and R. Smith (1995), 'Estimating long run relationships from dynamic heterogeneous panels', *Journal of Econometrics,* 68, 79–113.

Peersman G. and F. Smets (1998), 'The Taylor rule: A useful monetary policy guide for the ECB?', paper presented at the conference *'Monetary Policy of the ESCB: Strategic and Implementation Issues'*, Bocconi University, Milan, Italy.

Ramaswamy R. and T. Sloek (1997), 'The real effects of monetary policy in the European Union: what are the differences?' *IMF Working Paper No.* 97/160.

Rebucci, A. (2000), 'Estimating vector autoregressions with long heterogeneous panels', *mimeo,* International Monetary Fund, Washington DC..

Sala, L.(2000), 'Monetary transmission in the Euro area: a factor model approach', *mimeo,* ULB-ECARES.

Sims, C. and T. Zha (1998), 'Bayesian methods for dynamic multivariate models', *International Economic Review,* 39, 949-979.

Smets, F. (1997), 'Measuring monetary policy shocks in France, Germany and Italy: the role of the exchange rate', *BIS Discussion Paper.*

Theil, H. (1971), *Principles of Econometrics*, New York: Wiley.

Uhlig, H. (1999), 'What are the effects of monetary policy on output? Results from an agnostic identification procedure', *mimeo*, CentER, Tilburg.

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

Zha, T. (1999), 'Block recursion and structural vector autoregressions', *Journal of Econometrics* 90, 291-316.

Chapter 4

Testing Restrictions in Normal Data Models Using Gibbs Sampling

Lo más trágico no es ser mediocre pero inconsciente de esa mediocridad; lo más trágico es ser mediocre y saber que se es así y no conformarse con ese destino que, por otra parte (éso es lo peor) es de estricta justicia.

(Mario Benedetti, *La Tregua*)

4.1 Introduction

In these paper we consider the simple problem of testing the vector of restrictions $\mathsf{R}(\boldsymbol{\theta}) = 0$, where $\boldsymbol{\theta} \in \Theta$ is the unknown parameter vector of a model for the data Y, defined by a normal pdf, $\phi(Y \mid \boldsymbol{\theta})$. The aim is to form a posterior probability for the truth of the set of restrictions, conditional to the data. The paper can also be considered as a further illustration of the versatility and ease of practical implementation of the Gibbs sampler, a sampling-based approach proposed by Geman and Geman (1984) and popularized by Gelfand and Smith (1990) to calculate marginal posterior densities in complex hierarchical models. The setting of the problem and its solution are purely Bayesian, but the results are easily comparable (at least in terms of interpretation) with the classical approach to testing.

Traditionally, the comparison of two or more parametric (not necessarily nested) models in the Bayesian framework is based on posterior model probabilities. In the simplest case in which we have two models or hypotheses, H_0, H_1 with prior probabilities $p(H_0)$, $p(H_1)$, the statistic that is most frequently employed to compare H_0 and H_1 is the posterior odds (PO) ratio

$$\frac{p(H_o \mid y)}{p(H_1 \mid y)} = \frac{p(y \mid H_o)\, p(H_o)}{p(y \mid H_1)\, p(H_1)}$$

If the loss is one for choosing the incorrect model and zero for choosing the correct one, then we select model H_o if this ratio is greater than one.[1]

This way of comparing and eventually choosing between two models is feasible when all priors involved are informative. In fact, the marginal likelihood $m(y) = p(y \mid H_k)$ is generally obtained computing the integral

$$m(y) = \int p(\vartheta_k)\, p(y \mid \vartheta_k)\, d\vartheta_k \tag{4.1}$$

where ϑ_k denotes the vector of all the parameters of model k. In some very elementary cases this integral can be analytically tractable (Zellner, 1971, ch.10). However, when the dimension of the parameter vector increases, the integration can hardly be an easy task, and must be overcome with a Monte Carlo method. Chib (1995) developed an approach based on the simple fact that $m(y)$, by virtue of being the normalizing constant of the posterior density, can be written as

$$m(y) = \frac{f(y \mid \vartheta)\, p(\vartheta)}{p(\vartheta \mid y)}$$

where the numerator is the product of the sampling density and the prior, with all integrating constants included, and $p(\vartheta \mid y)$ is the posterior density of ϑ. For a given ϑ (the ML

[1] In fact the model with the highest posterior probability $p(H_k \mid D)$ must be chosen (and this rule is optimal, in the sense described by Zellner, 1971, pp.294–297), provided we can define a symmetric loss structure. For discussion and applications of other loss functions, see Schorfheide (2000).

estimate, for instance), the latter quantity can be estimated with the *Rao–Blackwellization* technique suggested by Gelfand and Smith (1990), using the Gibbs output, while the numerator is easily evaluated at the same ϑ chosen. In order to compute the marginal density $m(y)$, it is important that all integrating constants of the full conditional distribution of the Gibbs sampler be known.

Since non diffuse prior information affects posterior odds in both small and large samples, a special care must be exercised in representing the prior information to be employed in the analysis. In many situations a vague or diffuse prior information needs to be employed. When the prior information on the parameters is vague or diffuse, the posterior odds ratio cannot be calculated. In this case Lindley (1965) suggested a procedure that, for many problems leads to tests which are computationally equivalent to sampling-theory tests. This procedure uses a Bayesian confidence region. If we have a joint hypothesis about two or more parameters, say $\boldsymbol{\theta}$, a Bayesian "highest posterior density" confidence region for $\boldsymbol{\theta}$ is first obtained with a given probability content $1-\alpha$. If our hypothesis is for example $\boldsymbol{\theta} = \boldsymbol{\theta}_o$, where $\boldsymbol{\theta}_o$ is a given vector, we accept if $\boldsymbol{\theta}_o$ is contained in the confidence region and reject otherwise at the α level of significance.[2]

This procedure is appropriate only when prior information is vague or diffuse, otherwise it is important to take into account any prior knowledge. Consider a simple hypothesis $\boldsymbol{\theta} = \boldsymbol{\theta}_o$, where $\boldsymbol{\theta}_o$ is a value suggested by the theory. In this case, it is reasonable to believe that $\boldsymbol{\theta}_o$ is a more probable value for $\boldsymbol{\theta}$ than any other. Thus, a testing procedure that allows to incorporate non diffuse prior information is needed, and the comparison of alternative hypotheses might be based on the posterior odds ratio. In fact, as shown in Zellner (1971, p. 304), as sample size increases a "sampling theory test of significance can give results

[2] See Zellner, 1971, p. 298-302, for details. Notice that in most problems the interval (region) is numerically exactly the same as a sampling theory confidence interval (region) but is given an entirely different interpretation in the Bayesian approach.

differing markedly from those obtained from a calculation of posterior probabilities which takes account of non-diffuse prior and sample information". For large sample sizes, the paradox of obtaining a probability of $\boldsymbol{\theta} = \boldsymbol{\theta}_o$ close to one even in regions that would lead to rejection of the hypothesis $\boldsymbol{\theta} = \boldsymbol{\theta}_o$ can arise (Lindley's paradox).

The aim of this paper is to test the set of restrictions $\mathsf{R}(\theta) = 0$ in a complex hierarchical model with a procedure that avoids the computational difficulties of the PO ratio and could be used under diffuse and non diffuse prior information. The rationale of the approach is very simple, being based on the comparison between two distributions which are immediately obtained in the Gibbs sampler. One is the posterior distribution of $\boldsymbol{\theta}$ and the other is the posterior distribution of the parameter vector under the restriction. The degree of overlap of the two distributions provides a criterium to verify the restriction: the larger the distance between these two posterior distributions, the higher the (posterior) probability of rejecting the null. The idea is closer in spirit to Lindley's suggestion and can be considered as the Bayesian version of the classical Wald type tests. This similarity and the fact that the properties of the approach we propose are analyzed to a large extent using the sampling properties of the estimators involved, should make the approach attractive also to classical sampling-theory econometricians.

