

IMF Working Paper

Bayesian VARs: A Survey of the Recent Literature with an Application to the European Monetary System

Matteo Ciccarelli
and
Alessandro Rebucci

IMF Working Paper

Research Department

**Bayesian VARs:
A Survey of the Recent Literature with an Application to the European Monetary System**

Prepared by Matteo Ciccarelli and Alessandro Rebucci¹

Authorized for distribution by Tamim Bayoumi

May 2003

Abstract

The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

This paper reviews recent advances in the specification and estimation of Bayesian Vector Autoregressive models (BVARs). After describing the Bayesian principle of estimation, we first present the methodology originally developed by Litterman (1986) and Doan et al. (1984) and review alternative priors. We then discuss extensions of the basic model and address issues in forecasting and structural analysis. An application to the estimation of a system of time-varying reaction functions for four European central banks under the European Monetary System (EMS) illustrates how some of the results previously presented may be applied in practice.

JEL Classification Numbers: C11, C15, C32, E58

Keywords: Bayesian VAR, Gibbs sampling, Time-Varying Reaction Function, EMS

Author's E-Mail Addresses: arebucci@imf.org; matteo@merlin.fae.ua.es

¹ Universitat d'Alicant (Alicante, Spain) and IMF, respectively. Forthcoming in the "Rivista di Politica Economica". The authors are grateful to Gustavo Piga, two anonymous referees, Fabio Canova and Chris Gilbert for useful comments, and to Luigi Guiso for his suggestions and encouragement.

Contents	Page
I. Introduction.....	3
II. The Basics.....	4
A. BVAR as an Answer to the Overfitting Problem	4
B. Bayesian Estimation of VAR in A Nutshell	6
C. A Monte Carlo Simulation Method for Numerical Integration: The Gibbs Sampler	7
III. Alternative Prior Distributions	7
A. Empirical Bayes Estimation.....	8
The Minnesota Prior	8
The Diffuse Prior	10
Conjugate Prior Distributions.....	11
The Normal-Wishart Prior.....	11
The Normal-Diffuse Prior	12
B. Hierarchical Bayes Estimation	13
IV. Some Extensions.....	15
A. Time-Varying Coefficients.....	15
The Kalman Filter with a Minnesota Prior.....	15
A Hierarchical Time-Varying SUR Model	17
B. Non-Normal Data	19
C. Regime Switching, Nonlinear Models, and Beyond	21
V. Forecasting and Structural Analysis	22
A. Unconditional Forecasting	23
B. Structural Analysis	24
VI. An Application: Estimating a System of Reaction Functions	25
A. Specification	26
B. Identification	27
C. Estimation.....	29
D. Results	31
VII. Conclusions.....	35
Figures	
1. Reaction Function Residuals.....	36
2. Germany: Reaction Function Parameters1.	37
3. France: Reaction Function Parameters	38
4. Italy: Reaction Function Parameters	39
5. Spain: Reaction Function Parameters	40
6. Interest Rate Volatility.....	41
References.....	42

I. INTRODUCTION

Vector auto-regressive models (VARs) are a useful starting point for econometric modeling and a standard benchmark for the analysis of dynamic economic problems. In the original formulation of Sims (1972 and 1980), VAR users should rely mostly on the data to study the interaction among the economic variables of interest in order to avoid imposing “incredible restrictions.” This approach, however, may create two problems that have been widely discussed in the literature over the last three decades or so. First, leaving the lag structure the VAR model unrestricted to avoid imposing the *incredible restrictions* of traditional simultaneous equation models may result in *overfitting*: given their generous parameterization, VARs may suffer from a loss of degrees of freedom, which decrease geometrically with the number of variables and proportionally with the number of lags included, resulting in inefficient estimates. Second, by using reduced form VARs to avoid imposing identification restrictions on the contemporaneous causation among the variables of interest one may not go beyond a simple description of the data.

This paper reviews the efforts spent over the past several years to address the overfitting problem by adopting a Bayesian approach to specification and estimation of VARs. The Bayesian approach to estimation regards the true population structure as uncertain and does not assign too much “weight” on any particular value of the model parameters (e.g., *zero-restrictions on certain coefficients*). Instead, it takes this uncertainty into account in the form of a *prior* probability distribution over the model parameters. The degree of uncertainty represented by this prior distribution can then be altered by the information contained in the data, if the two sources of information are different. As long as the prior information is not too ‘vague’, it is altered only by the “signal” and not by the “noise” contained in the data sample, thus reducing the risk of overfitting. As a result, Bayesian VARs (BVARs) are known to produce better forecasts than reduced form VARs estimated in a classical way and, in some cases, even better than univariate autoregressive moving average processes or structural models.²

The choice of a prior distribution summarizing the researcher’s uncertainty over the model parameters is a crucial step in specifying a BVAR. As pointed out by Learner (1978), for most inference problems in economics, prior information matters *in the sense that two economists can legitimately make different inferences from the same data set*. In order to deal with this dependence of inference on prior information and to characterize the mapping from the set of alternative priors onto the corresponding inference, a method must be chosen to combine sample information with prior information, and a set of alternative priors must be considered. The Bayesian approach to estimation provides the needed method, while the range of alternative priors depends on the particular economic problem at hand.

We review the Bayesian estimation principle in the specific case of VARs in Section II. In Section III, we review the most relevant prior distributions that have been proposed in the literature, from the original and easily computed “Minnesota prior” to the most recent and computationally more demanding assumptions. In Section IV, we discuss how to extend the basic model to deal with

² See Canova (1995) for specific references.

(i) time-varying coefficients, (ii) nonnormal data and (iii) nonlinear relations. In Section V, we discuss basic issues in forecasting and structural analysis with BVARs. In addition, in Section VI, by drawing on our own applied work (Ciccarelli and Rebucci, 2002), we present an application of these methods to the estimation of a system of monetary reaction functions for four European central banks under the European Monetary System. This application uses some of the results previously discussed and illustrates how flexible the Bayesian approach is. Section VII concludes.

Our contribution, therefore, is to provide the reader with a road map to the general method and a few specific results that may be applied in a flexible manner for the empirical analysis of dynamic multivariate economic problems. The presentation in the paper reflects to some extent the broad chronological progress of research in this area, which in turn followed to some extent developments in computer technology permitting the estimation of increasingly more complex models at reasonable computing costs.³

Today's appearance of an increasing number of Bayesian applications, in particular, seems linked to enhanced and cheaper computing power. As it will evident from the rest of the paper, with statistically and economically richer prior assumptions than the original Minnesota prior, integration problems become easily intractable analytically and must be solved numerically. But the difficulties of implementing numerical integration have been substantially reduced by the recent advances in computer technology and in sampling-based methods. Quoting from Hsiao and others. (1999) therefore, *there is no excuse any longer for not considering Bayesian methods because of practical difficulties.*

II. THE BASICS

A. BVARs as an Answer to the Overfitting Problem

Consider a typical VAR,

$$Y_t = B_1 Y_{t-1} + B_2 Y_{t-2} + \dots + B_p Y_{t-p} + D z_t + \varepsilon_t, \quad t = 1, \dots, T \quad (1)$$

where Y_t is a $n \times 1$ vector of endogenous variables and ε_t is a $n \times 1$ vector of error terms independently, identically and normally distributed with variance-covariance matrix Σ , $\varepsilon_t \sim IIN(0, \Sigma)$, B_l ($l = 1, \dots, p$) and D are $n \times n$ and $n \times d$ matrices of parameters respectively, and z_t is a $d \times 1$ vector of exogenous variables.

Classical estimation of models like (1) may yield imprecisely estimated relations that fit the data well only because of the large number of variables included, a problem known as 'overfitting' in the literature. In fact, the number of parameters to be estimated, $n(np + d)$, grows geometrically with the number of variables (n) and proportionally with the number of lags (p). When the

³ Nonetheless, this survey does not provide specific indications on software implementation. The interested reader may consult Geweke (2000) for references to software freely available on the web for the implementation of some of the methods review in this paper. The application in Section VI of the paper has been implemented in RATS, which has a number of preprogrammed functions and procedures for Bayesian estimation of VARs.

number of variables on the right hand side of (1) is relatively high and the sample information is relatively loose, it is likely that the estimates are influenced by noise as opposed to signal, especially if the estimation method is designed to fit the sample data as closely as possible. When this is the case, it is recommendable to estimate models like (1) by imposing some restrictions to reduce the dimension of the parameter space. Therefore, the problem is to find restrictions that are as credible (in the sense of Sims, 1980) as possible.

A Bayesian approach to *VAR* estimation was originally advocated by Litterman (1980) as a solution to the ‘overfitting’ problem. The solution he proposed is to avoid overfitting without necessarily imposing exact zero restrictions on the coefficients. The researcher cannot be sure that some coefficients are zero and should not ignore their possible range of variation. A Bayesian perspective fits precisely this view. Without putting too much weight on certain values, one may think of this uncertainty over the exact value of the model’s parameters as a probability distribution for the parameter vector. The degree of uncertainty represented by this distribution can then be altered by the information contained in the data if the two sources of information are different. As long as the *prior* information is not too vague or non-informative, it should be altered only by the ‘signal’ and not by the ‘noise’ contained in the sample, thus reducing the overfitting risk.

More specifically, Litterman (1986) specifies his prior by appealing to three statistical regularities of macroeconomic time series data: (i) the trending behavior typical of macroeconomic time series; (ii) the fact that more recent values of a series usually contain more information on the current value of the series than past values and (iii) the fact that past values of a given variable contain more information on its current state than past values of other variables. If we apply these statistical regularities, (1) becomes a multivariate random walk.⁴ A Bayesian researcher instead can specify these regularities by assigning a probability distribution to the parameters in such a way that: (i) the mean of the coefficients assigned to all lags other than the first one is equal to zero; (ii) the variance of the coefficients depends inversely on the number of lags; and (iii) the coefficients of variable j in equation g are assigned a lower prior variance than those of variable g .

As we shall see below, these requirements can be expressed formally by introducing a vector of parameters (called hyperparameters), say $\Pi \equiv (\pi_1, \dots, \pi_H)$. For instance, π_1 may control the value of the mean of the first own lag coefficient, π_2 controls the variance of the lags of variable g in equation g , π_3 controls the variance of the lags of variable j in equation g , π_4 controls the speed of decreasing of the variance as the number of lags increase, π_5 controls the variance of the deterministic/exogenous part, and π_6 controls the overall degree of prior uncertainty. Therefore, the original problem of estimating $n(np + d)$ parameters can be converted in a problem of estimating just six “hyper-parameters.” As we will see, this is only one example, although most BVARs were originally specified by using this particular set of hyperparameters.

⁴ This specification has been criticized on the ground that the resulting estimation procedure might be affected by spurious regression problems as in the classical case, due to the presence of unit roots or stochastic trends in the data. Proponents of the Bayesian approach to VAR estimation, however, argue that this is not the case. See the special issue of the *Journal of Applied Econometrics* (1991, Vol. 6.) in which the Bayesian method of treating unit roots is thoroughly discussed, for a presentation of the arguments and the counterarguments.

Additional data characteristics can be easily accommodated by introducing different hyperparameters to control for other relevant information. Cointegrated systems, for instance, may be treated within the same framework by specifying, a priori, long-run relations among the variables of interest. As shown by Alvarez and Ballabriga (1994), however, a Minnesota-type prior with hyperparameters search performs well in the presence of cointegration. Alvarez and Ballabriga (1994) show also that adding long-run restrictions to the prior does not improve the small sample performance of the Bayesian estimation methods commonly applied to VAR models.

B. Bayesian Estimation of VARs in a Nutshell

To state more formally the Bayesian estimation principle, consider (1) rewritten in compact form as:

$$Y_t = X_t \beta + \varepsilon_t \quad t = 1, \dots, T, \quad (2)$$

where $X_t = (I_n \otimes W_{t-1})$ is $n \times nk$, $W_{t-1} = (Y'_{t-1}, \dots, Y'_{t-p}, z'_t)'$ is $k \times 1$, and $\beta = \text{vec}(B_1, B_2, \dots, B_p, D)$ is $nk \times 1$. The unknown parameters of the model are β and Σ .

Bayesian estimation of (2) is simple, in principle, and works as follows. Given the probability density function (*pdf*) of the data conditional on the model's parameters (the information contained in the data in the form of a likelihood function),⁵

$$L(Y | \beta, \Sigma) \propto |\Sigma|^{-T/2} \exp \left\{ -\frac{1}{2} \sum_t (Y_t - X_t \beta)' \Sigma^{-1} (Y_t - X_t \beta) \right\}, \quad (3)$$

and a joint prior distribution on the parameters, $p(\beta, \Sigma)$, the joint posterior distribution of the parameters conditional on the data is obtained through the Bayes rule,

$$\begin{aligned} p(\beta, \Sigma | Y) &= \frac{p(\beta, \Sigma) L(Y | \beta, \Sigma)}{p(Y)} \\ &\propto p(\beta, \Sigma) L(Y | \beta, \Sigma), \end{aligned}$$

noting that, by definition of conditional probability, the joint pdf of the data and the parameters, $p(\beta, \Sigma, Y)$, can be written as

$$\begin{aligned} p(\beta, \Sigma, Y) &= L(Y | \beta, \Sigma) p(\beta, \Sigma) \\ &= p(\beta, \Sigma | Y) p(Y), \end{aligned}$$

where \propto denotes 'proportional to'. Given $p(\beta, \Sigma | Y)$, the marginal posterior distributions conditional on the data, $p(\Sigma | Y)$ and $p(\beta | Y)$, can then be obtained by integrating out β and Σ from $p(\beta, \Sigma | Y)$, respectively. Finally, location and dispersion of $p(\Sigma | Y)$ and $p(\beta | Y)$ can be easily analyzed to yield point estimates of the parameters of interest and measures of precision, comparable to those obtained by using a classical approach to estimation.

⁵ We shall denote probability density functions with a 'p' when they are prior or posterior distributions of the parameters, and with a 'L' when a likelihood function.

C. A Monte Carlo Simulation Method for Numerical Integration: The Gibbs Sampler

In many applications, the analytical integration of $p(\beta, \Sigma | y)$ may be difficult or even impossible to implement. This problem, however, can often be solved by using numerical integration based on Monte Carlo simulation methods.

One particular method used in the literature to solve similar estimation problems to those discussed in this paper is the Gibbs sampler.⁶ The Gibbs sampler is a recursive Monte Carlo method which requires only knowledge of the full conditional posterior distribution of the parameters of interest, $p(\beta | \Sigma, y)$ and $p(\Sigma | \beta, y)$. Suppose Σ and β are scalars and that the conditional posterior distributions $p(\Sigma | \beta, y)$ and $p(\beta | \Sigma, y)$ are known. Then the Gibbs sampler starts from arbitrary values for $\beta^{(0)}$ and $\Sigma^{(0)}$, and samples alternately from the density of each element of the parameter vector, conditional on the values of the other element sampled in the previous iteration and the data. Thus, the Gibbs sampler samples recursively as follows:

$$\begin{aligned} \beta^{(1)} &\text{ from } p(\beta | \Sigma^{(0)}, y) \\ \Sigma^{(1)} &\text{ from } p(\Sigma | \beta^{(1)}, y) \\ \beta^{(2)} &\text{ from } p(\beta | \Sigma^{(1)}, y) \\ \Sigma^{(2)} &\text{ from } p(\Sigma | \beta^{(2)}, y) \\ &\vdots \\ \beta^{(m)} &\text{ from } p(\beta | \Sigma^{(m-1)}, y) \\ \Sigma^{(m)} &\text{ from } p(\Sigma | \beta^{(m)}, y) \\ &\text{and so on.} \end{aligned}$$

The vectors $\vartheta^{(m)} = (\beta^{(m)}, \Sigma^{(m)})$ form a Markov chain, and, for a sufficiently large number of iterations (say $m \geq M$), can be regarded as draws from the *true* joint posterior distribution. Given a large sample of draws from this *limiting* distribution, any posterior moment or marginal density of interest can then be easily estimated consistently with its corresponding sample average.

