Department of Economic Studies
University of Naples "Parthenope"

DES - dp

Discussion Paper

No.7/2007

Title: **Estimating Multi-country VAR models**

Author: * Fabio Canova, **Matteo Ciccarelli

Affiliation: * **Universitat Pompeu Fabra, Barcelona, Spain**
** **European Central Bank**

2007

Estimating Multi-country VAR models

Fabio Canova and Matteo Ciccarelli[1]

Abstract

This paper presents a method to estimate the coefficients, to test specification hypotheses and to conduct policy exercises in multi-country 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; 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 a MCMC routine. The transmission of certain shocks across countries is analyzed.

Key Words: Multi-country VAR, Markov Chain Monte Carlo methods, Flexible priors, International transmission.

JEL Classification nos: C3, C5, E5.

[1]This paper is forthcoming in the International Economic Review. We would like to thank three anonymous referess, F. Schorfheide (the editor), G. Amisano, T. Cogley, H. Van Djik, 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 comment and suggestions. The view expressed in this paper 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 paper has circulated under the title: Panel index VAR models: Specification, Estimation, Testing and Leading Indicators.

1 Introduction

There has been a growing interest over the last decade in using multi-country VAR models for applied macroeconomic analysis (see e.g. Canova and Marrinan 1998, Canova and De Nicolo' 2000, Del Negro and Obiols 2003, 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 multi-country setup differs from a multi-agent framework for several reasons. First, cross unit lagged interdependencies are likely to be important in explaining the dynamics of multi-country data. Second, heterogeneous dynamics are a distinctive feature of multi-country 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 non-standard. For example, the GMM estimator of Holtz Eakin, et al. (1988), 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, while 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 multi-country setups.

When dealing with multi-country 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 2006); 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 paper shows how to conduct inference in multi-country 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 which are common across units and variables (a "common" effect); variations which are specific to the unit (a "unit" effect), variations which are specific to a variable (a "variable" effect), etc. We complete the specifications using a hierarchical structure which 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 multi-country 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 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 which is suitable for a variety of policy purposes. For example, one can produce multi-step multi-country 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, while 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 while 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 multi-country 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 (i) a flexible coefficient factorization which renders estimation easy, (ii) a testing approach which makes model selection and inference tractable, (iii) a set of tools to conduct structural analyses and policy projection exercises.

The structure of the paper is a 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 multi-country VAR model we consider has the form:

$$y_{it} = D_{it}(L)Y_{t-1} + C_{it}(L)W_{t-1} + e_{it} \tag{1}$$

where $i = 1, ..., N;\ t = 1, ..., T;\ y_{it}$ is a $G \times 1$ vector of variables for each i, $Y_t = (y'_{1t}, y'_{2t}, \ldots y'_{Nt})'$, $D_{it,j}$ are $G \times GN$ matrices and $C_{it,j}$ are $G \times q$ matrices each j, W_t is a $q \times 1$ vector which 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 $D_t(L) \neq \Im \otimes \mathcal{D}_{it}(L)$ for some L, where $\Im$ is a $1 \times N$ vector with one in the *i-th* position and zero elsewhere. In words, if we stack the elements of $\mathcal{D}_{it,j}$ over i, we obtain a matrix which 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 δ_{it}^g be $k \times 1$ vectors containing, stacked, the G rows of the matrices D_{it} and C_{it}; define $\delta_{it} = (\delta_{it}^{1\prime}, \ldots, \delta_{it}^{G\prime})'$, and let $\delta_t = (\delta'_{1t}, \ldots, \delta'_{Nt})'$ be a $NGk \times 1$ vector. Whenever δ_{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 δ_t is factored as:

$$\delta_t = \sum_{f}^{F} \Xi_f \theta_{ft} + u_t \tag{2}$$

where $F << NGk$; θ_{ft} is a low dimensional vector, $\forall f$, Ξ_f are conformable matrices and u_t captures unmodelled and idiosyncratic variations present in δ_t. The typology of the factors θ_{ft} and the exact form of the Ξ_f's will become evident from the examples presented below.

Clearly, the choice of factorization is application and, possibly, sample dependent. While 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 while, when constructing indicators of GDP, one may want to specify, at least, a common, a country 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 $X_t = I_{NG} \otimes \mathbf{X}'_t$; where $\mathbf{X}_t = (Y'_{t-1},\ Y'_{t-2}, \ldots, Y'_{t-p},\ W'_t, \ldots, W'_{t-l})'$; set $\mathcal{X}_t \equiv X_t \Xi$;

$\Xi = [\Xi_1, \Xi_2, \Xi_3, \ldots, \Xi_F]$, $\zeta_t \equiv X_t u_t + E_t$, and let Y_t, E_t be $NG \times 1$ vectors, we can rewrite (1) as:

$$\begin{aligned} Y_t &= X_t \delta_t + E_t \\ &= X_t(\Xi\theta_t + u_t) + E_t \equiv \mathcal{X}_t \theta_t + \zeta_t \end{aligned} \tag{3}$$

In (3) we have reparametrized the original multi-country VAR so that the vector of endogenous variables depends on a small number of observable indices, $\mathcal{X}_{it}$, and the factors θ_{it} load on the indices. By construction, the $\mathcal{X}_{it}$'s are linear combinations of right hand side variables of the multi-country 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):

$$\begin{aligned} \begin{pmatrix} y_{11t} \\ y_{12t} \\ y_{21t} \\ y_{22t} \end{pmatrix} &= \begin{pmatrix} A_{1111} & A_{1112} & A_{1121} & A_{1122} \\ A_{1211} & A_{1212} & A_{1221} & A_{1222} \\ A_{2111} & A_{2112} & A_{2121} & A_{2122} \\ A_{2211} & A_{2212} & A_{2221} & A_{2222} \end{pmatrix} \begin{pmatrix} y_{11t-1} \\ y_{12t-1} \\ y_{21t-1} \\ y_{22t-1} \end{pmatrix} \\ &+ \begin{pmatrix} B_{1111} & B_{1112} & B_{1121} & B_{1122} \\ B_{1211} & B_{1212} & B_{1221} & B_{1222} \\ B_{2111} & B_{2112} & B_{2121} & B_{2122} \\ B_{2211} & B_{2212} & B_{2221} & B_{2222} \end{pmatrix} \begin{pmatrix} y_{11t-2} \\ y_{12t-2} \\ y_{21t-2} \\ y_{22t-2} \end{pmatrix} + \begin{pmatrix} e_{11t} \\ e_{12t} \\ e_{21t} \\ e_{22t} \end{pmatrix} \end{aligned} \tag{4}$$

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

Letting $i_1 = (1,1,1,1)'$, $i_2 = (1,1,0,0)'$ $i_3 = (0,0,1,1)'$, $i_4 = (1,0,1,0)'$ $i_5 = (0,1,0,1)'$,then

$$\delta = \begin{pmatrix} i_1 & 0 & 0 & 0 \\ 0 & i_1 & 0 & 0 \\ 0 & 0 & i_1 & 0 \\ 0 & 0 & 0 & i_1 \\ i_1 & 0 & 0 & 0 \\ 0 & i_1 & 0 & 0 \\ 0 & 0 & i_1 & 0 \\ 0 & 0 & 0 & i_1 \end{pmatrix} \theta_1 + \begin{pmatrix} i_2 & i_3 \\ i_2 & i_3 \\ i_2 & i_3 \\ i_2 & i_3 \\ i_2 & i_3 \\ i_2 & i_3 \\ i_2 & i_3 \\ i_2 & i_3 \end{pmatrix} \theta_2 + \begin{pmatrix} i_4 & i_5 \\ i_4 & i_5 \\ i_4 & i_5 \\ i_4 & i_5 \\ i_4 & i_5 \\ i_4 & i_5 \\ i_4 & i_5 \\ i_4 & i_5 \end{pmatrix} \theta_3 + \begin{pmatrix} i_1 & 0 \\ i_1 & 0 \\ i_1 & 0 \\ i_1 & 0 \\ 0 & i_1 \\ 0 & i_1 \\ 0 & i_1 \\ 0 & i_1 \end{pmatrix} \theta_4 + u \tag{5}$$

