Estimating Multi-country VAR models*

Fabio Canova
ICREA, Universitat Pompeu Fabra, CREI and CEPR
and
Matteo Ciccarelli†
European Central Bank

February 2006

Abstract

This paper describes a methodology 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 MCMC routine. The transmission of certain shocks across G7 countries is analyzed.

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

JEL Classification nos: C3, C5, E5.

*We would like to thank an anonimous referee, T. Cogley, H. Van Djik, D. Hendry, R. Paap, T.Trimbur, T. Zha and the seminar participants at Erasmus University, University of Stockholm, 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.

†Corresponding author: ECB, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany. Tel.: +496913448721, Fax:+496913446575, Email: matteo.ciccarelli@ecb.int.

Non-technical Summary

When dealing with multi-country data, the empirical literature has taken a number of short cuts. For example, it is typical to assume that in the dynamic specification slope coefficients are common across (subsets of the) units; that there are no interdependencies across units or that they can be summarized with a simple time and unit invariant index; that the structural relationships are stable over time; that asymptotics in the time series dimension apply; or a combination of all of these. None of these restrictions is appealing: short time series are the result, in part, of new definitions and the adaptation of international standards to data collection in developing countries; unit specific relationships may reflect difference in national regulations or policies; interdependencies results from world markets integration and time instabilities from evolving macroeconomic structures.

This paper shows how to conduct inference in multi-country VAR models featuring short time series 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, no classical estimation method is feasible. We take a flexible Bayesian viewpoint and weakly restrict the coefficient vector to depend on a low dimensional vector of time varying factors. These factors capture, for example, variations in the coefficients which are common across units and variables (a "common" effect); variations which are specific to the unit (a "fixed" effect) , variations which are specific to a variable (a "variable" effect), etc. Factors relating to lags and time periods, or capturing the extent of lagged interdependencies across units, can also be included. We complete the specifications using a hierarchical structure which allows for exchangeability in the fixed effects, and time variations in the law of motion of the factors and in the variance of their innovations.

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 coefficient 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 small number of loadings on certain combinations of the right hand side variables of the VAR. Therefore, despite its complex structure, the computation costs are small. Second, since the regressors of the SUR model are observable linear combinations of the right hand side variables of the VAR, the framework is

suitable for a variety of policy purposes. For example, one can produce multi-step multi-country leading indicators; conduct unconditional out-of-sample forecasting exercises; recursively estimate coincident indicators of world and national business cycles and examine their time variations; construct measures of core inflation or of potential output; and examine the propagation of certain shocks across countries.

Posterior distributions for the quantities of interest are obtained with Markov Chain Monte Carlo (MCMC) methods. We show how to use the output of the Gibbs sampler to compute responses to unexpected perturbations in the innovations of either the VAR or the loadings of one of the indices, 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 specification of the reparametrized SUR model. We also show how to quantify the importance of lagged interdependences, of unit specific dynamics and of time variations in the factors. This analysis is important since the inferential work of the investigator could be greatly simplified if some of the distinguishing features we have emphasized is absent from the data under consideration.

The methodology is used to model the dynamics of a vector of variables in the G-7 countries and to examine two issues which are important for policy makers: what are the effects of a US shock on GDP of the G-7 countries and what are the consequences of a persistent oil price increase on inflation in Euro area countries.

1 Introduction

Over the last decade, there has been a growing interest in using multi-county VAR models for applied macroeconomic analysis. This interest is due, in part, to the availability of higher quality data for a large number of countries and to advances in computer technology, which make the estimation of large scale models feasible in a reasonable time. Multi-country (or multi-sectors) VAR models arise in a number of fields and applications. For example, when studying the transmission of certain structural shocks across countries, it is desirable to have a model where cross country interdependencies are fully spelled out. Similarly, when examining issue related to income convergence and/or the evaluation of the effects of regional policies, it is necessary to have a framework that explicitly allows for spillover effects across regions, both of contemporaneous and lagged nature. Finally, questions about contagious of financial crises, spillover of exchange rate volatility or issues concerning the globalization of financial markets in advanced economies can be naturally studied in the framework of multi-country VARs. A multi-country setup differs from the multi-agent framework typically studied in applied microeconomics for several reasons. First, cross unit lagged interdependencies are likely to be important in explaining the dynamics of multi-country data, while this is not necessarily so for multi-agent data, especially once a (common) time effect is taken into account. Second, heterogeneous dynamics are a distinctive feature of multi-country time series data (see e.g. Canova and Pappa (2003) or Imbs et at. (2005)) while it is a less crucial feature in multi-agent, multi-period data. Third, while in multi-agent studies the number of cross sectional units is typically large and the time series is short, in multi-country studies 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. (2001), all of which are consistent as the cross section dimension becomes large, or the group estimator of Pesaran and Smith (1996), which is consistent as the time series dimension becomes large, are inapplicable. Finally, while with a large homogeneous cross section, estimation of time varying structures is feasible, the combination of heterogenous dynamics and moderately long time series 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; that there are no interdependencies across units or that they can be summarized with a simple time and unit invariant index; that the structural relationships are stable over time; that asymptotics in T apply; or a combination of all of these. None of these restrictions is appealing: short time series are the result, in part, of new definitions and the adaptation of international standards to data collection in developing countries; unit specific relationships may reflect difference in national regulations or policies; interdependencies results from world markets integration and time instabilities from evolving macroeconomic structures.

This paper shows how to conduct inference in multi-country VAR models featuring short time series 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, no classical estimation method is feasible. We take a flexible Bayesian viewpoint and weakly restrict the coefficient vector to depend on a low dimensional vector of time varying factors. These factors capture, for example, variations in the coefficients which are common across units and variables (a "common" effect); variations which are specific to the unit (a "fixed" effect) , variations which are specific to a variable (a "variable" effect), etc. Factors relating to lags and time periods, or capturing the extent of lagged interdependencies across units, can also be included. We complete the specifications using a hierarchical structure which allows for exchangeability in the fixed effects, and time variations in the law of motion of the factors and in the variance of their innovations.

