Panel index VAR models: Specification, Estimation, Testing and Leading Indicators

Fabio Canova *
Universitat Pompeu Fabra and CEPR
and
Matteo Ciccarelli
European Central Bank

First Draft, June 2002
This Revision, July 2003

Abstract

This paper proposes a method to conduct inference in panel VAR models with cross unit interdependencies and time variations in the coefficients. The approach can be used to obtain multi-unit forecasts and leading indicators and to conduct policy analysis in a multiunit setups. The framework of analysis is Bayesian and MCMC methods are used to estimate the posterior distribution of the features of interest. The model is reparametrized to resemble an observable index model and specification searches are discussed. As an example, we construct leading indicators for inflation and GDP growth in the Euro area using G-7 information.

Key Words: Panel VAR, Bayesian methods, Leading indicators, Markov Chain Monte Carlo methods

Jel Classification nos: C3, C5, E5

---

*We would like to thank 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.

1 Introduction

There has been a growing interest in using panel 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, in part, to advances in computer technology, which made the estimation of large scale models feasible in reasonable time. Problems concerning the transmission of shocks across countries, sectors or industries; issues related to convergence and to the evaluation of the effects of regional policies or the composition of portfolio of assets are naturally studied in this framework. Two characteristics distinguish macro panels from micro ones: first, cross unit interdependencies are likely to be more important in explaining the dynamics of the data in the former than in the latter, especially once a (common) time effect is taken into account. Second, while in micro panels the number of units is typically large and the time series short, in macro panels the number of units is generally limited and the time series dimension is of moderate size. These distinctive features make the inferential problem non-standard. So, for example, the GMM estimator of Holtz Eakin et al. (1988), the QML and a minimum distance estimators proposed by Binder, Hsiao and Pesaran (2001), all of which are consistent and asymptotically normal 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.

Regardless of which dimension is assumed to be asymptotically large, one is generally forced to impose strong restrictions to obtain estimates of the parameters of interest. For example, it is typical to assume that slope coefficients are common across units; that there are no interdependencies across units; that the structure is stable over time or a combination of all of these. None of these restrictions is appealing in macroeconomic frameworks: unit specific relationships may reflect differences in national regulations or policies; interdependencies are the results of world markets integration and time instabilities are the natural consequence of evolving economic structures. Recently, Canova and Ciccarelli (1999) proposed a framework which allows for unit specific dynamics and time variations in a panel VAR. Given the nature of the model, a hierarchical Bayesian approach is used to construct posterior estimates of the features of interest. Although the framework has appealing features, and its forecasting performance is good relative to more parsimoniously built candidates, the estimation process is computationally demanding whenever the structure of time variations is different across variables and units.

The last few years have also witnessed a renewed interest in using index models in macroeconomics - for example, to extract national and international business cycles or to capture the driving forces in APT models. Index models are based on the idea that the dynamics of a large number of macroeconomic series can be represented as the sum of low dimensional factors which are common to all (or a subset of the) units or variables, and of an orthogonal idiosyncratic residual. Static versions of one-index models have been used e.g. by Stock and Watson (1989) to construct coincident and leading indicators of economic activity and are routinely employed in statistical and government agencies. The static setup has been extended by Forni, Hallin, Lippi and Reichlin (FHLR) (2000) who allow for serial dependence in the index, by Otrok and Whiteman (1998) who study a Bayesian version of it, and by Stock and Watson (1998) and Marcellino, Stock and Watson (2003). Pesaran (2003) considers unobservable indices in dynamic cross sectional setups. Camba Mendez et al. (2001) provide a forecasting comparison of these models with VAR and BVAR. Despite remarkable progresses in the specification and estimation of these models, problems still remain. For example, in the FHLR approach estimates of the indices are functions not only of present and past dynamics but also of the future ones, therefore preventing their use for forecasting and policy purposes. Furthermore, all approaches but Otrok and Whiteman require a large cross sectional dimension for standard asymptotic theory to apply. Finally, time variations are not typically allowed for.

This paper develops a methodology for conducting inference in general macro panel VAR models. Because of interdependencies, unit specific dynamics and time variations in the coefficients, no classical estimation method is feasible. We take a Bayesian viewpoint and restrict the coefficient vector to have a low dimensional time varying factor structure. These factors are by construction orthogonal and capture, for example, variations in the coefficients which are common across units and variables (a "common" effect); variations across variables within a unit ("fixed" effects) or variations across units in a particular variable ("variable" effects). Factors relating to lags, time periods, or combinations of any of the above, can also be included. We complete the prior specifications using a hierarchical structure for the factors which allows for exchangeability across units, time variations and heteroskedasticity in the innovations driving the factors.

The factor structure on the coefficient vector allow us to transform a potentially overparametrized panel VAR into a parsimonious SUR model where the regressors are a set of observable indices, constructed using certain linear combinations of the right-hand-side variables of the VAR, and the loadings are the time varying factors. We derive posterior estimates for the unknowns using Markov Chain Monte Carlo (MCMC) methods. Posterior distributions of interesting functions of the loadings can be obtained as a by-product of the Monte Carlo routine. In particular, we show how to compute forecast revisions (generalized impulses) in response to unexpected perturbations in either the innovations of the VAR or in the loadings of one of the observable indices. These exercises are useful to trace out distributions of future scenarios following, e.g. disturbances to the systematic and the unsystematic part of policy reactions functions or to describe the effect of disturbances which commonly affects coefficients of one or more units.