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

$$y_{11t} = \theta_{11}\mathcal{X}_{1t} + \theta_{21}\mathcal{X}_{2t} + \theta_{22}\mathcal{X}_{3t} + \theta_{31}\mathcal{X}_{4t} + \theta_{32}\mathcal{X}_{5t} + \theta_{41}\mathcal{X}_{6t} + \theta_{42}\mathcal{X}_{7t} + \zeta_t \tag{6}$$

where $\mathcal{X}_{1t} = \sum_i \sum_g \sum_j y_{igt-j}$, $\mathcal{X}_{2t} = \sum_g \sum_j y_{1gt-j}$, $\mathcal{X}_{3t} = \sum_g \sum_j y_{2gt-j}$, $\mathcal{X}_{4t} = \sum_i \sum_g y_{i1t-j}$, $\mathcal{X}_{5t} = \sum_i \sum_g y_{i2t-j}$, $\mathcal{X}_{6t} = \sum_i \sum_g y_{igt-1}$, $\mathcal{X}_{7t} = \sum_i \sum_g y_{igt-2}$. Therefore, $\mathcal{X}_{1t}$ captures the information contained in the lags of all the variables of the model, $\mathcal{X}_{2t}$ ($\mathcal{X}_{3t}$) the information contained in the lags of the variables for country 1 (country 2), $\mathcal{X}_{4t}$ ($\mathcal{X}_{5t}$) the information contained in the lags of variable 1 (variable 2) and lags, while $\mathcal{X}_{6t}$ ($\mathcal{X}_{7t}$) 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

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

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

where β are structural parameters, $A(\beta), B(\beta), C(\beta)$ are time invariant matrices whose entries are nonlinear functions of β; y_{1t} is a state and y_{2t} a control, both of them are assumed to be scalar, for simplicity. The dimension of ϵ_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. (7) and (8) can be written as a structural VAR(1) model:

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

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

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

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

$$\delta = \begin{pmatrix} H_{11} \\ H_{12} \\ H_{21} \\ H_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \theta_1 + \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \theta_2 \equiv \Xi_1\theta_1 + \Xi_2\theta_2$$

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

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

and its SUR reparametrization is:

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

$$y_{2t} = \theta_{12}(y_{1t-1} + y_{2t-1}) + \theta_{22}y_{2t-1} + e_{2t} \quad (10)$$

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

$$\begin{aligned} y_{it} &= \alpha_{it} + T_t \qquad (1-\rho_t L)T_t = e_t \\ \alpha_{it} &= \alpha_i + v_{it} \qquad (1-\omega_i L)v_{it} = z_{it} \\ \alpha_i &= \alpha_0 + \epsilon \end{aligned} \tag{11}$$

where e_t is iid across t, v_{it} is iid across t and y_{it} is a $G \times 1$ vector for each $i = 1, 2, \dots, N$. This model has the following VAR representation

$$Y_t = \alpha_{0t}^* + A_t Y_{t-1} + B_t Y_{t-2} + \eta_t \tag{12}$$

$$= \alpha_{0t}^* + \delta_t X_t + \eta_t \tag{13}$$

where Y_t is a $NG \times 1$ vector each t, $\alpha_{0t}^* = diag\{(1-\omega_i)\}(1-\rho_t)\alpha_0, \;\; \eta_{it} = (1-\omega_i L)e_t + (1-\omega_i L)(1-\rho_t L)\epsilon + (1-\rho_t L)z_{it}$ while $A_{it} = \rho_t + \omega_i$ and $B_{it} = \rho_t \omega_i$. Therefore, an error component model generates a particular error structure in the VAR. Note that α_{0t}^* are time trends common to all the G variables for unit i. Suppose $\delta_t = [vec(A_t), vec(B_t)]$ is factored as

$$\delta_{tigj} = \Xi_1 \theta_{1t} + \Xi_2 \theta_{2i} + \Xi_3 \theta_{3g} + u_{tigj}^{\delta} \tag{14}$$

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

$$\alpha_{0it}^* = \Xi_4 \theta_{4it} + u_{jit}^{\alpha} \tag{15}$$

where θ_{4it} is a $NT \times 1$ vector. (14)-(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 θ_{4it} time or unit independent and exploit averages in the remaining dimensions to construct the appropriate regressors. Disregarding how α_{0t}^* is parametrized, the SUR model is

$$(Y_t - \alpha_{0t}^*) = \theta_{1t}\mathcal{X}_{1t} + \theta_{2i}\mathcal{X}_{2t} + \theta_{3g}\mathcal{X}_{3t} + \zeta_t$$

where $\mathcal{X}_{1t} = \Xi_1 X_t$ is a time index, $\mathcal{X}_{2t} = \Xi_1 X_t$ is a unit index, $\mathcal{X}_{3t} = \Xi_3 X_t$ is a variable index, and ζ_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 over-parametrization of the original multi-country VAR is dramatically reduced. In fact, in the resulting SUR model, estimation and specification searches are constrained only by the dimensionality of θ_t (δ_t is integrated out). A second advantage is that, given the MA nature of many $\mathcal{X}_{it}$, 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 θ_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 θ_{1t} captures information which is common to all the coefficients of the VAR, $\mathcal{X}_{1t}\theta_{1t}$ is an indicator for Y_t based on common information. Indicators containing other types of information can also be easily constructed. Since

$\mathcal{X}_{it}$ are predetermined, leading versions of these indicators can be obtained projecting θ_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) is restrictive and suggested the possibility to estimate the Ξ'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 Ξ_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 $\Xi_1 = (\sigma_1^{-1},\ldots,\sigma_1^{-1},\sigma_2^{-1},\ldots,\sigma_2^{-1},\ldots)$ where σ_1 and σ_2 measure the average standard deviation of the variables in countries 1 and 2. The idea of estimating the Ξ'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 Ξ_i as unknown and employs the factor models techniques described below. Given our emphasis on multi-country 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

$$(y_{t+1} - \alpha) = \gamma(L)(y_t - \alpha) + \beta(L)f_t + e_{t+1} \tag{16}$$

$$X_{it} = \lambda_i(L)f_t + u_t \tag{17}$$

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

$$(y_{t+1} - \alpha) = \gamma(L)(y_t - \alpha) + \beta F_t + e_{t+1} \tag{18}$$

$$X_t = \Lambda F_t + u_t \tag{19}$$

where $F_t = (f_t', \ldots, f_{t-q}')'$ is an $s \times 1$ vector, $s \leq (q+1)r$, the i-th row of Λ is $(\lambda_{i0}, \ldots \lambda_{iq})$ and $\beta = (\beta_0, \ldots, \beta_q)'$. While dynamic, (18)-(19) can be estimated with static principal components techniques: the loadings Λ are the first s eigenvalues of the $X'X$ matrix, where X is the $T \times N$ data matrix and $\hat{F} = \frac{X'\hat{\Lambda}}{N}$.

Since (18)-(19) is 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 capture the volatility of the data matrix X_t, while our indices extract comovements in series belonging to X_t. Second, ours indices are observable, while the factors in (18)-(19) are unobservable and need to be estimated with a data driven approach. Third, while 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)-(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

$$\phi(L)f_t = u_t \tag{20}$$

where $\phi(L)$ is assumed to be diagonal for each L and, typically, $corr(u_{jt}, u_{j't}) = 0$, $j = 1, \ldots, r$, 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 a UF model can be written as a VARMA. In fact, substituting (20) into (16) we have that

$$(I - \gamma(L)L)(y_{t+1} - \alpha) = \beta\,(L)\,\phi^{-1}(L)u_t + e_{t+1} \tag{21}$$

Hence, as long as $\phi(L)$ has a convergent representation, a VAR for y_t exist. Note that the error term has two components: one due to shocks to the common factors, 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 a 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, rather than 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:

$$\begin{aligned} \theta_t &= (I - \mathcal{C})\,\bar{\theta} + \mathcal{C}\theta_{t-1} + \eta_t & \eta_t &\sim (0, B) \\ \bar{\theta} &= \mathcal{P}\mu + \epsilon & \epsilon &\sim (0, \Psi) \end{aligned} \tag{22}$$

where $\bar{\theta}$ is the unconditional mean of θ_t; $\mathcal{P}, \mathcal{C}$ are known matrices; η_t and ϵ are mutually independent and independent of E_t and u_t; and $B = diag(\bar{B}_1, \dots \bar{B}_F)$. 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 Ω is a $NG \times NG$ 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 $\mathcal{C}$ is arbitrary, many patterns are allowed in the specification. While we treat $\mathcal{C}$ 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 θ_t is for all purposes satisfactory, we do not follow such an approach here.

Whenever $\mathcal{C} \neq I$, 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, while the block diagonality of B is needed to guarantee the identifiability of the factors.

We specify normal distributions for E_t, u_t, η and ϵ, but it is easy to allow for fat tails if aberrant or non-normal observations are presumed to be present. For example, we could let $(u_t \mid z_t) \sim \mathcal{N}(0, z_t(\Omega \otimes V))$ where $z_t^{-1} \sim \chi^2(\nu, 1)$, since, unconditionally, $u_t \sim t_\nu(0, \Omega \otimes V)$. 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 non-normal patterns in y_t, if this is needed.