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 coefficient 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 small number of loadings on certain combinations of the right hand side variables of the VAR. Thus, for example, in a model with G variables, N units and k coefficients each equation, a setup which requires the estimation of GNk, possibly time-varying parameters, our approach produces estimates of $1 + N + G$ loadings when a common, a unit and a variable specific vector of coefficient factors are specified. Therefore, despite its complex structure, the computation costs are small. In addition, if the loadings are time invariant and priors uninformative, OLS estimation, equation by equation, is everything that it is needed to obtain posterior distributions of the quantities of interest. Second, since the regressors of the SUR model are observable linear combinations of the right hand side variables of the VAR, the framework is suitable for a variety of policy purposes. For example, one can produce multi-step multi-country leading indicators (see Anzuini, et al. (2005); conduct unconditional out-of-sample forecasting exercises; recursively estimate coincident indicators of world and national business cycles and examine their time variations (see Canova, et al. (2003)); construct measures of core inflation or of potential output; and examine the propagation of certain shocks across countries.

The reparametrized multi-country VAR model resembles a classical factor model (see e.g. Stock and Watson (1999), Forni, et al.(2000), Pesaran (2003)). Nevertheless, several important differences need to be noticed. First, our starting point is a multi-country VAR with lagged interdependences, unit specific dynamics and time varying coefficients and our factorization is the results of flexible restrictions imposed on the coefficients of the model. Second, the regressors of the SUR model are observable unweighted combinations of lags of the VAR variables while in factors models they are estimated weighted combinations of the current endogenous variables. Therefore, their informational content is potentially different since, by construction, the regressors of our SUR model emphasize low frequency movements of the data, while those used in the factor literature do not usually have this feature. Third, estimates of the loadings obtained in classical factor models are asymptotically justifiable only if both T and NG are large, while here exact distributions are obtained for any T, N or G under assumptions about the distribution of the shocks. Finally, while our setup allows the estimation of time varying relationships, time varying loadings are not permitted in factor models estimated with standard classical (EM) techniques.[1] Therefore, they can only be used to answer a restricted set of questions which have policy interest.

Posterior distributions for the quantities of interest are obtained with Markov Chain Monte Carlo (MCMC) methods. We show how to use the output of the Gibbs sampler to compute responses to unexpected perturbations in the innovations of either the VAR or the loadings of one of the indices, 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 specification of the reparametrized SUR model. We also show how to quantify the importance of lagged interdependences, of unit specific

[1] Kim and Nelson (1998) and Otrok and Del Negro (2003) study time varying coefficients factor models, but employ Bayesian methods to estimate the unknowns.

dynamics and of time variations in the factors. This analysis is important since the inferential work of the investigator could be greatly simplified if some of the distinguishing features we have emphasized is absent from the data under consideration.

Canova and Ciccarelli (2004) proposed a structure to (unconditionally) forecast with multi-country VAR models which allows for unit specific dynamics and time variations. There the estimation process is computationally demanding since the structure of time variations is different across variables and units. Relative to that paper we innovate in three dimensions. First, we provide a flexible coefficient factorization which renders estimation easy. Second, we present a testing approach which makes model selection and inference tractable. Third, we provide 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 setup of the model. Section 3 describes estimation and inference. Section 4 deals with model selection. Section 5 shows how to compute impulse responses and conditional forecasts. In section 6 the methodology is used to model the dynamics of a vector of variables in the G-7 countries and to examine the transmission of certain shocks. 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 exogenous 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 exogenous variables. In (1), cross-unit lagged interdependencies exist whenever the matrix $D_{it}(L)$ can not be decomposed into $\mathfrak{I} \otimes \mathcal{D}_{it}(L)$ for some L, where $\mathfrak{I}$ is a $1 \times N$ vector with one in the *i-th* position and zero elsewhere, and $\mathcal{D}_{it,j}$ are $G \times G$ matrices each j. To see what this feature entails, consider a version of (1) with $N = 2$, $G = 2$, $p_1 = 2$, $q = 0$ of the form:

$$Y_t = \mathbf{D}_{t,1} Y_{t-1} + \mathbf{D}_{t,2} Y_{t-2} + e_t \tag{2}$$

where $Y_t = [y^1_{1t}; y^2_{1t}; y^1_{2t}; y^2_{2t}]'$, $e_t = [e^1_{1t}; e^2_{1t}; e^1_{2t}; e^2_{2t}]'$, and the matrices $\mathbf{D}_{t,j}$ contain $D_{it,j}$ stacked by i. Then, lagged cross units interdependencies appear if $\mathbf{D}_{t,1}$ or $\mathbf{D}_{t,2}$ are not block diagonal. The presence of this feature adds flexibility to the specification but it is costly: the number of coefficients, in fact, is greatly increased (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. While this latter feature may be of minor importance in multi-agent studies, it is crucial in macro setups where structural changes are relatively common. Let δ^g_{it} be $k \times 1$ vectors containing, stacked, the G rows of the matrices D_{it} and C_{it}; define $\delta_{it} = (\delta^{1\prime}_{it}, \ldots, \delta^{G\prime}_{it})'$, and let $\delta_t = (\delta'_{1t}, \ldots, \delta'_{Nt})'$ be a $NGk \times 1$ vector. Whenever δ_{it} varies with cross–sectional units in different time periods, it is impossible to estimate it using classical methods. To deal with this problem, the literature has employed various shortcuts: either it is assumed that the coefficient vector does not depend on the unit, apart from a time invariant fixed effect; that there are no interdependencies across units, that there are no time variations or a combination of all of these (see e.g. Chamberlain (1982), Holtz Eakin et al. (1988) or Binder et al. (2001)). None of these assumptions is appealing in our context. Instead, we assume that δ_t can be factored as:

$$\delta_t = \sum_f^F \Xi_f \theta_{ft} + u_t \tag{3}$$

where $F << NGk$; each θ_{ft} is a low dimensional vector, Ξ_f are conformable matrices and u_t captures unmodelled and idiosyncratic variations present in the coefficient vector.