The reparametrization of the VAR in terms of dynamic observable factors has a number of appealing features. First, it reduces the problem of estimating a large number of, possibly, unit specific and time varying coefficients for each equation into the problem of estimating a small number of loadings on certain combination 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 requires the estimation of $1 + N + G$ loadings when a common, a unit and a variable specific vector of factors are specified. This considerably reduces the computational burden and allows the estimation of large scale models in reasonable time. Second, since indices are constructed and estimated recursively in real time, they can be employed for a variety of policy and forecasting purposes. For example, one can construct a multi-step multi-country leading indicator of economic activity; produce real-time estimates of core inflation or the natural rate of unemployment; extract recursive estimates of world and national business cycles (see Canova, Ciccarelli and Ortega (2003)); or study the propagation of shocks across countries in various time periods. Third, since indices are observable and predetermined with respect to the endogenous variables, it is easy to design a statistic to determine how many should be included in a model. We propose a simple approach, based on predictive Bayes factors, which can be used for this purpose.

While our reparametrization leads us to estimate something similar to an index model, it is important to stress that our starting point is a panel VAR with interdependences and that our factorization is on the coefficients of the model. This should be contrasted with standard index models where indices are constructed directly from the variables entering the model. Also, because of our Bayesian setup, we can allow for time variations in the loadings - a feature which is not easily dealt with in standard index models - and for cross unit interdependencies - a possibility typically excluded in micro panel VARs. Finally, by construction, our observable indices dynamically span the interdependencies of the data.

The structure of the paper is a follows: the next section presents the setup of the model and the prior restrictions. Section 3 describes estimation and inference. Section 4 deals with measurement errors. Section 5 discusses the transformation of the panel VAR into an observable index model and a number of specification searches. Section 6 deals with generalized impulse responses. In Section 7 we apply the methodology to constructing leading indicators for inflation and GDP growth in the Euro area. Section 8 concludes.

2 A general framework

The panel 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 for each i, $Y_t = (y'_{1t}, y'_{2t}, \dots y'_{Nt})'$, $D_{it,j}$ are $G \times G$ matrices each j, $C_{it,j}$ are $G \times q$ matrices each j; W_t is a $q \times 1$ vector of exogenous variables, common to all i, and e_{it} is a $G \times 1$ vector of random disturbances. We assume p lags for the G endogenous variables and r lags for the q exogenous variables. In (1) we say that there are cross-unit lagged interdependencies whenever $D_{it}^{h,h'}(L) \neq 0$ for any $h' \neq h$ some L. To see what this feature entails, consider a version of (1) with $N = 2$, $G = 2$, $p = 2$, $q = 0$ of the form:

$$Y_t = D_{1t}Y_{t-1} + D_{2t}Y_{t-2} + e_t \tag{2}$$

where $Y_t = [y_{11t}; y_{12t}; y_{21t}; y_{22t}]'$ and $\mathrm{var}(e_t) = \Sigma_e$. Then, lagged cross units interdependencies appear whenever D_{1t} or D_{2t} is not block diagonal. The presence of lagged cross unit interdependencies adds flexibility to the specification but it is not without costs: the number of parameters is greatly increased (we have now $k = NGp + qr$ parameters in each equation); furthermore, the G variables entering the model must be the same for each i.

In (1) the dynamic relationships are allowed to be unit specific. Furthermore, the coefficients are allowed to vary over time. While this latter feature may be of minor importance in micro panels whenever T is short, it is crucial in macro setups where structural changes are more common. A flexible and parsimonious specification for the law of motion of the coefficients is specified below. We rewrite (1) as:

$$Y_t = X_t\delta_t + E_t \qquad E_t \sim N(0, \Omega) \tag{3}$$

where $X_t = I_{NG} \otimes \mathbf{X}_t'$; $\mathbf{X}_t = (Y_{t-1}',\ Y_{t-2}', \ldots, Y_{t-p}',\ W_t', \ldots, W_{t-l}')'$; $\delta_t = (\delta_{1t}', \ldots, \delta_{Nt}')'$ and $\delta_{it} = (\delta_{it}^{1\prime}, \ldots, \delta_{it}^{G\prime})'$. Here δ_{it}^g are $k \times 1$ vectors containing, stacked, the G rows of the matrices D_{it} and C_{it}, while Y_t and E_t are $NG \times 1$ vectors containing the endogenous variables and the random disturbances.