III. ALTERNATIVE PRIOR DISTRIBUTIONS

A second problem in implementing Bayesian estimation of (2) is the choice of a prior distribution for the model's parameters, $p(\beta, \Sigma)$. This choice is a fundamental step in the specification of the model. This section discusses several alternatives.

Many priors have been proposed in the literature, according to the specific economic problem, the sample data, and the way the parameters of the prior $p(\beta, \Sigma)$ are determined. These,

⁶ See Gelfand and others (1990) for a detailed discussion of the Gibbs sampler and Canova and Ciccarelli (2000), Ciccarelli and Rebucci (2001) and Hsiao and others (1999), among others, for some applications. The brief summary in the text is borrowed from this latter contribution. For a more general discussion of numerical Bayesian methods see also Geweke (2000) and Gelman and others (1995).

however, share common features. *Full Bayesian* estimation of (2) would require specifying a prior distribution also for the parameters of the prior $p(\beta, \Sigma)$ and then integrating them out of the posterior distributions, $p(\beta, \Sigma | Y)$. As noted, however, these integrations may be complicated or even impossible to implement. In addition, the specification of the prior *must* take some hyperparameters as given, eventually. The discussion of alternative priors, therefore, may be organized according to the way the problem of the dependence of the posterior moments on the unknown hyperparameters is solved.

One solution is to substitute estimates of the hyperparameters directly into the formulas for the mean and variance of the posterior distributions of the parameters of interest, sometime called *Empirical Bayesian Estimation*. These could be obtained in a previous stage by OLS, maximum likelihood, GMM, and other optimization criteria or informal methods such as the rules of thumb discussed above. Since uncertainty in the estimation of the hyperparameters is not accounted for by the substitution, the empirical Bayes approach will be only an approximation to the full Bayes approach, possibly approaching it in the limit, depending upon the statistical properties of the estimators used or the quality of the rules of thumb used. An empirical Bayesian estimate, may converge to a full Bayesian estimate if the estimator of the hyperparameters converges to its true value as the sample size increases. An alternative solution, sometime called *Hierarchical Bayesian Estimation* in the literature, is to incorporate into the model the prior distribution of the hyperparameters and then obtain the marginal densities of the parameters of interest by integrating out the hyperparameters from the joint posterior density numerically.

We are now going to discuss the empirical and the hierarchical Bayes approaches in turn.

A. Empirical Bayes Estimation

The Minnesota Prior

While the number of possibilities for a linear regression model as (2) is vast, a commonly used prior in the Bayesian *VAR* literature is that proposed by Litterman (1986), sometime called a Minnesota prior.

Consider the problem of estimating the $(k \times 1)$ vector β_g containing the parameters of the g^{th} equation of (2) when the variance of the error term ($\sigma_{g,g}^2$) is known. Litterman (1986) assumes that

$$p(\beta_g) = N(\bar{\beta}_g, \bar{\Omega}_g), \quad (4)$$

where $\bar{\beta}_g$ and $\bar{\Omega}_g$ denote the prior mean and variance-covariance matrix of β_g , respectively. The residual variance-covariance matrix, Σ , is assumed fixed and diagonal, $\sigma_{gg}^2 I_T$. Stacking the time observation of the g^{th} equation, (2) can be written as

$$Y_g = X\beta_g + \varepsilon_g \quad g \equiv 1, \dots, n;$$

where Y_g and ε_g are $T \times 1$ vectors, and X is the stacked version of X_t in (2).

Given the assumed independence of the error terms, the likelihood (3) is just the product of independent normal densities,

$$L(Y | \beta, \Sigma) \propto |\sigma_{g,g}^2|^{-T/2} \exp \left\{ -\frac{1}{2\sigma_{g,g}^2} (Y_g - X\beta_g)' (Y_g - X\beta_g) \right\}. \quad (5)$$

The posterior distribution of the parameters of interest, given by

$$p(\beta_g | Y) = p(\beta_g) L(Y | \beta, \sigma_{g,g}^2),$$

is proportional to:

$$\begin{aligned} & |\sigma_{g,g}^2|^{-T/2} |\bar{\Omega}_g|^{-T/2} \exp \left\{ -\frac{1}{2} \left[(\beta_g - \bar{\beta}_g)' \bar{\Omega}_g^{-1} (\beta_g - \bar{\beta}_g) + \frac{1}{\sigma_{g,g}^2} (Y_g - X\beta_g)' (Y_g - X\beta_g) \right] \right\} \\ & \propto \exp \left\{ -\frac{1}{2} \left[\left(\frac{1}{\sigma_{g,g}^2} (Y_g' Y_g - 2Y_g' X\beta_g + \beta_g' X' X \beta_g) + \beta_g' \bar{\Omega}_g^{-1} \beta_g - 2\bar{\beta}_g' \bar{\Omega}_g^{-1} \beta_g + \bar{\beta}_g' \bar{\Omega}_g^{-1} \bar{\beta}_g \right) \right] \right\} \\ & \propto \exp \left\{ -\frac{1}{2} \left[\left(\beta_g' \left(\frac{1}{\sigma_{g,g}^2} X' X + \bar{\Omega}_g^{-1} \right) \beta_g - 2 \left(\frac{1}{\sigma_{g,g}^2} X' Y_g + \bar{\Omega}_g^{-1} \bar{\beta}_g \right)' \beta_g \right) \right] \right\}. \end{aligned}$$

In fact, $|\sigma_{g,g}^2|^{-T/2}$ and $|\bar{\Omega}_g|^{-T/2}$ are constants with respect to the integration for β_g in the first proportionality statement above, while $Y_g' Y_g$ and $\bar{\beta}_g' \bar{\Omega}_g^{-1} \bar{\beta}_g$ are constants in the second proportionality relation. Now, completing the square in the last exponential, we obtain

$$p(\beta_g | Y) \propto \exp \left\{ -\frac{1}{2} \left[(\beta_g' - \tilde{\Omega}_g \tilde{\beta}_g)' \tilde{\Omega}_g^{-1} (\beta_g' - \tilde{\Omega}_g \tilde{\beta}_g) \right] \right\}$$

with

$$\tilde{\beta}_g = \tilde{\Omega}_g (\bar{\Omega}_g^{-1} \bar{\beta}_g + \sigma_{g,g}^{-2} X' Y_g),$$

and

$$\tilde{\Omega}_g = (\bar{\Omega}_g^{-1} + \sigma_{g,g}^{-2} X' X)^{-1}.$$

In other words, $p(\beta_g | Y) = N(\tilde{\beta}_g, \tilde{\Omega}_g)$, where once we know $\bar{\Omega}_g^{-1}$, $\bar{\beta}_g$, and $\sigma_{g,g}^{-2}$, we can take $\tilde{\beta}_g$ as a point estimate.

Several remarks are in order here. First, note that, given the assumptions made, there is prior and posterior independence between equations. This explains why the equations can be estimated separately. Second, as noted above, Σ is assumed fixed and diagonal here, with its diagonal elements obtained from the estimation of a set of univariate autoregressive models of order p , $AR(p)$. Third, $\bar{\beta}_g$ and $\bar{\Omega}_g$ are unknown and specified in terms of few hyper-parameters. Finally, by assuming an infinite dispersion of the prior distribution around its mean ($\bar{\Omega}_g^{-1} = 0$)—i.e., by assuming a diffuse prior in the terminology used in the literature—the posterior mean of β_g

becomes $\tilde{\beta}_g = (X'X)^{-1} X'Y_g$, which is the OLS estimator of $\tilde{\beta}_g$.⁷

Litterman (1986) then assigns numerical values to the hyperparameters of the model on the presumption that most macroeconomic time series are well represented by random walk processes, as discussed in the previous section. Specifically, he assumes that Π is a degenerate random variable on the assigned values with the following structure for the diagonal elements of $\bar{\Omega}_g$:

$$var(\bar{\beta}_g) = \begin{cases} \pi_6 \pi_2 / l^{\pi_4} & \text{for the } g^{\text{th}} \text{ lagged endogenous variable} \\ (\pi_6 \pi_3 / l^{\pi_4}) \sigma_{g,g} / \sigma_{j,j} & \text{for the } j^{\text{th}} \text{ lagged endogenous variable} \\ \pi_6 \pi_5 \sigma_{g,g} & \text{for deterministic and exogenous variables} \end{cases}.$$

Here, π_6 controls the overall prior tightness (or uncertainty); π_2 controls the tightness of own lags, while π_3 controls the tightness of own lags relative to the tightness of lags of the other variables in the equation; π_4 controls the lag-decay in the prior variance with $l = 1, \dots, p$ denoting the variable's lags; π_5 controls the degree of uncertainty on the coefficients of the deterministic and/or exogenous variables in equation g , while the factors $\sigma_{g,g}$ and $\sigma_{j,j}$ measure the scale of fluctuations in variables g and j taking the unit of measure of different variables into account. Finally, the mean vector is specified as $\bar{\beta}_g = (0, \dots, 0, \pi_1, 0, \dots, 0)$, where π_1 is in the g^{th} position and represents the prior mean of coefficient on the first lag of the endogenous variable in equation g .⁸

Obviously, in order to provide a closer approximation to a full Bayesian estimation, these hyperparameters may be estimated consistently rather than being assigned arbitrary values. For instance, the pdf of the data (5) can be written in terms of Π and then maximized—the approach that we shall take in our application reported in Section VI of the paper.

The Diffuse Prior

The two main features of the Minnesota prior, the posterior independence between equations and the fixed residual variance-covariance matrix, can be relaxed with the priors considered in this subsection and in the next one. In fact, it can be shown that (see Kadyala and Karlsson, 1997), with the (diffuse) prior distribution

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

the joint posterior distribution is given by

$$p(\beta, \Sigma | Y) = p(\beta | \Sigma, Y) p(\Sigma | Y),$$

where

$$p(\beta | \Sigma, Y) = N(\hat{\beta}_{ols}, \Sigma \otimes (X'X)^{-1}) \quad (6)$$

⁷ Thus, a diffuse prior may be interpreted as assigning zero probability weight to the prior information relative to the information contained in the data.

⁸ Note that by allowing π_6 to approach infinity the prior becomes diffuse, while if π_3 is set to zero, the prior defines a set of univariate AR(p) models.

and

$$p(\Sigma | Y) = iW \left((Y - X\hat{B}_{ols})' (Y - X\hat{B}_{ols}), T - k \right), \quad (7)$$

with X (of dimensions $T \times k$) and Y (of dimensions $T \times n$) denoting the matrix version of X_t and Y_t respectively, $B = [B'_1, \dots, B'_p, D']'$ (of dimensions $k \times n$), and $iW(Q, q)$ denoting an inverted Wishart distribution with scale matrix Q and degrees of freedom q .⁹ Further, integrating Σ out of the joint posterior distribution, the marginal posterior distribution of B (the matrix form of the parameter vector β), $p(B | Y)$, can be shown to be

$$\propto \left| (Y - X\hat{B}_{ols})' (Y - X\hat{B}_{ols}) + (B - \hat{B}_{ols})' X' X (B - \hat{B}_{ols}) \right|^{-T/2},$$

which is a generalized t-Student distribution with scales $(Y - X\hat{B}_{ols})' (Y - X\hat{B}_{ols})$ and $X'X$, mean \hat{B}_{ols} , and degrees of freedom $T - k$ (see Kadyiala and Karlsson, 1997, page 104).¹⁰

Conjugate Prior Distributions

A mathematically convenient way of solving the main drawbacks of the Minnesota prior is to use a *conjugate family* of prior distributions. *Conjugacy* is the property by which the posterior distribution follows the same parametric form as the prior distribution.¹¹ In what follows we describe two possibilities, the Normal-Wishart and the Normal-Diffuse prior.

The Normal-Wishart Prior To relax the assumption of a fixed and diagonal variance-covariance matrix of the residuals, a natural conjugate prior for normal data is the Normal-Wishart,

$$p(\beta | \Sigma) = N(\bar{\beta}, \Sigma \otimes \bar{\Omega}) \quad (8)$$

$$p(\Sigma) = iW(\bar{\Sigma}, \alpha) \quad (9)$$

i.e., the unconditional prior distribution of β will be normal with prior mean and variance $E(\beta) = \bar{\beta}$ and $V(\beta) = (\alpha - n - 1)^{-1} \bar{\Sigma} \otimes \bar{\Omega}$, respectively, where α denotes the degrees of freedom of the inverse-Wishart and satisfies $\alpha > n + 1$.

Given this prior assumption, the posterior distribution is obtained as (e.g. Kadyiala and Karlsson,

⁹ The Wishart distribution is the multivariate generalization of the gamma distribution. If $W \sim W(Q, q)$, where W is of dimensions $k \times k$, then its density is proportional to $|W|^{(q-k-1)/2} \exp(-\frac{1}{2} \text{tr}(Q^{-1}W))$. On the other hand, if $W^{-1} \sim W(Q, q)$, then W has the inverse-Wishart distribution. The inverse-Wishart is the conjugate prior distribution for the multivariate normal covariance matrix. More references for univariate and multivariate distributions and for sampling from them are Zellner (1971, Appendix B) and Gelman et al (1995, Appendix A).

¹⁰ This integration, however, can also be approximated numerically by first drawing Σ from the inverted Wishart (7), and then using these draws to simulate β from (6).

¹¹ Formally, conjugacy is defined as follows: if \mathfrak{S} is a class of sampling distributions $p(y | \theta)$, and \mathfrak{N} is a class of prior distributions for θ , then the class \mathfrak{N} is conjugate for \mathfrak{S} if $p(y | \theta) \in \mathfrak{S}$ for all $p(\cdot | \theta) \in \mathfrak{S}$ and $p(\cdot) \in \mathfrak{N}$. *Natural* conjugate priors arise by taking \mathfrak{N} to be the set of all densities having the same functional form as the likelihood (Gelman et al. 1995).

1997, page 104)

$$p(\beta \mid \Sigma, Y) = N(\tilde{\beta}, \Sigma \otimes \tilde{\Omega}) \quad (10)$$

$$p(\Sigma \mid Y) = iW(\tilde{\Sigma}, T + \alpha) \quad (11)$$

where

$$\tilde{\Omega} = (\bar{\Omega}^{-1} + X'X)^{-1}$$

$$\tilde{B} = \tilde{\Omega} (\bar{\Omega}^{-1} \bar{B} + X'X \hat{B}_{ols})$$

and

$$\begin{aligned} \tilde{\Sigma} = & \hat{B}'_{ols} X'X \hat{B}_{ols} + \bar{B}' \bar{\Omega}^{-1} \bar{B} + \bar{\Sigma} \\ & + (Y - X \hat{B}_{ols})' (Y - X \hat{B}_{ols}) - \tilde{B}' (\bar{\Omega}^{-1} + X'X) \tilde{B}. \end{aligned}$$

As before, integrating Σ out of the joint posterior, the marginal posterior distribution of B is a multivariate t-Student distribution, whose integration can be easily performed numerically (Kadiyala and Karlsson, 1997, page 104).