For the example considered in (2), $\delta_t = [vec(\mathbf{D}_{t,1}), vec(\mathbf{D}_{t,2})]$ is a 32×1 vector with typical element $\delta^{i,j,h}_t$ where i denotes the unit, j the variable, and h the lag. Then, one possible factorization of δ_t is

$$\delta^{i,j,h}_t = \theta_{1t} + \theta^i_{2t} + \theta^j_{3t} + \theta^h_{4t} + u^{i,j,h}_t$$

where θ_{1t} is scalar capturing common movements, $\theta_{2t} = (\theta^1_{2t},\ \theta^2_{2t})'$ is a 2×1 vector capturing coefficient movements which are unit specific, $\theta_{3t} = (\theta^1_{3t},\ \theta^2_{3t})'$ is a 2×1 vector capturing coefficient movements which are variable specific, $\theta_{4t} = (\theta^1_{4t},\ \theta^2_{4t})'$ is a 2×1 vector capturing coefficient movements which are lag specific while u_t is a 32×1 vector absorbing the remaining idiosyncratic variations. Here Ξ_1 is a 32×1 vector of ones, Ξ_2 and Ξ_3 are 32×2 vectors

$$\Xi_2 = \begin{bmatrix} \iota_1 & 0 \\ \iota_1 & 0 \\ 0 & \iota_2 \\ 0 & \iota_2 \end{bmatrix} \quad \Xi_3 = \begin{bmatrix} \iota_3 & 0 \\ \iota_3 & 0 \\ 0 & \iota_4 \\ 0 & \iota_4 \end{bmatrix}$$

where $\iota_1 = (1\;1\;0\;0\;1\;1\;0\;0)'$; $\iota_2 = (0\;0\;1\;1\;0\;0\;1\;1)'$ $\iota_3 = (1\;0\;1\;0\;1\;0\;1\;0)'$ $\iota_4 = (0\;1\;0\;1\;0\;1\;0\;1)'$, etc. Alternative factorizations can be obtained distinguishing, e.g. own vs. other variable/unit specific coefficients in θ_{2t} and θ_{3t}, etc.

Clearly, the choice of factorization is application and, possibly, sample dependent. While the choice of the number of factors is typically a-priori dictated by the needs of the investigation - 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 - there are situations where no a-priori information is available. A simple procedure to determine the dimension of F in these situations, which trades-off the fit of the model with the number of factors included, appears in section 4. Note also that in (3) 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 $\mathbf{X}_t = (Y_{t-1}',\ Y_{t-2}', \ldots, Y_{t-p}',\ W_t', \ldots, W_{t-l}')'$; and let Y_t and E_t be $NG \times 1$ vectors, we can rewrite (1) as:

$$\begin{aligned} Y_t &= X_t \delta_t + E_t \\ &= X_t(\Xi\theta_t + u_t) + E_t \equiv \mathcal{X}_t\theta_t + \zeta_t \end{aligned} \tag{4}$$

where $\mathcal{X}_t \equiv X_t\Xi$; $\Xi = [\Xi_1, \Xi_2, \Xi_3, \ldots, \Xi_F]$ and $\zeta_t \equiv X_t u_t + E_t$.

In (4) we have reparametrized the original multi-country VAR to have a structure where the vector of endogenous variables depends on a small number of observable indices, $\mathcal{X}_t$, and the coefficient factors, θ_t, load on the indices. By construction, $\mathcal{X}_t$ are particular combinations of right hand side variables of the multi-country VAR. For example, $\mathcal{X}_{1t}$ is a $NG \times 1$ vector with all entries equal to the sum of all regressors of the VAR; $\mathcal{X}_{2t} = \mathcal{X}_{2t}^{\dagger} \otimes \iota_G$ is a $NG \times N$ matrix where ι_G is a $G \times 1$ vector of ones and $\mathcal{X}_{2t}^{\dagger}$ is a $N \times N$ diagonal matrix with the sum of the lags of the variables belonging to a unit on the diagonal; etc. Furthermore, note that (i) the indices are correlated among each other by construction and that the correlation decreases as G or N or $p = max[p_1, p_2]$ increase; (ii) the sums are constructed equally weighting the lags of all variables; and (iii) each $\mathcal{X}_{it}$ is a one-sided moving average process of order p.

One advantage of the SUR structure in (4) is that the over-parametrization of the original multi-country VAR is dramatically reduced. In fact, estimation and specification searches are constrained only by the dimensionality of θ_t, not by the one of δ_t. A second advantage is that, given the moving average nature of $\mathcal{X}_t$, the regressors of (4) capture low frequency movements present in the lags of the original VAR. Since the parsimonious structure adopted averages out not only cross section but also time series noise, reliable and stable estimates of θ_t can potentially be obtained even in large scale models, and this makes the framework useful for medium term unconditional forecasting and policy analyses exercises. A third advantage of our reparametrization is that (4) has a useful economic interpretation. For example, $\mathcal{X}_{1t}\theta_{1t}$ is an indicator for Y_t based on the common information present in the lags of the VAR while $\mathcal{X}_{1t}\theta_{1t} + \mathcal{X}_{2t}\theta_{2t}$ is an indicator for Y_t based on the common and the country specific information present in the lags of the VAR. Indicators containing various type of information can therefore 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 $t-\tau,\ \tau = 1, 2, \ldots$.

If the loadings θ_t where independent of time, estimation of (4) would be easy: it would simply require regressing each element of Y_t on appropriate averages, adjusting estimates of the standard errors for the presence of heteroschedasticity. Regressions like these are typical in factor models of the type used by Stock and Watson (1999), Forni et al. (2000) and others. However, two differences are worth emphasizing. First, our indices are observable, as opposed to estimated; and can be recursively constructed as new data becomes available. Second, since our indices span the space of lagged interdependencies in models with unit specific dynamics, they can be used to examine the importance of both these features in the data.

We specify a flexible time varying structure on the factors of the form:

$$\theta_t = (I - \mathcal{C})\,\bar{\theta} + \mathcal{C}\theta_{t-1} + \eta_t \qquad \eta_t \sim (0, B_t) \tag{5}$$

$$\bar{\theta} = \mathcal{P}\mu + \epsilon \qquad \epsilon \sim (0, \Psi) \tag{6}$$

where $\bar{\theta}$ is the unconditional mean of θ_t; $\mathcal{P}, \mathcal{C}, \Psi$ are known matrices; η_t and ϵ are mutually independent and independent of E_t and u_t, and $B_t = diag(\bar{B}_1, \ldots \bar{B}_F) = \gamma_1 * B_{t-1} + \gamma_2 * \bar{B} \equiv \xi_t * \bar{B}$, where $B_0 = \bar{B}$, γ_1 and γ_2 are known, and $\xi_t = \gamma_1^t + \gamma_2 \frac{(1-\gamma_1^t)}{(1-\gamma_1)}$. 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.