Numerous specifications are nested in our model: for example, a factor is time invariant when $B_{it} = 0$ and the appropriate elements of $\mathcal{C}$ are set to zero; no exchangeability obtains when Ψ is large, exact pooling obtains when $\Psi = 0$, 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 heteroschedasticity. With a prior for $\bar{\theta}$, posterior estimates would be straightforward to construct.

When the θ_t's are time-varying, MCMC methods can be employed to construct their exact posterior distributions. The likelihood of the reparametrized SUR model is

$$\mathcal{L}(\theta, \Upsilon | Y) \propto \prod_t |\Upsilon_t|^{-1/2} \exp\left[-\frac{1}{2}\sum_t (Y_t - \mathcal{X}_t\theta_t)' \Upsilon_t^{-1} (Y_t - \mathcal{X}_t\theta_t)\right]$$

where $\Upsilon_t = \left(1 + \sigma^2 \mathbf{X}_t' \mathbf{X}_t\right)\Omega \equiv \sigma_t \Omega$. To calculate the posterior for the unknowns we need prior distributions for $\left(\mu,\ \Psi^{-1}, \Omega^{-1},\ \sigma^{-2},\ B^{-1}\right)$. 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 $p\left(\mu,\ \Psi^{-1}, \Omega^{-1},\ \sigma^{-2},\ B^{-1}\right) = p\left(\mu\right) p\left(\Psi^{-1}\right) p\left(\Omega^{-1}\right) p\left(\sigma^{-2}\right) \prod_f p\left(B_f^{-1}\right)$ where

$$\begin{aligned} p(\mu) &= \mathcal{N}(\bar{\mu}, \Sigma_\mu) \qquad\qquad p(\Psi^{-1}) = \mathcal{W}(z_0, Q_0) \\ p(\Omega^{-1}) &= \mathcal{W}(z_1, Q_1) \qquad\qquad p(\sigma^{-2}) = \mathcal{G}\left(\frac{a_1}{2}, \frac{a_2}{2}\right) \\ p(B_f^{-1}) &= \mathcal{W}(z_{2f}, Q_{2f}) \qquad\qquad f = 1, \ldots, F \end{aligned}$$

Here $\mathcal{N}(\cdot)$ stands for Normal, $\mathcal{W}(\cdot)$ for Wishart and $\mathcal{G}(\cdot)$ for Gamma distributions. The hyperparameters $(z_0, z_1, z_{2f}, a, b, vec(\bar{\mu}), vech(\Sigma_\mu), vech(Q_0, Q_1, Q_{2f}))$ are treated as fixed, where $vec(\cdot)$ $(vech(\cdot))$ denotes the column-wise vectorization of a rectangular (symmetric) matrix. Non-informative priors are obtained setting $a, b \to 0$, $Q_f^{-1} \to 0, \Sigma_\mu^{-1} \to 0$ 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 $Y^T = (Y_1, ..., Y_T)$ denote the data, $\psi = (\mu,\ \Psi^{-1}, \Omega^{-1},\ \sigma^{-2}, B_f^{-1}, \bar{\theta}, \{\theta_t\})$ the unknowns whose joint distribution needs to be found, and $\psi_{-\alpha}$ the vector of ψ excluding the parameter α. Let $\theta_{t-1}^* = (I - \mathcal{C})\,\bar{\theta} + \mathcal{C}\theta_{t-1}$ and $\tilde{\theta}_t = \theta_t - \mathcal{C}\theta_{t-1}$. Given Y^T, the conditional posteriors for the unknowns are:

$$
\begin{aligned}
\mu & \mid Y^T, \psi_{-\mu} \sim \mathcal{N}\left(\hat{\mu}, \hat{\Sigma}_\mu\right) \\
\Psi^{-1} & \mid Y^T, \psi_{-\Psi} \sim \mathcal{W}\left(z_0 + 1, \hat{Q}_o\right) \\
\Omega^{-1} & \mid Y^T, \psi_{-\Omega} \sim \mathcal{W}\left(z_1 + T, \hat{Q}_1\right) \\
B_f^{-1} & \mid Y^T, \psi_{-\bar{B}_f} \sim \mathcal{W}\left(T * \dim\left(\theta_t^f\right) + z_{2f}, \hat{Q}_{2f}\right) \\
\sigma^{-2} & \mid Y^T, \psi_{-\sigma^2} \propto (\sigma^{-2})^{-a_1 - 1} \exp\{a_2 \sigma^{-2}\} \times \mathcal{L}(\theta, \Upsilon | Y^T) \\
\bar{\theta} & \mid Y^T, \psi_{-\bar{\theta}} \sim \mathcal{N}\left(\widehat{\bar{\theta}}, \hat{\Psi}\right)
\end{aligned} \tag{23}
$$

where

$$
\begin{aligned}
\hat{\mu} &= \hat{\Sigma}_\mu \left(\mathcal{P}' \Psi^{-1} \bar{\theta} + \Sigma_\mu^{-1} \bar{\mu}\right); \\
\hat{\Sigma}_\mu &= \left(\mathcal{P}' \Psi^{-1} \mathcal{P} + \Sigma_\mu^{-1}\right)^{-1}; \\
\hat{Q}_o &= \left[Q_o^{-1} + \left(\bar{\theta} - \mathcal{P}\mu\right)\left(\bar{\theta} - \mathcal{P}\mu\right)'\right]^{-1}; \\
\hat{Q}_1 &= \left[Q_1^{-1} + \sum_t \left(Y_t - \mathcal{X}_t \theta_t\right) \sigma_t^{-1} \left(Y_t - \mathcal{X}_t \theta_t\right)'\right]^{-1}; \\
\hat{Q}_{2f} &= \left[Q_{2f}^{-1} + \sum_t \left(\theta_t^f - \theta_{t-1}^{*f}\right)\left(\theta_t^f - \theta_{t-1}^{*f}\right)'\right]^{-1}; \\
\widehat{\bar{\theta}} &= \hat{\Psi}\left[\Psi^{-1} \mathcal{P}\mu + (I - \mathcal{C})' \bar{B}^{-1} \sum_t \tilde{\theta}_t\right]; \\
\hat{\Psi} &= \left[\Psi^{-1} + (I - \mathcal{C})' \bar{B}^{-1} (I - \mathcal{C}) \sum_t 1\right]^{-1};
\end{aligned}
$$

θ_t^f refers to the f-th sub vector of θ_t, and $\dim\left(\theta_t^f\right)$ to its dimension.

The conditional posterior of $(\theta_1, ..., \theta_T \mid Y^T, \psi_{-\theta_t})$, 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

$$\begin{aligned}
\theta_{t|t} &= \theta^*_{t-1|t-1} + (R^*_{t|t-1}\mathcal{X}_t F^{-1}_{t|t-1})(Y_t - \mathcal{X}_t\theta_t) \\
R_{t|t} &= \left(I - (R^*_{t|t-1}\mathcal{X}_t F^{-1}_{t|t-1})\mathcal{X}_t\right)(R^*_{t-1|t-1} + \bar{B}) \\
F_{t|t-1} &= \mathcal{X}_t R^*_{t|t-1}\mathcal{X}'_t + \Upsilon_t
\end{aligned} \tag{24}$$

where $\theta^*_{t-1|t-1}$ and $R^*_{t-1|t-1}$ are, respectively, the mean and the variance covariance matrix of the conditional distribution of $\theta_{t-1|t-1}$. To obtain a sample $\{\theta_t\}$ from the joint posterior distribution $(\theta_1, ..., \theta_T \mid Y^T, \psi_{-\theta_t})$, the output of the Kalman filter is used to simulate θ_T from $\mathcal{N}(\theta_{T|T}, R_{T|T})$, θ_{T-1} from $\mathcal{N}(\theta_{T-1}, R_{T-1})$, and θ_1 from $\mathcal{N}(\theta_1, R_1)$, where $\theta_t = \theta_{t|t} + R_{t|t}R^{-1}_{t+1|t}\left(\theta_{t+1} - \theta_{t|t}\right)$, and $R_t = R_{t|t} - R_{t|t}R^{-1}_{t+1|t}R_{t|t}$. The recursions can be started choosing $R_{0|0}$ to be diagonal with elements equal to small values, while $\theta_{0|0}$ can be estimated in the training sample or initialized using a constant coefficient version of the model.

