ESTIMATING MULTICOUNTRY VAR MODELS*

BY FABIO CANOVA AND MATTEO CICCARELLI1

Universität Bern, ICREA-UPF, and CEPR, Switzerland; European Central Bank, Germany

This article presents a method to estimate the coefficients, to test specification hypotheses, and to conduct policy exercises in multicountry VAR models with cross-unit interdependencies, unit-specific dynamics, and time variations in the coefficients. The framework of analysis is Bayesian: A prior flexibly reduces the dimensionality of the model and puts structure on the time variations; Markov chain Monte Carlo (MCMC) methods are used to obtain posterior distributions; and marginal likelihoods to check the fit of various specifications. Impulse responses and conditional forecasts are obtained with the output of an MCMC routine. The transmission of certain shocks across countries is analyzed.

1. INTRODUCTION

There has been a growing interest over the last decade in using multicountry VAR models for applied macroeconomic analysis (see e.g., Canova and Marrinan, 1998; Canova and De Nicolo’, 2000; Del Negro and Obiols, 2001, among others). Problems concerning the transmission of shocks across countries, sectors, or industries; issues related to income convergence and the evaluation of the regional policies; and questions having to do with the composition of portfolio of assets, the contagion of financial crises, and globalization are naturally studied in this framework.

A multicountry setup differs from a multiagent framework for several reasons. First, cross-unit lagged interdependencies are likely to be important in explaining the dynamics of multicountry data. Second, heterogeneous dynamics are a distinctive feature of multicountry time series data (see e.g., Canova and Pappa, 2007). Third, the number of cross sectional units is generally limited and the time series dimension is of moderate size. These latter two features make the inferential problem nonstandard. For example, the GMM estimator of Holtz-Eakin et al. (1988),

* Manuscript received January 2006; revised March 2007; final revision February 2008.
1 We would like to thank three anonymous referees, F. Schorfheide (the editor), G. Amisano, T. Cogley, H. Van Dijk, D. Hendry, R. Paap, T. Trimbur, T. Zha and the seminar participants at Erasmus University, University of Stockholm, Universitat Autonoma de Barcelona, Tilburg University, the ECB, the Federal Reserve Bank of Atlanta; the Bank of Italy Conference “Monitoring Euro Area Business Cycle”; the Macro, Money and Econometric study group, London; the conference “Common Features” in Rio de Janeiro; the Meetings of the French Statistics Association, Bruxelles, for comments and suggestions. The view expressed in this article are exclusively those of the authors and not those of the European Central Bank. The first author gratefully acknowledge the financial support of CREI. An earlier version of this article has circulated under the title: Panel index VAR models: Specification, Estimation, Testing and Leading Indicators. Please address correspondence to: Matteo Ciccarelli, Department of DG Research, European Central Bank, Kaisenstrasse 29, Frankfurt am Main, Frankfurt 60311, Germany. Phone: +49 69 13 44 87 21. Fax: +49 69 13 44 65 75. E-mail: matteo.ciccarelli@ecb.int.

the QML, and a minimum distance estimators of Binder et al. (2005), all of which are consistent as the cross section dimension becomes large, or the mean group estimator of Pesaran and Smith (1996), which is consistent as the time series dimension becomes large, are inapplicable. Finally, although estimation of time-varying structures is feasible with a large homogeneous cross section, the combination of heterogenous dynamics and short cross sections makes it difficult to exploit cross sectional information to estimate time series variations in multicountry setups.

When dealing with multicountry data, the empirical literature has taken a number of short cuts and neglected some or all of these problems. For example, it is typical to assume that slope coefficients are common across (subsets of the) units (see e.g., Fatas and Mihov, 2006); that there are no lagged interdependencies across units (see Dees et al., 2005); that the structural relationships are stable over arbitrary samples and that asymptotics in $T$ apply (see Imbs et al., 2005); or a combination of all of these. None of these restrictions is appealing: Short time series are, in part, the result of new definitions and of the adaptation of international standards to data collection in developing countries; unit-specific relationships may reflect differences in national regulations or policies; interdependencies result from world markets integration and time instabilities from evolving macroeconomic structures.

This article shows how to conduct inference in multicountry VARs featuring time series of moderate length and, potentially, unit-specific dynamics, lagged interdependences and structural time variations. Since these last three features make the number of coefficients of the model large, we take a flexible Bayesian viewpoint to estimation, and weakly restrict the coefficients to depend on a low-dimensional vector of time-varying factors. These factors capture, for example, coefficient variations that are common across units and variables (a “common” effect); variations that are specific to the unit (a “unit” effect), variations that are specific to a variable (a “variable” effect), etc. We complete the specifications using a hierarchical structure that allows for time variations in the factors and exchangeability in the unit effects.

We employ Markov chain Monte Carlo (MCMC) methods to compute exact finite sample distributions of the quantities of interest and describe how MCMC draws can be used to compute responses to unexpected perturbations in the innovations of either the VAR or the factors, and conditional forecasting experiments, featuring displacements of certain blocks of variables from their baseline path—two exercises of great interest in policy circles. We employ the marginal likelihood to examine hypotheses concerning the importance of lagged interdependences and of time variations, and to evaluate other important specification choices.

The factor structure we employ effectively transforms the overparametrized multicountry VAR into a parsimonious SUR model, where the regressors are linear combinations of the right-hand-side variables of the VAR, the loadings are the time-varying factors, and the forecast errors feature a particular heteroschedastic structure. Such a reparametrization has, at least, two appealing features. First, it reduces the problem of estimating a large number of, possibly, unit-specific and time-varying coefficients into the problem of estimating a smaller number.
MULTICOUNTRY VAR MODELS

of loadings on certain combinations of the right-hand-side variables of the VAR. Therefore, computational costs are limited. Second, since the regressors of the SUR model are observable linear combinations of the right-hand-side variables of the VAR, we produce an estimable structure that is suitable for a variety of policy purposes. For example, one can produce multistep multicountry leading indicators (see Anzuini et al., 2005); recursively estimate coincident indicators of world and national business cycles and examine their time variations (see Canova et al., 2007); construct measures of medium-term core inflation or medium-term conditional and unconditional forecasts; or examine the propagation of shocks across countries (Caivano, 2006).

Our reparametrized model shares some similarities with those used in factor model literature (see e.g., Stock and Watson, 1989; Forni and Reichlin, 1998, or Otrok and Whiteman, 1998), but also has important differences. In fact, although the factor structure in this literature emerges from the desire to obtain the main drivers of the variability of a large set of variables, here it is the results of flexible restrictions imposed on the coefficients. As a consequence, the regressors of our SUR model are observable unweighted combinations of lags of the VAR variables capturing low-frequency comovements in the data whereas those in factor models are estimated weighted combinations of the current endogenous variables and are designed to best capture their variability.

Canova and Ciccarelli (2004) proposed a structure to forecast with multicountry VAR models, which allows for unit-specific dynamics and time variations. There the estimation process is computationally demanding since time variations are different across variables and units. Relative to that paper we innovate by providing (a) a flexible coefficient factorization that renders estimation easy, (b) a testing approach that makes model selection and inference tractable, (c) a set of tools to conduct structural analyses and policy-projection exercises.

The structure of the article is as follows: The next section presents the model; Section 3 discusses estimation and inference; Section 4 deals with model selection; and Section 5 with impulse responses and conditional forecasts. In Section 6, an application is presented. Section 7 concludes.

2. THE MODEL

The multicountry VAR model we consider has the form

\[ y_{it} = D_{it}(L)Y_{t-1} + C_{it}(L)W_{t-1} + e_{it}, \]

where \( i = 1, \ldots, N; t = 1, \ldots, T; y_{it} \) is a \( G \times 1 \) vector of variables for each \( i \), \( Y_t = (y_{1t}', y_{2t}', \ldots, y_{Nt}')' \), \( D_{it} \) are \( G \times GN \) matrices and \( C_{it} \) are \( G \times q \) matrices for each \( j \), \( W_t \) is a \( q \times 1 \) vector that may include unit-specific, time-invariant variables (for example, a vector of ones) or common variables (for example, oil prices), and \( e_{it} \) is a \( G \times 1 \) vector of random disturbances. We assume that there are \( p_1 \) lags for each of the \( G \) endogenous variables and \( p_2 \) lags for the \( q \) variables in \( W_t \). In (1), cross-unit lagged interdependencies exist whenever the matrix \( D_{it}(L) \neq I \otimes D_{ij}(L) \) for some \( L \), where \( I \) is a \( 1 \times N \) vector with one in the \( i \)th position and zero elsewhere.
words, if we stack the elements of $D_{it,j}$ over $i$, we obtain a matrix that is not block diagonal for at least one $j$. This feature adds flexibility to the specification but it is costly: The number of coefficients, in fact, is increased by a factor $N$ (we have $k = NGp_1 + qp_2$ coefficients in each equation). In (1), the dynamic relationships are allowed to be unit specific and the coefficients could vary over time. Let $\delta_{it}$ be $k \times 1$ vectors containing, stacked, the $G$ rows of the matrices $D_{it}$ and $C_{it}$; define $\delta_{it} = (\delta_{i1}, \ldots, \delta_{iG})^T$ be a $NG \times 1$ vector. Whenever $\delta_{it}$ is unrestricted, it is impossible to estimate it—there are more coefficients than data points. To solve this problem, we adopt a flexible structure where $\delta_{it}$ is factored as

$$
\delta_{it} = \sum_{f} \Xi_f \theta_{ft} + u_t,
$$

where $F << NGk; \theta_{ft}$ is a low-dimensional vector, $\forall f$, $\Xi_f$ are conformable matrices and $u_t$ captures unmodeled and idiosyncratic variations present in $\delta_{it}$. The typology of the factors $\theta_{ft}$ and the exact form of the $\Xi_f$’s will become evident from the examples presented below.

Clearly, the choice of factorization is application and, possibly, sample dependent. Although the selection of the type of factors is typically dictated by the needs of the investigation, its exact numbers is often a matter of choice. For example, in a cross-country study of business cycle transmissions, common and country-specific factors are probably sufficient although, when constructing indicators of GDP, one may want to specify, at least, a common, a country-specific, and a variable-specific factor. A simple procedure to determine the number of factors and to verify other specification choices, trading-off the fit of the model with the size of $F$, is in Section 4. Note also that in (2) all factors are permitted to be time-varying. Time invariant structures can be obtained via restrictions on their law of motion, as detailed below.