Intuitively, to permit time variations in the factors, we make them obey the flexible restrictions implied by (5) and (6). In (5) we have assumed an AR structure with time varying variances. Since the matrix $\mathcal{C}$ is arbitrary, the specification allows for general relationships. As shown in Canova (1993), the structure used in B_t imparts heteroschedastic swings in θ_t, which could be important in modelling the dynamics present, e.g., in financial variables, and nests two important special cases:

(a) no time variation in the factors, $\gamma_1 = \gamma_2 = 0$, and $\mathcal{C} = I$, and (b) homoschedastic variance $\gamma_1 = 0$ and $\gamma_2 = 1$. Cogley and Sargent (2005) have used a similar specification in a single country VAR framework. However, to capture conditional heteroschedasticity they set $B_t = \bar{B}$ and specify Ω to be a function of a set of stochastic volatility processes.

The matrix $\mathcal{P}$ allows the mean of the country factors to have an exchangeable structure. For example, if the unit specific factors are drawn from a distribution with common mean and there are, e.g. three units, two variables and three factors in (6), then:

$$\mathcal{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

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

Numerous interesting specifications are nested in our model: for example, time invariant factors are obtained by making B_t a reduced rank matrix and setting the appropriate elements of $\mathcal{C}$ to zero; no exchangeability obtains when Ψ is large and the factorization becomes exact if $\sigma^2 = 0$.

For the rest of the presentation 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 $(u_t \mid z_t) \sim N(0, z_t(\Omega \otimes V))$ where $z_t^{-1} \sim \chi^2(\nu, 1)$, and χ^2 is a chi-square with ν degrees of freedom and scale equal 1, since unconditionally, $u_t \sim t_\nu(0, \Omega \otimes V)$. Since the forecast errors of our SUR model display fat tail distributions even when all disturbances are normal (see section 3), this additional feature will not be considered here.

3 Inference

The model that needs to be estimated is composed of (4)-(6). While classical Kalman filter methods can be employed, we take a Bayesian approach to estimation for two reasons. First, our estimates are valid for any sample size, while classical estimates are only asymptotically justified. This is important in typical macroeconomic applications since T is either small or of moderate size. Furthermore, when T is short, shrewdly chosen priors can help to obtain economically meaningful estimates of the unknowns while this is hard with classical filtering techniques, unless additional restrictions are imposed.[2] Second, for large T, the likelihood of the data dominates the prior. Hence, our estimates asymptotically approach those obtained with classical methods. Clearly, gaussianity of the disturbances is necessary for efficient estimation in both frameworks.

The likelihood of the reparametrized SUR model (4) is

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

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

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

Here $N()$ stands for Normal, $W()$ for Wishart and $G()$ 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 below is unchanged by these modifications.

Despite the dramatic parameter reduction obtained with (4), the analytical computation of posterior distributions is unfeasible. However, a variant of the Gibbs sampler approach described, e.g., in Chib and Greenberg (1995) can also be used in our framework. Let $Y^T = (Y_1, ..., Y_T)$ denote the data, $\psi = \left(\mu,\ \Psi^{-1}, \Omega^{-1},\ \sigma^{-2},\ \bar{B}_f^{-1}, \{\theta_t\}, \bar{\theta}\right)$ 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

[2] Notice, however, that with small sample certain dogmatic features of the model might have an important effect on posterior inference and therefore a sensitivity analysis should always been performed when T is short. On the other hand, our following derivation of the posterior relies on the Normality assumption of the error term. This assumption could nevertheless be relaxed, though not at the cost of limiting an exact inference.

$\tilde{\theta}_t = \theta_t - \mathcal{C}\theta_{t-1}$. Given Y^T, the conditional posteriors for the unknowns are:

$$\begin{aligned}
\mu \quad &| \quad Y^T, \psi_{-\mu} \sim N\left(\hat{\mu}, \hat{\Sigma}_\mu\right) \\
\Psi^{-1} \quad &| \quad Y^T, \psi_{-\Psi} \sim W\left(z_0 + 1, \hat{Q}_o\right) \\
\Omega^{-1} \quad &| \quad Y^T, \psi_{-\Omega} \sim W\left(z_1 + T, \hat{Q}_1\right) \\
\bar{B}_f^{-1} \quad &| \quad Y^T, \psi_{-\bar{B}_f} \sim W\left(T * \dim\left(\theta_t^f\right) + z_{2f}, \hat{Q}_{2f}\right)
\end{aligned}$$

$$\begin{aligned}
\sigma^2 \quad &| \quad Y^T, \psi_{-\sigma^2} \propto (\sigma^{-2})^{-a-1} \exp\{b\sigma^{-2}\} \times \prod_t |\Upsilon_t|^{-0.5} \\
&\times \quad \exp\{-0.5 \sum_t (Y_t - \mathcal{X}_t\theta_t)' \Upsilon_t^{-1} (Y_t - \mathcal{X}_t\theta_t)\} \\
\bar{\theta} \quad &| \quad Y^T, \psi_{-\bar{\theta}} \sim N\left(\widehat{\bar{\theta}}, \hat{\Psi}\right)
\end{aligned} \tag{7}$$

where

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

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. We use here what Chib and Greenberg (1995) proposed for SUR models. In particular, given $\theta_{0|0}$ and $R_{0|0}$ the Kalman filter gives the recursions

$$\begin{aligned}
\theta_{t|t} &= \theta_{t-1|t-1}^* + (R_{t|t-1}^*\mathcal{X}_t F_{t|t-1}^{-1})\,(Y_t - \mathcal{X}_t\theta_t) \\
R_{t|t} &= \left(I - (R_{t|t-1}^*\mathcal{X}_t F_{t|t-1}^{-1})\mathcal{X}_t\right)(R_{t-1|t-1}^* + \xi_t\bar{B}) \\
F_{t|t-1} &= \mathcal{X}_t R_{t|t-1}^* \mathcal{X}_t' + \Upsilon_t
\end{aligned} \tag{8}$$

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 $\theta_{t-1|t-1}$. Subsequently, 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 $N(\theta_{T|T}, R_{T|T})$, then θ_{T-1} is simulated from $N(\theta_{T-1}, R_{T-1})$, and so on, until θ_1 is simulated from $N(\theta_1, R_1)$, with $\theta_t = \theta_{t|t} + R_{t|t} R^{-1}_{t+1|t} \left(\theta_{t+1} - \theta_{t|t}\right)$, 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 $(\sigma^2)^*$ is generated via $(\sigma^2)^* = (\sigma^2)^l + v$, 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 $(\sigma^2)^*$ to the kernel of the density of $(\sigma^2)^l$ and c^2 is chosen so that the acceptance rate is roughly 20-40 percent.