Note that working with this prior, it is necessary to assume $\bar{\beta}$, $\bar{\Omega}$, and $\bar{\Sigma}$ as known. However, a Minnesota-type of specification for these matrices could be adapted and used here. In practice, $\bar{\beta}$ is specified as dependent upon only one hyper-parameter that controls the mean of the first lag of the endogenous variable, $\bar{\Omega}$ is specified as a diagonal matrix according to the scheme given in the previous section, and the diagonal elements of $\bar{\Sigma}$ are set as $\bar{\sigma}_{gg} = (\alpha - n - 1) s_{gg}^2$ with s_{gg}^2 estimated from univariate AR(p) models. Finally, the prior degrees of freedom, α , must be chosen so as to ensure the existence of the prior variances of the regression parameters.

The Normal-Diffuse Prior A slightly different prior, which in addition to allowing for non-diagonal residual variance-covariance matrix also avoids the Normal-Wishart kind of restrictions on the variance-covariance matrix of β , is the Normal-Diffuse prior firstly introduced by Zellner (1971). The key assumption made by Zellner here is prior independence between β and Σ , where

$$p(\beta) = N(\bar{\beta}, \bar{\Omega}),$$

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

The combination of these priors with the data yields the following marginal conditional distributions

$$p(\beta \mid \Sigma, Y) = N(\tilde{\beta}, \tilde{\Omega}) \quad (12)$$

$$p(\Sigma^{-1} \mid \beta, Y) = W(\tilde{\Sigma}^{-1}, T) \quad (13)$$

with

$$\begin{aligned}\tilde{\beta} &= \tilde{\Omega} \left(\bar{\Omega}^{-1} \bar{\beta} + (\Sigma^{-1} \otimes X'X) \hat{\beta}_{ols} \right) \\ \tilde{\Omega} &= (\bar{\Omega}^{-1} + (\Sigma^{-1} \otimes X'X))^{-1} \\ \tilde{\Sigma} &= (Y - X\hat{B}_{ols})' (Y - X\hat{B}_{ols}) + (B - \hat{B}_{ols})' X'X (B - \hat{B}_{ols})\end{aligned}\tag{14}$$

and where $W(\tilde{\Sigma}^{-1}, T)$ denotes the inverse of iW .

Then the conditional posterior of β can be integrated analytically to obtain the marginal, but the form of the latter is complicated in the sense that large differences between the information contained in the prior and the likelihood can cause the posterior to become bimodal, and thus the posterior mean to have low posterior probability (see, again, Kadyiala and Karlsson, 1997, page 106). For this reason it is preferable to integrate numerically.

B. Hierarchical Bayes Estimation

A prior distribution for the hyper-parameters can be incorporated into the specification of the model in the context of the *hierarchical* version of Zellner's SUR model, as recently proposed by Chib and Greenberg (1995). In general, this kind of specification does not provide a closed form solution for estimation, and full Bayesian estimation is approximated by using numerical integration methods. The hierarchical setup is relatively straightforward to work with in a simulation context and provides a great deal of flexibility in modeling as it increasingly more widespread application shows.

Chib's hierarchical SUR model is specified as follows:

$$Y_t = X_t\beta + \varepsilon_t, \quad \varepsilon_t \sim N_n(0, \Sigma)\tag{15}$$

$$\beta = M_o\theta + \zeta, \quad \zeta \sim N_{nk}(0, D_o)\tag{16}$$

$$\theta = M_1\mu + \eta, \quad \eta \sim N_m(0, D_1)\tag{17}$$

where the $nk \times 1$ parameter vector β is related to a $m \times 1$ parameter vector θ , which in turn is modeled as a function of a $r \times 1$ parameter vector μ . If $r \leq m \leq nk$, then the VAR parameters are projected onto progressively lower-dimensional spaces. The literature refers to (15), (16) and (17) as the *first*, *second* and *third stage* of the hierarchy, respectively.

The following assumptions are made by Chib: (i) M_o , M_1 , μ and D_1 are known; (ii) the matrices Σ , D_o are unknown; (iii) the errors are mutually independent across the hierarchy, implying that Y_t is conditionally independent of θ , D_o . As a result, the likelihood of the data, again, is proportional to:

$$L(Y \mid \beta, \Sigma) \propto |\Sigma|^{-T/2} \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (Y_t - X_t\beta)' \Sigma^{-1} (Y_t - X_t\beta) \right\}.\tag{18}$$

The prior information can then be completed by assuming

$$p(\Sigma^{-1}, D_o^{-1}) = p(\Sigma^{-1}) p(D_o^{-1}),$$

with $p(\Sigma^{-1}) = W(\bar{\Sigma}, s)$ and $p(D_o^{-1}) = W(\bar{D}_o, \rho)$, where all the hyperparameters are assumed known. For instance, they may set in accordance to the Minnesota prior as before or more simply be estimated from data.

The posterior density of the parameters of interest $\psi = (\beta, \Sigma, \theta, D_o)$ is given by

$$p(\psi | Y) \propto L(Y | \beta, \Sigma) p(\psi),$$

where $p(\psi)$ is constructed from (16) and (17) (stages 2 and 3 of the hierarchy, respectively) and the prior distributions on the parameters. The Gibbs sampler can then run cycling through the conditional posterior distributions (19)–(22) below, which are in standard form (see Chib and Greenberg, 1995):

$$\beta | \psi_{-\beta}, Y \sim N(\tilde{\beta}, \tilde{\Omega}); \quad (19)$$

$$\Sigma^{-1} | \psi_{-\Sigma}, Y \sim W(\tilde{\Sigma}, s + T); \quad (20)$$

$$\theta | \psi_{-\theta}, Y \sim N(\Delta_1 (D_1^{-1} M_1 \mu + M_o' D_o^{-1} \beta), \Delta_1); \quad (21)$$

$$D_o^{-1} | \psi_{-D_o}, Y \sim W(\tilde{D}_o, \rho + 1); \quad (22)$$

where

$$\tilde{\beta} = \tilde{\Omega} \left(D_o^{-1} M_o \theta + \sum_t X_t' \Sigma^{-1} Y_t \right), \quad \tilde{\Omega} = \left(D_o^{-1} + \sum_t X_t' \Sigma^{-1} X_t \right)^{-1},$$

$$\tilde{\Sigma} = \left[\bar{\Sigma}^{-1} + \sum_{t=1}^T (Y_t - X_t \beta) (Y_t - X_t \beta)' \right]^{-1},$$

$$\Delta_1 = (D_1^{-1} + M_o' D_o^{-1} M_o)^{-1},$$

$$\tilde{D}_o = \left[\bar{D}_o^{-1} + \sum_{g=1}^n (\beta_g - \theta) (\beta_g - \theta)' \right]^{-1}.$$

The number of stages in the hierarchy could be more or less than the three discussed here. In the case of more stages, we essentially have to add more conditional posterior distributions to the Gibbs sampler. In the case of less stages, as for instance in the case of the Minnesota prior discussed in the previous section, we have to fix somewhat arbitrarily the value of the

hyperparameters of the omitted stage.

For example, the Minnesota prior—see equation (4) above—can be written as

$$\beta_g = M_g \theta + v_g, \quad v_g \sim N(0, \bar{\Omega}_g), \quad (23)$$

where $\bar{\beta}_g = M_g \theta$, $\theta = (\pi_1 \ 0)'$, and

$$M_g = \begin{bmatrix} 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix},$$

with the ones of the first column and the zeros of the second referring to the g^{th} row of the VAR. Therefore, the difference between the latter specification and the Chib's three-stages hierarchy is that θ and $\bar{\Omega}_g$ are either estimated or assumed to be known based on the Minnesota rules of thumb, while we give a proper prior on the hyper-parameters of the model in the three-stages hierarchy. Obviously, the hierarchical model is closer to a full Bayesian model than an empirical Bayesian one in the sense that it incorporates a more general prior assumption for the hyper-parameters.

IV. SOME EXTENSIONS

In this section we consider a few useful extensions of the models discussed thus far. In particular, we shall discuss models with time-varying coefficients (TVC), which have been shown to capture both nonlinearities and nonnormalities in the data, an alternative to the normal data model that captures outliers (or fat-tails) in the distribution of the error terms, and briefly touch upon regime switching models. The first two extensions are relevant for modeling financial data. The latter is more of interest in macroeconomic applications. The models discussed in this section clearly illustrate the flexibility of a Bayesian approach to estimation, which allows complicated estimation problems to be handled in a fairly simple way.

A. Time-Varying Coefficients

The Kalman Filter with a Minnesota Prior

A standard tool for the estimation of linear regression models with time-varying parameters is the Kalman Filter. The Kalman Filter is an algorithm for recursively updating a linear projection for a dynamic system represented in state-space form, given a set of initial values on the parameters describing its status.¹² An empirical Bayesian procedure for the estimation of a time-varying version of (2), equation-by-equation, with known variance-covariance matrix of the residuals as discussed in Section III.A, has been developed by Doan et al. (1984). Doan et al. suggest to

¹² See Hamilton (1994, Chapter 13) for more details on the Kalman Filter.

specify the law of motion of β_t as a first order autoregressive process represented in state-space form and to give a Minnesota-type prior on β_{t-1} , so that the Kalman Filter can be applied to update it for all t , recursively.

Consider the problem of estimating the entire parameter vector of (2) when this changes over time in all periods t (thus, we want to obtain a sequence of posterior distributions for $\beta_1, \beta_2, \dots, \beta_T$), and assume that β_t follows a stationary *VAR* process of order one, such as:

$$\beta_t = A\beta_{t-1} + \eta_t \quad \eta_t \mid \mathbf{l}_{t-1} \sim N(0, \Phi), \quad (24)$$

where

$$\beta_{t-1} \mid \mathbf{l}_{t-1} \sim N(\bar{\beta}_{t-1|\mathbf{l}_{t-1}}, \bar{\Omega}_{t-1|\mathbf{l}_{t-1}}),$$

with $\beta_{t-1}, \varepsilon_t$, and η_t independently distributed. Here, Σ , A , and Φ are assumed to be known, and \mathbf{l}_{t-1} denotes the information available at time $t - 1$. Thus, we are considering the following specification for (2):

$$\begin{aligned} Y_t &= X_t\beta_t + \varepsilon_t, & \varepsilon_t \mid \mathbf{l}_{t-1} &\sim N(0, \Sigma); \\ \beta_t &= A\beta_{t-1} + \eta_t, & \eta_t \mid \mathbf{l}_{t-1} &\sim N(0, \Phi); \\ \beta_{t-1} &\mid \mathbf{l}_{t-1} &\sim N(\bar{\beta}_{t-1|\mathbf{l}_{t-1}}, \bar{\Omega}_{t-1|\mathbf{l}_{t-1}}). \end{aligned} \quad (25)$$

Note that, conditional on \mathbf{l}_{t-1} , the prior distribution of β_t is

$$\beta_t \mid \mathbf{l}_{t-1} \sim N(\beta_{t|\mathbf{l}_{t-1}}^*, \Omega_{t|\mathbf{l}_{t-1}}^*),$$

where

$$\beta_{t|\mathbf{l}_{t-1}}^* = A\bar{\beta}_{t-1|\mathbf{l}_{t-1}} \text{ and } \Omega_{t|\mathbf{l}_{t-1}}^* = A\bar{\Omega}_{t-1|\mathbf{l}_{t-1}}A' + \Phi.$$

Substituting for β_t in the first line of (25), we obtain a state space model for β_t with observation equation given by

$$Y_t = X_t\beta_{t|\mathbf{l}_{t-1}}^* + X_t\eta_t + \varepsilon_t.$$

Now, it is possible to show (e.g., Ballabriga et al., 1998) that, under these assumptions, the marginal posterior distribution of β_t is given by:

$$\beta_t \mid \Sigma, \Phi, \mathbf{l}_{t-1}, Y_t \sim N(\bar{\beta}_{t|t}, \bar{\Omega}_{t|t}),$$

where $\bar{\beta}_{t|t}$ and $\bar{\Omega}_{t|t}$ are the one-period-ahead forecast of β_t and the variance-covariance matrix of its mean square error, respectively, calculated by the Kalman Filter as:

$$\begin{aligned} \bar{\beta}_{t|t} &= \beta_{t|\mathbf{l}_{t-1}}^* + \Omega_{t|\mathbf{l}_{t-1}}^* X_t' F_t (Y_t - X_t \beta_{t|\mathbf{l}_{t-1}}^*) \\ \bar{\Omega}_{t|t} &= \Omega_{t|\mathbf{l}_{t-1}}^* - \Omega_{t|\mathbf{l}_{t-1}}^* X_t' F_t X_t \Omega_{t|\mathbf{l}_{t-1}}^* \\ F_t &= (X_t \Omega_{t|\mathbf{l}_{t-1}}^* X_t' + \Sigma)^{-1}. \end{aligned} \quad (26)$$

To make this scheme operational, one needs a consistent estimate of Σ , values for the hyper-

parameters of the model (A and Φ), and initial values for the elements of $\Omega_{1|0}$ and $\beta_{1|0}$. As we saw in the time-invariant case discussed in the previous section, these may be obtained in many different ways.¹³

A Hierarchical Time-Varying SUR model

Parameter time-variation may be specified also in the context of the hierarchical framework proposed by Chib and Greenberg (1995), based on an important result previously obtained by Carter and Kohn (1994), by adding one stage to the hierarchy implicit in the model discussed in the previous section.¹⁴

In this case, parameter time-variation is specified as follows:

$$\begin{aligned} Y_t &= X_t \beta_t + \varepsilon_t, & \varepsilon_t &\sim IIN_n(0, \Sigma) \\ \beta_t &= M_o \theta_t + \zeta_t, & \zeta_t &\sim N_{nk}(0, D_o) \end{aligned} \quad (27)$$

$$\theta_t = \theta_{t-1} + \eta_t, \quad \eta_t \sim N_m(0, D_1), \quad (28)$$

where ε_t , ζ_t , and η_t are assumed to be independent. Compared to the model in section 3.2, it is now $\theta = \theta_t$, $\mu = \theta_{t-1}$ and $M_1 = I$. Then, the law of motion of θ_t is specified as a random walk without drift, but could be an $AR(1)$. By substituting (28) in (27) and rearranging the resulting expression we have:

$$\beta_t = \beta_{t-1} + v_t$$

with

$$v_t = M_o \eta_t + \zeta_t - \zeta_{t-1},$$

which has the same structure as (24) except that the error term now contains a moving average component of the first order, $MA(1)$.

Assuming that θ_o and D_1 are known and conditioning on $\{\theta_t\}_{t=0}^T$, the conditional posterior distribution of the remaining parameters can be easily derived. For instance, conditional on $\{\theta_t\}_{t=0}^T$, $\beta_t \sim N(M_o \theta_t, D_o)$, and its conditional posterior is obtained by combining this prior with the likelihood (18). Proceeding in this way, the following conditional distributions can be obtained:

$$\beta_t \mid \psi_{-\beta_t}, y \sim N(\tilde{\beta}_t, \tilde{\Omega}_t); \quad (29)$$

¹³ We apply this model to the estimation of a system of reaction functions for four European central banks under the EMS. See also Ballabriga et al. (1998) for a detailed presentation of the methodology.