If we let $X_t = I_{NG} \otimes X_t'$; where $X_t = (Y_{t-1}', Y_{t-2}', \ldots, Y_{t-p}', W_{t}', \ldots, W_{t-l}')$; set $X_t = X_t ; \Xi = [\Xi_1, \Xi_2, \Xi_3, \ldots, \Xi_F]$; $\zeta_t = X_t u_t + E_t$, and let $Y_t, E_t$ be $NG \times 1$ vectors, we can rewrite (1) as

$$
Y_t = X_t\delta_t + E_t
= X_t(\Xi_b \theta_t + u_t) + E_t = X_t' \theta_t + \zeta_t.
$$

In (3), we have reparametrized the original multicountry VAR so that the vector of endogenous variables depends on a small number of observable indices, $X_t', \zeta_t$, and the factors $\theta_{ft}$ load on the indices. By construction, the $X_t'$’s are linear combinations of right-hand-side variables of the multicountry VAR; are correlated among each other—the correlation decreases as $G$ or $N$ or $p = max[p_1, p_2]$ increase; and emphasize comovements across lagged variables.

2.1. Examples. To illustrate what our approach implies for different DGPs, we study three examples.
2.1.1. A two-country VAR. The first example we consider is a two-country, \( i = 2 \), two-variable \( g = 2 \), VAR(2)

\[
\begin{pmatrix}
  y_{11t} \\
  y_{12t} \\
  y_{21t} \\
  y_{22t}
\end{pmatrix} =
\begin{pmatrix}
  A_{1111} & A_{1112} & A_{1121} & A_{1122} \\
  A_{1211} & A_{1212} & A_{1221} & A_{1222} \\
  A_{2111} & A_{2112} & A_{2121} & A_{2122} \\
  A_{2211} & A_{2212} & A_{2221} & A_{2222}
\end{pmatrix}
\begin{pmatrix}
  y_{11t-1} \\
  y_{12t-1} \\
  y_{21t-1} \\
  y_{22t-1}
\end{pmatrix} +
\begin{pmatrix}
  B_{1111} & B_{1112} & B_{1121} & B_{1122} \\
  B_{1211} & B_{1212} & B_{1221} & B_{1222} \\
  B_{2111} & B_{2112} & B_{2121} & B_{2122} \\
  B_{2211} & B_{2212} & B_{2221} & B_{2222}
\end{pmatrix}
\begin{pmatrix}
  y_{11t-2} \\
  y_{12t-2} \\
  y_{21t-2} \\
  y_{22t-2}
\end{pmatrix} +
\begin{pmatrix}
  \epsilon_{11t} \\
  \epsilon_{12t} \\
  \epsilon_{21t} \\
  \epsilon_{22t}
\end{pmatrix}.
\]

Let \( \delta = (\text{vec}(A)'', \text{vec}(B)')' \) be the \( 32 \times 1 \) vector of parameters. We specify four factors for \( \delta, \) i.e., \( \delta_{k,g,i,j} = \theta_{1k} + \theta_{2i} + \theta_{3g} + \theta_{4j} \) where \( \theta_1 = (\theta_{11}, \ldots, \theta_{14}) \) is a \( 4 \times 1 \) vector defining the equation where a coefficient belongs, \( \theta_2 = (\theta_{21}, \theta_{22}) \) is a \( 2 \times 1 \) vector of country-specific factors, \( \theta_3 = (\theta_{31}, \theta_{32}) \) is a \( 2 \times 1 \) vector of variable-specific factors and \( \theta_4 = (\theta_{41}, \theta_{42}) \) is a \( 2 \times 1 \) vector of lag-specific factors. Letting

\[
\begin{pmatrix}
  i_1 \\
  0
\end{pmatrix} = (1, 1, 1, 1)' ,
\begin{pmatrix}
  i_2 \\
  0
\end{pmatrix} = (1, 1, 0, 0)' ,
\begin{pmatrix}
  i_3 \\
  0
\end{pmatrix} = (0, 0, 1, 1)' ,
\begin{pmatrix}
  i_4 \\
  0
\end{pmatrix} = (1, 0, 1, 0)' ,
\begin{pmatrix}
  i_5 \\
  0
\end{pmatrix} = (0, 1, 0, 1)' ,
\]

then

\[
\delta =
\begin{pmatrix}
  i_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & i_2 & i_3 & i_4 & i_5 & i_1 & 0 \\
  0 & i_1 & 0 & 0 & 0 & 0 & 0 & 0 & i_2 & i_3 & i_4 & i_5 & i_1 & 0 \\
  0 & 0 & i_1 & 0 & 0 & 0 & 0 & 0 & i_2 & i_3 & i_4 & i_5 & i_1 & 0 \\
  0 & 0 & 0 & i_1 & 0 & 0 & 0 & 0 & i_2 & i_3 & i_4 & i_5 & i_1 & 0 \\
  i_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & i_2 & i_3 & i_4 & i_5 & i_1 & 0 \\
  0 & i_1 & 0 & 0 & 0 & 0 & 0 & 0 & i_2 & i_3 & i_4 & i_5 & i_1 & 0 \\
  0 & 0 & i_1 & 0 & 0 & 0 & 0 & 0 & i_2 & i_3 & i_4 & i_5 & i_1 & 0 \\
  0 & 0 & 0 & i_1 & 0 & 0 & 0 & 0 & i_2 & i_3 & i_4 & i_5 & i_1 & 0
\end{pmatrix}
\]

which implies, for example, that the first equation of the VAR is reparametrized as

\[
\begin{pmatrix}
  y_{11t} \\
  y_{12t}
\end{pmatrix} =
\begin{pmatrix}
  A_{1111} & A_{1112} & A_{1121} & A_{1122} \\
  A_{1211} & A_{1212} & A_{1221} & A_{1222} \\
  A_{2111} & A_{2112} & A_{2121} & A_{2122} \\
  A_{2211} & A_{2212} & A_{2221} & A_{2222}
\end{pmatrix}
\begin{pmatrix}
  \theta_{11} \\
  \theta_{12} \\
  \theta_{21} \\
  \theta_{22}
\end{pmatrix}
\begin{pmatrix}
  y_{11t-1} \\
  y_{12t-1} \\
  y_{21t-1} \\
  y_{22t-1}
\end{pmatrix} +
\begin{pmatrix}
  \epsilon_{11t} \\
  \epsilon_{12t} \\
  \epsilon_{21t} \\
  \epsilon_{22t}
\end{pmatrix}.
\]

Therefore, \( X_{1t} \) captures the information contained in the lags of all the variables of the model, \( X_{2t} (X_{3t}) \) captures the information contained in the lags of the variables for country 1 (country 2), \( X_{4t} (X_{5t}) \) captures the information contained in the
lags of variable 1 (variable 2) and lags, whereas $X_6(X_7)$ captures the information contained in the first (second) lag, across countries and variables.

2.1.2. A DSGE model. Consider a log-linearized DSGE model of the form

$$y_{1t} = A(\beta) y_{1t-1} + B(\beta) \varepsilon_t$$ (7)

$$y_{2t} = C(\beta) y_{1t},$$ (8)

where $\beta$ are structural parameters, $A(\beta)$, $B(\beta)$, and $C(\beta)$ are time-invariant matrices whose entries are nonlinear functions of $\beta$; $y_{1t}$ is a state and $y_{2t}$ a control, both of them are assumed to be scalar, for simplicity. The dimension of $\varepsilon_t$ is typically smaller than the dimension of $y = [y_{1t}, y_{2t}]$ and there may be cross equations restrictions in the sense that $\beta_m, m = 1, 2, \ldots$ may appear in several of the entries of $A$, $B$, and $C$. Equations (7) and (8) can be written as a structural VAR(1) model

$$\begin{pmatrix} I \\ I \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} A(\beta) \\ 0 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} B(\beta) \\ 0 \end{pmatrix} \varepsilon_t$$

or, letting $D_1(\beta) = C(\beta)A(\beta)$ and $D_2(\beta) = C(\beta)B(\beta)$, as a factor model

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} A(\beta) \\ D_1(\beta) \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ D_2(\beta) \end{pmatrix} \varepsilon_t.$$

Consider a reduced-form VAR for $y_t = (y_{1t}, y_{2t})$ of the form $y_t = H_{y_t-1} + \varepsilon_t$ and assume that

$$\delta = \begin{pmatrix} H_{11} \\ H_{12} \\ H_{21} \\ H_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \theta_1 + \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \theta_2 = \Sigma_1 \theta_1 + \Sigma_2 \theta_2,$$

where $\theta_s$ has two components each $s = 1, 2$. Then the VAR is

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \theta_{11} + \theta_{21} & \theta_{11} \\ \theta_{12} & \theta_{12} + \theta_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

and its SUR reparametrization is

$$y_{1t} = \theta_{11}(y_{1t-1} + y_{2t-1}) + \theta_{21} y_{1t-1} + \varepsilon_{1t}$$ (9)

$$y_{2t} = \theta_{12}(y_{1t-1} + y_{2t-1}) + \theta_{22} y_{2t-1} + \varepsilon_{2t}.$$ (10)
Here \((y_{1t-1} + y_{2t-1})\) plays the role of a common index.

When \(H_{12}\) and \(H_{22}\) are zero, as the theory implies, \(\theta_{11} = 0\); \(-\theta_{22} = \theta_{12}\) and the model correctly recognizes that there is a factor of proportionality between the two types of equations of the system.