With the help of several simulation experiments, we find that this empirical method has very good properties in terms of power and size of the test, under different prior assumptions, and is competitive with the standard PO ratio both in small and in large samples. As the sample size increases, simulations do not seem to give rise to Lindley's paradox when prior information is vague or diffuse.

The paper is organized as follows. Section 2 describes the empirical approach. Section 3 discusses the design of the Monte Carlo study. In section 4 we analyze the properties of the test in terms of power and unbiasedness in several simulation experiments, under different assumptions on the prior information, and compare with PO ratio when informative priors are used. Section 5 concludes.

4.2 An empirical approach

In many circumstances it is reasonable to assume linearity. So, let the model be

$$y = X\theta + \varepsilon \tag{4.2}$$

where y is a vector of dimensions $n \times 1$, X is a $n \times k$ matrix of explanatory variables and ε is a vector of disturbances of dimensions $n \times 1$. Notice that under the assumption of linearity, several possible specification can be adapted. As a matter of fact, Eq. (1) can refer to both univariate and multivariate models; matrix X can contain lagged endogenous and exogenous variables; data can proceed from cross section, time series or panel analysis, dimensions changing accordingly in the specification (4.2).

Let us assume normality

$$\varepsilon \sim N\left(0, \Sigma_{\varepsilon}\right), \tag{4.3}$$

where Σ_{ε} is the error term variance-covariance matrix of dimensions $n \times n$, and model the population structure as

$$\theta \sim N\left(A_o\bar{\theta}, \Sigma_{\theta}\right) \tag{4.4}$$

where A_o is a known matrix of dimensions $k \times m$, relating the regression vector θ to a parameter vector $\bar{\theta}$ of dimensions $m \times 1$, possibly with $m \leq k$, and Σ_{θ} is the $k \times k$ variance-covariance matrix of the random vector θ.

Notice that this is a hierarchical model of the kind introduced by Lindley and Smith

(1972), whose applications abound in fields as different as educational testing (Rubin 1981), medicine (DuMouchel and Harris 1983), and economics (Hsiao et al., 1998).

A full implementation of the Bayesian approach is easily achieved – at least for the normal linear hierarchical model structure –using the Gibbs sampler. It requires the specification of a prior for Σ_ε, $\bar{\theta}$ and Σ_θ. Assuming independence, as it is customary, we may take the joint prior distribution

$$p\left(\bar{\theta}, \Sigma_\varepsilon^{-1}, \Sigma_\theta^{-1}\right) = p\left(\bar{\theta}\right) p\left(\Sigma_\varepsilon^{-1}\right) p\left(\Sigma_\theta^{-1}\right)$$

to have, for example, a normal–Wishart–Wishart form:

$$p\left(\bar{\theta}\right) = N\left(A_1\mu, C\right)$$

$$p\left(\Sigma_\varepsilon^{-1}\right) = W\left[\left(\sigma_\varepsilon S_\varepsilon\right)^{-1}, \sigma_\varepsilon\right]$$

$$p\left(\Sigma_\theta^{-1}\right) = W\left[\left(\sigma_\theta S_\theta\right)^{-1}, \sigma_\theta\right]$$

where A_1 is a known matrix of dimensions $m \times p$, relating the regression vector $\bar{\theta}$ to a parameter vector μ of dimensions $p \times 1$, possibly with $p \leq m$, while the hyperparameters μ, C, σ_ε, $S_\varepsilon, \sigma_\theta$, S_θ are assumed all known. The notation $W\left[\Omega, \omega\right]$ identifies a Wishart distribution with ω degrees of freedom and scale matrix Ω.

The unfeasible integrability of this model to get the posterior distributions of interest justifies the use of the Gibbs sampler. Typical inferences of interest in such studies include marginal posteriors for the population parameters θ or $\bar{\theta}$. Our purpose is to show how these inferences can be achieved by using the Gibbs sampling output in a very natural way.

In particular, let us concentrate our attention on $\bar{\theta}$. It is easy to show that the posterior distribution of $\bar{\theta}$ conditional on Σ_ε^{-1}, Σ_θ^{-1}, θ, y, is of the form

$$p\left(\bar{\theta} \mid \Sigma_\varepsilon^{-1}, \Sigma_\theta^{-1}, \theta, y\right) = N\left(\bar{\theta}^*, V^*\right) \tag{4.5}$$

where

$$\bar{\theta}^* = V^* \left[C^{-1} A_1 \mu + A_o' \Sigma_\theta^{-1} \theta\right] \quad (4.6)$$

$$V^* = \left(C^{-1} + A_o' \Sigma_\theta^{-1} A_o\right)^{-1} \quad (4.7)$$

Suppose now that we are interested in testing the set of linear restrictions

$$R\bar{\theta} = r \quad (4.8)$$

where R is a known matrix of dimensions $s \times m$, with $s \leq m$. From (4.5) we have the additional information that, conditional on $\Sigma_\varepsilon, \Sigma_\gamma^{-1}, \theta, y$, the quadratic form

$$q = \left[R\left(\bar{\theta} - \bar{\theta}^*\right)\right]' \left[RV^*R'\right]^{-1} \left[R\left(\bar{\theta} - \bar{\theta}^*\right)\right] \quad (4.9)$$

is distributed as a $\chi^2_{(s)}$. The marginal posterior distribution of this quantity can easily be obtained in the Gibbs sampling. It provides a rational for examining the posterior plausibility of the set of linear restrictions (4.8). As a matter of fact, according to (4.9), the probability that $R\bar{\theta}$ would equal r is related to the probability that, at each iteration of the Monte Carlo, a $\chi^2_{(s)}$ variable would assume the value

$$q_1 = \left[R\bar{\theta} - r\right]' \left[RV^*R'\right]^{-1} \left[R\bar{\theta} - r\right] \quad (4.10)$$

Therefore, the probability that a $\chi^2_{(s)}$ variable could exceed this magnitude represents the probability that the random variable $R\bar{\theta}$ might be as far from the posterior mean $R\bar{\theta}^*$ as is represented by the point $R\bar{\theta}^* = r$.

Provided we can obtain the empirical posterior distributions of q e q_1 , in order to construct a rejection region it is sufficient to compare these two distributions. The larger the distance between q and q_1, the greater is the probability, *a posteriori*, of rejecting the null.

Notice that, based on the comparison between (4.9) and (4.10), we are not testing the exact restriction (4.8), but rather the fact that $R\bar{\theta}$ is distributed *a posteriori* around r.[3]

It is immediate to see that the prior hyperparameters can be specified in such a way that they reflect vague initial information relative to that to be provided by the data. It is enough to assume, for example, an infinite uncertainty on the second stage of the hierarchy, by taking $C^{-1} = 0$. Under this prior assumption, (4.6) and (4.7) change accordingly without modifying the characteristics of the testing discussed above.

The idea behind the approach is basically the same as in the classical Wald test, where we compare two distributions: one under the null, which is asymptotically $\chi^2_{(s)}$; and the other under the alternative. The greater is the numerical value of the quadratic form where the set of restrictions has been substituted, the more likely this value belongs to the distribution under the alternative, which is a *non-central* $\chi^2_{(s)}$. Here (4.9) plays the role of the distribution under the null. The main difference is that this is an exact distribution whose posterior can be computed empirically and used to make probability assessments in a Bayesian fashion. On the other hand, the posterior distribution of (4.10) (and not just one value, as in the classical analysis) can also be computed and compared with (4.9). The greater is the distance between the two posterior distributions, the more likely the restriction we put is converting the reference distribution in a *non-central* one, and the more likely we reject the null.