¹⁴ See Canova and Ciccarelli (2000) and Ciccarelli and Rebucci (2002) for applications of this procedure to the prediction of turning points in the business cycle of G7 countries and to the analysis of the transmission mechanism of European monetary policy, respectively.

$$\Sigma^{-1} \mid \psi_{-\Sigma}, y \sim W\left(\tilde{\Sigma}, s + T\right); \quad (30)$$

$$D_o^{-1} \mid \psi_{-D_o}, y \sim W\left(\tilde{D}_o, \rho + T\right); \quad (31)$$

where

$$\tilde{\beta}_t = \tilde{\Omega}_t \left(D_o^{-1} M_o \theta_t + X_t' \Sigma^{-1} y_t \right), \quad \tilde{\Omega}_t = \left(D_o^{-1} + X_t' \Sigma^{-1} X_t \right)^{-1},$$

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

$$\tilde{D}_o = \left[\bar{D}_o^{-1} + \sum_{t=1}^T (\beta_t - M_o \theta_t) (\beta_t - M_o \theta_t)' \right]^{-1}.$$

The joint posterior distribution of $\{\theta_t\}$ conditional on the remaining parameters and the data can then be obtained in two steps as shown by Chib and Greenberg (1995, pp. 349-350). First, by initializing $\{\theta_t\} \mid Y, \psi_{-\theta}$ with the Kalman Filter, whose output $\left\{ \hat{\theta}_{t|t}, R_{t|t}, F_t \right\}_{t=0}^T$ is saved for each t :

$$\begin{aligned} \hat{\theta}_{t|t} &= \hat{\theta}_{t|t-1} + K_t \left(\beta_t - M_o \hat{\theta}_{t|t-1} \right), \\ R_{t|t} &= (I - K_t M_o) R_{t|t-1} \\ F_t &= R_{t|t} R_{t+1|t}^{-1}, \end{aligned} \quad (32)$$

where

$$\begin{aligned} \hat{\theta}_{t|t-1} &= \hat{\theta}_{t-1|t-1}, \\ R_{t|t-1} &= R_{t-1|t-1} + D_1, \\ \Gamma_{t|t-1} &= M_o R_{t|t-1} M_o' + D_o, \\ K_t &= R_{t|t-1} M_o' \Gamma_{t|t-1}^{-1}. \end{aligned}$$

And second, by sampling $p(\{\theta_t\} \mid Y, \psi_{-\theta})$ in reverse time order from

$$\begin{aligned} \theta_T &\sim N\left(\hat{\theta}_{T|T}, R_{T|T}\right) \\ \theta_{T-1} &\sim N\left(\hat{\theta}_{T-1|T}, R_{T-1|T}\right) \\ &\vdots \\ \theta_o &\sim N\left(\hat{\theta}_o, R_o\right), \end{aligned} \quad (33)$$

where

$$\begin{aligned}\hat{\theta}_t &= \hat{\theta}_{t|t} + F_t (\theta_{t+1} - \hat{\theta}_{t|t}) \\ R_t &= R_{t|t} - F_t R_{t+1|t} F_t'\end{aligned}$$

Note that with this algorithm, we simulate $\{\theta_t\}$ from the *joint* distribution $p(\theta_o, \theta_1, \dots, \theta_T \mid \psi_{-\theta})$, rather than from its full conditional distribution $\theta_t \mid \psi_{-\theta}$. The latter simulation strategy produces very slow convergence to the ergodic distribution because it requires adding $T + 1$ additional blocks to the Gibbs sampler, a problem exacerbated as the dimension of θ_t increases. In the former case, instead, only one additional block is introduced in the sampler.

B. Nonnormal Data

The Bayesian framework discussed thus far can be easily adapted to take into account the presence of outliers, or to model data coming from distributions with higher probability of extreme observations than under the normal distribution—sometimes called fat-tailed distributions in the empirical finance literature. Being able to accommodate outliers or fat-tails is particularly important in financial applications dealing with high frequency data, which are well known to depart from normality.

In a Bayesian framework, the presence of outliers can be accommodated by replacing the normal data model usually assumed by a fatter-tailed family of distributions for the error term and thus the data, such as the t-student, or by using mixture models.¹⁵ A nonnormal data specification can then be combined with alternative prior assumptions for the parameters of interest to obtain their posterior distributions along the line discussed thus far.

Unlike the classical analysis of outliers, the Bayesian analysis makes no distinction between methods that search for outliers—possibly removing them from the analysis—and robust estimation procedures that are not vulnerable to their presence. Bayesian researchers characterize outliers as observations coming from fat-tailed distributions, or high-variance periods in the context of mixture models of time series, rather than extreme observations from a normal distribution. As a result, these observations do not distort the point estimates of the population moments. In addition, such a modeling approach is compatible with many of the prior distributions for the parameters of interest seen in the previous sections, including the possibility of time-variation.

To present these ideas more formally, consider the same hierarchical model discussed in Section III.B, generalized to allow for fat-tails in the data as described by a t-student distribution for the

¹⁵ Mixture models are statistical model based on the combination of two or more distributions. They are frequently used in situations in which the measurement of a random variable is taken under two or more conditions, or where the population is known or assumed to consist of subpopulations that follow a different, simpler model. For more details see, for instance, Gelman et al. (1995).

vector of error terms:

$$Y_t = X_t \beta + \varepsilon_t, \quad \varepsilon_t \sim t_\nu(0, \Sigma) \quad (34)$$

$$\beta = M_o \theta + \zeta, \quad \zeta \sim N_{nk}(0, D_o) \quad (35)$$

$$\theta = M_1 \mu + \eta, \quad \eta \sim N_m(0, D_1) \quad (36)$$

where $t_\nu(\mu, \sigma^2)$ denotes a t-student distribution with location μ , scale σ , and degrees of freedom ν , $\nu \in (0, \infty)$, determining the shape of the distribution.¹⁶

As for the estimation of the model, the new specification requires only a small modification to what discussed in Section III.B. Specifically, the prior (37)-(38) in footnote must be added to the hierarchical scheme (34)-(36) and the conditional distribution of σ_t must be added to the Gibbs sampler. It is not difficult to show that, assuming ν is fixed, the conditional distributions becomes:

$$\beta \mid \psi_{-\beta}, Y \sim N(\tilde{\beta}, \tilde{\Omega}); \quad (39)$$

$$\Sigma^{-1} \mid \psi_{-\Sigma}, Y \sim W(\tilde{\Sigma}, s + T); \quad (40)$$

$$\theta \mid \psi_{-\theta}, Y \sim N(\Delta_1 (D_1^{-1} M_1 \mu + M_o' D_o^{-1} \beta), \Delta_1); \quad (41)$$

$$D_o^{-1} \mid \psi_{-D_o}, Y \sim W(\tilde{D}_o, \rho + 1); \quad (42)$$

$$\sigma_t \mid \psi_{-\sigma_t}, Y \sim \text{Inv-}\chi^2(\nu + T, s^2) \quad (43)$$

where

$$\tilde{\beta} = \tilde{\Omega} \left(D_o^{-1} M_o \theta + \sum_t X_t' (\sigma_t \Sigma)^{-1} Y_t \right), \quad \tilde{\Omega} = \left(D_o^{-1} + \sum_t X_t' (\sigma_t \Sigma)^{-1} X_t \right)^{-1},$$

¹⁶ The t-student distribution can be interpreted as mixture model, i.e., a mixture of normal distributions with a common mean and time-varying variance (in the case of a time series regression) distributed as scaled inverse chi-squared distribution. For instance,

$$\varepsilon_t \sim t_\nu(0, \Sigma)$$

is equivalent to

$$\varepsilon_t \sim N(0, \sigma_t \Sigma) \quad (37)$$

$$\sigma_t \sim \text{Inv-}\chi^2(\nu, 1). \quad (38)$$

Statistically, therefore, outliers may be interpreted as observations drawn from a distribution with higher variance. Hence, the variance of the error term is heteroschedastic and the degree of heteroschedasticity depends on the number of degrees of freedom ν . In fact, as ν approaches infinity, ε_t converges in distribution to $N(0, \Sigma)$, as $E(\sigma_t)$ tends to one and $V(\sigma_t)$ tends to zero.

$$\tilde{\Sigma} = \left[\bar{\Sigma}^{-1} + \frac{\sum_{t=1}^T (Y_t - X_t\beta) (Y_t - X_t\beta)'}{\sigma_t} \right]^{-1},$$

$$\Delta_1 = (D_1^{-1} + M_o' D_o^{-1} M_o)^{-1},$$

$$\tilde{D}_o = \left[\bar{D}_o^{-1} + \sum_{g=1}^n (\beta_g - \theta) (\beta_g - \theta)' \right]^{-1}$$

with

$$s^2 = [\nu + (Y_t - X_t\beta)' \Sigma^{-1} (Y_t - X_t\beta)] / \nu.$$

denoting the scale of the scaled inverse chi squared.

In practice, if the t-student aims at fitting a long-tailed distribution on a long series of observations, then it is generally appropriate to include the degrees of freedom as an unknown parameter to be estimated. If instead the t-student specification is chosen as a robust alternative to the normal to control for outliers, then ν can be fixed at an arbitrarily small value, but no smaller than prior understanding dictates (see Gelman et al., 1995, page 350 for more details on this).¹⁷

C. Regime Switching, Nonlinear Models, and Beyond

Regime-switching time series models are a popular device to model non-linear dynamic, following the seminal contribution of Hamilton (1989). Shifts between two or more regimes can be easily accommodated in the Bayesian framework discussed thus far.¹⁸ Canova (1993), however, has shown that nonlinearities, nonnormalities and conditional heteroschedasticity may also be modelled by a the kind of time-varying coefficient models (TVCs) discussed above. In particular, he shows that a TVC model nests specifications generally used to characterize typical departures from standard assumptions such as conditionally normal ARCH and ARCH-M models, or even Hamilton's regime switching effects. Interestingly, as we have seen in the previous subsections, a TVC specification is rather easily estimated in a Bayesian framework.

The flexibility of the Bayesian approach does not end here and the modeling possibilities are numerous. For instance, it is possible to specify a t-student prior for the *parameters* rather than the error terms, thereby allowing for outliers in the parameters, while continuing to use a normal data model. Sometimes, it is difficult to justify a priori the normality of the parameter vector, especially if the VAR includes equations referring to different decision units, as in the case of a panel VAR. Moreover, posterior inference is generally sensitive to the assumed prior, even when the model fits the data well. For this reason, it may be useful to check robustness by assuming

¹⁷ See Ciccarelli and Rebucci (2002) for an application of a non-normal, TVC model to the measurement of financial contagion. Similar models have been used by Sims (1999) and Cogley and Sargent (2002) to study monetary policy issues.

¹⁸ See Kim (1994) and Chib et al. (2001a) for univariate applications and Chib et al. (2001b) and Cogley and Sargent (2002), among others, for multivariate applications.

alternative specifications for the prior distribution of the parameter vector, such as a t-student as we discussed for the error terms.

Assume for instance that in (34)-(36) we change the assumption on the population structure as

$$\beta \mid \theta, \nu, D_o \sim t_\nu(M_o\theta, D_o).$$

As in previous subsection, the t -distribution can be written as a mixture of a normal and a scaled inverse- χ^2 as

$$\beta \mid \theta, \tau, D_o \sim N(M_o\theta, \tau D_o) \quad (44)$$

$$\tau \sim \text{Inv-}\chi^2(\nu, 1). \quad (45)$$

The conditional posterior of τ is then added to the Gibbs sampler as we added the conditional posterior of σ_t in the previous section.

V. FORECASTING AND STRUCTURAL ANALYSIS

In order to compute out of sample forecasts, impulse response functions (IRFs) and forecast error variance decompositions (FVDs), take the companion form of model (1) with no deterministic components but the constant:

$$\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{B}\mathbf{Y}_{t-1} + \mathbf{U}_t, \quad (46)$$

where

$$\mathbf{Y}_t = \begin{bmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p-1} \end{bmatrix}_{(np \times 1)}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(np \times 1)},$$

$$\mathbf{B} = \begin{bmatrix} B_1 & B_2 & \cdots & B_{p-1} & B_p \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{bmatrix}_{(np \times np)}, \quad \mathbf{U}_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(np \times 1)}.$$

Solving forward from period T , we can express the h -step ahead out-of-sample forecast at time T as:

$$\mathbf{Y}_{T+h} = \sum_{j=0}^{h-1} \mathbf{B}^j \boldsymbol{\mu} + \mathbf{B}^h \mathbf{Y}_T + \sum_{j=0}^{h-1} \mathbf{B}^j \mathbf{U}_{T+h-j} \quad h = 1, 2, \dots \quad (47)$$

Now, defining the $n \times np$ matrix $J = [I_n \ 0 \ \dots \ 0]$ as in Lütkepohl (1990, page 13) and using the

fact that $J'JU_t = U_t$ and that $JU_t = \varepsilon_t$, we obtain the h -step ahead, out-of-sample forecast of Y_{T-h} :

$$Y_{T+h} = JY_{T+h} = C_{h-1}\mu + JB^hY_T + \sum_{j=0}^{h-1} \Phi_j \varepsilon_{T+h-j}, \quad (48)$$

where $C_0 = I$, $C_i = I + \sum_{j=1}^i B_j C_{i-j}$ ($i = 1, 2, \dots$), $B_j = 0$ for $j > p$, and $\Phi_j = JB^j J'$.

Forecasting (conditional and unconditional), impulse response and variance decomposition analysis are natural application of (48), where the first two terms add up to the expected value of Y_{T+h} (i.e., $Y_T(h) = C_{h-1}\mu + JB^hY_T = E[Y_{T+h} | Y_T, \beta, \Sigma]$), while the last term is the forecast error ($Y_{T+h} - Y_T(h)$) with conditional variance equal to $\sum_{j=0}^{h-1} \Phi_j \Sigma \Phi_j'$.

A. Unconditional Forecasting

There are two different ways to forecast future realizations of the vector of variables of interest. If there are conditions or constraints on the future value of the variables or the shocks driving the model, the forecast produced by means of (48) is *conditional*; if there are no such conditions, it is an *unconditional forecast*. Here, we focus only on unconditional forecasts.¹⁹

The unconditional forecasting function ($Y_T(h)$) is given by the first two terms in (48),

$$Y_T(h) = C_{h-1}\mu + JB^hY_T. \quad (49)$$

Some algebra manipulations show that the h -step ahead forecasting function may also be written, recursively, as

$$Y_T(h) = \mu + B_1 Y_T(h-1) + \dots + B_p Y_T(h-p) \quad (50)$$

where $Y_T(j) = Y_{t+j}$ for $j \leq 0$. *Point forecasts* may now be computed in two ways.

First, by substituting in equation (50) estimates of μ and B_l ($l = 1, \dots, p$) obtained from the mean of the posterior distribution of β :

$$\hat{Y}_T'(h) = \hat{\mu} + \hat{B}_1 \hat{Y}_T(h-1) + \dots + \hat{B}_p \hat{Y}_T(h-p). \quad (51)$$

Alternatively, by defining $\vartheta = (\beta, \Sigma)$ and evaluating the integral

$$\hat{Y}_T''(h) = \int Y_T(h) \cdot p(\vartheta | Y) d\vartheta, \quad (52)$$

where $p(\vartheta | Y)$ denotes the posterior distribution of ϑ and the forecasting function is averaged by taking as weights the whole posterior density of the parameters and not just their posterior mean as in (51).

If instead one is interested in forecasting the whole density of $Y_t(h)$ and not one just “one

¹⁹ On conditional forecasting see, among others, Waggoner and Zha (1998). Kadyiala and Karlsson (1997) evaluate the forecasting performance of most of the specifications discussed in the previous section.

point", i.e., one is interested in *density forecasts*, the following integral may be evaluated either analytically or numerically by means of Gibbs sampler:²⁰

$$p(Y_t(h) | Y_T) = \int p(Y_t(h) | Y_T, \vartheta) p(\vartheta | Y) d\vartheta.$$

The procedure is relatively simple. For each draw of $\vartheta^{(j)}$, we draw $Y_t^{(j)}(h)$ from $p(Y_t(h) | Y_T, \vartheta^{(j)})$. After, say, M iterations of the Gibbs sampler these draws may be regarded as draws from $p(Y_t(h) | Y_T)$. A point forecast can then be obtained using the ergodic mean of the empirical distribution.