Since the conditional posterior of σ^2 is non-standard, a Metropolis step is needed to obtain draws for this parameter. We assume that a candidate $(\sigma^2)^\dagger$ is generated via $(\sigma^2)^\dagger = (\sigma^2)^l + v$, where v is a normal random variable with mean zero and variance 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 $(\sigma^2)^l$ and c^2 is selected to achieve a certain acceptance rate.

Draws from the posterior distributions can be obtained cycling through the conditional in (23)-(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 $\Upsilon_t = (1+\sigma^2 \mathbf{X}_t' \mathbf{X}_t)\Omega$, the prior distribution for the forecast error $\zeta_t = Y_t - \mathcal{X}_t \theta_t$ has the form $(\zeta_t | \sigma^2) \sim \mathcal{N}(0, \sigma_t \Omega)$. Therefore, unconditionally, ζ_t has a multivariate t distribution centered at 0, scale matrix proportional to Ω and ν_ζ degrees of freedom, and the innovations of (3) are endogenously allowed to have fat tails. To capture conditional heteroschedasticity in y_t, Cogley and Sargent (2005) specify Ω 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 which 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 δ_t is integrated out, it is not possible to easily account for them. One extreme possibility would be to specify a process for Δu_t, 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 $\frac{1}{L}[\sum_{\ell=\bar{L}+1}^{\bar{L}+L} \mathcal{G}(\psi^{\ell})]$ (the first $\bar{L}$ observations represent a burn-out sample discarded in the calculation). Predictive distributions for future y_{it}'s can be estimated using the recursive nature of the model and the conditional structure of the posterior. Let $Y^{t+\tau} = (Y_{t+1}, \dots, Y_{t+\tau})$, consider the conditional density of $Y^{t+\tau}$, given the data up to t, and a function $\mathcal{G}(Y^{t+\tau})$. Then

$$\mathcal{F}\left(\mathcal{G}(Y^{t+\tau}) \mid Y_t\right) = \int \mathcal{F}\left(\mathcal{G}(Y^{t+\tau}) \mid Y^t, \psi\right) p\left(\psi \mid Y^t\right) d\psi$$

and, e.g., forecasts for $Y^{t+\tau}$ can be obtained drawing $\psi^{(\ell)}$ from the posterior distribution and simulating the vector $Y^{\ell,t+\tau}$ from the density $\mathcal{F}\left(Y^{t+\tau} \mid Y_t, \psi^{(\ell)}\right)$. Turning point distributions can also be constructed by appropriately choosing $\mathcal{G}$. 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 multi-country VAR, or to verify general hypotheses on the type of indices to be included. To discriminate across models with different indices consider

$$\mathcal{L}(Y^t|M_h) = \int \mathcal{F}(Y^t|\psi_h, M_h)p(\psi_h|M_h)d\psi_h \tag{25}$$

which is the marginal likelihood for Y^t in a model with h indices. Here $p(\psi_h|M_h)$ is the prior density for ψ in model M_h and $\mathcal{F}(Y^t|\psi_h, M_h)$ the density of the data under the parameterization produced by M_h. (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 computation 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 like 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

$$\mathcal{B}_{hh'} \equiv \frac{\mathcal{L}(Y^t|M_h)}{\mathcal{L}(Y^t|M_{h'})} \tag{26}$$

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 ψ_h unrestricted and $\psi_{h'} = (....., \sigma^2 = 0, ...)$; or whether there are time variations in θ_t, setting $\bar{B}_f = b_f * I$, letting ψ_h be unrestricted and $\psi_{h'} = (....., b_f = 0, ...)$ 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.

Rather than 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 $\mathcal{B}_{h1}$ for each M_h. The posterior probability of model h is $p(M_h|Y^t) = \frac{a_h \mathcal{B}_{h1}}{\sum_{h=2}^{H} a_h \mathcal{B}_{h1}}$, where a_h are the prior odds for M_h, and model uncertainty can be accounted for weighting $\mathcal{G}(\psi_h)$ by $p(M_h|Y^t)$.

5 Dynamic analysis

Dynamic analysis is non-standard 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 a 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^t) and of the factors (θ^t), 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, while in the other the current shocks is set to its unconditional value (see also Gallant et al. 1996; Koop, et al. 1996). We condition on future shocks rather than 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. While 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. To formally define impulse responses we need some notation. The reparametrized SUR is:

$$\begin{aligned} y_t &= \mathcal{X}_t\theta_t + (E_t + X_t u_t) \\ \theta_t &= (I - \mathcal{C})(\mathcal{P}\mu + \epsilon) + \mathcal{C}\theta_{t-1} + \eta_t \end{aligned}$$

where $\theta_t = [\theta'_{1t}, \theta'_{2t}, \ldots, \theta'_{Ft}]'$, $\mathcal{X}_t = [\mathcal{X}_{1t}, \ldots, \mathcal{X}_{Ft}], \mathcal{X}_{it} = \Xi_i X_t,\ X_t = [Y_{t-1}, W_t]$. Let $\mathcal{U}_t = [(E_t + X_t u_t)', \eta'_t, \epsilon']'$ be the vector of reduced form shocks and $\mathcal{Z}_t = [H_t^{-1}(E_t + X_t u_t)', H_t^{-1}\eta'_t, H_t^{-1}\epsilon']'$ be the vector of structural shocks where $E_t = H_t v_t,\ H_t H'_t = \Omega$ so that $\text{var}(v_t) = I$ and $H_t = J * K_t$

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 while 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 Ω depend on t. Also, there may be situations where the covariance matrix of reduced form shocks is time invariant but the contemporaneous relationships of the structural model are time-varying.

Let $\mathcal{V}_t = (\Omega, \sigma^2, B_t, \Psi)$, let $\bar{\mathcal{Z}}_{j,t}$ be a particular realization of $\mathcal{Z}_{j,t}$ and $\mathcal{Z}_{-j,t}$ indicate the structural shocks, excluding the one in the $j-th$ component. Let $\mathcal{F}_t^1 = \{Y^{t-1}, \theta^t, \mathcal{V}_t, H_t, \mathcal{Z}_{j,t} = \bar{\mathcal{Z}}_{j,t}, \mathcal{Z}_{-j,t}, \mathcal{U}_{t+1}^{t+\tau}\}$ and $\mathcal{F}_t^2 = \{Y^{t-1}, \theta^t, \mathcal{V}_t, H_t, \mathcal{Z}_{j,t} = E\mathcal{Z}_{j,t}, \mathcal{Z}_{-j,t}, \mathcal{U}_{t+1}^{t+\tau}\}$ be two conditioning sets. Then responses to a shock at t in the $j-th$ component of $\mathcal{Z}_t$ are obtained as

$$IR(t, t+\tau) = E(Y_{t+\tau}|\mathcal{F}_t^1) - E(Y_{t+\tau}|\mathcal{F}_t^2) \qquad \tau = 1, 2, \ldots \tag{27}$$

To see what definition (27) involves rewrite the original VAR model (1) in a companion form

$$Y_{t+\tau} = A_{t+\tau} Y_{t+\tau-1} + C_{t+\tau} W_{t+\tau-1} + E_{t+\tau} \tag{28}$$

and let

$$\delta_{t+\tau} = \Xi[(I - \mathcal{C})(\mathcal{P}\mu + \epsilon) + \mathcal{C}\theta_{t+\tau-1} + \eta_{t+\tau}] + u_{t+\tau} \tag{29}$$

Here $\delta_{t+\tau} = [vec(A_{1t+\tau}), vec(C_{t+\tau})]$ and $A_{1t+\tau}$ is the first row of $A_{t+\tau}$. Taking $Y^{t-1} = (Y_{t-1}, Y_{t-2}, \ldots, W_{t-1}, W_{t-2}, \ldots,)$, $A^t = (A_t, A_{t-1}, \ldots), C^t = (C_t, C_{t-1}, \ldots)$ and $H_{t+\tau} = H_t \ \forall\tau$ as given, and solving backward we can write (28) and (29) we have

$$
\begin{aligned}
Y_{t+\tau} &= (\prod_{k=0}^{\tau} A_{t+\tau-k})Y_{t-1} + C_{t+\tau}W_{t+\tau-1} + \sum_{h=1}^{\tau}(\prod_{k=0}^{h-1} A_{t+\tau-k})C_{t+\tau-h}W_{t+\tau-h-1} \\
&+ H_{t+\tau}v_{t+\tau} + \sum_{h=1}^{\tau}(\prod_{k=0}^{h-1} A_{t+\tau-k})H_{t+\tau-h}v_{t+\tau-h} \qquad (30)
\end{aligned}
$$

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

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

$$
\begin{aligned}
IR(t, t+\tau) &= E_t[Y_{t+\tau}|Y^{t-1}, A^t, C^t, \mathcal{V}_t, H_t, \{\bar{v}_{jt+m}\}_{k=0}^m, \{v_{-jt+k}\}_{k=0}^m, \{v_{t+k}\}_{k=m+1}^{\tau}] \\
&- E_t[Y_{t+\tau}|Y^{t-1}, A^t, C^t, \mathcal{V}_t, H_t, \{v_{t+k}\}_{k=0}^{\tau}] \\
&= E_t[(\prod_{k=0}^{\tau-1} A_{t+\tau-k})^j H_t^j(\bar{v}_{jt} - Ev_{jt}) + (\prod_{k=0}^{\tau-2} A_{t+\tau-k})^j H_{t+1}^j(\bar{v}_{jt+1} - Ev_{jt+1}) + \ldots \\
&+ (\prod_{k=0}^{\tau-m-1} A_{t+\tau-k})^j H_{t+m}^j(\bar{v}_{jt+m} - Ev_{jt+m})] \qquad (32)
\end{aligned}
$$

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 η_t are correlated, both the sign and the size of the shocks matter - a shock in v_t may induce changes in A_t or C_t.