Whenever δ_t varies with cross–sectional units in different time periods, it is impossible to estimate it using classical methods. Two shortcuts are typically employed in the literature: it is assumed that the coefficient vector does not depend on the unit, apart from a time invariant fixed effect or that there are no interdependencies across units (see e.g. Chamberlain (1982), Holtz Eakin et al. (1988) or Binder et al. (2001)). Neither of these assumptions is appealing in our context. Instead, we assume that δ_t can be factored as:

$$\delta_t = \Xi_1\lambda_t + \Xi_2\alpha_t + \sum_{f=3}^{F} \Xi_f\rho_{f,t} + u_t \tag{4}$$

where Ξ_1 is a matrix of ones and zero of dimensions $NGk \times N_1 << N$; Ξ_2 is a matrix of ones and zeros of dimensions $NGk \times N$, and Ξ_f are conformable matrices. Here λ_t is a vector of common factors, α_t a vector of unit specific factors (the fixed effect), and $\rho_{f,t}$ a set of factors which could be indexed by the unit i, the variable g, the variable in a given equation m (independent of unit), the unit in a given equation s (independent of variable), the lag h or combinations of all of the above.

All the factors in (4) are assumed to be orthogonal to each other and allowed to be time varying. Time invariant structures can be obtained via restrictions on their law of motion, as detailed below. Also, while it is possible to decompose δ_t exactly, in practice only a few factors will be specified and the error term u_t will aggregate all the omitted terms.

Continuing with the previous example, rewrite (2) as in (3) where δ_t now is a 32×1 vector with typical element $\delta_{m,s,h,t}^{i,g}$. Then

$$\begin{aligned} \delta_{m,s,h,t}^{i,g} &= \lambda_t + \alpha_t^i + \rho_{1t}^g + \rho_{2,t}^m + \rho_{3,t}^s + \rho_{4,t}^h \\ &= \lambda_t + \alpha_t^i + u_{m,s,h,t}^g \end{aligned} \tag{5}$$

As it is clear from (5) each coefficient can be decomposed in several factors reflecting the position that the coefficient occupies in the system (unit, equation, variable, lag, etc.). Here, λ_t is a common factor, $\alpha_t = (\alpha_t^1,\ \alpha_t^2)'$ is a 2×1 vector of unit specific factors, $\rho_{1t} = (\rho_{1t}^1,\ \rho_{1t}^2)'$

is a 2×1 vector of variable specific factors, $\rho_{2,t} = (\rho^1_{2,t},\ \rho^2_{2,t})'$ is a 2×1 vector of variable specific factors in equation m, $\rho_{3,t} = (\rho^1_{3,t},\ \rho^2_{3,t})'$ is a 2×1 vector of unit specific factors in variable s and $\rho_{4,t} = (\rho^1_{4,t},\ \rho^2_{4,t})'$ is a 2×1 vector of lag specific factors across variables and equations. In the second line of (5), we have retained only the common and the unit specific factors and pooled the remaining ones into the error term.

While we interpret (4) as part of the prior, one is also free to see it as a part of the model. In the former case, the interest of the researcher or computational considerations dictate many factors are included. For example, in cross country study of business cycle transmission, a common and a country specific factors are probably sufficient while when constructing leading indicators of GDP in one country, one may want to specify at least a common, a country and a variables specific factors. In the latter case, the size of F can be statistically determined. We discuss this issue in section 5.

We let $\theta_t = [\lambda'_t, \alpha'_t, \rho'_{1,t}, \ldots, \rho'_{F,t}]$. Then (4) can be compactly written as

$$\delta_t = \Xi\theta_t + u_t \qquad u_t \sim N(0, \Sigma \otimes V) \tag{6}$$

where $\Xi = [\Xi_1,\ \Xi_2,\ \Xi_3, \ldots, \Xi_F]$, V is a $k \times k$ matrix and Σ a $NG \times NG$ matrix. We assume a hierarchical structure for θ_t of the form:

$$\theta_t = (I - \mathcal{C})\,\bar{\theta} + \mathcal{C}\theta_{t-1} + \eta_t \qquad \eta_t \sim N\,(0, B_t) \tag{7}$$

$$\bar{\theta} = \mathcal{P}\mu + \epsilon \qquad \epsilon \sim N(0, \Psi) \tag{8}$$

where $\bar{\theta}$ is the unconditional mean of θ_t. Furthermore, we assume that u_t, η_t, and ϵ are mutually independent and that $\mathcal{P}, \mathcal{C}$ are known. Here $\mathcal{C}$ is a full rank matrix, $\mathcal{P}$ a matrix which restricts (part of the) the means of θ_t's with an exchangeable structure. Finally we let

$$V = \sigma^2 I_k \tag{9}$$

$$B_t = \gamma_1 * B_{t-1} + \gamma_2 * \bar{B} = \xi_t * \bar{B} \tag{10}$$

$B_0 = \bar{B}$ with $\xi_t = \gamma_1^t + \gamma_2 \frac{(1-\gamma_1^t)}{(1-\gamma_1)}$ where $\bar{B} = diag(\bar{B}_1, \ldots \bar{B}_F)$, and γ_1 and γ_2 are known.