2.1.3. A variance component model. The model we consider here is of the form

\[
y_{it} = \alpha_{it} + T_{it} (1 - \rho_t L)T_{it} = e_{it}
\]

\[
\alpha_{it} = \alpha_i + v_{it} (1 - \omega_t L)v_{it} = \zeta_{it}
\]

\[
\alpha_i = \alpha_0 + \epsilon,
\]

where \(e_i\) is i.i.d. across \(t\), \(v_{it}\) is i.i.d. across \(t\), and \(y_{it}\) is a \(G \times 1\) vector for each \(i = 1, 2, \ldots, N\). This model has the following VAR representation:

\[
Y_t = \alpha^\ast_{0t} + A_t Y_{t-1} + B_t Y_{t-2} + \eta_t
\]

\[
= \alpha^\ast_{0t} + \delta_t X_t + \eta_t,
\]

where \(Y_t\) is a \(NG \times 1\) vector each \(t\), \(\alpha^\ast_{0t} = \text{diag}\{(1 - \omega_i)\}(1 - \rho_t)\alpha_0, \eta_{it} = (1 - \omega_t L)e_t + (1 - \omega_t L)(1 - \rho_t L)e + (1 - \rho_t L)\zeta_{it}\) whereas \(A_t = \rho_t + \omega_t\) and \(B_t = \rho_t \omega_t\). Therefore, an error component model generates a particular error structure in the VAR. Note that \(\alpha^\ast_{0t}\) are time trends common to all the \(G\) variables for unit \(i\). Suppose \(\delta_t = \text{vec}(A_t), \text{vec}(B_t)\) is factored as

\[
\delta_{itg} = \Xi_1 \theta_{1t} + \Xi_2 \theta_{2t} + \Xi_3 \theta_{3g} + u^\delta_{itg},
\]

where \(\theta_{1t}\) is a \(T \times 1\) vector of time effects (common to all \(g = \text{variable}, i = \text{country}, j = \text{lag}\)), \(\theta_{2t}\) is \(N \times 1\) vector of unit-specific effects (common to all \(j, g\)), \(\theta_{3g}\) is \(G \times 1\) vector of variables-specific effects (common to all \(j, i\)). As for \(\alpha^\ast_{0t}\) we assume

\[
\alpha^\ast_{0t} = \Xi_4 \theta_{4it} + u^\alpha_{0it},
\]

where \(\theta_{4it}\) is an \(NT \times 1\) vector. Equations (14) and (15) represent a version of the model of Canova and Ciccarelli (2004). Here the number of parameters to be estimated is \(NT + T + N + G\), which is still relatively large. To further reduce the dimensionality of the parameter vector one could make \(\theta_{4it}\) time- or unit-independent and exploit averages in the remaining dimensions to construct the appropriate regressors. Disregarding how \(\alpha^\ast_{0t}\) is parametrized, the SUR model is

\[
(Y_t - \alpha^\ast_{0t}) = \theta_{1t} \xi_1 + \theta_{2t} \xi_2 + \theta_{3g} \xi_3 + \zeta_t,
\]

where \(\xi_1 = \Xi_1 X_t\) is a time index, \(\xi_2 = \Xi_1 X_t\) is a unit index, \(\xi_3 = \Xi_3 X_t\) is a variable index, and \(\zeta_t\) is composite error whose variance depends on time, on the
unit, on the variable, and on the lag. Hence, the reparametrization maintains the original error component structure but somewhat reduces the dimensionality of the parameters space.

2.2. Discussion and Relationship with the Literature. One advantage of our flexible coefficient factorization is that the overparametrization of the original multicountry VAR is dramatically reduced. In fact, in the resulting SUR model, estimation and specification searches are constrained only by the dimensionality of $\theta_t$ ($\delta_t$ is integrated out). A second advantage is that, given the MA nature of many $X_{it}$, the regressors of (3) capture low-frequency comovements present in the lags of the VAR. Since the model averages out not only cross section but also time series noise, reliable and stable estimates of $\theta_t$ can potentially be obtained, and this makes the framework useful for a variety of medium-term policy analyses exercises. A third advantage is that (3) has some economic content. For example, if $\theta_{1t}$ captures information that is common to all the coefficients of the VAR, $X_{1t}\theta_{1t}$ is an indicator for $Y_t$ based on common information. Indicators containing other types of information can also be easily constructed. Since $X_{it}$ are predetermined, leading versions of these indicators can be obtained projecting $\theta_t$ on the information available at $t - \tau$, $\tau = 1, 2, \ldots$.

Some commentators have argued that the equal (and exogenous) weights that (2) imposes on the regressors of (3) are restrictive and suggested the possibility to estimate the $\Sigma_i$'s. Our structure is no more restrictive than the one used in related literature. Clearly, the equal weighting scheme is appropriate if all variables are measured in the same units (e.g., growth rates) and their variability is comparable; otherwise, preliminary transformations need to be used or the vector of $\Sigma_i$ appropriately scaled. For example, if the variability of the variables of country 1 is considerably larger than the variability of the variables of country 2, then one could specify $\Sigma_1 = (\sigma_1^{-1}, \ldots, \sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_2^{-1}, \ldots)$, where $\sigma_1$ and $\sigma_2$ measure the average standard deviation of the variables in countries 1 and 2. The idea of estimating the $\Sigma_i$'s is a bit foreign to our philosophy—the weights are a priori determined by the flexible factorization we use—but feasible if one directly starts from (3), treats $\Sigma_i$ as unknown, and employs the factor models techniques described below. Given our emphasis on multicountry VAR and the resulting observable SUR model, we do not pursue this idea further.

Our estimated specification has two types of advantages over single-country or two-country VARs. First, if the information is weak or the sample short, cross sectional information may help to get better estimates and smaller standard errors. Second, if the momentum that shocks induce across countries is the result of lagged interdependencies, our model will be able to capture it. Such pattern will instead emerge as "common shocks" in the other two frameworks.

How does our reparametrized SUR model compare with factor models? There are two types of factor models used in the literature. One is of the form

(16) $(y_{t+1} - \alpha) = \gamma(L)(y_t - \alpha) + \beta(L)f_t + e_{t+1}$
where $i = 1, \ldots, N$, $f_t$ is an $r \times 1$ vector of latent factors, $r < < N$, $N$ large, and $\gamma(L)$, $\beta(L)$, $\lambda_i(L)$ are one-sided polynomial in the lag operator. The so-called static version of the model, popularized by Stock and Watson (2002a,b), imposes the restriction that the latter two polynomials are of finite order (at most $q$ lags are allowed) and rewrite the model as

$$
(y_{t+1} - \alpha) = \gamma(L)(y_t - \alpha) + \beta F_t + e_{t+1}
$$

(18)

$$
X_t = \Lambda F_t + u_t,
$$

(19)

where $F_t = (f_t', f_{t-q}')'$ is an $s \times 1$ vector, $s \leq (q+1)r$, the $i$th row of $\Lambda$ is $(\lambda_{i0}, \ldots, \lambda_{iq})$ and $\beta = (\beta_0, \ldots, \beta_q)'$. Although dynamic, (18) and (19) can be estimated with static principal components techniques: The loadings $\Lambda$ are the first $s$ eigenvalues of the $X'X$ matrix, where $X$ is the $T \times N$ data matrix and $\hat{F} = X\hat{\Lambda}$. Since (18) and (19) are not nested into a VAR, comparison with our model is a bit difficult. To better highlight the relationship, set $\gamma(L) = 0$ and choose $X_t$ to be equal to the lags of $y_{t+1}$. Under these conditions, our indices differ from the factors produced by static principal components for several reasons. First, the latter captures the volatility of the data matrix $X_t$, whereas our indices extract comovements in series belonging to $X_t$. Second, ours indices are observable, whereas the factors in (18) and (19) are unobservable and need to be estimated with a data-driven approach. Third, although the factors obtained with principal component analysis are statistical in nature—and economic interpretations can be given only via identification devices—our indices have some direct economic interpretation. Fourth, our indices will be substantially smoother than the factors extracted with principal components techniques. Fifth, at least in their classical formulation, the law of motion of $f_t$ is never used in the estimation of the factors, time variations in the factor loading are difficult to deal with (see e.g., Stock and Watson, 2002a, p. 1170), and estimates enjoy good properties only if time variations are small—therefore excluding, e.g., smooth changes across regimes and/or volatility bursts. Finally, it is hard to map log-linearized solutions of DSGE models into (18) and (19). Therefore, the link between economic theory and empirical practice is less transparent.

The second type of factor models still assumes that $f_t$ is unobservable, but posits

$$
\phi(L) f_t = u_t,
$$

(20)

where $\phi(L)$ is assumed to be diagonal for each $L$ and, typically, $\text{corr}(u_{j\ell}, u_{j'\ell'}) = 0$, $j = 1, \ldots, r$, and $j \neq j'$. We will refer to this model as the unobservable factor (UF) model, which has been used, for example, by Stock and Watson (1989) among many others. Classical estimation of this model is somewhat more complicated as the Kalman filter needs to be used. Also, the EM algorithm typically used for this purpose is cumbersome when $N$ is large.
It is relatively easy to show that an UF model can be written as a VARMA. In fact, substituting (20) into (16) we have that

\[(I - \gamma(L)L)(y_{t+1} - \alpha) = \beta(L)\phi^{-1}(L)u_t + \epsilon_{t+1}.\]

(21)

Hence, as long as \(\phi(L)\) has a convergent representation, a VAR for \(y_t\) exists. Note that the error term has two components: one due to shocks to the common factors, and one due to the idiosyncratic shocks to the model. Because of this feature and because it is hard to separately identify \(\phi(L)\) and \(\beta(L)\), our indices and UF factors have little in common. Hence, when deciding between a SUR or an UF approach, one has to take a stand on whether (1) or (21) better represent the DGP of the data.

Bayesian versions of UF models have been estimated by Otrok and Whiteman (1998), Kim and Nelson (1998), Del Negro and Otrok (2006). The advantages of such an approach are multiple. The one more relevant here is that time variations in the coefficients can be dealt with within standard MCMC routines at no additional costs.

The SUR model we use has also some similarities with the models used by Pesaran (2003) and Pesaran et al. (2005) to model global interdependencies, even though the starting point, the underlying specification, and the estimation technique differ. In fact, in these papers the baseline specification is a traditional (micro) panel structure with unobservable common components in the error term, instead of a VAR; no time variations are allowed in the coefficients and no lagged interdependences are present; \(N\) is assumed to be large. In this setup, it is possible to obtain a consistent estimate of the common unobservable component by arithmetically averaging the dependent and the independent variables of the unit-specific regressions. Therefore, the estimated specification looks like a set of unrelated single-country VARs, where common factors are proxied by averages of the variables across countries. Our approach shares the idea of using arithmetic averages as regressors; it can be interpreted as an \(F\)-factor generalization of these authors’ approach, where each factor spans a difference space, when we allow for lagged interdependencies in the error term and for time-varying loading. Finally, our approach does not need a large \(N\) to work.