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

1. Choose a t, a τ and an J_t. Draw $\Omega^l = H_t^l(H_t^l)', (\sigma^2)^l$ from their posterior distribution and u_t^l from $\mathcal{N}(0, (\sigma^2)^l I \otimes H_t^l(H_t^l)')$. Compute $y_t^l = \mathcal{X}_t\theta_t + H_t\bar{v}_t + X_t u_t^l$.

2. Draw $\Omega^l = H_{t+1}^l(H_{t+1}^l)', (\sigma^2)^l, B_{t+1}^l, \Psi^l$. Draw η_{t+1}^l, ϵ^l from their posterior distribution. Use

the law of motion of the factors to compute $\theta^l_{t+1},\ l = 1, \ldots, L$ and the definition of Ξ to compute $\mathcal{X}_{t+1}$. Draw u^l_{t+1} from $N(0, (\sigma^2)^l I \otimes H^l_{t+1}(H^l_{t+1})')$ and compute $y^l_{t+1} = \mathcal{X}_{t+1}\theta^l_{t+1} + H_{t+1}\bar{v}_{t+1} + X_{t+1}u^l_{t+1}$, $l = 1, \ldots, L$.

3. Repeat step 2. and compute $\theta^l_{t+k}, y^l_{t+k},\ k = 2, \ldots \tau$.

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

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

so that

$$\begin{aligned} IR(t, t+\tau) &= E_t[\prod_{k=0}^{\tau}(\bar{A}_{t+1\tau-k} - A_{t+\tau-k})Y_{t-1} + \sum_{h=1}^{\tau}\prod_{k=0}^{h-1}(\bar{A}_{t+1\tau-k} - A_{t+\tau-k})C_{t+\tau-h}W_{t+\tau-h-1} \\ &+ \sum_{h=1}^{\tau}\prod_{k=0}^{h-1}(\bar{A}_{t+1\tau-k} - A_{t+\tau-k})H_{t+\tau-h}v_{t+\tau-h}] \end{aligned} \tag{33}$$

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 $\bar{W}_{t+k},\ k=0,1,\ldots,m$. Then the response of $Y_{t+\tau}$ is

$$IR(t,t+\tau) = E[(\prod_{k=0}^{\tau-2} A_{t+\tau-k})C_{t+1}(\bar{W}_{jt}-W_{jt}) + (\prod_{k=0}^{\tau-3} A_{t+\tau-k})C_{t+2}(\bar{W}_{jt+1}-W_{jt+1}) \quad (34)$$

$$+ \ldots + (\prod_{k=0}^{\tau-2-m} A_{t+\tau-k})C_{t+m+1}(\bar{W}_{jt+m}-W_{jt+m})] \quad (35)$$

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

1. Choose a t, a τ, a path $\{\bar{W}_{t+k}\}_{k=0}^{m}$. Draw $\Omega^l, (\sigma^2)^l$ from their posterior, draw $E_t^l + X_t u_t^l$ and compute y_t^l.

2. Draw $(B_t)^l, \Psi^l$ from their posterior distribution, draw η_{t+1}^l, ϵ^l and use the law of motion of the factors to draw $\theta_{t+1}^l,\ = 1,\ldots,L$ and the definition of Ξ to compute $\mathcal{X}_{t+1}$. Draw $E_{t+1}^l + X_{t+1}u_{t+1}^l$ and compute $y_{t+1}^l = \mathcal{X}_{t+1}\theta_{t+1}^l + (E_{t+1}^l + X_{t+1}u_{t+1}^l)$, $l=1,\ldots,L$.

3. Repeat steps 2. and compute $\theta_{t+k}^l, y_{t+k}^l,\ k=2,\ldots,\tau$.

4. Repeat steps 1.-3. setting $W_{t+k} = E(W_{t+k}),\ k=0,1,\ldots,m$, using the draws for the shocks in 1.-3.

Consider finally the case where the future path of a subset of Y_t's is fixed. For example, in a system with output growth, inflation and the nominal rate we would like to condition on a given path for the future interest rate. Partition $Y_t = A_tY_{t-1} + C_tW_{t-1} + E_t$ into two blocks, let $Y_{2t+k} = \bar{Y}_{2t+k}$ be the fixed variables and Y_{1t+k} those allowed to adjust. Then:

$$IR(t,t+\tau) = E[H_t^1(\prod_{k=0}^{\tau-1} A_{t+\tau-k})^1(\bar{v}_{2t}-v_{2t}) + H_{t+1}^1(\prod_{k=0}^{\tau-2} A_{t+\tau-k})^1(\bar{v}_{2t+1}-v_{2t+1}) \quad (36)$$

$$+ \ldots + H_{t+m}^1(\prod_{k=0}^{\tau-1-m} A_{t+\tau-k})^1(\bar{v}_{2t+m}-v_{2t+m})] \quad (37)$$

where $\bar{v}_{2t+k} = \bar{Y}_{2t+k} - A_{21t+k}Y_{1t-k-1} - A_{22t+k}Y_{2t-k-1} - C_{2t+k}W_{t+k-1}$ and the superscript 1 refers to the first row of the matrix. Hence, to compute this type of conditional forecasts we need to:

1. Partition $y_t = (y_{1t}, y_{2t})$, choose a t, and a path $\{y_{2t+k}\}_{k=0}^{\tau}$. Use the model to solve for the $\bar{v}_{2t}$ that gives $y_{2t} = \bar{y}_{2t}$ and back out the implied y_{1t}^l once draws for E_{1t}^l and u_t^l are made from their conditional posterior distribution. Draw η_{t+1}^l, ϵ^l, use the law of motion of the factors to obtain $\theta_{t+1}^l,\ l = 1, \ldots, L$ and the definition of Ξ to compute $\mathcal{X}_{t+1}$.

2. Use the model to solve for $\bar{v}_{2t+1}^l$ that gives $y_{2t+1}^l = \bar{y}_{2t+1}$ and back out the implied y_{1t+1}^l once draws for E_{1t+1}^l and u_{t+1}^l are made. Draw η_{t+2}^l, ϵ^l and use the law of motion of the factors to compute $\theta_{t+2}^l,\ l = 1, \ldots, L$ and the definition of Ξ to compute $\mathcal{X}_{t+2}$.

3. Repeat step 2. and compute $\theta_{t+k}^l, y_{t+k}^l,\ k = 2, 3 \ldots$.

4. Repeat steps 1.-3. setting $v_{2t+k} = E(v_{t+k}),\ \forall k$ using the draws for the shocks in 1.-3.

In step 2 of all algorithms we have implicitly assumed that selecting a path for the shocks does not alter the law of motion of the factors, nor it alters the beliefs about the true structural shocks (here H_t is kept fixed in the calculations). If this were not the case, an intermediate step, where a run of the Kalman filter updates the information about the factors, needs to be used.

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 US 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. Rather, we want to show how the tools we describe in the paper could be applied to questions which are of crucial interest for applied business cycle investigators in academics and central banks.

The last twenty 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 US 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 4 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*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 30×28 regression parameters plus 406 covariance parameters to be estimated at each t.