(6)-(10) describe the prior for the coefficients of the panel VAR. We have chosen to be general at this stage. Restricted specifications will be examined using the posterior distribution for the parameters. In (6) the form Σ is application dependent - for example, it could be the identity matrix, the Kronecker product of a unit specific and a variable specific matrices or a multiple of the variance covariance matrix of the innovations in (3). In (7) the factors evolve over time in a geometric fashion. When the eigenvalues of $\mathcal{C}$ are all less than one, this is simply an AR(1) with nonzero mean. If $\mathcal{C}$ is a circulant matrix, seasonal relationships could be modelled. In (8) the means of the factors are weakly linked across units. Thus, for example, if the unit specific factors are thought to be drawn from a distribution with common mean and there are, e.g. four units, two variables and three factors in (4), then:

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

The variance of the innovations in θ_t is allowed to be time varying to account for heteroskedasticity and other generic volatility clustering that may appear in several, or all, coefficients within and across units. The specification is flexible, builds on the one used by Canova (1993), and nests two important special cases: (a) no time variation in the factors, $\gamma_1 = \gamma_2 = 0$, and $\mathcal{C} = I$, and (b) homoskedastic variance $\gamma_1 = 0$ and $\gamma_2 = 1$. The spherical assumption (9) reflects the fact that factors are measured in common units, while the block diagonality of $\bar{B}$ is needed to guarantee the orthogonality across factors (which is preserved a-posteriori), and hence their identifiability.

Recently, Cogley and Sargent (2002) have used a specification similar to the above to study the changing dynamics in US inflation. However, to capture conditional heteroskedasticity in inflation they set $B_t = \bar{B}$ and specify Ω to be a function of a set of stochastic volatility processes. Our specification also produces conditional heteroskedasticity in the endogenous variables of the model and has two advantages. First, it retains conditional linearity, therefore making simulation of the posterior distributions substantially easier. Second, it captures a variety of non-normal patterns because of the interaction between changes in the law of motion of the coefficients and the evolution of the variables in the VAR.

Several specifications are nested in our general framework. For example, we can accommodate the case where some components of δ_t are a-priori independent of time, by making B_t a reduced rank matrix and setting the appropriate elements of $\mathcal{C}$ to zero. Thus if α_t is time invariant and three factors are used then $B_t = \textit{block-diag}[B_{1t}, 0, B_{3t}]$ and $\mathcal{C} = [\mathcal{C}_1, 0, \mathcal{C}_3]$. Furthermore, if exchangeability is not appropriate a-priori, we can choose Ψ to be large. Alternatively, if mean pooling is deemed necessary we can make $\Psi \to 0$. Finally, if enough factors are included, (6) can be made exact by setting $\sigma^2 = 0$.

Although we have specified normal distributions for all errors, we can account for aberrant observations or errors with fat-tailed distributions simply replacing one of these distribution with a family of longer-tailed distributions. For example, we could assume $u_t \mid h_t \sim N(0, h_t(\Sigma \otimes V))$ with $h_t \sim$Inv-χ^2 $(\nu, 1)$, where Inv-χ^2 is an inverted chi-squared with ν degrees of freedom and scale equal 1, in which case $u_t \sim t_\nu(0, \Omega \otimes V)$. As we will show below fat tail distributions in the errors of the model naturally emerge from our factor structure even assuming normality of all disturbances.

3 Inference

The likelihood of the data is proportional to

$$\begin{aligned}
& |\Omega|^{-T/2} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left(Y_t - X_t\delta_t\right)'\Omega^{-1}\left(Y_t - X_t\delta_t\right)\right\} \\
= \;& |\Upsilon|^{-T/2} \exp\left[-\frac{1}{2}\sum_{t}\left(Y_t - X_t\Xi\theta_t\right)'\left(\Upsilon\right)^{-1}\left(Y_t - X_t\Xi\theta_t\right)\right]
\end{aligned}$$

where $\Upsilon = \Omega + \sigma^2\Sigma\mathbf{X}_t'\mathbf{X}_t$.

To calculate the posterior distribution for the unknowns we need prior densities for $(\Omega^{-1},\ \Sigma^{-1}\ \mu,\ \Psi^{-1},\ \sigma^{-2},\ \bar{B}^{-1})$. Considerable simplifications in the calculations can be achieved if we assume $\Sigma = \Omega$, as it is done e.g. in the Minnesota prior (see Doan, Litterman and Sims (1984)) or in other standard priors (e.g. Kadiyala and Karlsson, 1997). In this case Υ reduces to $\Upsilon = (1 + \sigma^2\mathbf{X}_t'\mathbf{X}_t)\Omega \equiv \sigma_t\Omega$. Intuitively, this assumption induces a correlation between the innovations of the model and those of the coefficients. Hence, innovations in the panel VAR may alter the dynamics of the model. Having this feature is important in forecasting contexts, as we will show in section 7.

We assume that an investigator has available a "training sample" which can be used to "estimate" prior features of the model or that observations on similar units provide information on how the hyperparameters of the model are likely to behave. This assumption is not restrictive: when no such information is available or when a researcher is interested in minimizing the impact of prior choices on the posterior, one simply need to appropriately modify the expressions for the posterior moments presented below. Furthermore, for large T, the posterior will be invariant to the specification of the prior distribution.