There are several ways of measuring this distance, beside the graphical overlap. The

---

[3] The test of the exact restriction can be conducted instead by constructing the quadratic form

$$q_2 = \left[r - R\bar{\theta}^*\right]' \left[RV^*R'\right]^{-1} \left[r - R\bar{\theta}^*\right].$$

In a Bayesian set up like the one described above, previous works (see Hsiao et al., 1998, for references) have shown that the estimates of the average coefficients ($\bar{\theta}^*$) have a very reduced bias, even in a dynamic panel data model. Therefore, it is very likely that, when the null is true, the distance $\left[r - R\bar{\theta}^*\right]$ would be much lower than $\left[R\bar{\theta} - r\right]$ in the same metric $\left[RV^*R'\right]^{-1}$, hence leading to a much lower number of rejections, given the size of the test. Since several simulation experiments (not shown) confirmed this finding, we prefer to base our reasoning on the comparison between q and q_1.

simplest one can be based on a test on the means of the distributions of q and q_1. More sophisticated nonparametric methods can concern the comparison of the cumulative distribution functions (cdf) of q and q_1 (*Kolmogorov-Smirnov Goodness-of-Fit test*), as well as of the percentiles of the empirical posterior density functions of the two quantities (*one-sample sign test*).

Notice that this framework can be adapted to non linear restrictions as well. Concretely, assume the following null hypothesis

$$\Phi\left(\bar{\theta}\right) = r$$

where $\Phi\left(\bar{\theta}\right)$ is a vector of non linear function of $\bar{\theta}$. The method can be accomplished by linearizing the function $\Phi\left(\bar{\theta}\right)$, for example, around the conditional posterior mean of $\bar{\theta}$ with a Taylor expansion approximated at the first order

$$\Phi\left(\bar{\theta}\right) \simeq \Phi\left(\bar{\theta}^*\right) + \nabla\Phi\left(\bar{\theta}^*\right)'\left(\bar{\theta} - \theta^*\right)$$

where $\nabla\Phi\left(\bar{\theta}^*\right)$ is the gradient of $\Phi\left(\bar{\theta}\right)$ computed at $\bar{\theta}^*$. The quadratic forms (4.9) and (4.10) then becomes respectively

$$q = \left[\Phi\left(\bar{\theta}\right) - \Phi\left(\bar{\theta}^*\right)\right]'\left(\nabla\Phi\left(\bar{\theta}^*\right)' V^* \nabla\Phi\left(\bar{\theta}^*\right)\right)^{-1}\left[\Phi\left(\bar{\theta}\right) - \Phi\left(\bar{\theta}^*\right)\right]$$

$$q_1 = \left[\Phi\left(\bar{\theta}\right) - r\right]'\left(\nabla\Phi\left(\bar{\theta}^*\right)' V^* \nabla\Phi\left(\bar{\theta}^*\right)\right)^{-1}\left[\Phi\left(\bar{\theta}\right) - r\right]$$

and the reasoning follows as before.

4.3 The Monte Carlo Study

In order to analyze the statistical properties of the testing procedure we take the following data generating process for each observation

$$y_{it} = \alpha_i + \rho_i y_{it-1} + \varepsilon_{it} \tag{4.11}$$

with $i = 1, ..., N$ and $t = 1, ..., T$.

We assume that the disturbances are generated from

$$\begin{aligned} \varepsilon_{it} &\sim N\left(0, \sigma_i^2\right) \\ E\left(\varepsilon_{it}\varepsilon_{js}\right) &= 0, \quad i \neq j,\ t \neq s \end{aligned} \tag{4.12}$$

and

$$\sigma_i^2 \sim IG\left(\frac{v}{2}, \frac{\delta}{2}\right) \tag{4.13}$$

where $IG\left(\frac{v}{2}, \frac{\delta}{2}\right)$ denotes an inverted gamma distribution with shape v and scale δ.

Random coefficients are obtained from the joint distribution

$$\begin{pmatrix} \alpha_i \\ \rho_i \end{pmatrix} \sim N\left[\begin{pmatrix} \bar{\alpha} \\ \bar{\rho} \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \sigma_{\alpha\rho} \\ \sigma_{\rho\alpha} & \sigma_\rho^2 \end{pmatrix}\right] \tag{4.14}$$

This set up (Eq. (4.11) through (4.14)) can easily be written in terms of (4.2)-(4.4). In particular $\theta = (\theta_1, ..., \theta_N)'$, $\theta_i = (\alpha_i, \rho_i)'$, $A_o = (\mathbf{I}_2, ..., \mathbf{I}_2)'$, $\bar{\theta} = (\bar{\alpha}, \bar{\rho})$, $X = diag\,(X_1..., X_N)$, with $X_i = (x_{i1}, ..., x_{iT})'$ and the matrix $\Sigma_\varepsilon = diag\left(\sigma_1^2, ..., \sigma_N^2\right)$. This means that in terms of model (4.2)-(4.4), we have: $n = NT$, $k = 2N$, $m = 2$.

This model specification can be seen as a dynamic heterogeneous panel data model, where i denotes the cross sectional dimension, whereas t is the time dimension. In a recent paper, Hsiao et al. (1998) show that a hierarchical Bayesian approach, like the one considered in the previous section, performs reasonably well in the estimation of dynamic panel data models relatively to other traditional methods, in the presence of coefficient heterogeneity across sectional units, especially in small samples. Fixed effect or instrumental variable estimators, neglecting the coefficient heterogeneity, are biased and inconsistent, the degree of inconsistency being a function of the degree of coefficient heterogeneity and the extent of serial correlation in the regressors.

These points motivate to some extent the choice of the data generating process in these notes. If we believe that data behave as (4.11) and perhaps we need to make inference on the mean of the coefficients, then we might need to estimate the model in a hierarchical Bayes fashion.

In the benchmark simulations the hyperparameters v and δ are set equal to 6 and 1, respectively, while $\bar{\rho} = 0.4$, $\bar{\alpha} = 0.6$, $\sigma_{\alpha}^2 = \sigma_{\rho}^2 = 0.025$, and $\sigma_{\rho\alpha} = -0.00625$. This choice implies that data are generated form a stationary process (ρ_i lies inside the unit interval) with low variability ($v = 6$ and $\delta = 1$ implies that the mean and the standard deviation of σ_i^2 are both equal to 0.25) and with population parameters showing low heterogeneity. Three departures from this benchmark situations are analyzed. First, because the degree of heterogeneity can be important in the estimation of this model, in most simulation experiments we also take $\sigma_{\alpha}^2 = 0.25$, $\sigma_{\rho}^2 = 0.05$. Second, the case of near non-stationarity is also considered by setting $\bar{\alpha} = 0.1$ and $\bar{\rho} = 0.9$. Though in principle there is no need to restrict data to be stationary, we have been more cautious both in increasing the variance of ρ_i and the true parameter $\bar{\rho}$, because the y series become explosive with simulated data when ρ_i lies outside the unit interval, even with small T. So when it happens that ρ_i lies outside the unit interval we generate ρ_i from a truncated normal distribution, by truncating the distribution to the unit interval. The problem is that when coefficients are generated with such a restriction, the prior distribution must be different and the derivation of the Bayes estimators should take this into account. Given the relative complexity, we decided not to pursue this adjustment on the prior, because in any case it is interesting to see how the test performs without the adjustment. Finally, the case of higher variability of the y_i series is considered by setting $v = 4.2$, and $\delta = 2$. This choice implies that the mean and the standard deviation of σ_i^2 are approximately equal to 1.0 and 3.0 respectively, values much greater than the benchmark ones.

The number of cross sectional units is $N = 10, 20$ in all simulations with $T = 150$, while the number of time data points is $T = 10, 20$ in all simulations in which $N = 50$. The first combination may be typical in a "macro" data field, whereas the second combination is more typical in a "micro" panel data set. For each sectional units $T + 50$ data points are generated starting from $y_{i0} \sim Uniform(-0.5, 2.)$. The first 50 observations are then dropped in order to reduce the dependency on the initial conditions.