2.3. Completing the Model. We assume that the factors evolve according to a general law of the form

\[\theta_t = (I - C)\bar{\theta} + C\theta_{t-1} + \eta_t, \quad \eta_t \sim (0, B),\]

\[\bar{\theta} = P\mu + \epsilon, \quad \epsilon \sim (0, \Psi),\]

where \(\bar{\theta}\) is the unconditional mean of \(\theta_t; P, C\) are known matrices; \(\eta_t\) and \(\epsilon\) are mutually independent and independent of \(E_t\) and \(u_t\); and \(B = diag(B_1, \ldots, B_F)\).
Furthermore, we let $E_t \sim (0, \Omega)$, and $u_t \sim (0, \Omega \otimes V)$, where $V = \sigma^2 I_k$ is a $k \times k$ matrix and $\Omega$ is an $NG \times NG$ matrix.

The intuition behind this specification is simple: To permit time variations, we make the factors obey the stochastic restrictions implied by (22). In the first equation of (22), we have assumed a general AR structure: Since the matrix $C$ is arbitrary, many patterns are allowed in the specification. Although we treat $C$ as fixed, it is possible to make it function of a small set of hyperparameters whose posterior can be jointly obtained with the one of the other parameters. Given that such a choice adds to the computational costs and that a near random walk specification for $\theta_t$ is for all purposes satisfactory, we do not follow such an approach here.

Whenever $C \neq I$, the second equation in (22) links the unconditional mean of the certain factors in an exchangeable fashion. In particular, if a vector country-specific factors is present, the specification implies that they will have the same mean and variance. This permits some degree of pooling, which can help to improve the precision of the estimates.

The spherical assumption on $V$ reflects the fact that factors are measured in common units, whereas the block diagonality of $B$ is needed to guarantee the identifiability of the factors.

We specify normal distributions for $E_t$, $u_t$, $\eta_t$, and $\epsilon_t$, but it is easy to allow for fat tails if aberrant or nonnormal observations are presumed to be present. For example, we could let $(u_t | z_t) \sim N(0, z_t(\Omega \otimes V))$ where $z_t^{-1} \sim \chi^2 (v, 1)$, since, unconditionally, $u_t \sim t_{\chi}(0, \Omega \otimes V)$. As it will be clear from the next section, the forecast errors of our SUR model already display fat tail distributions even when all disturbances are normal. Hence, this extension will not be considered here. Further complication, allowing, for example, for skewness in the errors, or for time variations in the variance of shocks to the factors, are easy to introduce (see Canova, 1993, or Fernandez and Steel, 1998). All of these additions go in the direction of capturing nonnormal patterns in $y_t$, if this is needed.

Numerous specifications are nested in our model: For example, a factor is time invariant when $B_t = 0$ and the appropriate elements of $C$ are set to zero; no exchangeability obtains when $\Psi_t$ is large, exact pooling obtains when $\Psi_t = 0$, and the factorization becomes exact when $\sigma^2 = 0$.

3. INFERENCE

If $\theta_t = \theta \forall t$, estimation of (3) is easy: It only requires regressing each element of $Y_t$ on appropriate averages, adjusting estimates of the standard errors for the presence of heteroscedasticity. With a prior for $\theta$, posterior estimates would be straightforward to construct.

When the $\theta_t$'s are time-varying, MCMC methods can be employed to construct their exact posterior distributions. The likelihood of the reparametrized SUR model is
\begin{align*}
\mathcal{L}(\theta, \Psi | Y) \propto \prod_i |\Upsilon_i|^{-1/2} \exp \left[ -\frac{1}{2} \sum_i \left( Y_i - \lambda_i \theta_i \right)^\top \Upsilon_i^{-1} \left( Y_i - \lambda_i \theta_i \right) \right],
\end{align*}

where \( \Upsilon_i = (1 + \sigma^2 X_i \Sigma X_i)^\top \Omega \equiv \sigma \Omega \). To calculate the posterior for the unknowns we need prior distributions for \((\mu, \Psi^{-1}, \Omega^{-1}, \sigma^{-2}, B^{-1})\). Let data run from \((-\tau, T)\), where \((-\tau, 0)\) is a “training sample” used to estimate features of the prior. When such a sample is unavailable or when a researcher is interested in minimizing the impact of prior choices, it is sufficient to modify the expressions for the prior moments, as suggested below.

We let \( p(\mu, \Psi^{-1}, \Omega^{-1}, \sigma^{-2}, B^{-1}) = p(\mu) p(\Psi^{-1}) p(\Omega^{-1}) p(\sigma^{-2}) \prod_f p(B^{-1}_f) \) where

\begin{align*}
p(\mu) &= \mathcal{N}(\hat{\mu}, \Sigma_p) \quad p(\Psi^{-1}) = \mathcal{W}(z_0, Q_0) \\
p(\Omega^{-1}) &= \mathcal{W}(z_1, Q_1) \quad p(\sigma^{-2}) = \mathcal{G} \left( \frac{a_1}{2}, \frac{a_2}{2} \right) \\
p(B^{-1}_f) &= \mathcal{W}(z_{2.f}, Q_{2.f}) \quad f = 1, \ldots, F.
\end{align*}

Here \( \mathcal{N}(\cdot) \) stands for Normal, \( \mathcal{W}(\cdot) \) for Wishart, and \( \mathcal{G}(\cdot) \) for Gamma distributions. The hyperparameters \((z_0, z_1, z_{2.f}, a, b, \text{vec}(\hat{\mu}), \text{vech}(\Sigma_p), \text{vech}(Q_0, Q_1, Q_{2.f}))\) are treated as fixed, where \( \text{vec}(\cdot) (\text{vech}(\cdot)) \) denotes the column-wise vectorization of a rectangular (symmetric) matrix. Noninformative priors are obtained setting \( a, b \to 0, Q^{-1}_j \to 0, \Sigma^{-1}_\mu \to 0, \) and \( Q_i \to 0, i = 0, 1 \). The form of the conditional posterior distributions we present below is unchanged by these modifications.

Despite the dramatic parameter reduction obtained with (3), the analytical computation of posterior distributions is unfeasible. However, a variant of the Gibbs sampler approach can be used in our framework. Let \( Y^T = (Y_1, \ldots, Y_T) \) denote the data, \( \psi = (\mu, \Psi^{-1}, \Theta^{-1}, \sigma^{-2}, B^{-1}, \hat{\theta}, (\theta_i)) \) the unknowns whose joint distribution needs to be found, and \( \psi_{-\alpha} \) the vector of \( \psi \) excluding the parameter \( \alpha \). Let \( \theta^*_{-\alpha} = (I - C)\hat{\theta} + \hat{C}\theta_{-1} \) and \( \hat{\theta}_i = \hat{\theta}_i - C\theta_{i-1} \). Given \( Y^T \), the conditional posteriors for the unknowns are

\begin{align*}
\mu | Y^T, \psi_{-\mu} &\sim \mathcal{N}(\hat{\mu}, \hat{\Sigma}_\mu) \\
\Psi^{-1} | Y^T, \psi_{-\Psi} &\sim \mathcal{W}(z_0 + 1, \hat{Q}_0) \\
\Omega^{-1} | Y^T, \psi_{-\Omega} &\sim \mathcal{W}(z_1 + T, \hat{Q}_1) \\
B^{-1}_f | Y^T, \psi_{-Bf} &\sim \mathcal{W}(T * \dim(\theta^*_f) + z_{2.f}, \hat{Q}_{2.f}) \\
\sigma^{-2} | Y^T, \psi_{-\sigma^2} &\propto (\sigma^{-2})^{-a_1 - 1} \exp[a_2 \sigma^{-2}] \times \mathcal{L}(\hat{\theta}, \Psi | Y^T) \\
\hat{\theta} | Y^T, \psi_{-\theta} &\sim \mathcal{N}(\hat{\theta}, \hat{\Psi}),
\end{align*}

(23)
MULTICOUNTRY VAR MODELS

where

\[
\tilde{\mu} = \hat{\Sigma}_\mu \left( P' \Psi^{-1} \bar{\theta} + \Sigma_{\mu}^{-1} \tilde{\mu} \right);
\]

\[
\hat{\Sigma}_\mu = \left( P' \Psi^{-1} P + \Sigma_{\mu}^{-1} \right)^{-1};
\]

\[
\hat{Q}_1 = \left[ Q_{1}^{-1} + \sum_t (Y_t - \hat{\chi}_t \hat{\theta}_t) \sigma_t^{-1} (Y_t - \hat{\chi}_t \hat{\theta}_t)' \right]^{-1};
\]

\[
\hat{Q}_2 = \left[ Q_{2}^{-1} + \sum_t (\theta_t^\star - \hat{\theta}_t^\star) (\theta_t^\star - \hat{\theta}_t^\star)' \right]^{-1};
\]

\[
\hat{\theta} = \hat{\Psi} \left[ \Psi^{-1} \bar{\theta}_s + (I - C)' \bar{B}^{-1} \sum_t \tilde{\theta}_t \right];
\]

\[
\hat{\Psi} = \left[ \Psi^{-1} + (I - C)' \bar{B}^{-1} (I - C) \sum_t 1 \right]^{-1};
\]

\[\theta_t^\star\] refers to the \( f \)th subvector of \( \theta_t \), and \( \text{dim}(\theta_t^\star) \) to its dimension.