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

$$\begin{aligned}
p(\Omega^{-1}) &= W(z_1, Q_1) \\
p(\mu) &= N(\bar{\mu}, \Sigma_\mu) \\
p(\Psi^{-1}) &= W(z_0, Q_0) \\
p(\sigma_t^{-1}) &= G\left(\frac{\zeta}{2}, \frac{\zeta s_t}{2}\right) \\
p(\bar{B}_f^{-1}) &\propto W(z_{2f}, Q_{2f}) \qquad f = 1, \ldots F
\end{aligned} \tag{11}$$

and where $s_t^{-1} = E(\sigma_t^{-1})$. Here N stands for Normal; W for Wishart and G for Gamma distributions. The hyperparameters $(z_0, z_1, z_{2f}, \zeta,\ vec(\bar{\mu}), vech(\Sigma_\mu),\ vech(Q_0, Q_1, Q_{2f}))$ are assumed to be known or estimated from the data where $vec\,(\cdot)$ ($vech\,(\cdot)$) denotes the columnwise vectorization of a rectangular (symmetric) matrix.

The assumed distributions imply, among other things, that the prior distribution for the forecast error $v_t = Y_t - X_t\Xi\theta_t = E_t + X_t u_t$ has the form $(v_t|\sigma_t) \sim N(0, \sigma_t\Omega)$ where

$\sigma_t \sim \text{Inv-}\chi^2(\zeta, s_t)$. Therefore innovations are endogenously allowed to have fat tails (v_t is distributed as a multivariate t centered at 0, with scale matrix proportional to Ω and ζ degrees of freedom).

Given the large number of parameters, the analytical integration needed to obtain posterior distributions is unfeasible. Therefore, we integrate numerically using the Gibbs sampling. This method is particularly useful in our framework since it only requires knowledge of the conditional posterior distribution of the unknowns, which we can calculate analytically given (11). Let $Y^T = (Y_1, ..., Y_T)$ denote the data, $\psi = (\Omega^{-1}, \bar{\theta}, \mu, \bar{B}_f^{-1}, \sigma_t^{-1}, \Psi^{-1}, \{\theta_t\}_t)$ the unknowns whose joint distribution needs to be found and $\psi_{-\varkappa}$ the vector of ψ excluding the parameter $\varkappa$. Furthermore, 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 $\left(\{\theta_t\}_{t=0}^T, Y^T\right)$, the conditional distribution for the unknowns are:

$$\Omega^{-1} \mid Y^T, \psi_{-\Omega} \sim W\left(\hat{z}_1, \hat{Q}_1\right);$$

$$\bar{\theta} \mid Y^T, \psi_{-\bar{\theta}} \sim N\left(\widehat{\bar{\theta}}, \hat{\Psi}\right);$$

$$\mu \mid Y^T, \psi_{-\mu} \sim N\left(\hat{\mu}, \hat{\Sigma}_\mu\right);$$

$$\Psi^{-1} \mid Y^T, \psi_{-\Psi} \sim W\left(\hat{z}_o, \hat{Q}_o\right);$$

$$\bar{B}_f^{-1} \mid Y^T, \psi_{-\bar{B}_f} \sim W\left(\hat{z}_{2f}, \hat{Q}_{2f}\right);$$

$$\sigma_t^{-1} \mid Y^T, \psi^*_{-\sigma_t} \sim G\left(\frac{\hat{\zeta}}{2}, \frac{\mathcal{R}}{2}\right).$$

where expressions for $\hat{z}_0, \hat{z}_1, \hat{z}_{2f}, \hat{Q}_0, \hat{Q}_1, \hat{Q}_{2f}, \widehat{\bar{\theta}}, \hat{\Psi}^{-1}, \hat{\mu}, \hat{\Sigma}_\mu, \hat{\zeta}, \mathcal{R}$ are in the appendix.

The conditional posterior of $(\theta_1, ..., \theta_T \mid Y^T, \psi_{-\theta_t})$, can be obtained with the Kalman filter. In particular, given $\theta_{0|0}$ and $R_{0|0}$ we have

$$\begin{aligned}
\hat{\theta}_{t|t} &= \hat{\theta}^*_{t|t-1} + K_t\left(Y_t - X_t\Xi\theta_t\right) \\
R_{t|t} &= \left(I - K_tX_t\Xi\right)R^*_{t|t-1} \\
K_t &= R^*_{t|t-1}X_t\Xi F^{-1}_{t|t-1} \\
F_{t|t-1} &= \left(X_t\Xi\right)R^*_{t|t-1}\left(X_t\Xi\right)' + \Upsilon
\end{aligned} \tag{12}$$

where $\hat{\theta}^*_{t|t-1} = \hat{\theta}^*_{t-1|t-1}$ and $R^*_{t|t-1} = R^*_{t-1|t-1} + \xi_t\bar{B}$, and $\hat{\theta}^*_{t-1|t-1}$ and $R^*_{t-1|t-1}$ are, respectively, the mean and the variance covariance matrix of the conditional distribution of $\theta^*_{t-1|t-1}$. Draws for θ_t are made from $N(\hat{\theta}_{t|t}, R_{t|t})$ if recursive estimates are needed or from $N(\hat{\theta}_{t|T}, R_{t|T})$ if smoothed estimates are preferred. The recursions can be started by choosing $R_{0|0}$ to be diagonal with elements equal to small values, while $\theta_{0|0}$ can be initialized by running a constant coefficient version of the model.