The number of replications chosen for the Monte Carlo is 100 in all cases, while the number of replications used for the Gibbs sampling is 2500, after discarding the first 500, when $N = 10$, 20 with $T = 150$, and 1500, after discarding the first 500, when $T = 10$, 20 in all cases in which $N = 50$. Without loss of generality, the null hypothesis chosen is $H_o : \bar{\alpha} + \bar{\rho} = 1$ when $T = 150$. In this case the restriction matrix is $R = \begin{bmatrix} 1 & 1 \end{bmatrix}$, and $r = 1$. When $N = 50$ and T is smaller the null hypothesis is simply $H_o : \bar{\alpha} = 0.6$ (or $H_o : \bar{\alpha} = 0.1$ when the true $\bar{\alpha}$ is 0.1). Here trivially $R = \begin{bmatrix} 1 & 0 \end{bmatrix}$, and $r = 0.6$ (or 0.1). The reason for different restrictions according to the sample sizes is simple. When the dimension of the time series is high, the mean coefficients $\bar{\alpha}$ and $\bar{\rho}$ are estimated with much greater precision than in the case of small T. On the contrary, when the time series size of each cross section is small, relatively to N, the parameter $\bar{\rho}$ are usually estimated with a downward bias, whereas and as a consequence of this, the estimate of $\bar{\alpha}$ is upward biased. This means that the sum $\bar{\alpha} + \bar{\rho}$ is still giving approximately 1, and, as a result, the properties of the testing approach would be indistinguishable from those in the case of $T = 150$ and $N = 20$.

We also briefly comments on the properties of the approach in the case of non linear restrictions. The null hypothesis here is $H_o : \bar{\alpha}\bar{\rho} = 0.24$ (or $H_o : \bar{\alpha}\bar{\rho} = 0.09$) in all cases analyzed.

The procedure for the Monte Carlo experiments includes the following steps: (i) generate the data according to Eq. (4.11)-(4.14) and the numerical values of the hyperparameters presented above; (ii) estimate initially the model using the mean group estimator[4] and subsequently use these estimation results to initialize the Gibbs sampling; (iii) run the Gibbs sampling to get the marginal posterior of interest, in particular the posterior distributions of $\bar{\theta}$, σ_i^2, Σ_θ, q, and q_1. Steps (i)-(iii) are then repeated 100 times.

To analyze the properties of the testing procedure, we pay attention to several aspects. For each set of the Monte Carlo simulation we consider 20 departures form the true parameters to be able to compute the power function and to test the distance between the posterior distributions of q and q_1. Specifically, maintaining fixed the true value of $\bar{\rho}$ (0.4, or 0.9), we consider 10 progressively different values of $\bar{\alpha}$ above and below its value (0.6, or 0.1). Because the results are pretty much the same, we only show the 10 departures above $\bar{\alpha}$. Concretely the Monte Carlo si performed assuming the true $\bar{\alpha}_j$ progressively equal to $\bar{\alpha}_{j-1} + 0.2$, with $j = 1, .., 11$, and $\bar{\alpha}_0 = 0.6$ or 0.1. For each case j, the estimated values of the parameters are averaged over 100 and so are the distributions of q and q_1. In this way, for each j we are able to: (i) evaluate the performance of the hierarchical Bayes estimation under different prior assumptions; (ii) compare the means of the distributions of q and q_1; (iii) compare the entire distributions of these quantities testing the nulls of equal cdf and equal percentiles of the respective empirical density functions; (iv) get a flavor on the size and the unbiasedness of the test; (v) compute the power function in a classical sampling-theory fashion; (vi) compare this approach with the standard PO ratio, whenever

[4] For the definition and the properties of the mean group estimator see Pesaran and Smith, (1995). The authors show that in the context of dynamic heterogeneous panel data models, this is a consistent estimator. Hsiao et. al (1998) then prove the asymptotic equivalence between the full Bayesian and the mean group estimator.

possible.

The experiment is performed by assuming the following general prior information

$$p\left(\bar{\theta}, \sigma_i^2, \Sigma_\theta^{-1}\right) = p\left(\bar{\theta}\right) p\left(\sigma_i^2\right) p\left(\Sigma_\theta^{-1}\right)$$

with

$$p\left(\bar{\theta}\right) = N\left(\mu, C\right)$$

$$p\left(\sigma_i^2\right) = IG\left[\frac{\phi}{2}, \frac{\tau^2\phi}{2}\right]$$

$$p\left(\Sigma_\theta^{-1}\right) = W\left[\left(\sigma_\theta S_\theta\right)^{-1}, \sigma_\theta\right]$$

The simulations explained above are then repeated for most cases under an informative and *non informative prior on* $\bar{\theta}$. Table 1 resumes the Monte Carlo design and in Table 2 the values of the hyperparameters of the prior chosen in each subcase are reported.

In the case of non-diffuse or informative prior two further subcases are analyzed, according to the values given to the hyperparameter vector μ. Specifically, in one case we take, for each $j > 1$, $\mu = (0.6,\ 0.4)'$ (or $\mu = (0.1,\ 0.9)'$), while in the other the vector is the true one corresponding to j. We consider the former as a way of putting more weight on the null, and the latter as a way of assigning more weight to the alternative hypothesis.

The comparison between our approach and the standard PO ratio is possible only when the prior is informative. In this case the PO ratio is computed using the technique suggested by Chib (1995), as surveyed in section 1.

4.4 Results

Tables 3-9 present the simulation results. The posterior estimates, a comparison of the distributions of q and q_1 and the comparison between this approach and the PO ratio are reported.

In table 3 we show the posterior mean estimates of the parameters of the model and of the quantities q and q_1. The first column refers to the corresponding column in table 1, while the second column gives the true $\bar{\alpha}$. Parameter $\bar{\alpha}$ is estimated quite precisely when $T = 150$, with a bias that falls within the range of 0 to 40%, both in the informative and in the non informative case. The bias increases in small samples ($T = 10$, or $T = 20$) and in some cases (particularly when data show high variability and the degree of coefficient heterogeneity is high – cases 11, 12, 19) it exceeds 100%. As one would expect, the issue is more serious when the prior is diffuse (cases 12 and 19). The characteristics of the bias in the estimation of $\bar{\rho}$ are similar, though the bias seems to be more reduced with respect to the estimation of the constant, falling within the range of 2,5 to 50% in all cases analyzed. This performance of the Bayes estimator is not very surprising in view of the fact that all the estimation results are derived conditional on initial y_{io}. Previous studies (e.g. Blundell and Bond, 1996) have outlined that the bias due to ignoring initial observation may be quite significant in sampling approaches, when the time series dimension is small. Roughly speaking, our results seem to replicate the features obtained in Hsiao et al. (1998), though they are not directly comparable because of the different specification of the data generating process.[5]

Another feature which confirms the findings of previous studies is the upward bias in the estimation of the posterior elements of the matrix Σ_θ. As discussed in Hsiao et al., these results may depend upon the choice of the scale matrix S_θ, as well as the actual degree of coefficient heterogeneity. Our choice of S_θ and σ_θ has followed previous studies on typical examples of the Gibbs sampling applications (Gelfand et al., 1990, among others). To check the sensitivity of the results we have tried different choices, according to the sample size

[5] In Hsiao et al. data are generated from a model which does not include the constant term, while consider the presence of a stationary explicative variable.

and the degree of coefficient heterogeneity in the data generating process. In the cases of low heterogeneity and large samples, the Swamy (1971) estimate of Σ_θ seems to give better performances in terms of posterior estimates of the elements of this matrix. The estimation of Σ_θ is given by

$$\hat{\Sigma}_\theta = \frac{1}{n}\sum_i \left(\hat{\theta}_i - \frac{1}{n}\hat{\theta}_i\right)\left(\hat{\theta}_i - \frac{1}{n}\hat{\theta}_i\right) - \frac{1}{n}\sum_i \hat{\sigma}_i^2 \left(X_i' X_i\right)^{-1}$$

where $\hat{\sigma}_i^2 = \hat{\varepsilon}_i' \hat{\varepsilon}_i / (T - k)$, and the hats "^" denote OLS estimation for each cross sectional units.

On the contrary, when the degree of heterogeneity is high and the sample is small (especially in the time dimension), the choice described in table 2 performs better. In both cases, the choice of the scale matrix seems to affect only the posterior estimates of the matrix and sometimes the posterior estimates of the other parameters, but not the results on the properties of the testing procedure, which is our main concern.