B. Structural Analysis

Computation of unconditional forecasts is related to the calculation of IRFs and FVDs. But while the reduced form is all we need for unconditional forecasting, structural VAR analysis requires the solution of a joint estimation-identification problem. However, if the system is exactly identified, the analysis remains relatively simple.²¹

For instance, take the model with time-invariant parameters and any of the prior assumptions discussed above. Any such model may be seen as the reduced form of the a structural form in which:

$$\Sigma^{-1} = A' A,$$

where A is an $n \times n$ non-singular matrix of contemporaneous correlations, $\varepsilon_t = A^{-1}v_t$, and v_t is assumed normally and independently distributed with $E[v_t | Y_{t-s}, s > 0] = 0$ and $E[v_t v_t' | Y_{t-s}, s > 0] = I$, for all t . Equation (48) can then be re-written as

$$Y_{T+h} = JY_{T+h} = C_{h-1}\mu + JB^h Y_T + \sum_{j=0}^{h-1} \Psi_j v_{T+h-j} \quad (53)$$

where the $n \times n$ matrix $\Psi_j = \Phi_j A^{-1}$ is the matrix of the j^{th} -period-ahead impulse responses, while the forecast error variance is then given by $\sum_{j=0}^{h-1} \Psi_j \Psi_j' = \sum_{i=1}^n \sum_{j=0}^{h-1} \Phi_j a_i a_i' \Phi_j'$ where a_i is the i -th column of A^{-1} .

Probability distributions of the responses to an impulse in the k -th structural shock may be computed by making random draws for the k -th column of Ψ_j for periods $j = 0, \dots, h$, as explained for unconditional forecasting. The mean response and the percentiles are then used to summarize the posterior distribution of these statistics. Similarly a distribution for the contribution of the i -th innovation to the forecast error variance of the h -step ahead forecast can be obtained by making

²⁰ Density forecast is of great help when the researcher wants to compute turning point probabilities, for instance, as in Canova and Ciccarelli (2001).

²¹ See Amisano and Giannini (1997) on the general issue of identification in VARs. For a more complete Bayesian treatment of the calculation of impulse response functions in structural VAR models see Koop (1992). For the case of over-identification in a Bayesian framework see Zha (1999) and Sims and Zha (1998).

draws of $\sum_{j=0}^{h-1} \Phi_j \mathbf{a}_i \mathbf{a}_i' \Phi_j'$ as a proportion of the total forecast error variance.

In the case of overidentification, the mapping between Σ^{-1} and A is not one-to-one. This means that we cannot obtain the posterior distribution of A while computing the posterior distribution of Σ , given the chosen identifying scheme. In case of overidentification, as explained in Sims and Zha (1998), a joint prior distribution on the parameters governing *contemporaneous and lagged* interdependence between the variables of interest must be specified, and the posterior distribution of the parameters governing contemporaneous correlations becomes nonstandard. In this case, one needs to take a second-order Taylor expansion around the maximum of the likelihood to obtain the posterior distribution of these parameters.²²

VI. AN APPLICATION: ESTIMATING A SYSTEM OF REACTION FUNCTIONS

In this section, by drawing on the work reported by Ciccarelli and Rebucci (2002), we apply some of the results previously reviewed to the estimation of a system of monetary reaction functions of the type discussed and estimated by Clarida, Gali, and Gertler (1997).

Researchers estimate central bank reaction functions either to investigate the behavior of the monetary authorities or to obtain measures of the expected and unexpected component of monetary policy. In the empirical monetary policy literature, examples of use of estimated reaction functions for both purposes abound. For instance, in a seminal contribution, Clarida and others (1997) estimate the reaction function of the largest advanced economies' central banks to compare their behavior; Dornbusch and others (1998) estimate a system of reaction functions for European central banks to compare the transmission mechanism of monetary policy across these countries; Sims (1999) and Sargent and Cogley (2002) revisit the U.S. postwar monetary policy history by estimating a reaction function for the US Federal Reserve with methods similar to those reviewed in this paper; finally, Ciccarelli and Rebucci (2002) use the system of reaction functions presented in this section to investigate how the transmission mechanism of European monetary policy has changed over time.

More specifically, we consider four central banks in the European Monetary System (EMS)—Germany, France, Italy, and Spain—which are the four largest economies currently in the European Monetary Union (EMU), accounting for about 80 percent of the euro-area GDP.²³ As these countries' commitment to the EMS, and hence their monetary policy regime, has changed over time, we let also the model parameters change over time. In addition, as short-term interest rates may reflect not only monetary policy actions and surprises but also exogenous shocks to exchange rate risk premia and money demand shocks not fully accommodated by the money supply, we assume that the error terms are t-distributed.

Therefore, we estimate a time-varying coefficient system for nonnormal data by adapting the

²² See Zha (1999, page 300) for details.

²³ There is nothing in our empirical framework that would prevent us to include more than four countries except additional computing costs.

models discussed in Sections IV.A and VI.B. Specification, identification and estimation of the particular econometric model used are presented in the first three subsections. The estimation results are reported and discussed in the following one.

A. Specification

The behavior of the four European central banks considered is modeled by the following time-varying structural VAR:

$$A_t(L) R_t = B_t(L) W_t + D_t + U_t, \quad (54)$$

where $R_t = [r_t^1, \dots, r_t^4]'$ is a (4×1) vector of monetary policy instruments, $W_t = [w_t^1, \dots, w_t^4]'$ is a (4×1) vector of monetary policy final objectives and exogenous variables, $A_t(L)$ and $B_t(L)$ are time-varying polynomial matrices in the lag operator L , with lag length p_1 and p_2 respectively, and D_t is a (4×1) vector of constants. Here, $U_t = [u_t^1, \dots, u_t^4]'$ is a (4×1) vector of monetary policy shocks such that:

$$\begin{aligned} E[U_t | Z_{t-s}] &= 0, \text{ for all } t \text{ and } s > 0; \\ E[U_t U_t' | Z_{t-s}] &= I, \text{ for all } t \text{ and } s > 0; \\ E[U_t U_s'] &= 0, \text{ for all } t \neq s, \end{aligned} \quad (55)$$

where Z_t contains lagged R_t and contemporaneous and lagged W_t , and E denotes the expectation operator and that.

A short-term interest rate is assumed to be the monetary policy instrument. Consistent with the specification of a standard VAR for the analysis of monetary policy in open economy, in w_{it} we include contemporaneous and lagged inflation (π), output (y), and lagged nominal exchange rate (e) in percent deviation from its target (π^* , y^* , e^* , respectively). In addition, we include a contemporaneous and lagged index of commodity prices (cp), the U.S. federal fund's rate (r_t^{US}) and the lagged value of a broad monetary aggregate (m). Commodity prices and the US money market rate are included to control for external shocks, while monetary aggregates are widely believed to have been monitored closely by most European central banks throughout the period considered. Thus,

$$w_{it} = [(\pi_{i,t} - \pi_i^*), (y_{i,t} - y_i^*), (e_{i,t-1} - e_i^*), m_{i,t-1}, cp_t, r_t^{US}]' \text{ for } i = 1, \dots, 4.$$

All data used are from the *International Financial Statistics* database of the IMF. As a proxy for the short-term interest rate, following Bernanke and Mihov (1997) and Clarida, Gali, and Gertler (1997), we use the money market rate. Output is measured by an industrial production index. Inflation is measured by the annual change in the consumer price index. We use the bilateral exchange rate vis-à-vis the deutsche mark (DM) for France, Italy, and Spain, and the DM/U.S. dollar rate for Germany. The commodity price index and the monetary aggregate are entered in first-difference. The monetary aggregate chosen is a seasonally adjusted M3 series.²⁴ The targets

²⁴ All variables are transformed in natural logarithm so that estimated coefficients can be

variables π^* and y^* are the fitted values of a linear regression of the actual variables ($\pi_{i,t}$ and $y_{i,t}$) on a constant and a linear trend and a constant and a quadratic trend, respectively, while e^* is the central parity vis-a-vis the DM.²⁵

The specification chosen imposes very few a priori restrictions on the system of reaction functions: all parameters in $A_t(L)$ and $B_t(L)$ are unrestricted and can vary over time, including those governing the contemporaneous causation among short-term interest rates.²⁶ Leaving $B_t(L)$ unrestricted allows the behavior of the central banks considered to change during the sample period, letting the data reveal which objectives they were actually pursuing in each period. For instance, the Spanish peseta joined the EMS only in 1989, the Italian lira has been floating more or less freely from September 1992 to November 1996, and the fluctuation bands of all three currencies vis-à-vis the DM have changed several times during the period considered. Even the Bundesbank's focus might have shifted away from strictly domestic objectives, after the early years of the German unification, in the run up to the EMU. It is evident that, if the policy changes and possible outliers are not accounted for by the estimated parameters of the system of reaction functions, they will end up in the estimated residuals, thereby potentially undermining their interpretation of well behaved (i.e., white noise) policy innovations as assumed in (55).

Leaving $A_t(L)$ unrestricted allows for lagged interdependence among short term interest rates of different countries as well as for varying degrees of interest rate smoothing over time. In addition, as we shall see below, our prior assumptions on the matrix of contemporaneous relations among the interest rates considered (i.e., the matrix $A_t(0)$) allow for the possibility that there are exchange rate risk premium shocks and money demand shocks not fully accommodated by money supply possibly resulting in large outliers in U_t .

Nonetheless, we do impose some lag-length restriction by choosing p_1 and p_2 based on the Schwarz (BIC) criterium and ex post misspecification tests on the estimated residuals. While the BIC criterion suggests to set $p_1 = 6$ and $p_2 = 1$, we find that our residuals pass all misspecifications tests using $p_1 = 2$ and $p_2 = 1$ (see below). Therefore, to save computing time, the final results are based on the shorter lag structure.

B. Identification

Identification of (54) may be achieved through exclusion restrictions on the coefficient matrix $A_t(0)$. The specific scheme used exploits the Bundesbank's presumed leading role under the

interpreted as elasticities. For interest rates we take the natural logarithm of the gross rates.

²⁵ The exchange rate gap for Italy and Spain is set to zero before the Spanish peseta joined the EMS and during the period in which the Italian lira was floating following the 1992 ERM crisis. We compute the inflation and output target also by taking deviations from the German inflation rate and by using the HP filter for the output series finding similar results available on request.

²⁶ Note however that the relative tightness of the prior distribution given on the elements of $A_t(L)$ and $B_t(L)$ distinguishes between own and other countries' monetary policy instruments (the endogenous variables), between instruments and objectives and between own and other countries' objectives (the exogenous variables).

EMS and the relative economic size of other countries.²⁷ More 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. We then assume that French and Italian interest rates affect contemporaneously the Spanish rate without being affected by it. This is plausible given that Spain's GDP was considerably smaller than that of France and Italy during much of the period considered. (Spain also joined the EMS only in 1989.) Finally, we assume that the impact on France of an increase in interest rates in Italy is the same as the impact on Italy of an increase in French rates.²⁸

Formally, we need six restrictions on $A_t(0)$ to identify the model. The assumptions above provide the six restrictions that identify the model exactly and translate into the following block recursive structure for $A_t(0)$:

$$A_t(0) = \begin{bmatrix} A_{11}(0) & 0 & 0 \\ A_{21}(0) & A_{22}(0) & 0 \\ A_{31}(0) & A_{32}(0) & A_{33}(0) \end{bmatrix}_t \quad (56)$$

where $A_{11}(0)$, $A_{31}(0)$, and $A_{33}(0)$ are scalars, $A_{21}(0)$ and $A'_{32}(0)$ are 2×1 , and $A_{22}(0)$ is 2×2 . The leader-follower behavior presumably characterizing the EMS imposes three zero restrictions in the first row of this matrix. The smaller size of Spain relative to that of France and Italy allows to impose two more zero restrictions in the last column of this matrix, while the last restriction is obtained by imposing symmetry on $A_{22}(0)$.

The structural VAR (54), therefore, can be rewritten as:

$$\begin{aligned} & \begin{bmatrix} A_{11}(0) & 0 & 0 \\ A_{21}(0) & A_{22}(0) & 0 \\ A_{31}(0) & A_{32}(0) & A_{33}(0) \end{bmatrix}_t \begin{pmatrix} R_{1t} \\ R_{2t} \\ R_{3t} \end{pmatrix} = \\ & \begin{bmatrix} A_{11}(L) & A_{12}(L) & A_{13}(L) \\ A_{21}(L) & A_{22}(L) & A_{23}(L) \\ A_{31}(L) & A_{32}(L) & A_{33}(L) \end{bmatrix}_t \begin{pmatrix} R_{1t-1} \\ R_{2t-1} \\ R_{3t-1} \end{pmatrix} + \\ & + \begin{bmatrix} B_{11}(L) & B_{12}(L) & B_{13}(L) \\ B_{21}(L) & B_{22}(L) & B_{23}(L) \\ B_{31}(L) & B_{32}(L) & B_{33}(L) \end{bmatrix}_t \begin{pmatrix} W_{1t} \\ W_{2t} \\ W_{3t} \end{pmatrix} + D_t + \begin{pmatrix} U_{1t} \\ U_{2t} \\ U_{3t} \end{pmatrix} \quad (57) \end{aligned}$$

where R_{1t} , W_{1t} and U_{1t} are the German monetary policy instrument, objectives and shock, respectively, R_{2t} , W_{2t} and U_{2t} (R_{3t} , W_{3t} , and U_{3t}) are the vectors containing the same variables for France and Italy (and Spain).

²⁷ See Giavazzi and Giovannini (1988) and Kenen (1995) on the Bundesbank's presumed leading role in the EMS from the mid-1980s onward.

²⁸ See Amisano and Giannini (1997, pages 166-67) for an example of identification by means of symmetry as assumed here.

C. Estimation

Bayesian estimation of (57) exploits its block recursive structure. Following Zha (1999), let k_j and G_j be the total number of right-hand-side variables per equation and the total number of equations in block j of (57), respectively, where the same set of variables enter the equations of each block j . If we pre-multiply (57) by the (4×4) matrix

$$A_{t,d}^{-1}(0) = \begin{bmatrix} A_{11}^{-1}(0) & \mathbf{0} & 0 \\ \mathbf{0} & A_{22}^{-1}(0) & \mathbf{0} \\ 0 & \mathbf{0} & A_{33}^{-1}(0) \end{bmatrix}_t,$$

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

$$R_{jt} = Z_{jt}\delta_{jt} + v_{jt} \quad j = 1, 2, 3 \text{ for all } t. \quad (58)$$

Here, $Z_{jt} = \text{diag} [Z_{1,t}^j, Z_{2,t}^j, \dots, Z_{G_j,t}^j]$ denotes a $(G_j \times k_j G_j)$ diagonal matrix whose elements are the $(1 \times k_j)$ vectors $(Z_{g,t}^j)$ containing all contemporaneous (in our case R_{1t} in block 2 and 3 and R_{2t} in block 3) and lagged endogenous variables, exogenous and deterministic variables of equation g in block j for $g = 1, \dots, G_j$; $\delta_{jt} = [\delta_{1,t}^j, \delta_{2,t}^j, \dots, \delta_{G_j,t}^j]$ denotes a $(k_j G_j \times 1)$ vector whose $(k_j \times 1)$ element $(\delta_{g,t}^j)$ contains the parameters of equation g in block j (for $g = 1, \dots, G_j$); and $v_{jt} = A_{t,jj}^{-1}(0) U_{jt}$.