Convergence of the Gibbs sampler to the true invariant distribution is somewhat standard since the model (3) with the factor structure (6) is a time-varying SUR model with serially correlated errors (see e.g. Chib and Greenberg (1995)). Convergence in these setups typically occurs under a set of mild conditions (for example see Geweke (2000)).

Inference on any continuous function $\mathcal{G}(\psi)$ of the unknowns can be easily constructed using the output of the Gibbs sampler and the ergodic theorem. For example $E(\mathcal{G}(\psi)) = \int \mathcal{G}(\psi)p(\psi|Y)d\psi$ can be approximated using $\frac{1}{L}[\sum_{\ell=L+1}^{L+\bar{L}} \mathcal{G}(\psi^{\ell})^{-1}]^{-1}$. Predictive distributions for future y_{it}'s can be estimated using the recursive nature of the model and the simple conditional structure of (3). In particular, let $Y^{t+\tau} = (Y_{t+1}, \ldots, Y_{t+\tau})$. To compute forecasts and turning points we need to construct

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

where $\mathcal{F}\left(\mathcal{G}(Y^{t+\tau}) \mid Y_t, \psi\right)$ is the conditional density of the function $\mathcal{G}$ of future Y's, given ψ. Forecasts can be obtained drawing $\psi^{(\ell)}$ from the posterior distribution and simulating the vector $Y^{\ell,t+\tau}$ from the density $\mathcal{F}\left(Y^{t+\tau} \mid Y_t, \psi^{(\ell)}\right)$. $\left\{Y^{\ell t+\tau}\right\}_{\ell=L+1}^{L+\bar{L}}$ constitutes a sample, from which we can compute moments and function of interest. For example, a point estimate of the forecast is the ergodic average $\hat{Y}^{t+\tau} = \bar{L}^{-1}[\sum_{\ell=L+1}^{L+\bar{L}}(Y^{\ell t+\tau})^{-1}]^{-1}$ or the median of the distribution; its numerical variance can be estimated using $var\left(\hat{Y}^{t+\tau}\right) = \bar{L}^{-1}\left[\mathcal{Q}_o + \sum_{s=1}^{r}\left(1-\frac{s}{r+1}\right)\left(\mathcal{Q}_s+\mathcal{Q}_s'\right)\right]$ where $\mathcal{Q}_s = \bar{L}^{-1}\left[\sum_{\ell=s+1+L}^{L+\bar{L}}(\left[Y_{\ell t+\tau} - \hat{Y}_{t+\tau}\right]\left[Y_{\ell t+\tau} - \hat{Y}_{t+\tau}\right]')^{-1}\right]^{-1}$ and interdecile ranges can be obtained ordering the draws for each $t+\tau$. Turning point distributions can be obtained by appropriately defining the function $\mathcal{G}$. Impulse response profiles can also be computed using these forecasts. We describe their calculation in some details in section 6.

As mentioned the prior distributions we use are informative. Uninformative structures can be obtained setting $\zeta \to 0$, $Q_f^{-1} \to 0, \Sigma_\mu^{-1} \to 0$. It is easy to see that the form of the conditional posterior distributions is unchanged if non-informative priors are preferred. However, posterior moments now reflect only sample information.

4 Measurement Error

An issue of crucial importance in examining cross-sections of time series is the one of measurement error. In macro panels measurement error may emerge because of the uneven quality of data across units or because of different definitions of the same quantity in different units. For example, since the establishment of the European Central Bank, the harmonized CPI has substituted national CPI measures to reduce cross country biases in the measurement of price indices. Measurement error can be easily allowed for in our framework. Let y_{it} and W_t be unobservable and let instead $y_{it}^{+} = y_{it} + u_{it}^{y}$ and $W_t^{+} = W_t + u_t^{w}$ be available, where

u_{it}^j, $j = y, w$ are serially uncorrelated and uncorrelated with y_{it} and W_t. Then (3) implies

$$\tilde{E}_t = E_t + U_t^y - \mathbf{U}_t \delta_t \tag{13}$$

where U_t^y is the stacked vector of measurement errors in y_{it}, $\mathbf{U}_t = I_{NG} \otimes U_t'$ and $U_t = (u_{t-1}^{y'}, \ u_{t-2}^{y'}, \ldots, u_{t-p}^{y'}, \ u_t^{w'}, \ldots, u_{t-q}^{w'})'$. The presence of serially uncorrelated measurement error makes the model a panel VARMA. Hence, if measurement error is deemed important, one has two alternatives: (i) specify a long lag length for the VAR so that at least the dominant elements of the MA components are accounted for; (ii) impose a particular MA structure on the error of (3). Here we discuss this second strategy.