We assume that the coefficient vector δ_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 θ_{1t} a 2×1 vector of common factors, one for euro area variables and one for the rest of the world, so that $\Xi_{11t} = \sum_{US,JP,CA,UK}\sum_g\sum_j y_{igt-j}$, $\Xi_{12t} = \sum_{GE,IT,FR}\sum_g\sum_j y_{igt-j}$, θ_{2t} is a 7×1 vector of country specific factors and $\Xi_{2it} = \sum_g\sum_j y_{igt-j}$, $i = 1, \ldots, 7$; θ_{3t} is a 4×1 vector of variable specific factors and $\Xi_{3gt} = \sum_i\sum_j y_{igt-j}$, $g = 1, \ldots, 4$. We also set $\mathcal{C} = I$. Hence, $\theta_t = (\theta'_{1t}, \theta'_{2t}, \theta'_{3t})'$ is 13×1 vector and the estimated model is

$$\begin{aligned} y_t &= \mathcal{X}_{1t}\theta_{1t} + \mathcal{X}_{2t}\theta_{2t} + \mathcal{X}_{3t}\theta_{3t} + \zeta_t \\ \theta_t &= \theta_{t-1} + \eta_t \end{aligned} \tag{38}$$

Since our sample is relatively short, no training sample is available to tune the prior up. To minimize the influence of our prior choices we select relatively loose but proper priors and set $p(b_i^{-1}) = \mathcal{G}(5, 0.5)$, $i = 1, 2, 3$ and $p(\Omega^{-1}) = \mathcal{W}((z_1\Omega_{OLS})^{-1}, z_1)$, where Ω_{OLS} is the OLS estimate of the Ω obtained on a fixed coefficient version of the model, and the degrees of freedom are chosen to approximately match the sample size, i.e., $z_1 = ng + 50$. We set $\theta_{0|0}$ to be equal to the OLS estimate obtained on the time invariant version of the model, and set $R_{0,0}$ 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 5 different specification. In all of them the coefficient factorization is exact, i.e. $\sigma^2 = 0$, since specifications which which 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 θ_{2t} or the variable specific components θ_{3t}, 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 θ_t 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 θ_t) evaluated at 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.[2] 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.

Table 1. Log Marginal Likelihood of models

Method	M_1	M_2	M_3	M_4	M_5
Chib's ML	-1200	-1236	-2908	-1548	-1579
nse	230	245	578	374	330
Harmonic Mean	-1589	-1636	-1627	-1608	-1619
nse	7	8	6	8	8
Max Loglike	-1530	-1617	-1610	-1580	-1595
nse	13	17	12	16	15
Parameters	409	409	406	408	408

Note: The number of parameters is equal to free elements in B + free elements in Ω.

The ranking of the models differ across methods. With Chib's measure, the basic model (M1) is clearly preferred, while the model which excludes time variations is clearly the worst. On the other hand, since a model with no variables 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) while 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 which excludes interdependences is now worse than a model which excludes time variations.

[2] As it is well known (Kass and Raftery, 1995), the harmonic mean converges almost surely to the correct value but does not generally satisfy a Gaussian central limit theorem. The measure can therefore be unstable, but it has proven to provide reliable estimates (Newton and Raftery, 1994). We prefer a simple harmonic mean to a modified one (e.g. Gelfand and Day, 1994) because for high-dimensional problems, it is hard to find an appropriate modification function and results can be poor (e.g. Chib, 1995).

Two important points needs to be made here. First, while 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.[3] Second, the size of the drop in the marginal likelihood obtained with Chib for model M_3 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 like ours is sorely needed.

In sum, it appears that a factorization of the coefficient vector which includes three factors and allows for no idiosyncratic component summarizes the information present in the multi-country VAR reasonably well. Lagged interdependencies, unit specific dynamics and (small) time variations also appear to be important features of our multi-country VAR. In the following exercises, we therefore use M_1 as our specification.

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 US real shocks on the GDP of other countries. We construct such a shock by making US variables contemporaneously causally prior to the other G-6 countries. Within the US block, we make employment growth and output growth jointly increase 1 percent for one period, while the dynamics of the other two variables are unrestricted. Figure 1 presents the median responses together with a 68 percent posterior band obtained with information up to 2000:4.

[3] This instability is probably the direct consequence of the point made by Neal (1999). We thank one of the referees for pointing out this problem to us. Similar instability problems were also experienced by Osiewalski and Pipien (2004) in different models.

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 US and country 2 one of the other six countries. Shocks are identified in the same way as in the multi-country 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 θ_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 1.0e-09.

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 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 comovements, the responses of a multi-country VAR should be smoother than those of a two-country VAR. Figure 1 indicates that at least the first prediction is satisfied: while 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, UK, and Canada), those of the multi-country model are more informative about the features of transmission. For example, there clearly is an Anglo-Saxon cycle (peak responses of GDP in UK and Canada are contemporaneous and almost of the same magnitude as in the US); 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 can not 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. 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.

Figure 1 HERE

(Responses of GDP growth to a shock to the growth rate of real US variables)

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. 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 percent 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 4 quarters after the 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.

Figures 2 and 3 HERE

(Responses of inflation to an oil price growth shock)

Finally, the estimated model can be used to compute a variety of measures which are of interest for policymakers. Figure 4 presents the time profile for the posterior 68 percent band for a coincident measure of world inflation, constructed as $CVLI_t^{\pi} = X_{1t}\theta_{1t} + (X_{3t}\theta_{3t})^{\pi}$. 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.

Figure 4 HERE

(A coincident measure of global inflation)

7 Conclusions

This paper develops an approach to conduct inference in time-varying coefficient multi-country VAR models with lagged cross unit interdependencies and unit specific dynamics. We take a Bayesian viewpoint to estimation 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 which permits exchangeability, time variations and heteroschedasticity in the innovations in the factors.

The factor structure on the coefficients allows us to transform an overparametrized 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 Markov Chain Monte Carlo 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 a MA structure, they emphasize low frequency comovements in the lags of the VAR variables.

The tools described in this paper 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 multi-country 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 paper.

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.

Authors' affiliations:

Fabio Canova: Universität Bern, ICREA-UPF, and CEPR, Switzerland

Matteo Ciccarelli: European Central Bank, Germany

References

[1] Anzuini, A., M. Caivano, F., Lamorgese, and F. Venditti, "A Leading Indicator for Inflation and GDP Growth in Italy and the Euro Area," manuscript, Bank of Italy, 2005

[2] Binder, M., C. Hsiao and H. Pesaran, "Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration," *Econometric Theory* 21 (2005), 795-837.

[3] Bos, C., "A Comparison of Marginal Likelihood Computational Methods," Discussion Paper TI 2002-084/4, Tinbergen Institute, 2002.

[4] Caivano, M., "The Transmission of Monetary and Technology Shocks between the US and the Euro Area," manuscript, Bank of Italy, 2006.

[5] Canova, F., "Modelling and Forecasting Exchange Rates Using a Bayesian Time-Varying Coefficient Model," *Journal of Economic Dynamics and Control* 17 (1993), 233-262.

[6] Canova, F. and M. Ciccarelli, "Forecasting and Turning Point Prediction in a Bayesian Panel VAR Model," *Journal of Econometrics* 120 (2004), 327-359.

[7] Canova, F., M. Ciccarelli, and E. Ortega, "Similarities and Convergence of G-7 Cycles," *Journal of Monetary Economics* 54 (2007), 850-878.

[8] Canova, F. and G. De Nicolo', "Stock Returns, Term Structure, Real Activity and Inflation: An International Perspective," *Macroeconomic Dynamics* 4 (2000), 343-372.

[9] Canova, F. and Gambetti, L., "Structural Changes in the US Economy: Is there a Role for Monetary Policy?" manuscript, Universitat Pompeu Fabra, 2007.

[10] Canova, F. and J. Marrinan, "Sources and Propagations of Output Shocks: Common Shocks or Transmission?" *Journal of International Economics* 42 (1998), 133-167

[11] Canova, F. and E. Pappa, "Price Differential in Monetary Unions: The Role of Fiscal Shocks," *Economic Journal* 107 (2007), 713-737.

[12] Ciccarelli, M. and A. Rebucci, "Measuring Contagion Using a Bayesian Time-Varying Coefficient Model," *Journal of Financial Econometrics* 5 (2007), 285-320.

[13] Chib, S., "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association* 90 (1995), 1313-1321.

[14] Chib, S. and E. Greenberg, "Hierarchical Analysis of SUR Models with Extensions to Correlated Serial Errors and Time-Varying Parameter Models," *Journal of Econometrics* 68 (1995), 409-431.

[15] Cogley, T. and T. Sargent, "Drift and Volatilities: Monetary Policy and Output in Post WWII US," *Review of Economic Dynamics* 8 (2005), 275-308.