The conditional posterior of \( (\theta_1, \ldots, \theta_T | Y^T, \psi_{-\theta}) \), can be obtained with a run of the Kalman filter and of a simulation smoother as in Chib and Greenberg (1995). In particular, given \( \theta_{0|0} \) and \( R_{0|0} \), the Kalman filter gives the recursions

\[
\theta_{t|t} = \theta_{t-1|t-1} + \left( R_{t|t-1}^{-1} \chi_t F_{t|t-1}^{-1} \right) (Y_t - \hat{\chi}_t \hat{\theta}_t);
\]

\[
R_{t|t} = (I - (R_{t|t-1}^{-1} \chi_t F_{t|t-1}^{-1}) \chi_t) (R_{t-1|t-1}^{-1} + \bar{B})
\]

\[
F_{t|t-1} = \chi_t R_{t|t-1}^{-1} \chi_t' + \Gamma_t;
\]

where \( \theta_{t-1|t-1}^\star \) and \( R_{t-1|t-1}^\star \) are, respectively, the mean and the variance covariance matrix of the conditional distribution of \( \theta_{t-1|t-1} \). In order to obtain a sample \( \{ \theta_t \} \) from the joint posterior distribution \( (\theta_1, \ldots, \theta_T | Y^T, \psi_{-\theta}) \), the output of the Kalman filter is used to simulate \( \theta_T \) from \( N(\theta_{T|T}, R_{T|T}) \), \( \theta_{T-1} \) from \( N(\theta_{T-1}, R_{T-1}) \), and \( \theta_1 \) from \( N(\theta_1, R_1) \), where \( \theta_t = \theta_{t|t} + R_{t|t} R_{t+1|t}^{-1} (\theta_{t+1|t} - \theta_{t|t}) \), and \( R_t = R_{t|t} - R_{t|t} R_{t+1|t}^{-1} R_{t+1|t} \). The recursions can be started choosing \( R_{0|0} \) to be diagonal with elements equal to small values, whereas \( \theta_{0|0} \) can be estimated in the training sample or initialized using a constant coefficient version of the model.

Since the conditional posterior of \( \sigma^2 \) is nonstandard, a Metropolis step is needed to obtain draws for this parameter. We assume that a candidate \( (\sigma^2)^\star \) is generated via \( (\sigma^2)^\star = (\sigma^2)^\dagger + v \), where \( v \) is a normal random variable with mean zero and variance \( \bar{c}^2 \). The candidate is accepted with probability equal to the ratio of the kernel of the density of \( (\sigma^2)^\dagger \) to the kernel of the density of \( (\sigma^2)^\star \) and \( \bar{c}^2 \) is selected to achieve a certain acceptance rate.
Draws from the posterior distributions can be obtained cycling through the conditional in (23) and (24) after an initial set of draws is discarded. Checking for convergence of the algorithm to the true invariant distribution is somewhat standard, given the structure of the model. Convergence, in fact, only requires the algorithm to be able to visit all partitions of the parameter space in a finite number of iterations (for example, see Geweke, 2000).

Our choice of making $E_t$ and $u_t$ correlated, an assumption also used in the Minnesota prior (see Doan et al., 1984) and in other priors (e.g., Kadiyala and Karlsson, 1997), allows conjugation between the prior and the likelihood, avoids identification issues, and greatly simplifies the computation of the posterior. Furthermore, it provides an interesting interpretation for the errors of the model. In fact, since $\Upsilon_t = (1 + \sigma^2 X_t' X_t) \Omega$, the prior distribution for the forecast error $\xi_t = Y_t - X_t' \theta_t$ has the form $(\xi_t | \sigma^2) \sim \mathcal{N}(0, \sigma_t \Omega)$. Therefore, unconditionally, $\xi_t$ has a multivariate $t$ distribution centered at 0, scale matrix proportional to $\Omega$ and $\nu_t$ degrees of freedom, and the innovations of (3) are endogenously allowed to have fat tails. In order to capture conditional heteroskedasticity in $y_t$, Cogley and Sargent (2005) specify $\Omega$ to be a function of a set of stochastic volatility processes. The above discussion shows that a similar result can be equivalently obtained with a simpler set of assumptions. We regard our specification more appealing on another count: Since shocks to the model may alter its dynamics, there is, built-in, an endogenous adaptive scheme that allows coefficients to adjust when breaks in the relationships occur.

The regressors of the SUR model are correlated, but the presence of correlation (even of extreme form) does not create problems in identifying the loading as long as the priors are proper (see e.g., Ciccarelli and Rebucci, 2007), which is the case in our setup.

While we have assumed that $u_t$ is serially uncorrelated, it is conceivable that this is not the case. General patterns of serial correlation are not allowed in our specification: Since $\delta_t$ is integrated out, it is not possible to easily account for them. One extreme possibility would be to specify a process for $\Delta u_t$, specify difference (3) and estimate the resulting model. This choice does not seem to be sensible when the variables of the VAR are measured in growth rates, as it is the case for the specification used in Section 6.

Posterior distributions for any continuous function $\mathcal{G}(\psi)$ can be obtained using the output of the MCMC algorithm and the ergodic theorem. For example, $E(\mathcal{G}(\psi)) = \int \mathcal{G}(\psi) p(\psi | Y) d\psi$ can be approximated using $\frac{1}{L} \sum_{l=1}^{L} \mathcal{G}(\psi_l)$ (the first $L$ observations represent a burn-out sample discarded in the calculation). Predictive distributions for future $y_t$'s can be estimated using the recursive nature of the model and the conditional structure of the posterior. Let $Y_{t+\tau} = (Y_{t+1}, \ldots, Y_{t+\tau})$, consider the conditional density of $Y_{t+\tau}$, given the data up to $t$, and a function $\mathcal{G}(Y_{t+\tau})$. Then

$$\mathcal{F} \left( \mathcal{G}(Y_{t+\tau}) | Y_t \right) = \int \mathcal{F} \left( \mathcal{G}(Y_{t+\tau}) | Y_t, \psi \right) p(\psi | Y_t) d\psi$$
and, e.g., forecasts for $Y^{t+\tau}$ can be obtained drawing $\psi^{(l)}$ from the posterior distribution and simulating the vector $Y^{t+\tau}$ from the density $F(Y^{t+\tau} | Y_t, \psi^{(l)})$.

Turning point distributions can also be constructed by appropriately choosing $G$. Impulse responses and conditional forecasts can be obtained with the same approach as detailed in Section 5.

4. MODEL SELECTION

Although we have assumed that the choice of the type of factors in (2) depends on the nature of the problem, one may be interested in having a method to statistically determine the number of indices needed to capture the heterogeneities present across time, units, and variables in the multicountry VAR, or to verify general hypotheses on the type of indices to be included. In order to discriminate across models with different indices consider

$$L(Y^s | M_h) = \int F(Y^s | \psi_h, M_h) p(\psi_h | M_h) d\psi_h,$$

which is the marginal likelihood for $Y^s$ in a model with $h$ indices. Here $p(\psi_h | M_h)$ is the prior density for $\psi$ in model $M_h$ and $F(Y^s | \psi_h, M_h)$ is the density of the data under the parameterization produced by $M_h$. Equation (25) is conceptually simple, but can be evaluated analytically only in few elementary cases. More often, it is intractable and must be computed by numerical methods, using the output of the MCMC sampler, as suggested by Newton and Raftery (1994), Gelfand and Dey (1994), or Chib (1995). Given the complexity of our model, these numerical computations are not entirely straightforward. As an alternative, one can rely on asymptotic (normal) approximations to (25), for example Laplace’s method—which takes a second-order expansion of (25) around the mode—or the Schwarz criterion—which expands (25) around the maximum-likelihood estimator. Since in hierarchical models such as the one we propose, asymptotic normality might not be a sensible approximation, it is probably a good idea to compute alternative measures of marginal likelihood before taking decisions about the size of $h$.

Once the marginal likelihood is obtained for any model $h$, the Bayes factor,

$$B_{hh'} = \frac{L(Y^s | M_h)}{L(Y^s | M_{h'})},$$

can be used to decide whether $M_h$ or $M_{h'}$ fits the data better. Since marginal likelihoods can be decomposed into the product of one-step ahead prediction errors, pairs of models are compared using their one-step ahead predictive record. Also, since the marginal likelihood implicitly discounts the performance of models with a larger number of indices, (26) directly trades off the predictive record with the dimensionality of the model.

With (26) it is also possible to conduct useful specification searches. For example, it is possible to examine whether the factorization in (2) is exact, letting $\psi_h$ unrestricted and $\psi_{h'} = (\ldots, \sigma^2 = 0, \ldots)$; or whether there are time variations in
θ_t, setting \( \bar{B}_f = b_f \ast I \), letting \( \psi_h \) be unrestricted and \( \psi_{h'} = (\ldots, b_f = 0, \ldots) \) for some \( f \). Support for the presence of interdependencies is obtained, on the other hand, by comparing the marginal likelihoods of the unrestricted model and that of a vector of country-specific TVC-VARs.

Instead of examining hypotheses on the structure of the model, one may want to incorporate model uncertainty directly into posterior estimates. Let \( M_1 \) be the model with one index and \( M_h \) the model with \( h \) indices, \( h = 2, \ldots, H \), and suppose we have computed the Bayes factor \( B_h \) for each \( M_h \). The posterior probability of model \( h \) is

\[
p(M_h | Y_t) = \frac{a_h B_h}{\sum_{h=2}^{H} a_h B_h},
\]

where \( a_h \) are the prior odds for \( M_h \), and model uncertainty can be accounted for weighting \( G(\psi_h) \) by \( p(M_h | Y_t) \).

5. DYNAMIC ANALYSIS

Dynamic analysis is nonstandard in our SUR model, because of the specification of the error term and the time variations potentially present in the coefficients. Hence, we describe in details how to produce statistics useful for academics and policymakers.

5.1. Recursive Unconditional Forecasts. Given the information at time \( t \), unconditional forecasting exercises only require the computation of the predictive distribution of future observations. In some applications recursive unconditional forecasts are needed, in which case the predictive density of future observations has to be constructed for every \( t = \bar{t}, \ldots, T \) once recursive estimates of \( p(\psi_h | Y_t) \) are computed. These recursive distributions are straightforward to obtain (we need to run an MCMC for every \( t \)) and, although computationally demanding, they are feasible on available machines.

5.2. Impulse Responses. Impulse responses are generally computed as the difference between two realizations of \( y_{t+\tau} \), \( \tau = 1, 2, \ldots \), which are identical up to time \( t \), but one assume that between \( t + 1 \) and \( t + \tau \) a one time impulse in the \( j \)th component of \( e_{t+\tau} \) occurs only at time \( t + 1 \), and the other that no shocks take place at all dates between \( t + 1 \) and \( t + \tau \).

In a model with time-varying coefficients such an approach is inadequate since it disregards that between \( t + 1 \) and \( t + \tau \), structural coefficients may also change. Therefore, our impulse responses are obtained as the difference between two conditional expectations of \( y_{t+\tau} \). In both cases, we condition on the history of the data \( (Y') \) and of the factors \( (\theta') \), the parameters of the law of motion of the coefficients, and all future shocks. However, in one case we condition on a random draw for the current shocks, whereas in the other the current shocks is set to its unconditional value (see also Gallant et al., 1993; Koop et al., 1996). We condition on future shocks instead of integrating them out because, computationally, such a choice gives more stable responses, even though, in practice, this makes standard error bands larger than in the case where future shocks are integrated out.