Draws from the posterior distributions can be obtained cycling through the conditional in (7)-(8) 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), 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 $(\zeta_t | \sigma^2) \sim N(0, \sigma_t \Omega)$. 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 (4) are endogenously allowed to have fat tails. Since with this feature, shocks to the model may alter its dynamics, there is a built-in an endogenous adaptive scheme which allows coefficients to adjust when breaks in the relationships are present.

While the regressors of the SUR model are correlated, 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 (2003)), which is the case in our setup.

Posterior distributions for any continuous function $\mathcal{G}(\psi)$ of the unknowns 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 $Y^{t+\tau} = (Y_{t+1}, \ldots, Y_{t+\tau})$, consider the conditional density of $Y^{t+\tau}$, given the data up to t, and a function $\mathcal{G}(Y^{t+\tau})$. Then

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

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 $\mathcal{F}\left(Y^{t+\tau} \mid Y_t, \psi^{(\ell)}\right)$. $\left\{Y^{\ell t+\tau}\right\}_{\ell=\bar{L}+1}^{\bar{L}+L}$ constitutes a sample, from which we can compute a location measure - e.g. $\hat{Y}^{t+\tau} = L^{-1}[\sum_{\ell=\bar{L}+1}^{\bar{L}+L}(Y^{\ell t+\tau})]$ or $Y^{t+\tau,50}$; and a dispersion measure - $var\left(\hat{Y}^{t+\tau}\right) = L^{-1}\left[\mathcal{Q}_o + \sum_{s=1}^{r}\left(1 - \frac{s}{r+1}\right)(\mathcal{Q}_s + \mathcal{Q}_s')\right]$, where $\mathcal{Q}_s = L^{-1}\left[\sum_{\ell=s+1+\bar{L}}^{L+\bar{L}}\left(Y_{\ell t+\tau} - \hat{Y}_{t+\tau}\right)\left(Y_{\ell t+\tau} - \hat{Y}_{t+\tau}\right)'\right]$ or various interdecile ranges. 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 factors in (3) is dictated by 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 VAR, etc., especially when there are no a-priori reasons to choose one decomposition over another. It is easy to design a diagnostic to discriminate across models with different indices. Let

$$\mathcal{L}(Y^t|M_h) = \int \mathcal{F}(Y^t|\psi_h, M_h)p(\psi_h|M_h)d\psi_h \tag{9}$$

be 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. (9) can be easily computed using the output of the Gibbs sampler, as suggested by Chib (1995), or using the modified harmonic mean approach of Gelfand and Dey (1993), for any model M_h. Then the Bayes factor

$$\mathcal{B}_{hh'} \equiv \frac{\mathcal{L}(Y^t|M_h)}{\mathcal{L}(Y^t|M_{h'})} \tag{10}$$

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, (10) directly trades off the predictive record with the dimensionality of the model.

When the two specifications are nested, that is, when $\psi = (\psi_1, \psi_2)$ and $\psi_2 = \bar{\psi}_2$ is the restriction of interest, if $p(\psi_1|M_h) = \int p(\psi_1, \psi_2|M_{h'})d\psi_2$ and ψ_1 and ψ_2 are independent, Bayes factor is $\mathcal{B}_{h,h'} = \frac{p(\bar{\psi}_2|M_{h'})}{p(\bar{\psi}_2|Y^t, M_{h'})}$ (see Kass and Raftery (1995)), which only requires the prior and the posterior of the model with h' indices.

With this form of the Bayes factor it is possible to conduct several specification searches. For example, it is possible to examine whether the factorization in (5) is exact, i.e. whether there are no idiosyncratic elements in the coefficients, letting $\psi_2 = \sigma^2$ and $\bar{\psi}_2 = 0$; or whether there are time variations in θ_t, setting $\bar{B}_f = b_f * I$, $\psi_2 = b_f$ some f, and $\bar{\psi}_2 = 0$. Posterior support for the presence of interdependencies is obtained, on the other hand, comparing the marginal likelihoods of the unrestricted model and that of a vector of country specific VARs with time varying coefficients.

Rather than examining hypotheses on the structure of the model, one may want to incorporate model uncertainty into posterior estimates. Let M_1 be the model with one index and M_h the model with h indices, $h = 2, \dots H$, 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

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 $t = \bar{t}, \dots T$ once recursive estimates of $p(\psi_h|Y^t)$ are computed. These recursive distributions are straightforward to obtain (we only need to run a MCMC for every t) and only require computer memory. Since in models with about 30 variables one complete run of the MCMC routine takes about 45 minutes on a high speed PC, recursive computation of posterior distributions are computational demanding but feasible on available machines.

5.2 Impulse responses

The computation of impulse responses in a model with time varying coefficients is non-standard. 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 assumes 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. 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 the first case we condition on a random draw for the current shocks, while in the second the current shocks is set to its unconditional value.

To formally define impulse responses we need some preliminary notation. Recall that the reparametrized multi-country model VAR 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 $\mathcal{Z}_t = [H_t^{-1}(E_t + X_t u_t)', H_t^{-1}\eta'_t, H_t^{-1}\epsilon']'$ 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 shocks of the model. For example, 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.