[16] Dees, S., F. Di Mauro, H. Pesaran, and V. Smith, "Exploring the International Linkages of the Euro Area: A Global VAR Analysis," Working Paper No. 1425, CESifo, 2005.

[17] 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), 404-433.

[18] Del Negro, M. and C. Otrok, "Dynamic Factor Models with Time-Varying Parameters: Measuring the Evolution of the European Business Cycle," mimeo, Federal Reserve Bank of Atlanta, 2006.

[19] Doan, T., R. Litterman and C. Sims, “Forecasting and Conditional Projection using Realistic Prior Distributions,” *Econometric Reviews* 3 (1984), 1-100.

[20] Fatas, A. and I. Mihov, “Restricting Fiscal Policy Discretion: The Case of US States,” *Journal of Public Economics* 90 (2006), 101-117.

[21] Fernandez, C. and M. Steel, “On Bayesian Modelling of Fat Tails and Skewness,” *Journal of the American Statistical Association* 93 (1998), 359-373.

[22] Gallant, R., P. Rossi and G. Tauchen, “Nonlinear Dynamic Structures,” *Econometrica* 61 (1993), 871-907.

[23] Gelfand, A. and D. Dey, “Bayesian Model Choice: Asymptotics and Exact Calculations,” *Journal of the Royal Statistical Society*, Ser. B 56 (1994), 501-514.

[24] Geweke, J., “Simulation Based Bayesian Inference for Economic Time Series,” in Mariano, R., Shuermann and M. Weeks, eds., *Simulation Based Inference in Econometrics: Methods and Applications* (Cambridge: Cambridge University Press, 2000), 255-299.

[25] Holtz–Eakin D., W. Newey and H. Rosen, “Estimating Vector Autoregressions with Panel Data,” *Econometrica* 56 (1988), 1371-1395.

[26] Kadiyala, R. and S. Karlsson, “Numerical Methods for Estimation and Inference in Bayesian VAR Models,” *Journal of Applied Econometrics* 12 (1997), 99-132.

[27] Kass, R. and A. Ratfery, “Bayes Factors,” *Journal of the American Statistical Association* 90 (1995), 773-795.

[28] Kim, C. J. and C. Nelson, "Business Cycle Turning Points: A New Coincident Index and tests of Duration Dependence Based on a Dynamic Factor Model with Regime Switching," *Review of Economics and Statistics* 80 (1998), 188-201.

[29] Koop, G., H. Pesaran and S. Potter, "Impulse Response Analysis in Non-Linear Multivariate Models," *Journal of Econometrics* 74 (1996), 119-147.

[30] Imbs, J., H. Mumtaz, M. Ravn, and H. Rey, "PPP Strikes Back: Aggregation and the Real Exchange Rate," *Quarterly Journal of Economics* CXX (2005), 1-43.

[31] Neal R., "Erroneous Results in Marginal Likelihood from the Gibbs Output," mimeo, University of Toronto, 1999.

[32] Newton, M. A., and A. E. Raftery, "Approximate Bayesian Inference by the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society*, Ser. B 56 (1994), 3-48.

[33] Osiewalski, J., and M. Pipien, "Bayesian Comparison of Bivariate ARCH-Type Models for the Main Exchange Rates in Poland," *Journal of Econometrics* 123 (2004), 371-391.

[34] Otrok, C. and C. Whiteman, "Bayesian Leading Indicators: Measuring and Predicting Economic Conditions in Iowa," *International Economic Review* 39 (1998), 997-1014.

[35] Pesaran, H., "Estimation and Inference in Large Heterogeneous Panels with Cross Section Dependence," Working Paper No. 0305, University of Cambridge, 2003.

[36] Pesaran, H., T. Schuerman, and S. Weiner, "Modelling Regional Interdependencies using a Global Error Correction Macroeconomic Model," *Journal of Business and Economic Statistics* 22 (2005), 129-162.

[37] Pesaran, H. and R. Smith, "Estimating Long Run Relationships for Dynamic Heterogenous Panels," *Journal of Econometrics* 68 (1996), 79-113.

[38] Stock, J. and M. Watson, "New Indexes of Coincident and Leading Economic Indicators," in O. Blanchard and S. Fisher, eds., *NBER Macroeconomic Annual* (1989), 352-394, .

[39] Stock, J. and M. Watson, "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association* 97(2002a), 1167-1179.

[40] Stock, J. and M. Watson, "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics* 20 (2002b), 147-162.

Figure 1. Responses of GDP growth to a 1% shock to the growth rate of US real variables

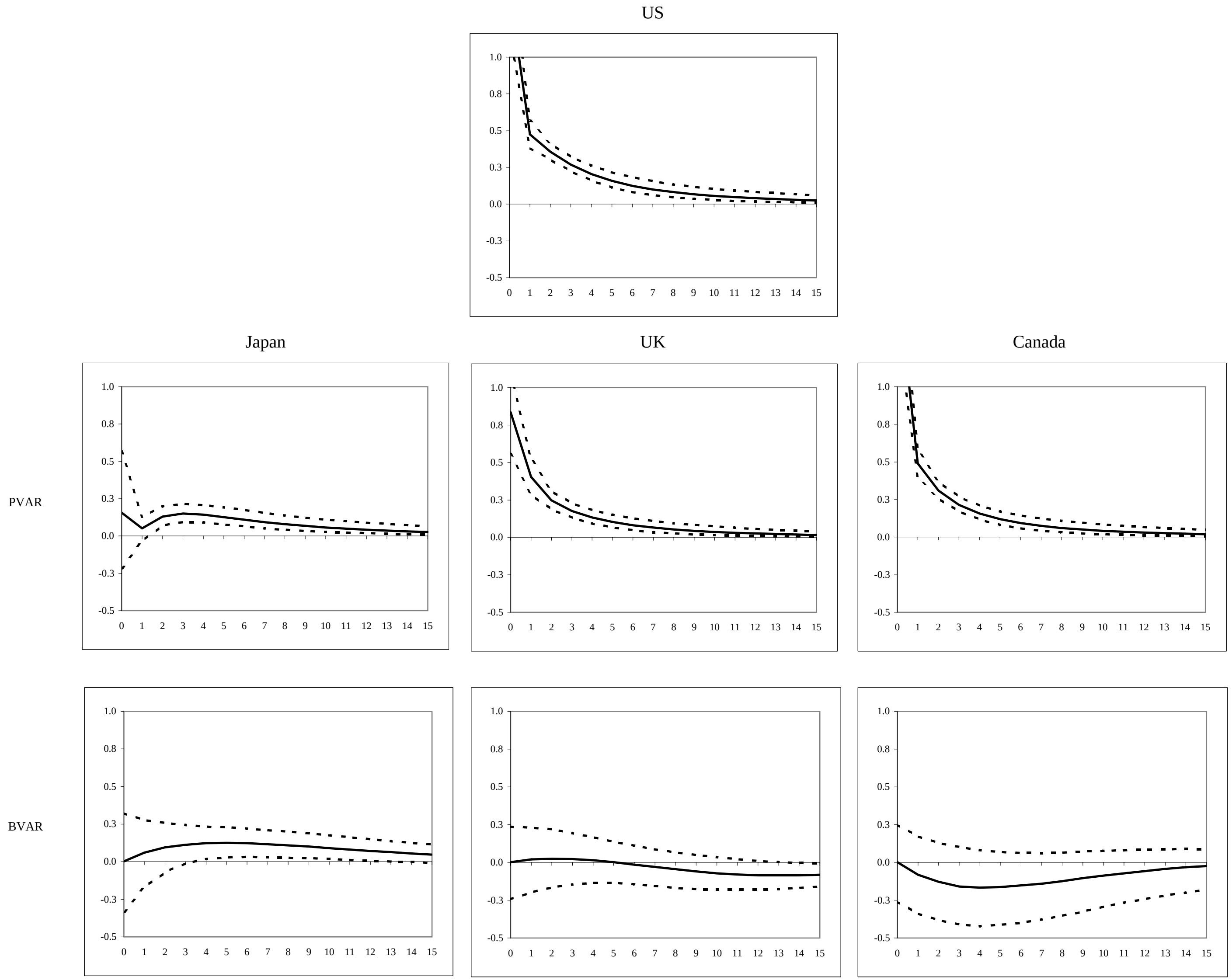

Figure 1 (cont'd). Responses of GDP growth to a 1% shock to the growth rate of US real variables