The last three columns of table 3 report the estimated average posterior mean of the variance of the error term, which does not show serious biases in all cases analyzed, and the estimated posterior means of the distributions of q and q_1. In all cases under discussion, except two concerning the nonlinear restriction (17 and 19), the mean of q is not statistically different from the mean of a chi–square with one degree of freedom (not shown). This result is more general and applies not only to the posterior mean of q but also to its entire empirical posterior distribution, whose draws in all cases analyzed (with the exception of case 17 and 19) *are statistically indistinguishable from those of a* $\chi^2_{(1)}$. This is not surprising, provided the model specification is based on natural conjugate priors. However, this finding is not strictly necessary for the assessment of the goodness of the testing procedure. As a matter of fact, the empirical posterior density of q is our reference distribution, independently of its exact shape. In the non linear restriction, when data are generated from a close-to-non-stationary model with high variability (cases 17 and 19), approximating at the first order the Taylor expansion is probably not enough to get a posterior chi-square for q with the right degrees of freedom. Notwithstanding the comparison between q and q_1 is still possible. As remarked above, this point represents the main difference with the classical hypotheses setting where the comparison must be conducted between a single value of the distribution under the restriction and a critical value of a standard distribution to which the former should asymptotically converge under the null.

Table 4 tests the equality of the posterior means of q and q_1. The test is a two sample Wilcoxon test and the corresponding p-value is reported.[6] For each case, the table presents only two of these probabilities. The first (case a) tests the equality when the null is true, whereas the second (case b) reports the p-value under the first rejected null, when the null is false. The corresponding column j gives the iteration number in the departures from the assumed true value of $\bar{\alpha}$ (see previous section). Hence the ideal situation in all cases would be to accept when the null is true and start rejecting for low values of j, i.e., small departures from the null. Clearly, the cases in which the test rejects the equality of q and q_1 when the null is true ($j = 1$) would reveal a bias in the testing procedure. The tables has two sides. The left-hand side refers to the estimation under a non-informative prior, while the right-hand side considers an informative prior. In the latter case, two subcases are analyzed: one in which more weight is given to the null and the other where more weight is given to the alternative, as explained in the previous section.

A rough look of the table reveals that in most cases the distributions of q and q_1 seem to share the same locations when $j = 1$. The only exceptions concern the cases where the

[6] This is a non-parametric technique used to test whether two sets of observations come from the same distribution. The alternative hypothesis is that the observations come from distributions with identical shape but different locations. Although a standard two-sampled t-test produced the same results, we prefered not to use it because it assumes that the observations come from Gaussian distributions, which is not the case here.

degree of coefficient heterogeneity is high or the sample size is small (case 2 and 16 in the non informative case, and case 4 in the informative one). The high heterogeneity seems to be crucial when the cross sectional dimension is small relative to the time dimension. This conclusion is easily achieved from the comparison of case 2 and case 6 in the non informative prior and from the comparison of cases 4 (non informative) and 8 (informative). When the prior is informative the high degree of coefficient heterogeneity does not seem important (case 2) unless data are generated from a close-to-non stationary process with a high variability (case 4). Finally, when the cross sectional dimension increases (cases 9 to 15), the high heterogeneity, non-stationarity and high variability do not affect any longer the equality of the means of q and q_1 at $j = 1$, though in case of small time dimension with high heterogeneity (case 10, informative) and in three out of the four non linear cases (16, 17 and 19) the p-values would reveal a statistical difference at the 10% level of confidence.

The means of the two quantities start to be statistically different at most when $j = 3$ in all cases. As one would expect, this event is more frequent when the model is estimated under an informative prior when more weight is given to the null, especially when the degree of coefficient heterogeneity is low.

In order to have a better idea about the posterior shape of the quantities q and q_1, tables 5 and 6 compare not only the posterior means but the entire distributions. Both tables are organized as table 4. Concretely, in table 5 we compare the posterior densities of q and q_1 testing the equality of the respective 5, 25, 75 and 95 percentiles. The reported p-value is the one calculated in the so called *one-sample sign test* and is based on an exact binomial distribution.[7] The null hypothesis is $H_o : \xi_p = \xi$, where ξ_p is the p-th percentile of the posterior density of q_1 and ξ is the value taken by the corresponding percentile of q. In table

[7] For a simple description, see for example Mood et al. (1974), p. 514, 515.

6 the equality of the cdf of the two quantities is examined by means of the Kolmogorov-Smirnov goodness of fit test.[8] The p-values can be considered as a measure of the distance between the two distributions. Again, as for table 4 the first p-values reported (case *a*) are computed under the null, while the second ones (case *b*) represent the first rejection after departing from the null. The last column of table 6 provides an idea about the power of the test. Concretely, if we cannot reject the equality under the null, the distributions of q and q_1 overlap. In this case, using a classical terminology, we would say that the power coincides with the size. From the first rejection on, the power is greater than the size (ideally, it is equal to 1). The interpretation is the same as discussed above. The posterior distribution of q is a reference distribution, i.e., the one which in a classical analysis would be tabulated. The larger is the distance between q and q_1, the higher the probability that the more likely values of q_1 fall in the tale of the less likely values of q, leading to a rejection. Both in table 5 and in table 6 the p-values are compared with a significance level of 0.05.

The values reported in these two tables tend to confirm what discussed above for table 4. In particular, the only cases in which the test seems to be biased are those in which the degree of parameter heterogeneity is high (case 2, non informative and case 4 informative), or the cross sectional dimension is small (case 10, informative). Under a non informative prior, when the cross sectional size increases, the bias disappears, even with a small time dimension. In the non linear restriction case the test seems to show more serious problems, as the low p-values indicate (cases 16, 17, and 19). When the prior is informative the test is clearly biased when coefficients are highly heterogeneous (case 10) and data show non stationarity and high variability (case 4). In all other cases, the performance of the testing

[8] This statistic is used to test whether two sets of observations could reasonably have come from the same distribution. This test assumes that the samples are random samples, the two samples are mutually independent, and the data are measured on at least an ordinal scale. In addition, the test gives exact results only if the underlying distributions are continuous. See Mood et al. (1974, p. 508-511) for more details.

procedure seems quite good and its power function si close to an ideal one, being equal to the size for those values of θ corresponding to the null hypothesis and greater than the size (ideally equal to 1) for those θ corresponding to the alternative.[9] As commented before for table 4, the restriction to be tested converts the distribution of q_1 in a non-central one with respect to the reference distribution q at most when $j = 3$. We interpret this finding as a strong signal that the testing approach shows a good power function.

The performance of the test can be evaluated also on a sampling-theory base. Tables 7 and 8, for example, report the size and the power function of the test as in a classical analysis. Concretely, we can compute the power function calculating, at each iteration of the Gibbs sampling, the $Prob\left(\chi_1^2 \geq q_1\right)$, and then counting the number of times of this probability being less or equal to 0.05, the significance level chosen. After repeating the previous steps 100 times, the power function can be taken as the average of these probabilities. The size of the test would just be the power function when the null is true. By using the 100 iteration of the Monte Carlo, table 7 reports more precisely 4 percentiles of the "distribution" of the size over the draws. The two tables refer only to the non informative, low and high heterogeneity cases with $N = 10$, and $N = 20$, (cases 1,2, and 5,6). The results of the two tables confirms the findings discussed above with some caveats. In particular, the test seems unbiased, in the sense that, on average, the probability of rejecting the null is greater or equal than the size for all the values of $\bar{\alpha}$ considered. Moreover, for N = 20, the power is almost one for relatively low values of j. If instead we use a more precise definition of unbiasedness such that, if $\pi(\theta)$ is our power function and the null $H_o : \theta \in \Theta_o$ is to be

---

[9] Here "size" means the significance level we should consider if we used the distribution of q as the reference distribution to which a given value of q_1 (the mean or the median, for example) would be compared in a classical analysis.

tested against the alternative $H_1 : \theta \in \Theta_1$, the test is unbiased if and only if

$$\sup_{\theta \in \Theta_o} \pi(\theta) \leq \inf_{\theta \in \Theta_1} \pi(\theta)$$

then, it turns out that over the 100 iteration of the Monte Carlo, the $\inf_{\theta \in \Theta_1} \pi(\theta)$ start to be larger than the $\sup_{\theta \in \Theta_o} \pi(\theta)$ only when $j = 3$. Notice also that when the degree of coefficient heterogeneity is high the percentage of rejections when the null is true is always greater than the level of significance chosen (0.05). In our opinion, these caveats simply suggest to be cautious in the use of a sampling-theory evaluation of the performance of a Bayesian approach.