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. Define $\mathcal{F}_t^1 = \{Y^{t-1}, \theta^t, \mathcal{V}_t, H_t, \mathcal{Z}_{j,t} = \bar{\mathcal{Z}}_{j,t}, \mathcal{Z}_{-j,t}, \mathcal{U}_{t+1}^{t+\tau}\}$ 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 in the $j-th$ component of $\mathcal{Z}_t$ are obtained as

$$IR(t, t+\tau) = E(Y_{t+\tau}|\mathcal{F}_t^1) - E(Y_{t+\tau}|\mathcal{F}_t^2) \qquad \tau = 1, 2, \ldots \tag{11}$$

While (11) resembles the impulse response function suggested by Gallant et al. (1996), Koop et al. (1996) and Koop (1996), three important differences need to be noted. First, our responses are history dependent but state independent - histories are not random variables. Second, we condition on the future path of the reduced form shocks. We do this for two reasons: responses are easier to compute this way and produce numerically more stable distributions; when the model has constant coefficient responses to structural shocks generated by (11) correspond to the standard ones. Note that our impulse responses display larger variability than those of Gallant et al.(1996) and Koop et al. (1996) since future shocks are averaged out. Third, since $\theta_{t+\tau}$ is integrated out, we concentrate attention on time differences which depend on the history of y_t and θ_t but not on the size of the sample. Note also that, contrary to the case of constant coefficient models, the sign and the size of certain elements of $\mathcal{Z}_{jt}$ affect the responses of the system.

To see what definition (11) involves rewrite the original VAR model (1) in a companion form

$$Y_{t+\tau} = A_{t+\tau} Y_{t+\tau-1} + C_{t+\tau} W_{t+\tau-1} + E_{t+\tau} \tag{12}$$

and let

$$\delta_{t+\tau} = \Xi[(I - \mathcal{C})(\mathcal{P}\mu + \epsilon) + \mathcal{C}\theta_{t+\tau-1} + \eta_{t+\tau}] + u_{t+\tau} \tag{13}$$

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 $Y^{t-1} = (Y_{t-1}, Y_{t-2}, \dots,$ $W_{t-1}, W_{t-2}, \dots,),$ $A^t = (A_t, A_{t-1}, \dots), C^t = (C_t, C_{t-1}, \dots)$ and $H_{t+\tau} = H_t \ \forall \tau$ as given, and solving backward we can write (12) as

$$\begin{aligned} Y_{t+\tau} &= (\prod_{k=0}^{\tau} A_{t+\tau-k}) Y_{t-1} + C_{t+\tau} W_{t+\tau-1} + \sum_{h=1}^{\tau} (\prod_{k=0}^{h-1} A_{t+\tau-k}) C_{t+\tau-h} W_{t+\tau-h-1} \\ &+ H_{t+\tau} v_{t+\tau} + \sum_{h=1}^{\tau} (\prod_{k=0}^{h-1} A_{t+\tau-k}) H_{t+\tau-h} v_{t+\tau-h} \end{aligned} \tag{14}$$

while solving backward (13) we have

$$\delta_{t+\tau} = \Xi(I - \mathcal{C})(\mathcal{P}\mu + \epsilon) \sum_{k=0}^{\tau} \mathcal{C}^k + \Xi \mathcal{C}^{\tau+1} \theta_{t-1} + \Xi \sum_{k=0}^{\tau} \mathcal{C}^k \eta_{t+\tau-k} + u_{t+\tau} \tag{15}$$

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

$$\begin{aligned} IR(t,t+\tau) &= E_t[Y_{t+\tau}|Y^{t-1},\delta^t,\mathcal{V}_t,H_t,\{\bar{v}_{jt+m}\}_{k=0}^m,\{v_{-jt+k}\}_{k=0}^m,\{v_{t+k}\}_{k=m+1}^{\tau}] \\ &- E_t[Y_{t+\tau}|Y^{t-1},\delta^t,\mathcal{V}_t,H_t,\{v_{t+k}\}_{k=0}^{\tau}] \\ &= E_t[(\prod_{k=0}^{\tau-1} A_{t+\tau-k})^j H_t^j(\bar{v}_{jt}-Ev_{jt}) + (\prod_{k=0}^{\tau-2} A_{t+\tau-k})^j H_{t+1}^j(\bar{v}_{jt+1}-Ev_{jt+1}) + \ldots \\ &+ (\prod_{k=0}^{\tau-m-1} A_{t+\tau-k})^j H_{t+m}^j(\bar{v}_{jt+m}-Ev_{jt+m})] \end{aligned} \tag{16}$$

where the superscript j refers to the j-th column of the matrix. It is easy to see that, when $\delta_t = \delta, \forall t$, (16) 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 since, e.g. a shock in v_t induces changes in δ_t.

Given (16), responses in our reparametrized model can be computed as follows

1. Choose a t, a τ and an J_t. Draw $\Omega^l = H_t^l(H_t^l)'^2)^l, (B_t)^l, \Psi^l$ from their posterior distribution and u_t^l from $N(0,(\sigma^2)^l I \otimes H_t^l(H_t^l)')$. Compute $y_t^l = \mathcal{X}_t\theta_t + H_t\bar{v}_t + X_t u_t^l$.
2. Use the law of motion of the coefficients to compute θ_{t+1}^l, $l=1,\ldots,L$ after drawing $\eta_{t+1}^l, u_{t+1}^l, \epsilon^l$ and the definition of Ξ to compute $\mathcal{X}_{t+1}$. Draw u_{t+1}^l from $N(0,(\sigma^2)^l I\otimes H_t^l(H_t^l)')$ and compute $y_{t+1}^l = \mathcal{X}_{t+1}\theta_{t+1}^l + H_t\bar{v}_{t+1} + X_{t+1}u_{t+1}^l$, $l=1,\ldots,L$.
3. Repeat step 2. and compute $\theta_{t+k}^l, y_{t+k}^l,\ k=2,\ldots\tau$.
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.

Shocks to the law of motion of the factors can be computed in the same way. A shock in $\eta_t = \bar{\eta}$ lasting one period implies from (15) that

$$E(\bar{\delta}_{t+\tau} - \delta_{t+\tau}) = \Xi\sum_{k=0}^{m} H_{t+k}\mathcal{C}^k(\bar{\eta}_{t+\tau-k} - E\eta_{t+\tau-k}) \tag{17}$$

so that

$$IR(t,t+\tau) = E_t[\prod_{k=0}^{\tau}(\bar{A}_{t+1\tau-k} - A_{t+\tau-k})Y_{t-1} + \sum_{h=1}^{\tau}\prod_{k=0}^{h-1}(\bar{A}_{t+1\tau-k} - A_{t+\tau-k})C_{t+\tau-h}W_{t+\tau-h-1} \tag{18}$$

$$+ \sum_{h=1}^{\tau}\prod_{k=0}^{h-1}(\bar{A}_{t+1\tau-k} - A_{t+\tau-k})H_{t+\tau-h}v_{t+\tau-h}] \tag{19}$$

Therefore, responses to shocks in the factors can also be easily computed using the output of the Gibbs sampler routine.