In our model, one has two potential types of impulses, one to the variables of the system and one to the factors. Although the former have the standard
interpretation, the latter can be used, for example, to represent shocks to particular structural coefficients, e.g., a shock that reduces the sensitivity of some the variables to world conditions. In order to formally define impulse responses, we need some notation. The reparametrized SUR is

\[ y_t = X_t \theta_t + (E_t + X_t u_t) \]

\[ \theta_t = (I - C)(P \mu + \epsilon) + C \theta_{t-1} + \eta_t, \]

where \( \theta_t = [\theta'_1, \theta'_2, \ldots, \theta'_s]' \), \( X_t = [X'_1, \ldots, X'_s] \), \( X_{jt} = \Xi_t X_t, X_t = [Y_{t-1}, W_t] \).

Let \( U_t = [(E_t + X_t u_t)', \eta'_t, \epsilon_t]' \) be the vector of reduced-form shocks and \( Z_t = [H_t^{-1}(E_t + X_t u_t)', H_t^{-1} \eta'_t, H_t^{-1} \epsilon_t]' \) be the vector of structural shocks where \( E_t = H_t \eta_t, H_t H_t' = \Omega \) so that \( \text{var}(v_t) = I \) and \( H_t = J \times K_t \) where \( K_t K_t' = I \) and \( J \) is a matrix that orthogonalizes the VAR shocks.

In our setup a Choleski system is obtained setting \( K_t = I, \forall t \) and choosing \( J \) to be lower triangular whereas more structural identification schemes are obtained letting \( J \) be an arbitrary square root matrix and \( K_t \) a matrix implementing certain theoretical restrictions. Note also that we have allowed the identification matrix \( K_t \) to be time-varying. We do this because, in certain applications where recursive estimation is used, estimates of \( \Omega \) depend on \( t \). Also, there may be situations in which the covariance matrix of reduced-form shocks is time invariant but the contemporaneous relationships of the structural model are time-varying.

Let \( \mathcal{V}_t = (\Omega, \sigma^2, B, \Psi) \), let \( \tilde{Z}_{jt} \) be a particular realization of \( Z_{jt} \) and \( Z_{jt} \) indicate the structural shocks, excluding the one in the \( j \)th component. Let \( \mathcal{F}^1_t = \{Y_t^{-1}, \theta', \mathcal{V}_t, H_t, Z_{jt} = \tilde{Z}_{jt}, Z_{jt}, \tilde{U}_{jt} \} \) and \( \mathcal{F}^2_t = \{Y_t^{-1}, \theta', \mathcal{V}_t, H_t, Z_{jt} = \tilde{Z}_{jt}, Z_{jt}, \tilde{U}_{jt} \} \) be two conditioning sets. Then responses to a shock at \( t \) in the \( j \)th component of \( Z_t \) are obtained as

\[ IR(t, t + \tau) = E(Y_{t+\tau} | \mathcal{F}^1_t) - E(Y_{t+\tau} | \mathcal{F}^2_t) \quad \tau = 1, 2, \ldots. \]

In order to see what definition (27) involves, rewrite the original VAR model (1) in a companion form

\[ Y_{t+\tau} = A_{t+\tau} Y_{t+\tau-1} + C_{t+\tau} W_{t+\tau-1} + E_{t+\tau} \]

and let

\[ \delta_{t+\tau} = \Xi[(I - C)(P \mu + \epsilon) + C \theta_{t+\tau-1} + \eta_{t+\tau}] + u_{t+\tau}. \]

Here \( \delta_{t+\tau} = [\text{vec}(A_{1t+\tau}), \text{vec}(C_{1t+\tau})] \) and \( A_{1t+\tau} \) is the first row of \( A_{t+\tau} \). Taking \( Y_t^{-1} = Q_1 \) \( (Y_{t-1}, Y_{t-2}, \ldots, W_{t-1}, W_{t-2}, \ldots), A' = (A_t, A_{t-1}, \ldots), C' = (C_t, C_{t-1}, \ldots) \), and
H_{t+\tau} = H_t \forall \tau$ as given, and solving backward we can write (28) and (29) we have

\begin{equation}
Y_{t+\tau} = \left( \prod_{k=0}^{\tau} A_{t+\tau-k} \right) Y_{t-1} + C_{t+\tau} W_{t+\tau-1} + \sum_{h=1}^{\tau-1} \left( \prod_{k=0}^{h-1} A_{t+\tau-k} \right) C_{t+\tau-h} W_{t+\tau-h-1} \nonumber \\
+ H_{t+\tau} v_{t+\tau} + \sum_{h=1}^{\tau-1} \left( \prod_{k=0}^{h-1} A_{t+\tau-k} \right) H_{t+\tau-h} v_{t+\tau-h} \nonumber 
\end{equation}

(31) \[ \delta_{t+\tau} = \Xi(I - C)(P \mu + \epsilon) \sum_{k=0}^{\tau} C^k + \Xi C^{\tau+1} \theta_{t-1} + \Xi \sum_{k=0}^{\tau} C^k \eta_{t+\tau-k} + u_{t+\tau}. \]

Consider first the case of a $(m+1)$-period impulse in the $j$th component of $\nu_t$, i.e., $v_{j,t+k} = \bar{v}_{j,t+k}$ whereas $v_{-j,t+k}, k = 0, 1, \ldots, m$ and $v_{i+\tau'} \forall m' > m$ are unrestricted. Then

\begin{equation}
IR(t, t+\tau) = E_t[Y_{t+\tau} | Y^{t-1}, A', C', \nu_t, H_t; \{v_{j+k}\}_{k=0}^m, \{v_{-j+k}\}_{k=0}^m] \nonumber \\
- E_t[Y_{t+\tau} | Y^{t-1}, A', C', \nu_t, H_t; \{v_{j+k}\}_{k=0}^m] \nonumber \\
= E_t \left[ \left( \prod_{k=0}^{\tau-1} A_{t+\tau-k} \right)^{\bar{v}_{j,t}} H_t^i (\bar{v}_{j,t} - E v_{j,t}) + \left( \prod_{k=0}^{\tau-2} A_{t+\tau-k} \right)^{\bar{v}_{j,t}} H_t^i (\bar{v}_{j,t} - E v_{j,t}) + \ldots + \left( \prod_{k=0}^{\tau-m} A_{t+\tau-k} \right)^{\bar{v}_{j,t}} H_t^i (\bar{v}_{j,t} - E v_{j,t}) \right] \nonumber \\
\times H_t^i (\bar{v}_{j,t+1} - E v_{j,t+1}) + \ldots + \left( \prod_{k=0}^{\tau-m} A_{t+\tau-k} \right)^{\bar{v}_{j,t}} H_t^i (\bar{v}_{j,t} - E v_{j,t}) \nonumber \\
\times H_t^i (\bar{v}_{j,t+m} - E v_{j,t+m}) \right],
\end{equation}

where the superscript $j$ refers to the $j$th column of the matrix. It is easy to see that, when $A_t = A, C_t = C, \forall t$, (32) reduces to standard impulse responses and that when $E_t$ and $\eta_t$ are correlated, both the sign and the size of the shocks matter—a shock in $\nu_t$ may induce changes in $\nu_t$ or $C_t$.

Given (27), responses in our SUR model can be computed as follows:

1. Choose $t$, $\tau$, and $J_t$. Draw $\Omega_t = H_t^i (H_t^i)'$, $(\sigma^2)^i_t$ from their posterior distribution and $u_t^i$ from $N(0, \sigma^2_t I \otimes H_t^i (H_t^i)')$. Compute $y_t^i = \chi_t \theta_t + H_t \bar{v}_t + X_t u_t^i$.
2. Draw $\Omega_t = H_{t+1}^i (H_{t+1}^i)'$, $(\sigma^2)_t, B_t^i, \Psi_i$. Draw $u_{t+1}^i, \epsilon_i$ from their posterior distribution. Use the law of motion of the factors to compute $\theta_{t+1}^i, l = 1, \ldots, L$ and the definition of $\Xi$ to compute $\chi_{t+1}$. Draw $u_{t+1}^i$ from $N(0, \sigma^2_t I \otimes H_{t+1}^i (H_{t+1}^i)'$) and compute $y_{t+1}^i = \chi_{t+1} \theta_{t+1}^i + H_{t+1} \bar{v}_{t+1} + X_{t+1} u_{t+1}^i, l = 1, \ldots, L$.
3. Repeat step 2 and compute $\theta_{t+k}^i, y_{t+k}^i, k = 2, \ldots, \tau$. 

---

46
45
43
42
41
40
39
38
4. Repeat steps 1–3 setting \( v_{t+k} = E(v_{t+k}) \), \( k = 0, \ldots, m \) using the draws for the shocks in 1–3.

Responses to structural shocks to the law of motion of the factors can be computed in the same way. An impulse in \( \eta_t = \bar{\eta} \) lasting \((m + 1)\) periods implies from (31) that

\[
E(\delta_{t+\tau} - \delta_{t+\tau}) = \Xi \sum_{k=0}^{m} H_{t+k}C^k(\bar{\eta}_{t+\tau-k} - E\eta_{t+\tau-k})
\]

so that

\[
IR(t, t + \tau) = E \left[ \prod_{k=0}^{\tau-1} (A_{k+t+\tau-k} - A_{k+t+\tau-k})Y_{t+1} \right]
\]

\[
\times C_{t+\tau-k}W_{t+\tau-h-1} + \left( \prod_{k=0}^{\tau-1} (A_{k+t+\tau-k} - A_{k+t+\tau-k})H_{t+\tau-h}v_{t+\tau-h} \right).
\]

5.3. **Conditional Forecasts.** There are two types of conditional forecasts one can compute in our model: Those involving displacement of the exogenous variables \( W_t \) from their unconditional path, and those involving a particular path for a subset of the endogenous variables. Both types of conditional forecasts can be constructed using the output of the Gibbs sampler routine.