We make the following prior assumption on v_{jt} and $A_{t,jj}^{-1}(0)$:

$$\begin{aligned} v_{jt} &\sim N(0, \Upsilon_{t,jj}), \\ E[v_{it}v_{jt}' | Z_{t-s}] &= 0 \quad \text{for } i \neq j, \text{ all } t, \text{ and } s > 0 \end{aligned}$$

with

$$\Upsilon_{t,jj} = A_{t,jj}^{-1}(0) A_{t,jj}^{-1}(0)' \text{ and } A_{t,jj}^{-1}(0) = \sqrt{\sigma_{jt}} \mathbf{A}_{jj}^{-1}(0).$$

Thus,

$$\Upsilon_{t,jj} = \sigma_{jt} \Sigma_{jj} \text{ with } \Sigma_{jj} = \mathbf{A}_{jj}^{-1}(0) \mathbf{A}_{jj}^{-1}(0)'.$$

As explained in Section IV.B, this is equivalent to assume that $v_{jt} \sim t_\nu(0, \Sigma_{jj})$, which in turn is the same as assuming that

$$v_{jt} \mid \sigma_{jt} \sim N(0, \sigma_{jt} \Sigma_{jj}) \quad (59)$$

$$\sigma_{jt} \sim \text{Inv-}\chi^2(\nu_j, 1). \quad (60)$$

This assumption is economically plausible because it allows for the possibility that there are other shocks in addition to money supply shocks, such as exchange risk premium and money demand shocks not fully accommodated by monetary authorities, that may possibly result in large outliers in U_t potentially distorting our estimates.²⁹

²⁹ By imposing an autoregressive structure on σ_{jt} (or its log) instead of assuming a simple t -student distribution for the error term we could specify a stochastic volatility model such

The other prior assumptions on the model's parameters generalize those introduced by Zellner (1971, Chapter 8) to take into account the presence of time-varying coefficients: a time-varying Minnesota prior (e.g. Doan, Litterman, and Sims, 1984) for the slope coefficients (δ_{jt}) is combined with a diffuse prior on the variance-covariance matrix of the residuals (Σ_{jj}) and the assumption made for the time variation of the error term (σ_{jt}) with prior independence.³⁰ Thus:

$$p(\delta_{jt}, \Sigma_{jj}, \sigma_{jt}) = p(\delta_{jt}) p(\Sigma_{jj}) p(\sigma_{jt}), \quad (61)$$

where

$$\begin{aligned} p(\Sigma_{jj}) &\propto |\Sigma_{jj}|^{-(G_j+1)/2}, \\ \delta_{jt} &= P_j \delta_{jt-1} + (I - P_j) \delta_{0j} + \eta_{jt} \quad \eta_{jt} \sim N(0, \Phi_j), \\ \sigma_{jt} &\sim Inv - \chi^2(\nu_j, 1). \end{aligned}$$

Here, P_j is a $(G_j k_j \times G_j k_j)$ matrix governing the law of motion of δ_{jt} , δ_{0j} is the unconditional mean of δ_{jt} , Φ_j governs the time variation of δ_{jt} , and η_{jt} is assumed to be independent from v_{jt} .

Now, denoting with R_j^T the sample data for each block j of (58), the pdf of the data, $L(R_j^T | Z_{jt}, \delta_{jt}, \Sigma_{jj}, \sigma_{jt})$, conditional on the exogenous variables, the initial observation, R_{j0} , and the parameters of the model (δ_{jt} and Σ_{jj}) is proportional (\propto) to:

$$|\Sigma_{jj}|^{-T/2} \exp \left[-\frac{1}{2} \sum_t (R_{jt} - Z_{jt} \delta_{jt})' (\sigma_{jt} \Sigma_{jj})^{-1} (R_{jt} - Z_{jt} \delta_{jt}) \right]. \quad (62)$$

The posterior distribution of Σ_{jj}^{-1} conditional on the entire history of δ_{jt} for $t = 0, \dots, T$ (denoted $\{\delta_{jt}\}$), σ_{jt} , and the data, is easily obtained combining (62) with $p(\Sigma_{jj})$ as the following Wishart distribution with T degrees of freedom and scale matrix S (Zha, 1999, p. 299):

$$\Sigma_{jj}^{-1} | \{\delta_{jt}\}, \sigma_{jt}, R_j^T, R_{j0} \sim W(T - G_j - 1, S), \quad (63)$$

where

$$S = \left[\sum_t (R_{jt} - Z_{jt} \delta_{jt}) \sigma_{jt} (R_{jt} - Z_{jt} \delta_{jt})' \right]^{-1}.$$

The joint posterior distribution of $\{\delta_{jt}\}$ conditional on Σ_{jj} , σ_{jt} and the data is obtained as discussed in section 4.A. Specifically,

$$\delta_{jt} | \Sigma_{jj}, \sigma_{jt}, R_j^T, R_{j0} \sim N(\bar{\delta}_{jt|t}, \bar{\Omega}_{jt|t}), \quad (64)$$

where $\bar{\delta}_{jt|t}$ and $\bar{\Omega}_{jt|t}$ are one-period-ahead forecast of δ_{jt} and the variance-covariance matrix of its

as that used by Cogley and Sargent (2002) or Uhlig (1992). However, the prior assumption chosen is sufficient to model temporary, non persistent shifts in the conditional variance of v_{jt} .

³⁰ See the *Normal-diffuse* case above.

mean square error, respectively, calculated by the Kalman Filter as in (26).

The *conditional* posterior distribution of σ_{jt} is similar to (43):

$$\sigma_t \mid \{\delta_{jt}\}, \Sigma_{jj}, R_j^T, R_{j0} \sim \text{Inv-}\chi^2(\nu_j + T, s_j^2) \quad (65)$$

where

$$s_j^2 = \frac{[\nu_j + (R_{jt} - Z_{jt}\delta_{jt})' \Sigma_{jj}^{-1} (R_{jt} - Z_{jt}\delta_{jt})]}{\nu_j}.$$

Given (63), (64) and (65), and initial values for P_j , Φ_j , $\hat{\Omega}_{j0}$, and $\hat{\delta}_{j0}$, the *marginal* posterior distributions of Σ_{jj}^{-1} , $\{\delta_{jt}\}$ and σ_{jt} can then be obtained from the Gibbs sampler by drawing alternately from these conditional distributions. Here, note that, given the marginal posterior distribution of σ_{jt} and Σ_{jj} , we can recover the posterior distribution of $A_{t,jj}(0)$, and thus also the posterior distribution of the structural residual U_t , since the matrices $A_{jj}(0)$ are exactly identified and thus linked to Σ_{jj} by a one-to-one mapping.

As explained in Section III.A, we define P_j , Φ_j , $\hat{\Omega}_{j0}$, $\hat{\delta}_{j0}$ in terms of few hyperparameters and then maximize the likelihood of the data as a function of this smaller parameter set to obtain numerical values that are then fed into the Gibbs sampler. More specifically, $\hat{\Omega}_{j0}$ and $\hat{\delta}_{j0}$ are exactly as $\bar{\Omega}_g$ and $\bar{\beta}_g$ in section 3.1.1, while the matrices P_j and Φ_j are defined as:

$$P_j = \text{diag}(P_{j1}, \dots, P_{jG_j})$$

$$\Phi_j = \text{diag}(\Phi_{j1}, \dots, \Phi_{jG_j}) \hat{\Omega}_{j0}$$

where $P_{jg} = \text{diag}(\pi_{8,g})$ are $(k_j \times k_j)$ matrices with $\pi_{8,g}$ controlling the coefficient of the law of motion of each element of δ_{jg} , and $\Phi_{jg} = \text{diag}(\pi_{7,g})$ are $(k_j \times k_j)$ matrices with $\pi_{7,g}$ controlling the variance around these values actually introduced in the model, for $g = 1, \dots, G_j$. Finally, for all j we set ν_j equal 5.³¹

Given the values of the model hyperparameters, (32) is run. Then the Gibbs sampler starts iterating, switching between (63), (64) and (65) and taking the estimated values of π_1, \dots, π_8 as given. The Gibbs sampler iterates 5000 times yielding 4000 draws from the joint and marginal posterior distributions of the parameters of interest after discarding the first 1000 draws. All the numerical integrations and the statistics reported are based on these last 4000 draws.

D. Results

In this section we report the estimation results: the estimated reaction function residuals and parameters derived from (57) for each country considered. The posterior median of the estimated residuals are reported in Figure 1. The posterior median of the estimated parameters together with the first and third quartile for each estimated reaction function are reported in Figures 2-5, while

³¹ Note that as the first and third blocks of the model contain only one equation, (63) becomes an inverted gamma for $j = 1$.

Figure 6 reports the posterior median of the error variance. The sample period is from January 1980 to December 1998, but the reported results are from February 1981 to December 1998 because 24 observations are used to initialize the estimation procedure. In Table 1 we also report the estimated hyperparameters (with the same notation is as in Section 3.A).

Table 1. Estimated Hyperparameters

	Germany	France	Italy	Spain
π_1	0.9782	0.9922	0.9328	0.9814
π_2	1.0	1.0	1.0	1.0
π_3	0.3039	0.2075	47.0987	0.0668
π_4	6.8168	10.659	6.3644	0.9099
π_5	363.376	8.89e-003	377848.0	136.187
π_6	0.2183	2.7086	1.19e-005	2.4612
π_7	4.52e-007	4.61e-009	3.22e-008	9.36e-006
π_8	0.9849	0.9429	0.9952	0.9673

The estimated hyperparameters are fed into the Gibbs sampler, affecting the prior assumptions actually used in estimation, and thus also our posterior estimates of the model residuals and parameters. Therefore, a few remarks are in order regarding the estimated hyperparameters in Table 1. First, without loss of generality, to economize on free hyperparameters, we assume that the tightness on the coefficients of the lagged endogenous variables (π_2) is equal to 1 (as suggested by Doan, 2000, Chapter 10). Second, time variation, as controlled by π_7 , does not appear to be a major feature of the data, as will also be discussed below. Third, the prior mean of the coefficient on the first lagged endogenous variable (π_1) is estimated to be close to one for all countries, as expected. Higher order lags, decay rather quickly in the case of Germany, France and Italy, but less so in the case of Spain (π_4). Fifth, overall parameter uncertainty, as controlled by π_6 , seems higher in the case of Spain and France than Italy and Germany, meaning that the posterior distributions should be more concentrated in the latter cases. Finally, we note that π_8 is very close to one for all countries considered, meaning that the process for the coefficient vector δ_{jt} is close to a random walk.

The estimated residuals may reflect potential model misspecifications and the model's goodness of fit. As we can see from Figure 1, they appear remarkably well behaved for all countries considered: there are essentially no outliers and there is also little or no evidence of serial autocorrelation and/or heteroschedasticity.

Table 2. White Noise Test Statistics

	Germany	France	Italy	Spain
Sample Mean	0.0053 (0.937)	0.0062 (0.927)	-0.0301 (0.659)	-0.0044 (0.948)
Cum.Period.	0.0816 (0.0981)	0.0665 (0.0981)	0.0617 (0.0981)	0.0997 (0.0981)
Q(4)	6.2035 (0.184)	1.4762 (0.831)	2.3753 (0.667)	5.3435 (0.254)
Q(8)	7.1921 (0.516)	8.3869 (0.397)	6.5074 (0.591)	12.186 (0.143)
Q(12)	7.8025 (0.800)	10.430 (0.578)	16.4072 (0.173)	23.8513 (0.021)
Arch(2)	2.7708 (0.250)	3.5593 (0.169)	7.1603 (0.029)	33.4121 (0.000)

This is also borne out by a battery of standard test statistics reported in Table 2. Table 2 reports summary and test statistics on the null hypothesis that the estimated posterior median of the residuals follow a white noise process. The first two lines report the sample mean and the p-value for the null hypothesis that this is zero, respectively. The second two lines report a Kolmogorov-Smirnov statistics for the Durbin's (1969) cumulated periodogram test.³² The following six lines report the Ljung-Box's statistics for the null hypothesis of absence of serial correlation of order higher than specified with their respective p-values. Finally, the last two lines report Engle's test for the null of absence of autoregressive conditional heteroschedasticity of second order together with their p-values in brackets.

As we can see from this table, the null hypothesis of zero mean cannot be rejected for all countries, and the absence of heteroschedasticity and autocorrelation is also clearly rejected by all statistics for Germany and France. The results for Spain, and to a lesser extent also for Italy, show some presence of ARCH effects and autocorrelation of order higher than 12, though. Nonetheless, compared to other studies one notable difference is the absence of large outliers in these residuals corresponding to the 1992 EMS crisis and the subsequent periods of financial turbulence documented by Favero and Giavazzi (2002), among others. This is because these episodes have been captured by the time-varying interest rate volatility introduced through our prior specification, whose posterior medians are reported in Figure 6.

We can now focus on the estimated model parameters. Figure 2 reports the posterior median of selected parameters of the Bundesbank's reaction function together with a band containing 50 percent of the posterior distribution. Interestingly, the estimated parameters appear rather stable

³² This test cumulates the periodogram (i.e., the squared Fourier transform) over frequencies 0 to π and scales it so as to have an end-value equal one. If the series examined is white noise, its cumulated periodogram should differ only marginally from the theoretical periodogram of a white noise, which is a straight line. Concretely, the table reports the maximum gap between the actual and theoretical cumulated periodograms together with the approximate rejection limit at 5 percent significance level in brackets.

over time, except for the structural break around the time of German unification, which does not even appear in all coefficients. This suggests that no major behavioral change took place in the Bundesbank reaction function in the run up to the monetary union (EMU).

These estimation results conform well to standard views of the behavior of the German central bank: they show a high degree of interest rate smoothing or persistence, relatively low weight on other European countries' targets and a relatively large weight on the domestic inflation target, with some weight also attached to U.S. variables. The coefficient of the German own lagged interest rate remains close to 1 throughout the sample period. Among domestic objectives, the inflation gap is by far the most important variable, although its effect is not estimated very precisely. The U.S. interest rate and the DM/U.S. dollar exchange rate also have notable impacts.³³ The coefficient on the U.S. federal funds rate, in particular, is comparable in size to that of the domestic inflation target. The coefficients of other European countries' targets are generally not significantly different from zero, except for the output gaps and the Italian exchange rate gap. But their magnitudes are quite small compared to other objectives and slightly declining over time. The coefficients of foreign inflation gaps are clearly insignificant statistically.

The parameters of the reaction function of the Bank of France also conform well to what one would expect for a follower country in the EMS (Figure 3): the German policy interest rate is the most important variable, and exchange rate targets have a smaller (and slightly declining) but significant weight. Domestic and other countries' inflation and output targets, instead, have weights which are either very small or not significant statistically. Only monetary growth appears to have had a stronger role, but this declined markedly over time.

The results for Spain, and to a lesser extent Italy, are consistent with the view that the Bank of Spain and the Bank of Italy were less constrained than the central banks of other European countries by the EMS (Figure 4 and 5). The Bank of Spain, in particular, appears to have been the least constrained among the countries considered. Both reaction functions show a much smaller and insignificant weight attached to the German contemporaneous rate and somewhat larger weights on output gaps compared to France. Interestingly, as in the case of France, the coefficient of the German exchange rate appears to have declined significantly after 1992. The reaction function of the Bank of Spain, in particular, also shows more instability, consistent with Spain's later entry in the EMS. The weight attached to monetary growth in Italy (Spain) is also smaller (larger) than that in the French reaction function. The weight on the domestic output gap affects Spanish short-term interest rates to a much greater extent than in France or even Italy, with an impact comparable to that of the German interest rate. Finally, interest rate persistence is the smallest in Spain.