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 $\bar{W}_{t+k},\ k = 0, 1, \ldots, m$. Defining the response of $Y_{t+\tau}$ as the difference between the conditional expectations of $Y_{t+\tau}$ under the two different paths for W_{t+k} we have

$$IR(t, t+\tau) = E[(\prod_{k=0}^{\tau-2} A_{t+\tau-k})C_{t+1}(\bar{W}_{jt} - W_{jt}) + (\prod_{k=0}^{\tau-3} A_{t+\tau-k})C_{t+2}(\bar{W}_{jt+1} - W_{jt+1}) \quad (20)$$

$$+ \ldots + (\prod_{k=0}^{\tau-2-m} A_{t+\tau-k})C_{t+m+1}(\bar{W}_{jt+m} - W_{jt+m})] \quad (21)$$

Therefore, to compute conditional forecasts of this type in our 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 $\theta_{t+1}^l, l = 1, \ldots, L$ 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 $Y_t = A_t Y_{t-1} + C_t W_{t-1} + E_t$ into two blocks, let $Y_{2t+k} = \bar{Y}_{2t+k}$ be the fixed variables and Y_{1t+k} those allowed to adjust. Then we have that

$$IR(t, t+\tau) = E[H_t^1(\prod_{k=0}^{\tau-1} A_{t+\tau-k})^1(\bar{v}_{2t} - v_{2t}) + H_{t+1}^1(\prod_{k=0}^{\tau-2} A_{t+\tau-k})^1(\bar{v}_{2t+1} - v_{2t+1}) \quad (22)$$

$$+ \ldots + H_{t+m}^1(\prod_{k=0}^{\tau-1-m} A_{t+\tau-k})^1(\bar{v}_{2t+m} - v_{2t+m})] \quad (23)$$

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 second 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. Draw η_{t+1}^l, ϵ^l and use the law of motion of the factors to obtain $\theta_{t+1}^l,\ l = 1, \dots, L$ 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 and use the law of motion of the factors to compute $\theta_{t+2}^l,\ l = 1, \dots, L$ and the definition of Ξ to compute $\mathcal{X}_{t+2}$.

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

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

6 The transmission of shocks in G-7 countries

In this section we show how to use multi-country VAR models to examine two issues which are important for policy makers: what are the effects of a US shock on GDP of the G-7 countries and what are the consequences of a persistent oil price increase on inflation in Euro area countries.

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 suggests that a time varying specification will be probably useful in modelling cross country interdependencies.

For each of the G-7 countries we use 4 endogenous variables (real GDP growth, CPI inflation, employment growth, and rent inflation) and one predetermined ones (the growth rate of an oil price index). Besides GDP growth and CPI inflation, which are the focus of attention in this section, 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. Five lags of the endogenous variables, a constant and two lags of the predetermined variable are used. Therefore, each equation has k=7*4*5+2=1=143 coefficients and there are 28 equations in the system. The estimation sample covers the period 1980:1-2000:4, and the exercises we conduct are conditional on estimates obtained with the information up to 2000:4. Since a training sample is not available, we set $\Psi = \mathcal{C} = 0$, $\mathcal{P} = I$, $a = 10$, $b = 1$, choose $\zeta_t = 1$, $\bar{B} = b_i * I$ and let $p(b_i) = G(0.5, 5), i = 1, 2, \dots, F$. We initialize $\bar{\theta}$ with a sequential OLS on the time invariant version of the model and set σ^2 to the average estimated variances of NG AR(p) models.

The vector δ_t is decomposed into three factors: a 2×1 vector of common factors, θ_{1t} - one for the Euro area and one for the rest of the world; a 7×1 vector of country specific factors, θ_{2t}; and a 4×1 vector of variable specific factors, θ_{3t}, Hence, $\theta_t = (\theta'_{1t}, \theta'_{2t}, \theta'_{3t})'$ is 13×1 vector.

We produce 55,000 iterations of the MCMC routine starting from arbitrary initial conditions. Runs of 50 elements are drawn 1100 times and the last observation of the final 1000 runs was used for inference. We checked convergence recursively calculating the first two moments of the posterior of the parameters using 300, 500, 1000 draws and found that convergence was sufficiently easy to achieve and obtained with about 500 draws.

In order to preliminary search across possible specification, we have computed the marginal likelihood for 6 models: M_0 is our benchmark specification; M_1 is a model with an exact factorization of δ_t; in M_2 there are no interdependencies; in M_3 there are no time variations; finally, in M_4 and M_5 we exclude the country and the variable specific components θ_{2t} and θ_{3t}, respectively. Marginal likelihoods are computed following Chib (1995), treating both θ_t and σ_t as latent vectors.

Table 1. Log Marginal Likelihood of models

sample	M_0	M_1	M_2	M_3	M_4	M_5
1980-2000	7336.3	10594.2	9377.3	-1799.1	10404.7	9117.9

Several important aspects of table 1 are worth emphasizing. First, M_1, the model with an exact factorization of δ_t is preferred. Second, a model with lagged cross-country interdependencies (M_1) is superior to a model without interdependencies (M_2). Hence, there is important information in the lags of the variables which is neglected when standard static factor approaches are used. Third, consistently with Del Negro and Otrok (2003), time variations in the loadings are extremely useful in tracking the dynamics of the data (compare M_1 and M_3). As a matter of fact, M_3 has the lowest marginal likelihood of all the specifications we consider. Finally the marginal likelihood of a model with three indices is always higher than the marginal likelihood of a model with only two indices, regardless of whether the two indices capture world and variable specific factors (M_4), or world and country specific factors (M_5).

Figure 1, which plots the evolution of the posterior mean and the centered posterior 68% band for θ_{2t}, supports the idea that country specific factors are important. Excluding relevant cases, the seven factors are small but significant. Furthermore, they appear to add important variations at certain dates (see e.g. the German factor at the time of German unification).

In conclusion, lagged interdependencies, unit specific dynamics and time variations appear to be important features of our VAR. Furthermore, a factorization of the coefficient vector which includes three factors and allows for no idiosyncratic component summarizes the information present in the VAR reasonably well.