Finally, table 9 reports a comparison between the procedure proposed and the standard P.O. ratio test. The first four columns of the table are the same as in table 6 (informative). In the last two columns the percentage of negative values of the $log(PO)$ over the Monte Carlo simulations is reported, together with the benchmark ($j = 1$) and the first j in which the average posterior $\log(PO)$ starts to be negative.

A couple of comments are in order. First, when $j = 1$, the average PO ratio is greater than one in all cases considered and hence it always selects the null hypothesis against the alternative, whereas the empirical procedure proposed here has some problem when the degree of heterogeneity is high or the data are non stationary (cases 4 and 10) as discussed above. Notwithstanding, when the time dimension is small and the degree of heterogeneity is high or the data are generated from a close to non stationary process with high variability, the percentage of negative $\log(PO)$ is quite high (cases 10, 11, 14 and 15). If we interpret this percentage as the equivalent of the significance level in a sampling-theory test, this result indicates that in these cases the PO ratio would produce too many rejections of the null when it is true and hence that it could be biased. On the contrary, in the same cases (especially 14 and 15) our procedure accepts without doubts the null when it is true as the high p-values of the test $F(q) = F(q_1)$ reveal.

The second important thing to notice is that the minimum j at which the average $\log(PO)$ starts to be negative is 3, whereas the proposed procedure starts rejecting the null when it is false *at most* when $j = 3$. This means that in most situations considered our q–test may be more powerful than the PO ratio, though one must be cautious with such a conclusion provided we are not sure about the comparison of the sizes of the two testing approaches.

In summary, the few Monte Carlo experiments tend to indicate that the procedure proposed in this paper seems to perform fairly well under different behaviors of the data and the vector of coefficients and different prior assumptions. As already remarked, this good performance is based on estimation results which have been obtained conditional on initial observations y_{io} and, in some cases, generating the autoregressive coefficient from a truncated normal distribution without modifying its prior distribution. We believe that following the suggestions of Sims (1998) of using a proper likelihood function for $(y_{io}, ..., y_{iT})$ and modifying the prior assumption without necessarily restricting the model only to the stationary case cannot worsen the findings obtained here.

4.5 Conclusions

In this paper we have discussed a simple way of verifying restrictions in complex hierarchical normal data models using the output of the Gibbs sampling in a natural way. The procedure has the advantage that can be used under informative and non informative priors on the parameters of interest and does not require the estimation of two models, one with and the other without the restriction to be tested. In a sense, we could say that this procedure stays to the PO ratio test as, in the classical analysis, the Wald test stays to the Likelihood ratio test. This parallel and the similarity of interpretation should make the method appealing also to sampling theory econometricians. The limited Monte Carlo experience seems to indicate that under different behaviors of the data and different prior assumptions, the procedure has good properties and is competitive with the standard PO ratio approach, besides being computationally easier in the kind of models considered here and more useful when the prior is diffuse.

Table 1. Design of the Monte Carlo study

	T	N	α	ρ	$\sigma(\alpha)$	$\sigma(\rho)$	v	δ	prior	null
Linear										
1	150	10	0,6	0,4	0,025	0,025	6,0	1,0	i/ni	$\alpha+\rho=1$
2	150	10	0,6	0,4	0,25	0,05	6,0	1,0	i/ni	$\alpha+\rho=1$
3	150	10	0,1	0,9	0,025	0,025	4,2	2,0	ni	$\alpha+\rho=1$
4	150	10	0,1	0,9	0,25	0,05	4,2	2,0	i	$\alpha+\rho=1$
5	150	20	0,6	0,4	0,025	0,025	6,0	1,0	i/ni	$\alpha+\rho=1$
6	150	20	0,6	0,4	0,25	0,05	6,0	1,0	i/ni	$\alpha+\rho=1$
7	150	20	0,1	0,9	0,025	0,025	4,2	2,0	i	$\alpha+\rho=1$
8	150	20	0,1	0,9	0,25	0,05	4,2	2,0	ni	$\alpha+\rho=1$
9	10	50	0,6	0,4	0,025	0,025	6,0	1,0	i/ni	$\alpha=0.6$
10	10	50	0,6	0,4	0,25	0,05	6,0	1,0	i/ni	$\alpha=0.6$
11	10	50	0,1	0,9	0,025	0,025	4,2	2,0	i	$\alpha=0.1$
12	10	50	0,1	0,9	0,25	0,05	4,2	2,0	ni	$\alpha=0.1$
13	20	50	0,6	0,4	0,025	0,025	6,0	1,0	i/ni	$\alpha=0.6$
14	20	50	0,6	0,4	0,25	0,05	6,0	1,0	i/ni	$\alpha=0.6$
15	20	50	0,1	0,9	0,25	0,05	4,2	2,0	i	$\alpha=0.1$
nonlinear										
16	150	20	0,6	0,4	0,025	0,025	6,0	1,0	ni	$\alpha\rho=0.24$
17	150	20	0,1	0,9	0,025	0,025	4,2	2,0	ni	$\alpha\rho=0.09$
18	10	20	0,6	0,4	0,025	0,025	6,0	1,0	ni	$\alpha\rho=0.24$
19	10	20	0,1	0,9	0,025	0,025	4,2	2,0	ni	$\alpha\rho=0.09$

Note: "i" = informative; "ni" = non-informative

Table 2. Prior hyperparameters in the Monte Carlo

		informative	non informative
B	T = 150	C = diag(4.0) S(θ) = diag(3.0) $\sigma(\theta) = 4.0$ $\phi = 0.3,\ \tau = 3.0$	C^(-1) = 0 S(θ) = diag(3.0) $\sigma(\theta) = 2.0$ $\phi = 0.0$
	N = 50	C = diag(1.0) S(θ) = diag(10, 1.0) $\sigma(\theta) = 10.0$ $\phi = 0.3,\ \tau = 3.0$	C^(-1) = 0 S(θ) = diag(10, 1.0) $\sigma(\theta) = 2.0$ $\phi = 0.0$
DB	T = 150	C = diag(4.0) S(θ) = diag(5.0) $\sigma(\theta) = 4.0$ $\phi = 0.3,\ \tau = 3.0$	C^(-1) = 0 S(θ) = diag(5.0) $\sigma(\theta) = 2.0$ $\phi = 0.0$
	N = 50	C = diag(1.0) S(θ) = diag(20, 1.0) $\sigma(\theta) = 10.0$ $\phi = 0.3,\ \tau = 3.0$	C^(-1) = 0 S(θ) = diag(20, 1.0) $\sigma(\theta) = 2.0$ $\phi = 0.0$

Note: B = Benchmark; DB = departures from B

Table 3. Posterior estimates of the mean parameters

Informative

	α	α^	ρ^	σ(α)	σ(ρ)	σ(α,ρ)	σ(ε)	q	q1
1	0,6	0,62	0,39	0,81	0,80	-0,0115	0,25	1,00	1,02
2	0,6	0,61	0,39	0,99	0,80	-0,0150	0,25	1,00	1,08
4	0,1	0,11	0,80	1,07	0,80	0,0007	0,97	1,00	1,20
5	0,6	0,61	0,39	0,84	0,84	-0,0587	0,24	1,00	1,03
6	0,6	0,62	0,39	0,85	0,84	-0,0052	0,26	1,00	1,05
7	0,1	0,13	0,83	0,81	0,77	-0,0031	0,99	1,00	1,07
9	0,6	0,75	0,23	0,87	0,72	-0,0646	0,31	1,00	1,12
10	0,6	0,73	0,22	2,07	1,22	-0,0909	0,32	1,00	1,14
11	0,1	0,21	0,54	1,08	0,98	-0,1129	1,06	1,00	1,03
13	0,6	0,65	0,31	1,00	0,72	-0,0349	0,27	1,00	1,09
14	0,6	0,70	0,31	0,88	0,85	-0,0182	0,27	1,00	1,00
15	0,1	0,16	0,64	1,11	0,73	-0,0235	1,01	1,00	1,02