Let $\tilde{E}_t = \varphi \kappa_t$ where φ is a $r \times 1$ vector, r is the length of the MA components and $\kappa_t \sim N(0, I)$. Then, as in Chib and Greenberg (1995), define $y_t^* = y_t - \sum_{i=1}^r \varphi_i y_{t-i}^*$, $x_t^* = x_t - \sum_{i=1}^r \varphi_i x_{t-i}^*$ with $y_s^* = x_s^* = 0$ if $s < 0$ and $\upsilon_{tj} = -\sum_{i=1}^r \theta_i \upsilon_{t-i,j} + \theta_{t+j-1}$ where $\upsilon_{sj} = 0$ if $s < 0$. With this transformation the model is:

$$Y_t^* = X_t^* \delta_t + \sum_{i=0}^{r-1} \upsilon_{ti} \kappa_{-i} + \kappa_t \tag{14}$$

or in matrix form

$$Y^* = X^* \delta + \Upsilon \varpi + \kappa \tag{15}$$

where $\varpi = (\kappa_0, \kappa_{-1}, \ldots \kappa_{-r+1})'$. The addition of ϖ and φ to the set of conditioning variables leaves the conditional posterior of ψ unchanged. The posterior distribution of ϖ conditional on $(\theta_t, \varphi, y^*, x^*)$ can be found by rewriting (15) as $\bar{y} = Y^* - X^* \delta = \Upsilon \varpi + \kappa$. Finally, with normal prior $\varphi \sim N(\bar{\varphi}, \mathcal{R}^{-1})$, the conditional posterior kernel is $\varsigma(\theta) \prod_{t=1}^T \exp\{-0.5 \kappa_t^2\} \times \exp\{-0.5\,(\varphi - \bar{\varphi})' \ \mathcal{R}\,(\varphi - \bar{\varphi})\}$ where $\varsigma(\theta)$is the density of the first r observations. Sampling from this posterior requires a MH step within the Gibbs algorithm, but not further complications. As a candidate density for φ one could take $\exp\{-0.5\left(\varphi - \varphi^\dagger\right)' \mathcal{R}^\dagger \left(\varphi - \varphi^\dagger\right)\}$ where $\varphi^\dagger$ and $\mathcal{R}^\dagger$are nonlinear least square estimates of φ and $\mathcal{R}$.

5 An interpretation

It turns out that (3) endowed with the factor structure (6) is equivalent to an observable index model. In fact substituting (6) into (3) one obtains

$$Y_t = \mathcal{W}_t \lambda_t + \mathcal{A}_t \alpha_t + \sum_{f=2}^{F} \mathcal{Z}_{f,t} \rho_{f,t} + v_t \tag{16}$$

where $\mathcal{W}_t = X_t \Xi_1, \mathcal{A}_t = X_t \Xi_2$ and $\mathcal{Z}_{f,t} = X_t \Xi_f$.

In (16) the $NG \times 1$ vector of endogenous variables depends on a vector of common time indices $\mathcal{W}_t$, on a vector of unit specific indices $\mathcal{A}_t$, and on a set of $\mathcal{Z}_{f,t}$ vectors indexed by variables, lags, unit, etc. These indices are particular combinations of right hand side variables of the panel VAR, while $\lambda_t, \alpha_t, \rho_{f,t}$ play the role of loadings and measure the impact that different linear combinations of the regressors have on the current endogenous variables.

As (16) makes it clear, if the loadings $\lambda_t, \alpha_t, \rho_{f,t}$ were constant, estimates could be obtained by regressing Y_t on appropriate averages. Roughly speaking this is what the index models of Forni and Reichlin (1998), Stock and Watson (1998) and Forni et al. (2000) do. But while Forni and Reichlin use cross sectional averages, Stock and Watson and Forni et al. use principal components, we average over the right hand side variables of the VAR. Therefore, our indices are observable, they can be recursively constructed as new data becomes available and involve only information coming from predetermined and exogenous variables.

The machinery developed in section 3 shows how regressions on these averages can be computed when coefficients are time varying and cross-unit interdependencies are allowed for. Intuitively, to permit time variations, our setup forces estimated loadings to obey particular restrictions - e.g. that the responses to changes in the unit specific index are the same for every variable in the unit. Since time variations are difficult to account for in index models, our approach provides an intuitive way to incorporate them in such frameworks.

One advantage of the factorization (16) is that the over-parametrization of the original panel VAR is dramatically reduced. This has two implications. First, estimation and the specification searches described below are constrained only by the dimensionality of θ_t not by the one of δ_t. That it is to say, our approach is feasible even when NGk is large or the number of degrees of freedom in the original panel VAR is small. In the application of section 7, the model has 146 coefficient in each of the 28 equations. However, since θ_t is a 13×1 vector estimation is feasible and fast on available PCs. Second, since a parsimonious structure is adopted noise is averaged out and reliable estimates (and forecasts) can be obtained even in large scale models.