Overall, these results show how difficult it would have been to choose a restricted and yet uniform econometric specification to describe the different behavior of the central banks considered. The fact that we find relatively "clean" estimated residuals then confirms that most of the behavioral differences across countries and over time are well captured by the adopted econometric

³³ Note that an increase in the exchange rate gap denotes a depreciating movement.

specification for the system of reaction functions studied.

VII. CONCLUSIONS

In this paper we have reviewed recent developments in the literature on Bayesian VARs and have discussed how to apply some of these results with an application to the estimation of a system of monetary reaction functions. Starting from the general Bayesian principle applied to the concrete case of VAR estimation, we have described several prior distributions and some useful extensions of the basic model, including nonnormal data, nonlinear models, and time-varying coefficients models. In all cases analyzed, we have provided expressions for the posterior distributions, either analytically or numerically computed. Besides its interesting retrospective economic content, the application presented showed how flexible the Bayesian approach might be in dealing with complex dynamic economic problems.

Figure 1. Reaction Function Residuals
(Standardized, January 1981-December 1998)

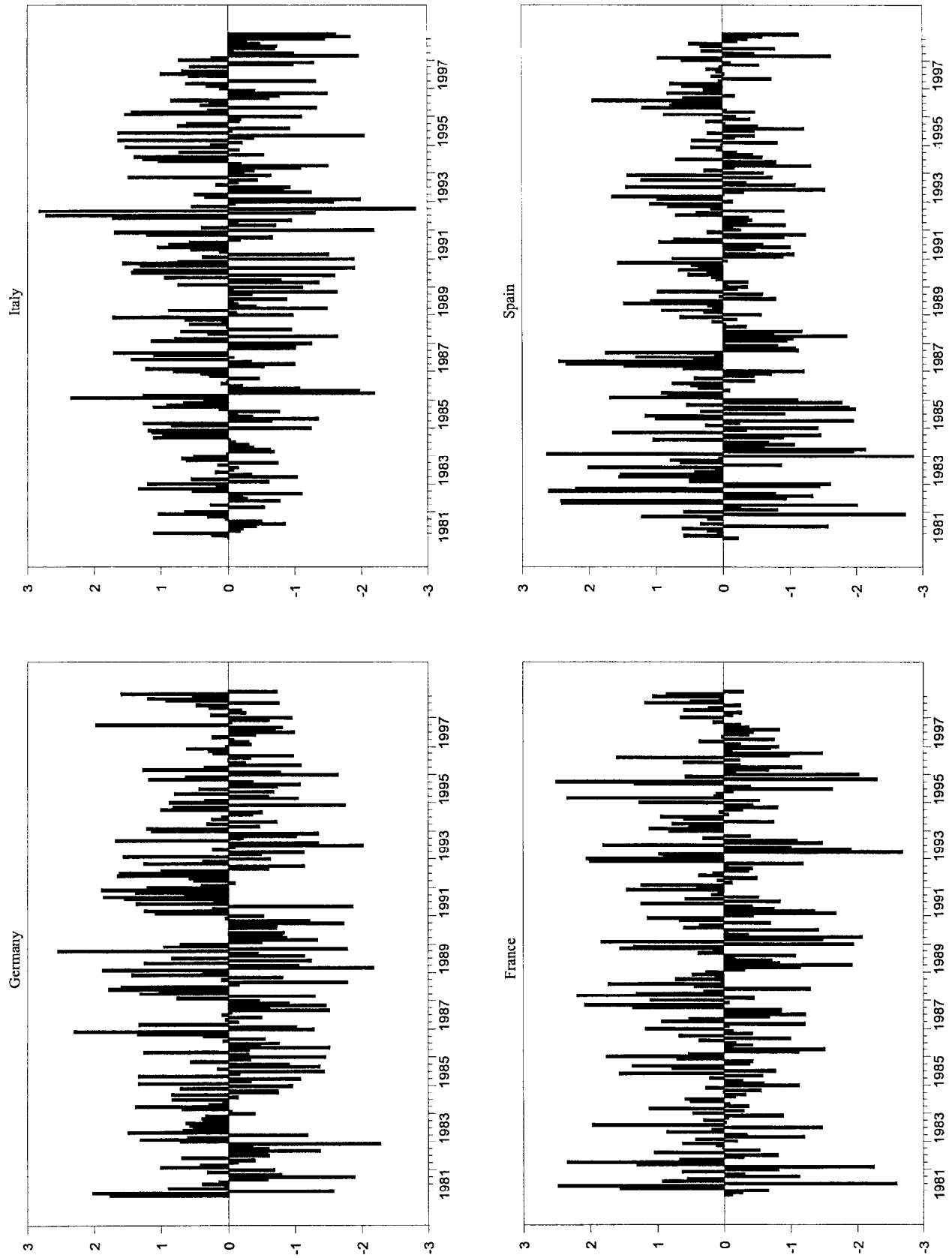


Figure 2. Germany: Reaction Function Parameters
(Elasticities, January 1981-December 1998)

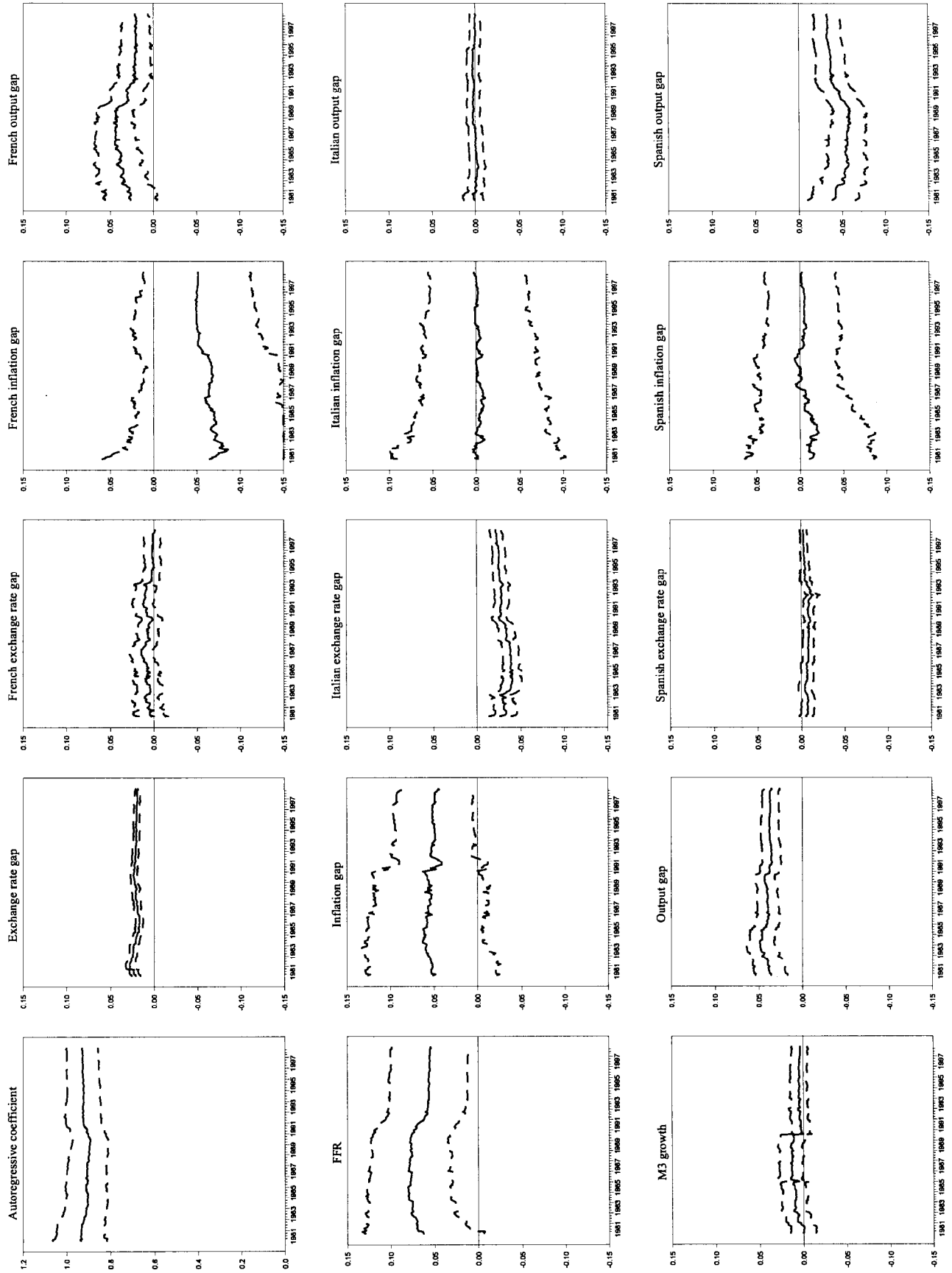


Figure 3. France: Reaction Function Parameters
(Elasticities, January 1981-December 1998)

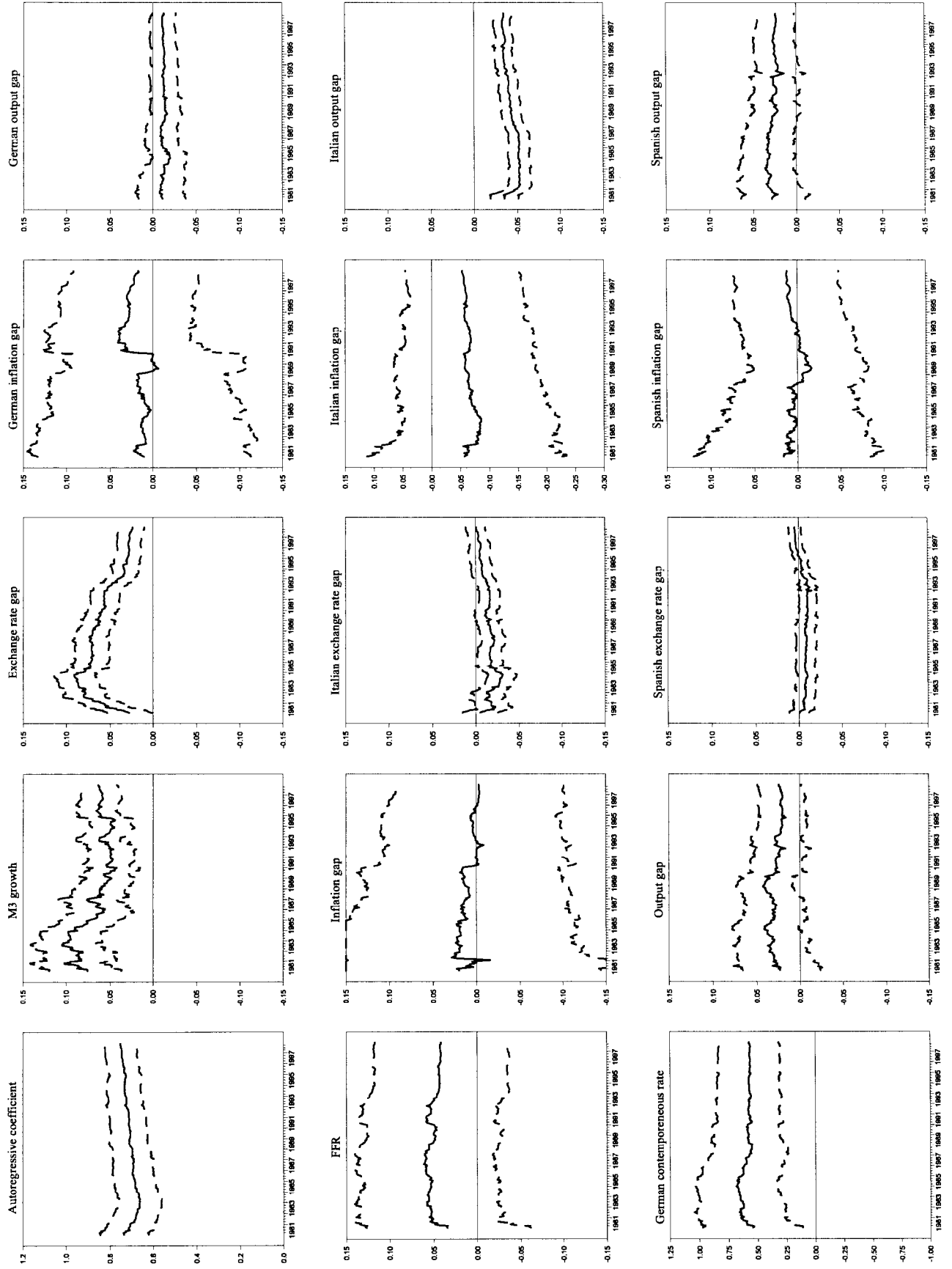


Figure 4. Italy: Reaction Function Parameters
(Elasticities, January 1981-December 1998)

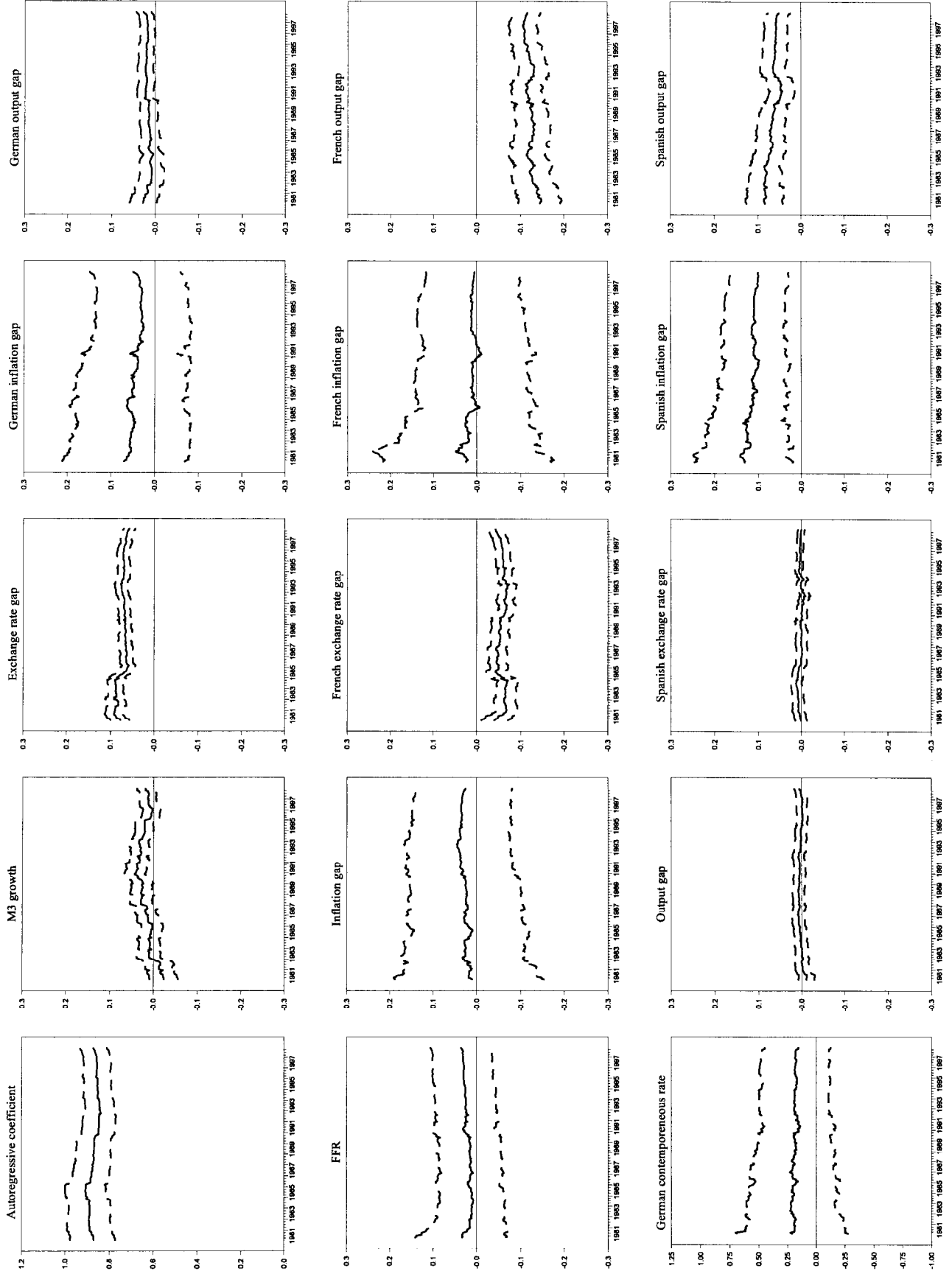


Figure 5. Spain: Reaction Function Parameters
(Elasticities, January 1981-December 1998)