Consider first displacing the exogenous variables from their expected future path for \( m + 1 \) periods. Call the new path \( W_{t+k}, k = 0, 1, \ldots, m \). Then, the response of \( Y_{t+\tau} \) is

\[
IR(t, t + \tau) = E \left[ \prod_{k=0}^{\tau-2} A_{k+t+\tau-k} \right] C_{t+1}(\bar{W}_{jt} - W_{jt}) + \left( \prod_{k=0}^{\tau-3} A_{k+t+\tau-k} \right)
\]

\[
\times C_{t+2}(W_{jt+1} - W_{jt+1}) + \cdots + \left( \prod_{k=0}^{\tau-2-m} A_{k+t+\tau-k} \right)
\]

\[
\times C_{t+m+1}(W_{jt+m} - W_{jt+m}) \right].
\]

Therefore, to compute conditional forecasts of this type in our SUR model we need to

1. Choose \( t, \tau \), and a path \( \{\bar{W}_{t+k}\}_{k=0}^{m} \). Draw \( \Omega^l, (\sigma^2)^l \) from their posterior, draw \( E^l + X_lu^l \) and compute \( \gamma^l \).
2. Draw \( (B^l), \psi^l \) from their posterior distribution, draw \( \bar{\eta}^l_{t+1}, \epsilon^l \) and use the law of motion of the factors to draw \( \bar{\theta}^l_{t+1}, l = 1, \ldots, L \), and the definition of \( \Xi \) to compute \( X^l_{t+1} \). Draw \( E^l_{t+1} + X^l_{t+1}u^l_{t+1} \) and compute \( \gamma^l_{t+1} = X^l_{t+1}\bar{\theta}^l_{t+1} + (E^l_{t+1} + X^l_{t+1}u^l_{t+1}) \), \( l = 1, \ldots, L \).
we want to show how the tools we describe in the article could be applied to the two problems of oil price change and US real shock. By no means we intend to be exhaustive about these two problems. Instead, we want to show how the tools we describe in the article could be applied in the context of the G-7 countries.

6. THE TRANSMISSION OF SHOCKS IN G-7 COUNTRIES

This section shows how one can use our setup to examine two issues of economic interest: What are the effects of a U.S. real shock on the GDP of G-7 countries, and what are the consequences of an unexpected oil price change on inflation in euro area countries. By no means we intend to be exhaustive about these two problems.
to questions that are of crucial interest for applied business cycle investigators in academics and central banks.

The last 20 years have witnessed an increased globalization of world economies. Given the current high level of integration in the G-7, inflation and economic activity in the euro area are closely related not only to those of the United States but also of the other industrialized countries. Therefore, it makes sense to try to exploit cross-sectional information to construct probability distributions of various scenarios. Furthermore, the evolutionary nature of the relationship, documented e.g., in Del Negro and Otrok (2006) among others, suggests that a time-varying specification will probably be more useful than arbitrarily selecting fixed subsamples, as it is often done in the literature.

For each of the G-7 countries, we use four endogenous variables (real GDP growth, CPI inflation, employment growth, and rent inflation) and a predetermined one (the growth rate of an oil price index). GDP growth is measured using Eurostat real GDP at 1995 prices, employment by the OECD index of total employment, inflation and rent inflation using GDP and housing rental deflators (again from Eurostat), and the variables are scaled by their standard deviation. Oil prices are obtained from the IMF Financial statistic series. For all variables, growth rates are computed quarter-on-quarter and annualized. Besides GDP growth and CPI inflation, which are the focus our attention here, the other two endogenous variables have been selected because they have considerable in-sample predictive power for output growth and inflation across countries. We exclude monetary variables from the specification as they do not seem to have predictive power for inflation or output growth once lags of these variables are included. We use one lag of the endogenous variables, a constant and one lag of the predetermined variable. Since in the SUR model, regressors average over lags of the endogenous variables, the exact number of lags does not matter in our exercises.

Each equation of the VAR has \( k = 7 \times 4 + 1 + 1 = 30 \) coefficients and there are 28 equations in the system. The estimation sample covers the period 1980:1-2000:4. Therefore, without restrictions, there would be a total of \( 30 \times 28 \) regression parameters plus 406 covariance parameters to be estimated at each \( t \).

We assume that the coefficient vector \( \delta_t \) in (2) depends on three factors, and that the factorization is exact, i.e., \( \delta_t = \Xi_1 \theta_{1t} + \Xi_2 \theta_{2t} + \Xi_3 \theta_{3t} \). Here \( \theta_{1t} \) is a \( 2 \times 1 \) vector of common factors, one for euro area variables and one for the rest of the world, so that \( \Xi_{1t} = \sum_{US,JP,CA,UK} \sum_{j} y_{igt-j} \), \( \Xi_{12t} = \sum_{UK,FR} \sum_{j} y_{igt-j} \), \( \theta_{2t} \) is a \( 7 \times 1 \) vector of country-specific factors and \( \Xi_{2jt} = \sum_{g} \sum_{j} y_{igt-j} \), \( i = 1, \ldots, 7; \theta_{3t} \) is a \( 4 \times 1 \) vector of variable-specific factors and \( \Xi_{3gt} = \sum_{t} \sum_{j} y_{igt-j} \), \( g = 1, \ldots, 4 \). We also set \( C = I \). Hence, \( \theta_t = (\theta_{1t}', \theta_{2t}', \theta_{3t}')' \) is \( 13 \times 1 \) vector and the estimated model is

\[
\begin{align*}
y_t &= X_1 \theta_{1t} + X_2 \theta_{2t} + X_3 \theta_{3t} + \xi_t \\
\theta_t &= \theta_{t-1} + \eta_t.
\end{align*}
\]

Since our sample is relatively short, no training sample is available to tune the prior up. In order to minimize the influence of our prior choices we select relatively loose but proper priors and set \( p(h^{-1}_i) = G(5, 0.5) \), \( i = 1, 2, 3 \) and
\[ p(\Omega^{-1}) = \mathcal{N}((z_1 \Omega_{OLS}^{-1}, z_1), \text{where } \Omega_{OLS} \text{ is the OLS estimate of the } \Omega \text{ obtained on a fixed coefficient version of the model, and the degrees of freedom are chosen to approximately match the sample size, i.e., } z_1 = ng + 50. \text{ We set } \theta_{00} \text{ to be equal to the OLS estimate obtained on the time-invariant version of the model, and set } R_{00} \text{ to the average estimated variances of NG AR(p)'s.} \]

We produce 3,000,000 iterations of the MCMC routine starting from arbitrary initial conditions. Runs of 600 elements are drawn 5000 times and the last observation of the final 4000 is used for inference. We checked convergence recursively calculating the first two moments of the posterior of the parameters using 500, 1000, 2000 draws and found that convergence was sufficiently easy to achieve and obtained with about 1000 draws. We have also experimented with different combinations of runs and chains, keeping the total number of iterations fixed. Results appear to be robust to this choice.

Our basic model has several bells and whistles. Therefore, prior to conducting the exercises we are interested in, we want to check whether all the features we consider are really necessary to model the available data. For this reason, we have computed the marginal likelihood for five different specification. In all of them the coefficient factorization is exact, i.e., \( \sigma^2 = 0 \), since specifications that do not impose this restrictions fit the data worse. \( M_1 \) is our benchmark model specification. The remaining four models impose additional restrictions on \( M_1 \). Specifically, \( M_2 \) excludes from \( M_1 \) international lagged interdependencies; \( M_3 \) is a model with no time variations in the coefficients, i.e., \( \text{var}(\eta_t) \equiv B = 0 \); \( M_4 \) and \( M_5 \) modify \( M_1 \) by excluding either the country-specific component \( \theta_2 \) or the variable-specific components \( \theta_3 \), respectively.

Since, as we have mentioned in Section 4, different methods to compute marginal likelihoods have advantages and drawbacks, and it is empirically unclear which method to prefer (see e.g., Bos, 2002), Table 1 presents results obtained using three different approaches: Chib’s calculation from the Gibbs output (Chib, 1995), a harmonic mean estimator (Newton and Raftery, 1994), and the Schwarz approximation. In the first method, since we treat \( \theta_i \) as a latent variables, and given the assumptions we have made, we only need one additional set of Gibbs sampling iterations to obtain the estimate. The second method averages over all draws the concentrated likelihood (after integrating out the latent vector \( \theta_i \)) evaluated at

**Table 1**

<table>
<thead>
<tr>
<th>Method</th>
<th>( M_1 )</th>
<th>( M_2 )</th>
<th>( M_3 )</th>
<th>( M_4 )</th>
<th>( M_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chib’s ML nse</td>
<td>-1200</td>
<td>-1236</td>
<td>-2908</td>
<td>-1548</td>
<td>-1579</td>
</tr>
<tr>
<td>Harmonic Mean nse</td>
<td>-1589</td>
<td>-1636</td>
<td>-1627</td>
<td>-1608</td>
<td>-1619</td>
</tr>
<tr>
<td>Max Loglike nse</td>
<td>-1530</td>
<td>-1617</td>
<td>-1610</td>
<td>-1580</td>
<td>-1595</td>
</tr>
<tr>
<td>Parameters</td>
<td>409</td>
<td>409</td>
<td>406</td>
<td>408</td>
<td>408</td>
</tr>
</tbody>
</table>

Note: The number of parameters is equal to free elements in \( B \) + free elements in \( \Omega \).
each draw of the posterior. In the last method, we report the log of the maximum likelihood across draws along with the number of parameters estimated in each model. Note that, because all models have approximately the same number of parameters, the Schwarz criterion ranking should resemble the ranking obtained from the simple maximized likelihood. Numerical standard errors (nse), computed using 10 different runs of the Gibbs sampler for each of the models, are also presented.

The ranking of the models differ across methods. With Chib’s measure, the basic model (M1) is clearly preferred, whereas the model that excludes time variations is clearly the worst. On the other hand, since a model with no variable-specific factors is considerably worse than a model with no country-specific factors, one can conclude that the dynamics of the endogenous variables across countries are similar (so the “world” factor largely suffices) whereas the dynamics of, e.g., output growth and inflation are fairly different within countries. The harmonic mean estimator and the Schwarz criterion roughly maintains the same relative ranking across models, even though a model that excludes interdependences is now worse than a model that excludes time variations.