Non informative

	α	α^	ρ^	σ(α)	σ(ρ)	σ(α,ρ)	σ(ε)	q	q1
1	0,6	0,61	0,39	0,84	0,83	-0,0142	0,26	1,00	1,02
2	0,6	0,61	0,38	1,00	0,80	-0,0108	0,25	1,01	1,18
3	0,1	0,14	0,80	0,86	0,81	-0,0049	0,98	0,99	1,04
5	0,6	0,62	0,38	0,81	0,81	-0,0004	0,26	1,00	1,04
6	0,6	0,60	0,39	0,20	0,30	-0,0260	0,25	1,00	1,02
8	0,1	0,12	0,75	1,24	0,50	0,0041	0,99	1,00	1,03
9	0,6	0,81	0,21	2,89	0,80	-0,0688	0,32	1,00	1,07
10	0,6	0,86	0,21	2,26	1,13	-0,1036	0,31	1,02	1,10
12	0,1	0,37	0,49	1,14	2,82	-0,1011	1,12	1,00	1,07
13	0,6	0,72	0,31	2,28	0,75	-0,0303	0,28	1,00	1,06
14	0,6	0,74	0,30	2,46	1,07	-0,0395	0,28	1,00	1,03
16	0,6	0,62	0,38	0,82	0,81	-0,0067	0,25	1,00	1,08
17	0,1	0,15	0,80	1,69	0,75	-0,0040	0,96	1,13	1,15
18	0,6	0,79	0,21	1,06	0,91	-0,0775	0,31	1,00	1,00
19	0,1	0,41	0,54	1,42	0,92	-0,0444	1,00	1,29	1,54

Table 4. Testing equality of the posterior means of q and q1

Non informative

		p-value	j
1	a	0,2372	1
	b	0,0000	2
2	a	0,0143	1
	b	0,0000	2
3	a	0,2597	1
	b	0,0012	2
5	a	0,9726	1
	b	0,0000	2
6	a	0,6577	1
	b	0,0000	2
8	a	0,1774	1
	b	0,0000	3
9	a	0,1149	1
	b	0,0000	2
10	a	0,1340	1
	b	0,0003	2
12	a	0,1286	1
	b	0,0000	2
13	a	0,2985	1
	b	0,0000	2
14	a	0,2634	1
	b	0,0428	2
16	a	0,0426	1
	b	0,0000	2
17	a	0,0873	1
	b	0,0002	2
18	a	0,7550	1
	b	0,0000	2
19	a	0,0742	1
	b	0,0009	2

Informative

		p-value (1)	j	p-value (2)	j
1	a	0,9160	1	0,9160	1
	b	0,0000	3	0,0006	2
2	a	0,2538	1	0,2538	1
	b	0,0000	3	0,0000	2
4	a	0,0098	1	0,0098	1
	b	0,0000	2	0,0000	2
5	a	0,7760	1	0,7760	1
	b	0,0000	2	0,0000	2
6	a	0,1515	1	0,1515	1
	b	0,0000	2	0,0000	2
7	a	0,7864	1		
	b	0,0000	2		
9	a			0,8843	1
	b			0,0000	2
10	a	0,0764	1	0,0764	1
	b	0,0000	2	0,0000	2
11	a	0,3834	1		
	b	0,0000	3		
13	a	0,2774	1	0,2774	1
	b	0,0000	3	0,0000	2
14	a	0,8564	1		
	b	0,0000	2		
15	a	0,4034	1		
	b	0,0000	3		

Notes:

1. The test used is a Wilcoxon Two-Sample t-Test
2. In all cases except 17 and 19 we accept the null hypothesis that the mean of q is equal to the mean of a chi-square with 1 degree of freedom
3. "a" is the case in whichα is the benchmark (see the corresponding j); "b" is the first departur from the benchmark where the means of q and q1 start to be significatively different
4. "p-value(1)" is the p-value when more weight is given to the null
   "p-value(2)" is the p-value when more weight is given to the alternative

Table 5. Testing equality of the 5, 25, 75 and 95 percentiles of the posterior densities of q and q1

Non Informative

		p-value				j
		5	25	75	95	
1	a	0,1994	0,1796	0,5020	0,1365	1
	b	0,0006	0,0000	0,0000	0,0000	3
2	a	0,2191	0,0000	0,0000	0,0000	1
	b	0,0400	0,0000	0,0000	0,0000	2
3	a	0,2191	0,4235	0,3527	0,3555	1
	b	0,0400	0,0000	0,0000	0,0000	3
5	a	0,6785	0,6983	0,7656	0,8590	1
	b	0,0014	0,0000	0,0000	0,0000	2
6	a	0,0854	0,5313	0,9525	0,1726	1
	b	0,0000	0,0000	0,0000	0,0000	2
8	a	0,4065	0,7205	0,0042	0,3428	1
	b	0,0000	0,0000	0,0000	0,0000	3
9	a	0,9528	0,8347	0,1140	0,0854	1
	b	0,0020	0,0000	0,0000	0,0000	2
10	a	0,1558	0,1704	0,3401	0,0379	1
	b	0,0438	0,0000	0,0000	0,0000	2
12	a	0,2374	0,8815	0,0790	0,0541	1
	b	0,0000	0,0000	0,0000	0,0000	2
13	a	0,7220	0,8347	0,3401	0,7220	1
	b	0,0813	0,0318	0,0000	0,0000	2
14	a	0,4768	0,6764	0,9762	0,5533	1
	b	0,0515	0,0001	0,0000	0,0000	3
16	a	0,7672	0,0145	0,0002	0,1381	1
	b	0,0004	0,0000	0,0000	0,0000	2
17	a	0,0477	0,6334	0,8815	0,5533	1
	b	0,0379	0,0122	0,0000	0,0000	3
18	a	0,5940	0,9800	0,9900	0,8900	1
	b	0,0579	0,0011	0,0000	0,0000	2
19	a	0,4768	0,0008	0,3711	0,0659	1
	b	0,0053	0,0050	0,0000	0,0000	2

Informative

		p-value (1)				j	p-value (2)				j
		5	25	75	95		5	25	75	95	
1	a	0,4086	0,6946	0,9632	0,7134	1	0,4086	0,6946	0,9632	0,7134	1
	b	0,0000	0,0000	0,0000	0,0000	3	0,1995	0,0583	0,0000	0,0000	2
2	a	0,9634	0,8354	0,7118	0,7830	1	0,9634	0,8354	0,7118	0,7830	1
	b	0,0000	0,0000	0,0000	0,0000	3	0,0045	0,0000	0,0000	0,0000	2
4	a	0,0089	0,0000	0,0000	0,0000	1	0,0089	0,0000	0,0000	0,0000	1
	b	0,0089	0,0000	0,0000	0,0000	1	0,0089	0,0000	0,0000	0,0000	1
5	a	0,4768	0,7885	0,2967	0,3137	1	0,4768	0,7885	0,2967	0,3137	1
	b	0,0374	0,0000	0,0000	0,0000	2	0,0974	0,0000	0,0000	0,0000	2
6	a	0,8125	0,0400	0,6764	0,5533	1	0,8125	0,0400	0,6764	0,5533	1
	b	0,0579	0,0564	0,0000	0,0000	2	0,0076	0,0000	0,0000	0,0000	2
7	a	0,3741	0,3252	0,7656	0,5533	1					
	b	0,0679	0,0000	0,0000	0,0000	2					
9	a						0,9528	0,6123	0,7656	0,6785	1
	b						0,0237	0,0000	0,0000	0,0000	2
10	a	0,9056	0,0046	0,0157	0,0127	1	0,9056	0,0046	0,0157	0,0127	1
	b	0,3428	0,0017	0,0000	0,0000	2	0,0045	0,0000	0,0000	0,0000	2
11	a	0,3741	0,5313	0,4383	0,1726	1					
	b	0,3741	0,0122	0,0000	0,0000	3					
13	a	0,5940	0,2835	0,6983	0,8125	1	0,5940	0,2835	0,6983	0,8125	1
	b	0,0108	0,0000	0,0000	0,0000	3	0,0152	0,0000	0,0000	0,0000	2
14	a	0,5147	0,9900	0,7656	0,4768	1					
	b	0,0108	0,0038	0,0001	0,0000	2					
15	a	0,5533	0,3872	0,5915	0,9056	1					
	b	0,0974	0,0491	0,0000	0,0000	3					