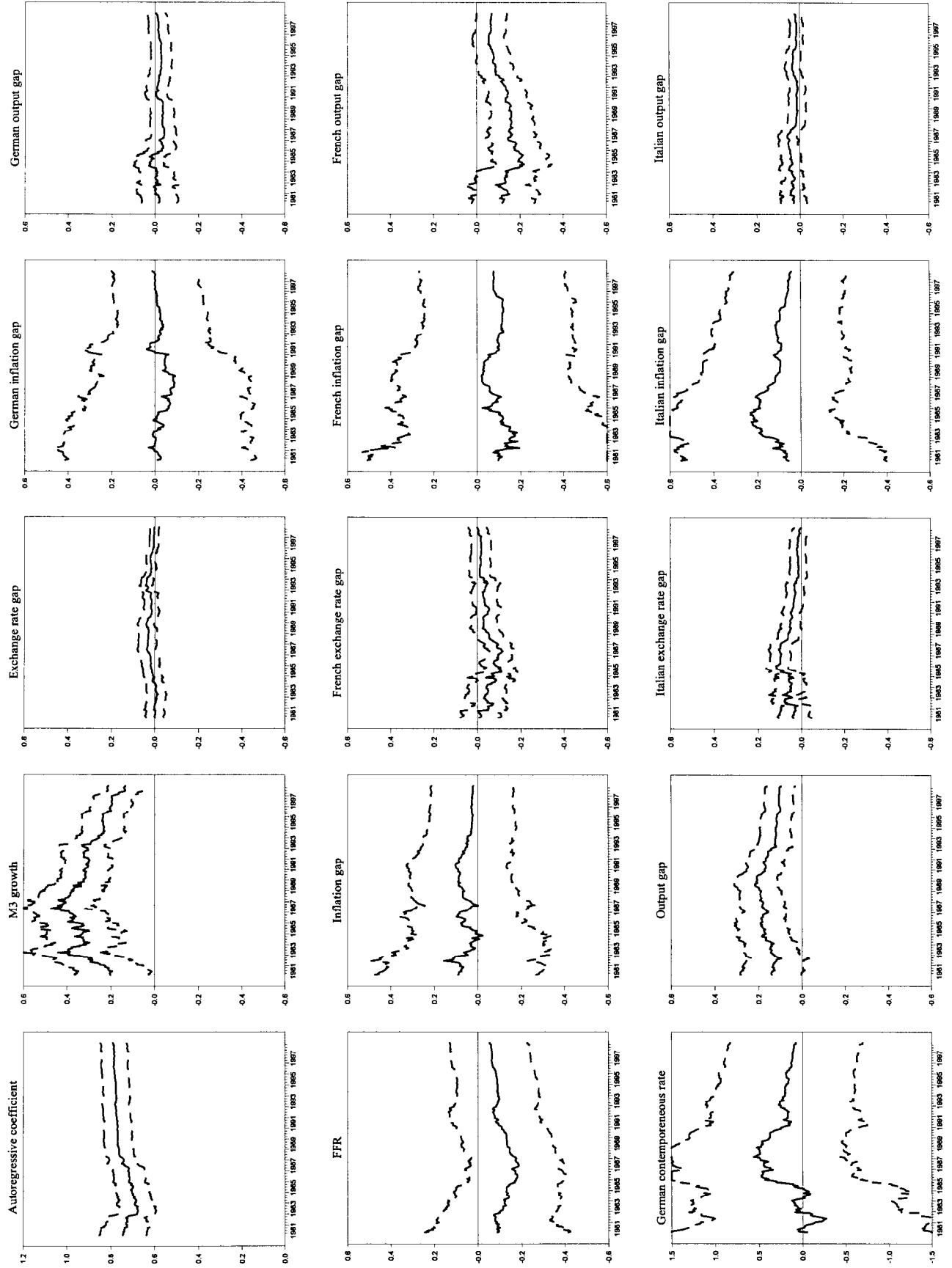
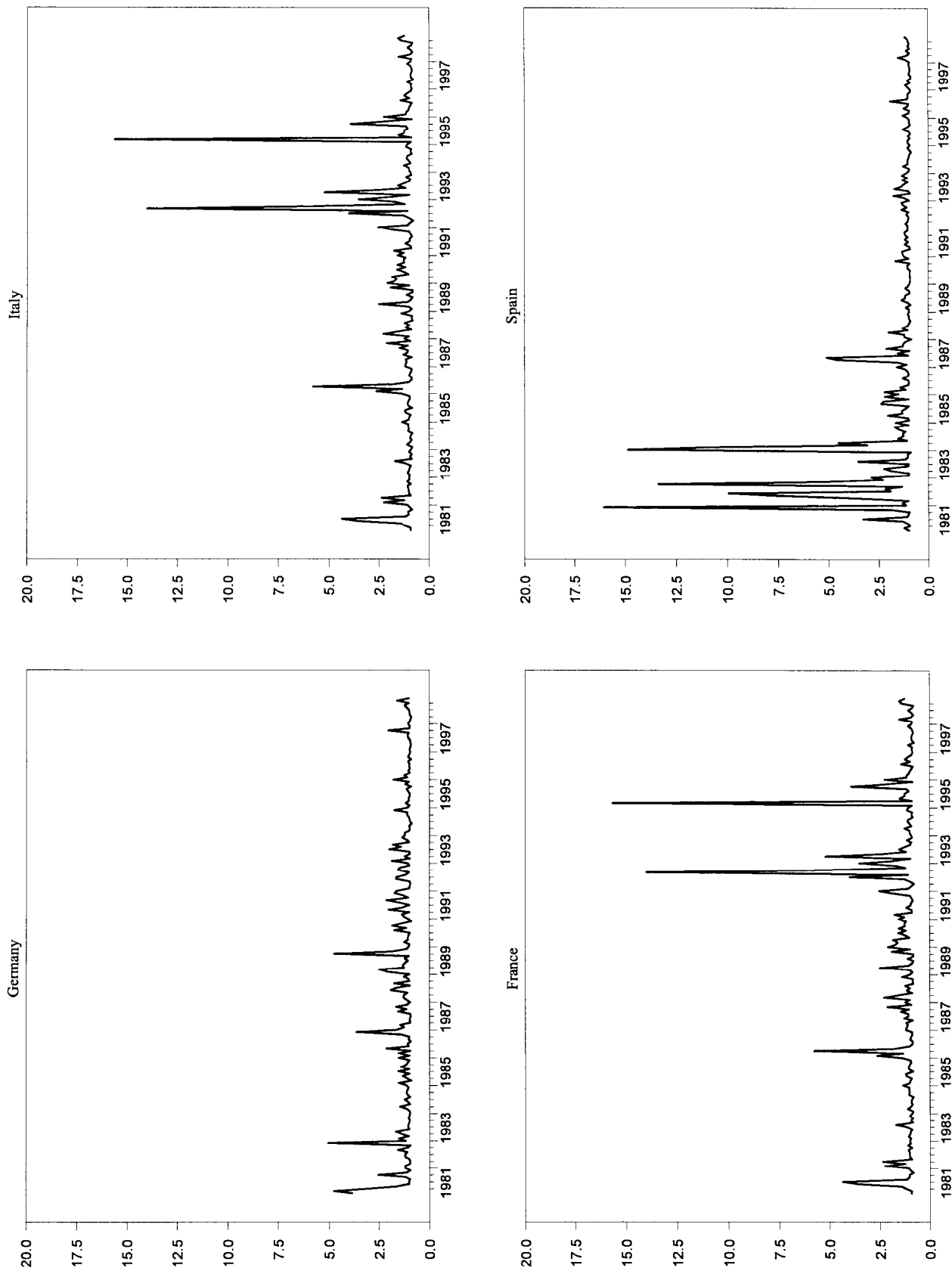


Figure 6. Interest Rate Posterior Volatility
(January 1981-December 1998)



References

- Alvarez J. L. and F.C. Ballabriga, 1994, "BVAR Models in the Context of Cointegration: A Monte Carlo Experiment," Bank of Spain Working Paper No. 9405 (Madrid).
- Amisano G. and C. Giannini, 1997, *Topics in Structural VAR Econometrics* (Berlin: Springer Verlag).
- Ballabriga F. C., M. Sebastian, and J. Vallés, 1998, "European Asymmetries," *Journal of International Economics*, Vol. 48, pp. 233-53.
- Bernanke B., and I. Mihov, 1997, "What Does the Bundesbank Target?" *European Economic Review*, Vol. 41, pp. 1025-53.
- Canova F., 1993, "Modelling and Forecasting Exchange Rates with a Bayesian Time-Varying Coefficient Model," *Journal of Economic Dynamic and Control*, Vol. 17, pp. 233-61.
- , 1995, "Vector Autoregressive Models: Specification, Estimation, Inference, and Forecasting," in *Handbook of Applied Econometrics Macroeconomics*, ed. by M.H. Pesaran and M. Wickens (Oxford: Blackwell).
- , and M. Ciccarelli, 2000, "Forecasting and Turning Point Prediction in a Bayesian Panel VAR Model," UPF Working Paper No. AD 2000-05; (Also forthcoming in the *Journal of Econometrics*).
- Carter C. K., and R. Kohn, 1994, "On Gibbs sampling for state space models," *Biometrika*, Vol. 81, No. 2, pp. 541-53.
- 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*, Vol. 68, No. 2, pp. 339-60.
- Chib S., F. Nardari and N. Shepard, 2001a, "MCMC for Stochastic Volatility Models" (unpublished; Oxford: Nuffield College).
- , 2001b, "Analysis of High Dimensional Multivariate Stochastic Volatility Models" (unpublished; Oxford, Nuffield College).
- Ciccarelli M. and A. Rebucci, 2001, "The Transmission Mechanism of European Monetary Policy: Is There Heterogeneity: Is It Changing over Time?," Bank of Spain Working Paper No. 0115 (Madrid) (Also issued as IMF Working Paper No. 02/54).

- , 2002, “Measuring Shift-Contagion with a Time-Varying Coefficient Model” (unpublished; Washington: International Monetary Fund and University of Alicante, Alicante, Spain).
- Clarida R., J. Gali, and M. Gertler, 1997, “Monetary Policy Rules in Practice. Some International Evidence,” *European Economic Review*, Vol. 42, pp. 1033-67.
- Cogley T. and T.J. Sargent, 2002, “Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.” (unpublished; Arizona State University).
- Doan T., R. Litterman, and C. Sims, 1984, “Forecasting and Conditional Projections Using Realist Prior Distributions,” *Econometric Review*, Vol. 3, No. 1, pp. 1-100.
- Doan T., 2000, *Rats manual, Version 5*.
- Dornbusch R., C. Favero, and F. Giavazzi, 1998, “A Red Letter Day,” CEPR Discussion Paper No. 1804. Published also as “The Immediate Challenge for the European Central Bank,” *Economy Policy* Vol. 29 pp. 17-64.
- Durbin J., 1969, “Tests for Serial Correlation in Regression Analysis Based on the Periodogram of Least Squares Residuals,” *Biometrika*, Vol. 59, pp. 1-16.
- Engle R., 1982, “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation,” *Econometrica*, Vol. 50, pp. 987-1008.
- Favero C., and F. Giavazzi, 2002, “Is the International Propagation of Financial Shocks Non-Linear? Evidence from the ERM” *Journal of International Economics*, Vol. 57, No.1, pp. 231-46.
- 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*, Vol. 85, pp. 972-85.
- Gelman A., J.B. Carlin, H.S. Stern, and D.B Rubin, 1995, *Bayesian Data Analysis* (London: Chapman & Hall).
- Geweke J., 2000, “Simulation Based Bayesian Inference for Economic Time Series,” in *Simulation Based Inference in Econometrics: Methods and Applications*, ed. by Mariano, R. Shermann, and M. Weeks (Cambridge: Cambridge University Press).
- Giavazzi F. and A. Giovanni, 1988, “The Role of Exchange-Rate Regime in a Disinflation,” in *The European Monetary System*, ed. by F. Giavazzi, S. Micossi, and M. Miller (Cambridge: Cambridge University Press).

Hamilton J.D., 1989, "A New Approach to the Economic Analysis of Non-Stationary Time Series and the Business Cycle, *Econometrica*, No. 57, Vol. 2, pp. 357-84.

———, 1994, *Time Series Analysis* (Princeton: Princeton University Press).

Hsiao C., M. H. Pesaran, and A. K. Tahmiscioglu, 1999, "Bayes Estimation of Short Run Coefficients in Dynamic Panel Data Models," in *Analysis of Panels and Limited Dependent Variable Models: In Honor of G.S. Maddala*, ed. By C. Hsia and others (Cambridge: Cambridge University Press).

Kadiyala K. R., and S. Karlsson, 1997, "Numerical Methods For Estimation and Inference in Bayaesian VAR Models," *Journal Applied Econometrics*, Vol. 12, pp. 99-132.

Kenen P.B., 1995, *Economic and Monetary Union in Europe: Moving Beyond Maastricht* (Cambridge: Cambridge University Press).

Kim S., 1994, "Dynamics Linear Models with Markov Switching," *Journal of Econometrics*, Vol. 60, pp. 1-22.

Koop G., 1992, "Aggregate Shocks and Macroeconomic Fluctuations: A Bayesian Approach," *Journal of Applied Econometrics*, Vol. 7, pp. 395-411.

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. Dissertation (Minneapolis).

———, 1986, "Forecasting with Bayesian Vector Autoregressions: Five Years of Experience," *Journal of Business and Economic Statistics*, Vol. 4, pp. 25-38.

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

Sims C., 1972, "Money, Income and Causality," *American Economic Review*, No. 62, Vol. 1, pp. 540-53.

———, 1980, "Macroeconomics and Reality," *Econometrica*, Vol. 48, pp. 1-48.

———, and T. Zha, 1998, "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, Vol. 39, pp. 949-79.

———, (1999), "Drifts and Breaks in Monetary Policy" (unpublished; Princeton University)

- Uhlig H., (1997), "Bayesian Vector Autoregressions with Stochastic Volatilities", *Econometrica*, Vol. 65, No. 1., pp. 59-73.
- Waggoner, D.F., and T. Zha, 1998, "Conditional Forecasts in Dynamic Multivariate Models," *Federal Reserve Bank of Atlanta*, Working Paper pp. 98-22.
- Zha T. 1999, "Block Recursion and Structural Vector Autoregressions," *Journal of Econometrics*, Vol. 90, pp. 291-316.
- Zeller A., 1971, *An Introduction to Bayesian Inference in Econometrics* (New York: Wiley).