[natheight=5.361in, natwidth=9.6963in, height=4.5674in, width=6.4307in]
C:/swp50/temp/graphics/count$_{loading1_{1.pdf}}$
Figure 1. Evolution of country loadings over time

First, we consider the effect of a US real shocks on the GDP of other countries. We construct such a shock by making US variables contemporaneously casually prior. Within the US block, employment growth and output growth increase by one unit for one period, while the other two variables change according to the domestic correlation matrix. Figure 2 presents the median responses together with a 68 percent posterior band. Three features of the figure are worth emphasizing. First, responses are relatively smooth despite the large number of VAR parameters because of the moving average nature of our indices. Second, there appears to be a significant Anglo-Saxon real cycle with Canadian and UK GDP growth responding significantly and instantaneously to the US shock. The response is instantaneously significant also in Japan and, to a much smaller extent, in Germany. Third, the peak response in Italy and France is delayed by at least one period, suggesting that transmission to these two countries takes time an probably occurs via Germany. Finally, impulses typically dissipate very quickly: except for Italy and France, all response bands include zero two quarters after the shocks.

[natheight=3.7325in, natwidth=3.8086in, height=1.7063in, width=2.0081in]
C:/swp50/temp/graphics/japan$_{2.pdf}$
Japan

[natheight=3.6971in, natwidth=3.8181in, height=1.7063in, width=2.0081in]
C:/swp50/temp/graphics/uk$_{3.pdf}$
UK

[natheight=3.6737in, natwidth=3.8181in, height=1.7063in, width=2.0081in]
C:/swp50/temp/graphics/canada$_{4.pdf}$
Canada

[natheight=3.6971in, natwidth=3.8181in, height=1.7063in, width=2.0081in]
C:/swp50/temp/graphics/germany$_{5.pdf}$
Germany

[natheight=3.6971in, natwidth=3.8181in, height=1.7063in, width=2.0081in]
C:/swp50/temp/graphics/France$_{6.pdf}$
France

[natheight=3.6737in, natwidth=3.8181in, height=1.7063in, width=2.0081in]
C:/swp50/temp/graphics/Italy$_{7.pdf}$
Italy

Figure 2. 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 30 percent higher than its 2000:4 level and the increase lasts for four quarters. Figure 3 reports the posterior median and the posterior 68% band for inflation responses in Germany, Italy and France. Responses in the three countries look different in magnitude and timing. All countries have a delayed peak reaction. However, the peak of German inflation occurs after 3 to 4 quarters after the shock has died out, whereas for Italy and France the peak response is at quarter 2 and the reaction remains significantly positive only for 3 to 4 quarters, roughly the length of the increase. Interestingly, the average magnitude of the reaction of French inflation is lower than the one obtained in Italy and Germany.

[natheight=2.879in, natwidth=3.8086in, height=1.7063in, width=2.0081in]
C:/swp50/temp/graphics/IRUTFG00$_{8.pdf}$
Germany

[natheight=2.879in, natwidth=3.8086in, height=1.7063in, width=2.0081in]
C:/swp50/temp/graphics/IRWBIS00$_{9.pdf}$
France

[natheight=2.8599in, natwidth=3.8086in, height=1.7063in, width=2.0081in]
C:/swp50/temp/graphics/IRWBJP01$_{10.pdf}$
Italy

Figure 3. 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. In Figure 4 we present the time profile for the posterior 68% band for a coincident measure of potential world output growth, constructed as $CVLI_t^{GDP} = (\mathcal{X}_{1t}\theta_{1t} + \mathcal{X}_{3t}\theta_{3t})^{GDP}$. Three features are worth emphasizing. First, cyclical movements of potential output roughly correspond to those of actual output. Second, there is a marked and significant difference in the level of potential output growth in the 1990's as compared to the end of 1980's. The decline is driven by both Japanese and the Euro area variables. Third, our measure of potential output starts declining significantly at the beginning of 2000.

[natheight=5.6291in, natwidth=10.2057in, height=2.8504in, width=5.2702in]
C:/swp50/temp/graphics/potential1$_{11.pdf}$
Figure 4. Potential output growth

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 allows for 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 indices are predetermined with respect to the endogenous variables, it is easy to construct Bayes factors to select the number of indices or to examine the model specification to be used.

The tools described in this paper can be applied to a number of interesting problems. For example, Canova, et al. (2003) have used a multi-country VAR to extract world and national business cycles while Anzuini, et al. (2005) use a multi-country VAR structure to construct coincident and leading indicators for inflation and output growth in Italy and the Euro area. 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 far from being computationally demanding (one full run of the MCMC routine for the example of section 6 takes about 45 minutes). Therefore, the approach is at least competitive with existing alternatives.

References

[1] Anzuini, A. Caivano, M., Canova, F., Lamorgese, A. and Venditti, F. (2005), A leading indicator for Inflation and GDP growth in Italy and the Euro area, Bank of Italy, manuscript.

[2] Binder, M., C. Hsiao and H. Pesaran (2001), Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration, University of Maryland, manuscript.

[3] Canova, F. (1993), Modelling and Forecasting Exchange Rates using a Bayesian Time varying coefficient model, *Journal of Economic Dynamics and Control*, 17, 233-262.

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

[5] Canova, F., M. Ciccarelli and E. Ortega (2003), Similarities and Convergence of G-7 cycles, forthcoming, *Journal of Monetary Economics*.

[6] Canova, F. and Pappa, E. (2003), Price differential in monetary unions: the role of fiscal shocks, forthcoming, *Economic Journal*.

[7] Ciccarelli, M and Rebucci, A. (2003) Measuring Contagion using a Bayesian Time varying coefficient model, ECB Working paper.

[8] Chamberlin, G. (1983), Panel Data, in Griliches, Z. and M. Intrilligator (eds.), *The Handbook of Econometrics, II*, North Holland.

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

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

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

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

[13] Del Negro, M. and Otrok, C. (2003) Dynamic Factor Models with Time-Varying Parameters: Measuring the evolution of the European Business Cycle, Federal Reserve Bank of Atlanta, manuscript.

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

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

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

[17] Geweke, J. (2000) 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.

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

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

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

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

[22] Kim, and Nelson, C. (1998) 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(3), 188-201.

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

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

[25] Pesaran, H. (2003), Estimation and Inference in Large Heterogeneous Panels with Cross Section Dependence, University of Cambridge, working paper 0305.

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

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