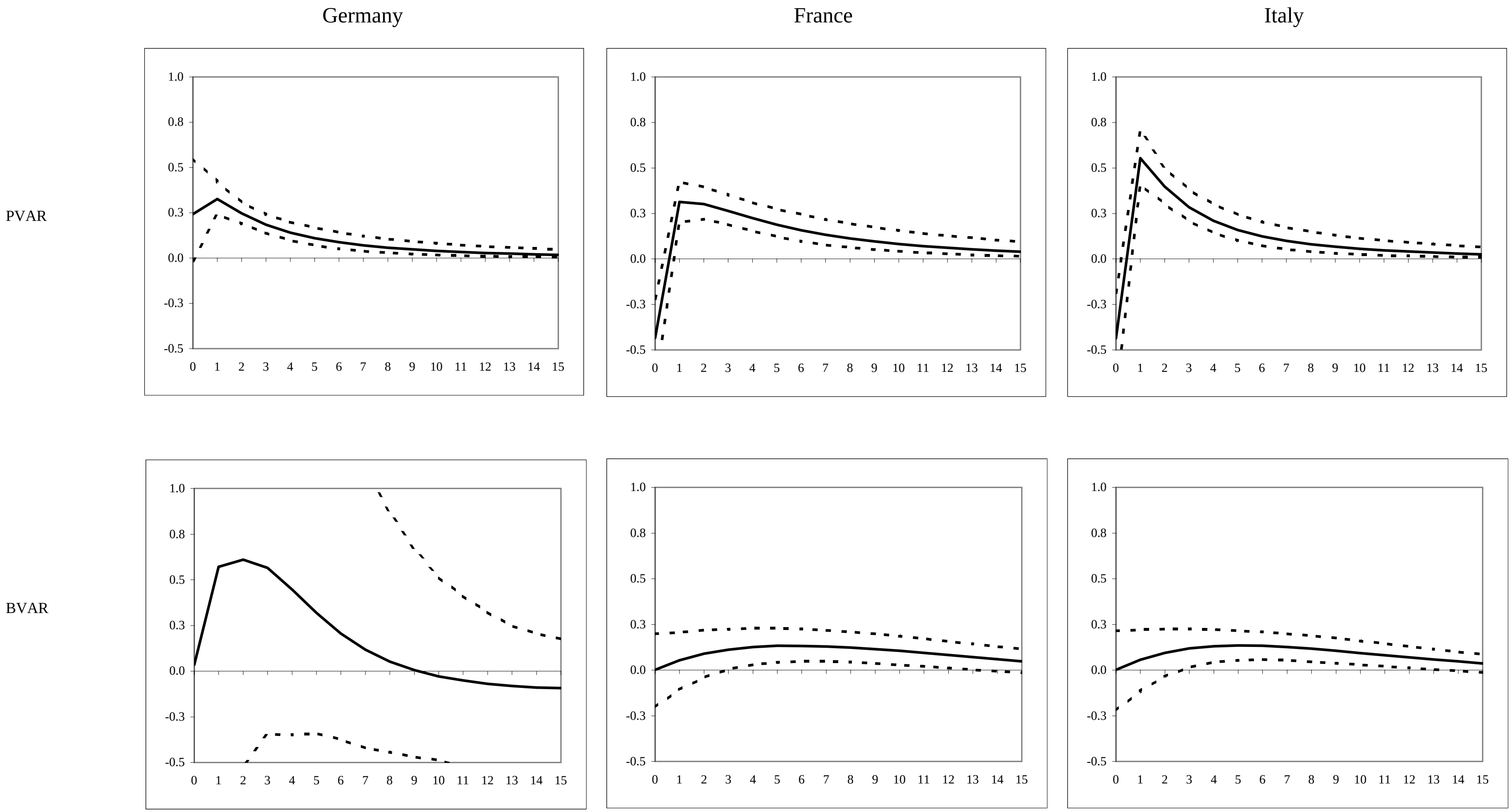

Figure 2. Oil shock

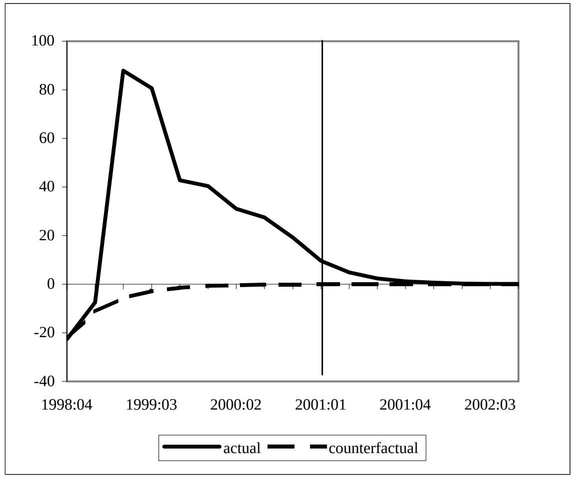

Figure 3. Forecast of inflation conditional on a shock to oil price growth

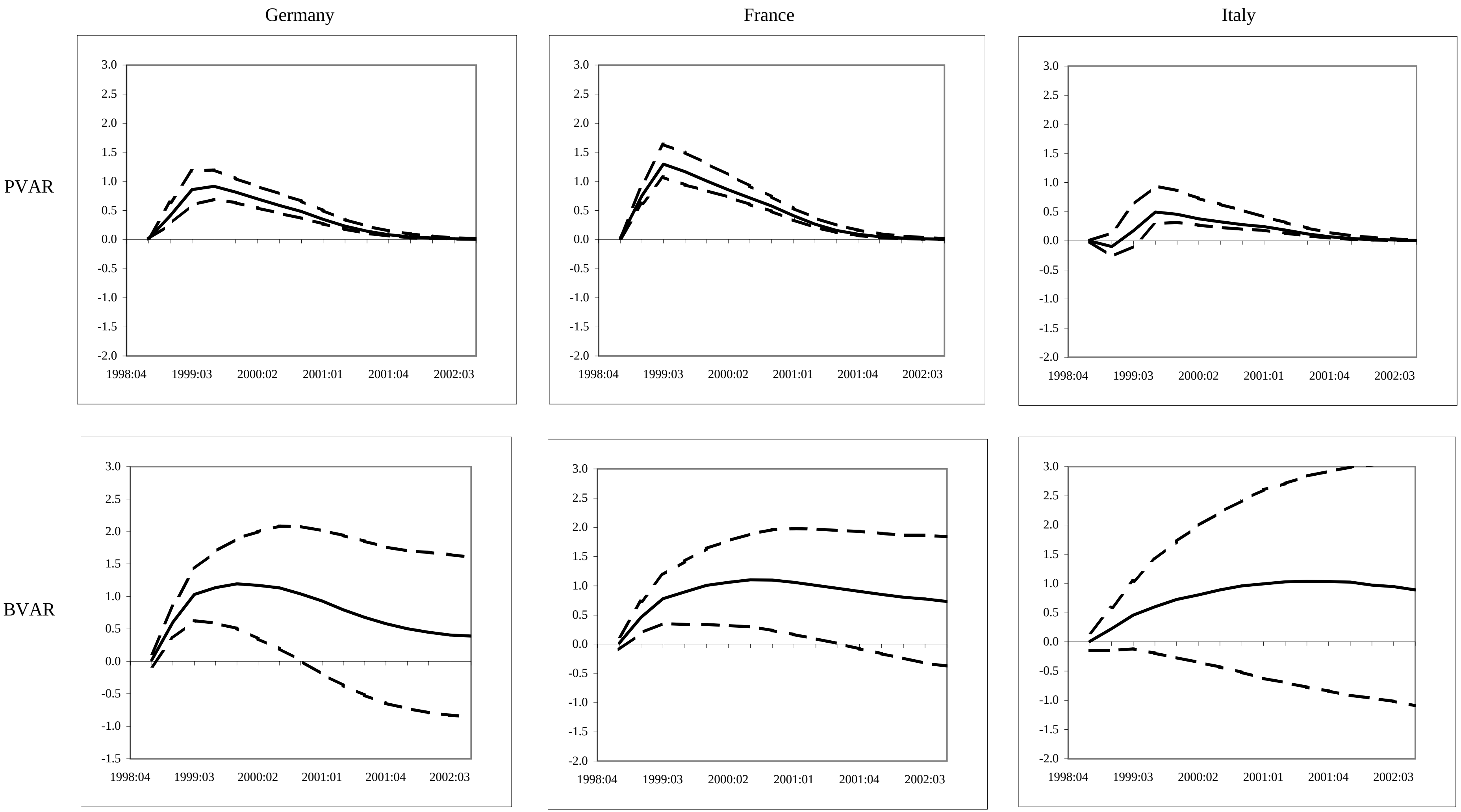

Figure 4. A coincident measure of Global Inflation