Notes:

1. The test used is a *one-sample sign test.*

2. In all cases, except 17 and 19 we accept the null hypothesis that the mean of q is equal to the mean of a chi-square with 1 degree of freedom

3. "a" is the case in which α is the benchmark (see the corresponding j)
"b" is the *first departure from the benchmark when at least two quantiles of q and q1 start to be significatively different*

4. "p-value(1)" is the p-value when more weight is given to the null
"p-value(2)" is the p-value when more weight is given to the alternative

Table 6. Testing equality of the cdf of q and q1

Non informative

		p-value	j	power
1	a	0,5923	1	size
	b	0,0000	2	1
2	a	0,0068	1	greater than size
	b	0,0000	2	1
3	a	0,8189	1	size
	b	0,0000	3	1
5	a	0,9117	1	size
	b	0,0000	2	1
6	a	0,5474	1	size
	b	0,0000	2	1
8	a	0,2008	1	size
	b	0,0000	3	1
9	a	0,5474	1	size
	b	0,0000	2	1
10	a	0,1336	1	size
	b	0,0000	2	1
12	a	0,3126	1	size
	b	0,0000	2	1
13	a	0,8909	1	size
	b	0,0051	2	1
14	a	0,9803	1	size
	b	0,0000	3	1
16	a	0,0051	1	greater than size
	b	0,0000	2	1
17	a	0,0998	1	size
	b	0,0065	2	1
18	a	0,8590	1	size
	b	0,0000	2	1
19	a	0,0941	1	size
	b	0,0023	2	1

Informative

		p-value	j	power	p-value	j	power
1	a	0,8600	1	size	0,8600	1	size
	b	0,0295	2	1	0,0000	3	1
2	a	0,8826	1	size	0,8826	1	size
	b	0,0000	3	1	0,0000	2	1
4	a	0,0000	1	greater than size	0,0000	1	greater than size
	b	0,0000	1	1	0,0000	1	1
5	a	0,9601	1	size	0,9601	1	size
	b	0,0000	2	1	0,0000	2	1
6	a	0,3349	1	size	0,3349	1	size
	b	0,0000	2	1	0,0000	2	1
7	a	0,8909	1	size			
	b	0,0000	2	1			
9	a				0,9713	1	size
	b				0,0000	2	1
10	a	0,0289	1	greater than size	0,0289	1	greater than size
	b	0,0000	2	1	0,0000	2	1
11	a	0,1579	1	size			
	b	0,0002	3	1			
13	a	0,7601	1	size	0,7601	1	size
	b	0,0000	3	1	0,0013	2	1
14	a	0,9713	1	size			
	b	0,0002	2	1			
15	a	0,7001	1	size			
	b	0,0000	3	1			

Notes:

1. The test used is the Kolmogorov-Smirnov

2. In all cases except 17 and 19 we accept the null hypothesis that the cdf of q is equal to the cdf of a chi-square with 1 degree of freedom

3. "a" is the case in which α is the benchmark (see the corresponding j)
"b" is the first departure from the benchmark where the cdf of q and q1 start to be significatively different

Table 7. Classical size. Quantiles. Diffuse case

		5	25	75	95
n = 10	low	0,0448	0,0496	0,0556	0,0605
	high	0,0468	0,0539	0,08	0,1153
n = 20	low	0,042	0,0487	0,0587	0,0688
	high	0,046	0,0527	0,0837	0,1331

Note: "low" = *Low heterogeneity;* "high" = *high heterogeneity*

Table 8. Classical Power. Diffuse case

		n = 10		n = 20	
j	α	low	high	low	high
1	0,6	0,05	0,07	0,05	0,06
2	0,8	0,08	0,09	0,16	0,17
3	1	0,17	0,17	0,49	0,38
4	1,2	0,31	0,29	0,78	0,69
5	1,4	0,46	0,43	0,94	0,87
6	1,6	0,63	0,56	0,99	0,97
7	1,8	0,75	0,70	1,00	0,99
8	2	0,85	0,79	1,00	1,00
9	2,2	0,90	0,86	1,00	1,00
10	2,4	0,94	0,91	1,00	1,00
11	2,6	0,96	0,94	1,00	1,00

Note: "low" = *Low heterogeneity;* "high" = *high heterogeneity*

Table 9. Comparison with the P.O. Ratio

			B		E
		F(q) = F(q1)	j	% log(PO)<0.0	j
1	a	0,860	1	4,8	1
	b	0,030	2	52,2	4
2	a	0,883	1	5,9	1
	b	0,000	3	51,9	4
4	a	0,000	1	4,6	1
	b	0,000	1	55,4	7
5	a	0,960	1	2,6	1
	b	0,000	2	50,4	4
6	a	0,335	1	3,5	1
	b	0,000	2	59,4	4
7	a	0,891	1	3,4	1
	b	0,000	2	54,1	5
10	a	0,029	1	15,3	1
	b	0,000	2	53,6	4
11	a	0,158	1	23,1	1
	b	0,000	3	50,1	4
13	a	0,760	1	2,4	1
	b	0,000	2	59,9	5
14	a	0,971	1	11,6	1
	b	0,000	2	55,2	5
15	a	0,700	1	20,9	1
	b	0,000	3	52,4	3

Notes:

1. "a" is the case in whichα is the benchmark (see the corresponding j
"b" is the first departure from the benchmark where
the cdf of q and q1 start to be different (column B)
and where the log(PO) averaged over the MC draws starts to be negative (column E)

REFERENCES

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

Blundell, R. and S. Bond (1996), 'Initial conditions and moment restrictions in dynamic panel data models', *mimeo.*

Chib, S. (1995), 'Marginal likelihood from the Gibbs sampling output', *Journal-of-the-American-Statistical-Association,* 90, 1313-1321.

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

DuMouchel, W.H. and J.E Harris (1983), 'Bayes methods for combining the results of cancer studies in humans and other species' (with discussion) *Journal of the American Statistical Association,* 78, 293-315.

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.

Gelfand, A.E., S.E. Hills, A. Racine-Poon, and A.F.M. Smith (1990), 'Illustration of Bayesian inference in normal data models using Gibbs sampling', *Journal of the American Statistical Association,* 85(412), 972-985.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin (1995), *Bayesian Data Analysis,* London: Chapman & Hall.

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.

Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu (1997), 'Bayes estimation of short run coefficients in dynamic panel data models', *mimeo,* Cambridge University.

Lindley, D.V. (1965), *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 2. Inference.* Cambridge: Cambridge University Press.

Lindley, D.V. and A.F.M. Smith (1972), 'Bayes estimates for the linear model' (with discussion), *Journal of the Royal Statistical Society,* Ser. B, 34, 1-41.

Mood, A.M, F.A. Graybill, and D.C. Boes (1974), *Introduction to the Theory of Statistics,* Singapore: McGraw-Hill.

Pesaran, M.H. and R.Smith (1995), 'Estimating long run relationships from dynamic heterogeneous panels', *Journal of Econometrics,* 68, 79–113.

Rubin, D.B. (1981), 'Estimation in parallel randomized experiments', *Journal of Educational Statistics,* 6, 377-401.

Schorfheide, F. (2000), 'Loss function based evaluation of DSGE Models' forthcoming in *Journal of Applied Econometrics.*

Sims, C. (1998), 'Using a likelihood perspective to sharpen econometric discourse: three examples', *mimeo.*

Swamy, P.A.V.B. (1971), *Statistical Inference in Random Coefficient Regression Models,* Berlin: Springer and Verlag.

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