Two important points need to be made here. First, although one may find it surprising that the marginal-likelihood values estimated with the three criteria are so different, one should also notice that the numerical standard errors around Chib’s estimates are quite large, indicating that this estimate is much more volatile and probably less reliable than the other two. Second, the size of the drop in the marginal likelihood obtained with Chib for model M3 is also quite surprising. One might guess that the estimated posterior distribution obtained is extremely imprecise and could be due to the fact that without time variations in the coefficients, the model is essentially regressing volatile variables on slow-moving ones. Hence, further work on the properties of Chib’s estimator of the marginal likelihood in complex hierarchical models such as ours is sorely needed.

In sum, it appears that a factorization of the coefficient vector that includes three factors and allows for no idiosyncratic component summarizes the information present in the multicountry VAR reasonably well. Lagged interdependencies, unit-specific dynamics and (small) time variations also appear to be important features of our multicountry VAR. In the following exercises, we therefore use M1 as our specification.

In order to show how dynamic analysis can be undertaken in our model and the advantages/disadvantages one can obtain with our setup relative to, for example, single-country or two-country VARs, we first consider the effect of a U.S. real
shocks on the GDP of other countries. We construct such a shock by making the U.S. variables contemporaneously causally prior to the other G-6 countries. Within the U.S. block, we make employment growth and output growth jointly increase 1% for one period, whereas the dynamics of the other two variables are unrestricted. Figure 1 presents the median responses together with a 68% posterior band obtained with information up to 2000:4. We also report the results obtained by running six two-country TVC-BVAR(1) with time-varying coefficients and a Litterman prior where country 1 is always the United States and country 2 one of the other six countries. Shocks are identified in the same way as in the multicountry VAR. Therefore, apart from using cross sectional information, the setup of two models is identical.

As Section 5.3 mentioned, one has to make assumptions to compute responses in TVC models. In particular, one needs to decide whether the loadings are affected by the shock or not. In the latter case, one would use the law of motion of the loadings to predict their development over the forecast horizon. In the former
case a learning process, where estimates of $\theta_t$ are updated as $y_{t-1}$ changes, could be used. Figure 1 presents responses using the law of motion of the loadings since results appears to be more stable than in the other case. With our random walk assumption this is equivalent to freeze the loadings at their end-of-sample values. The amount of in-sample time-variation is very important to have sensible impulse responses and forecasts in general. In our experiments we use a tight prior on the time variation, which is obtained by assuming for each $b_i$ a prior mean of $0.000001$ and a standard deviation of $1.0 \times 10^{-9}$.

As we have pointed out, one should expect two types of differences in the responses obtained with the two models: First, since the sample is short and the number of coefficients to be estimated is large, we should expect standard errors around the impulse responses to be less precisely estimated in the two-country VAR. Second, since the regressors of our model emphasize low-frequency co-movements, the responses of a multicountry VAR should be smoother than those of a two-country VAR. Figure 1 indicates that at least the first prediction is satisfied: Although responses in a two-country BVAR model are poorly estimated and often leave open the question of whether there is any transmission across countries (see, e.g., the responses of Germany, United Kingdom, and Canada), those of the
multicountry model are more informative about the features of transmission. For example, there clearly is an Anglo-Saxon cycle (peak responses of GDP in United Kingdom and Canada are contemporaneous and almost of the same magnitude as in the United States); European responses are positive but typically lagged, except for Germany, with French GDP responding somewhat more persistently than German and Italian GDP; the response of Japan is lagged but relatively small.

Note also that responses in the Panel VAR and in the two-country VAR die out at a similar rate but display different magnitudes.

Since our identification scheme has little economic content, we do not give responses any structural interpretation. In particular, we cannot say what is the reason for the asymmetric response across blocks of countries, whether policy matters or not, and whether the shock we consider is a technological improvement. In order to do this, a more structural identification scheme and a different set of variables needs to be considered.

Using the same logic of Pesaran and Smith (1996), one may suspect that our estimates display some kind of bias because of the way information is pooled in the stochastic model. This suspicion is incorrect for two reasons. First, pooling is stochastic and the amount of pooling is endogenously selected. Second, stochastic pooling has a long tradition in panel data and there is no evidence that such a procedure produced information-processing biases in reasonable experimental designs.

Next, we consider the response of inflation in the three European countries when the growth rate of the oil price index is set to zero for 16 periods from 1998:1 to 2000:4. Since this is a period where the growth rate of the oil price index was strongly positive, such an experiment mimic what would have happened if the boost in oil prices would not have occurred. The design of our experiment is illustrated in Figure 2. The shock is given by the difference between the actual and the counterfactual growth rates, where the latter assumes that the growth rate of the oil price index goes to zero at a gradual pace. In order to avoid a sudden drop to zero after 2000:4 and to allow for a more complete dynamics, we use data until 2002:4 in the exercise. On this additional sample, we assume that the growth rate of oil continues to gradually lessen the difference with the counterfactual path after the shock. Note that this is one type of conditional forecasting exercises that Central Banks routinely conduct in the quarterly assessment of current and future economic conditions. The major difference here is that we do this in the framework of a model with cross-country interdependences and allow for time-varying structure.

Figure 3 reports the posterior median and the posterior 68% band for inflation responses in Germany, France, and Italy. For comparison, we also report the responses obtained from a single-country BVAR(1) where the growth rate of oil is predetermined and we allow for time variations in the coefficients and a Litterman prior. Once again, the difference between the two sets of responses is due only to the use of cross-country information.

Responses in the three countries look different both in terms of magnitude and timing. The responses of German and French inflation are significant immediately after the shock, whereas Italian inflation is significant only four quarters after the
MULTICOUNTRY VAR MODELS

Figure 2
OIL PRICE SHOCK: ACTUAL (SOLID LINE) AND COUNTERFACTUAL (DASHED LINE)

Figure 3
FORECAST OF INFLATION CONDITIONAL ON A SHOCK TO OIL PRICE GROWTH (SOLID LINES), AND 68% CONFIDENCE BANDS (DASHED LINES)
shock. In general, it appears that oil price increases had moderately large and persistent effects on inflation of the three major EU countries. In comparison, the responses estimated with single-country VARs are more persistent but less significant (especially in the case of Italian inflation) as the bands tend to blow up as the horizon increase, suggesting that there is little information in the data about the likely direction of inflation changes.

Finally, the estimated model can be used to compute a variety of measures that are of interest to policymakers. Figure 4 presents the time profile for the posterior 68% band for a coincident measure of world inflation, constructed as $CVLI_t = X_1 \theta_1 + (X_3 \theta_3)^\pi$. Two points can be made. First, the bands are tight reflecting the usefulness of the cross sectional information. Second, the dynamics of our measure seem to match the conventional wisdom about the local trends present in the inflation rates over the period.

7. CONCLUSIONS

This article develops an approach to conduct inference in time-varying coefficient multicountry VAR models with lagged cross-unit interdependencies and unit-specific dynamics. We take a Bayesian viewpoint to estimate and restrict the coefficients to have a low-dimensional time-varying factor structure. We complete the specifications using a hierarchical prior for the vector of factors that permits exchangeability, time variations, and heteroscedasticity in the innovations in the factors.

The factor structure on the coefficients allows us to transform an over-parametrized VAR into a parsimonious SUR model where the regressors are observable linear combinations of the right-hand-side variables of the VAR, and the loadings are the time-varying coefficient factors. We derive posterior distributions
for the vector of loadings using MCMC methods. We show how to construct unconditional forecasts, responses to impulses in interesting structural shocks and conditional forecasts, using the output of the MCMC routine.

The reparametrization of the VAR has a number of appealing features. First, it reduces the problem of estimating a large number of, possibly, unit-specific and time-varying coefficients into the problem of estimating a small number of loadings on certain combinations of the right-hand side variables of the VAR. Second, since the regressors of the model are observable, the model can be employed recursively for a variety of policy purposes. Third, since some indices features an MA structure, they emphasize low-frequency comovements in the lags of the VAR variables.

The tools described in this article have been applied to a number of interesting problems (see e.g., Canova et al., 2007; Anzuini et al., 2005; and Caivano, 2006). For instance, the construction of measures of core inflation and of the natural rate of unemployment in multicountry settings, the study of the transmission of monetary policy shocks across economic areas and sectors; and the construction of portfolios of assets in different geographical regions can all be studied within the general framework presented in this article.

To conclude, one should mention that the procedure is computationally feasible on modern computers: One full run of the MCMC routine for the example of Section 6 takes about 45 minutes. Therefore, the approach is competitive with existing alternatives.

REFERENCES


DEL NEGRO, M., and F. OBIOLS, “Has Monetary Policy been so Bad that it is Better to Get Rid of It: The Case of Mexico,” Journal of Money, Credit and Banking 33 (2001), 404–33.


NEAL, R., “Erroneous Results in Marginal Likelihood from the Gibbs Output,” mimeo, University of Toronto, 1999.


MULTICOUNTRY VAR MODELS


QUERIES

Q1 Author: Please provide the spelled out form of “VAR.”

Q2 Author: Del Negro and Obiols (2003) has been changed to Del Negro and Obiols (2001) so that it matches reference list. Please confirm that is correct.

Q3 Author: Holtz Eakin et al. (1988) has been changed to Holtz-Eakin et al. (1988) so that it matches the reference list. Please confirm that this is correct.

Q4 Author: Please provide the spelled out form of “GMM” and “QML.”

Q5 Author: Dees et al. (2006) has been changed to Dees et al. (2005) so that it matches reference. Please confirm that is correct.

Q6 Author: Please provide the spelled out form of “SUR.”

Q7 Author: Forni and Reichlin (1998) has not been included in the Reference List, please supply full publication details.

Q8 Author: Please provide the spelled out form of “DSGE.”

Q9 Author: Please provide the spelled out form of “MA.”

Q10 Author: Please provide the spelled out form of “VARMA.”

Q11 Author: Gallant et al. (1996) has been changed to Gallant et al. (1993) so that it matches reference list. Please confirm that is correct.

Q12 Author: “Here $\delta_{t+1} = \ldots$” The meaning of this sentence is not clear; please rewrite or confirm that the sentence is correct.

Q13 Author: Kass and Raftery (1995) has been changed to Kass and Ratfery (1995) so that it matches reference list. Please confirm that is correct.
Q14  Author: Canova and Gambetti (2007) has not been cited in the text. Please indicate where it should be cited; or delete from the Reference List and renumber the References in the text and Reference List.

Q15  Author: Please provide volume number in Imbs et al. (2005).