A second advantage of our reparametrization is that it provides a method to automatically construct multi-unit leading indicators of economic activity. In fact, a leading indicator for Y_t based on the common information is $CLI_t = \mathcal{W}_t\lambda_t$; a vector of leading indicators based on the common and unit specific information is $CULI_t = \mathcal{W}_t\lambda_t + \mathcal{A}_t\alpha_t$; a vector of indicators based on the common and variable specific information is $CVLI_t = \mathcal{W}_t\lambda_t + \mathcal{Z}_{1t}\rho_t$; finally, a vector of leading indicators based on the common, unit specific and variable specific information can be constructed as $CUVLI_t = \mathcal{W}_t\lambda_t + \mathcal{A}_t\alpha_t + \mathcal{Z}_{1t}\rho_t$.

Five additional points needs to be emphasized. First, single-step and multi-steps leading indicators can be easily obtained from (16). For example, one can construct medium term measures of core inflation, potential output and the natural rate of unemployment using multi-unit information available at each t. Second, since posterior estimates of the loadings are obtained with the Kalman filter, the timing of the relationship among variables is maintained. Hence, these leading indicators can be used to conduct a number of real time experiments. Third, there is no need to preliminary categorize variables in leading, coincident and lagging: for example, all variables in the VAR enter the construction of the leading indicator based on the common information. By appropriately averaging, we therefore considerably robustify the selection of leading indicators (much in the same spirit as Granger's (2001) robust predictors). Fourth, contrary to index setups, our approach works even when series are non-stationary and the posterior distributions for leading indicators we derive are meaningfully even when both N and T are small. Finally, with uninformative priors, our leading indicators are identical to those produced in a frequentist framework.

Although we have setup the problem so that the factors in (4) are chosen a-priori by the investigator, one may be interested in statistically determining the number of indices necessary to capture heterogeneities in the coefficients across time, units and variables. It is easy to design a out-of-sample predictive diagnostic to discriminate across models. For this purpose consider the predictive Bayes factors,

$$\mathcal{B} \equiv \frac{\mathcal{L}(Y^{t+\tau}|M_h)}{\mathcal{L}(Y^{t+\tau}|M_{h+1})} \tag{17}$$

where

$$\mathcal{L}(Y^{t+\tau}|M_h) = \int \mathcal{F}(Y^{t+\tau}|\psi_h, M_h)p(\psi_h|M_h)d\psi_h \tag{18}$$

is the predictive density for $Y^{t+\tau}$ of model with h indices (M_h). Here $p(\psi_h|M_h)$ is the prior density for ψ in model h and $\mathcal{F}(Y^{t+\tau}|\psi_h, M_h)$ the density of future data under the parameterization given by M_h. Since predictive densities can be decomposed into the product of one-step ahead prediction errors, model h can be evaluated against model $h+1$ using its out-of-sample prediction record. When the two specifications are nested, that is $\psi^* = (\psi_1, \psi_2)$ and $\psi_2 = \bar{\psi}_2$ is the restriction of interest, if $\mathcal{L}(\psi_1|M_h) = \int \mathcal{L}(\psi_1, \psi_2|M_{h+1})d\psi_2$ and ψ_1 and ψ_2 are independent, then (17) reduces to $\mathcal{B} = \frac{\mathcal{L}(\psi_1|M_{h+1})}{\mathcal{L}(\psi_1|Y_t,\ M_{h+1})}$ (see Kass and Raftery (1995)) which requires only estimates of the model with $h+1$ indices.

The predictive density of model h can be easily computed using the output of the Gibbs sampling. To do so, draw θ_{ht}^{ℓ} from the updating distribution (12), construct forecast $Y_{t+\tau}^{\ell}$ for each horizon τ, compute the prediction errors at each step and for each draw and average across draws. The numerator and the denominator of (17) can be computed using $\frac{1}{\bar{L}}[\sum_{\ell=L+1}^{\bar{L}+L} \mathcal{L}(Y^{t+\tau}|\psi_j^{\ell})^{-1}]^{-1}$ where ψ^{ℓ} is the ℓ-th draws for the parameters of model $j = h, h+1$. As $\bar{L} \to \infty$, $\frac{1}{\bar{L}}[\sum_{\ell=L+1}^{\bar{L}+L} \mathcal{L}(Y^{t+\tau}|\psi_h^{\ell})^{-1}]^{-1} \to \mathcal{L}(Y^{t+\tau}|M_h)$.

Various other specification searches can be easily conducted. For example, it is possible to check whether the factorization in (6) is exact or not. As seen above, the conditional posterior distribution of σ_t is of inverted gamma type and $E(\sigma_t^{-1}) = 1$ if $\sigma^2 = 0$. Therefore, if posterior draws are centered around one, there is evidence in favour of $\sigma^2 = 0$.

One way of formally evaluating the closeness of σ^2 to zero is to construct the ratio $\mathcal{S} = \frac{P(\sigma^2 \leq \epsilon|y)P(\sigma^2 > \epsilon|y)}{P(\sigma^2 \leq \epsilon)P(\sigma^2 > \epsilon)}$ where the numerator is computed using posterior draws and the denominator using prior draws. A similar approach can be used also to examine the posterior support, e.g., for time variations in θ_t or the importances of interdependencies in the model. For example, in the former case and setting, e.g., $\bar{B}_f = b_f * I$, time variations are significant if the posterior draws of b_f are large relative to prior draws.

Instead of sequentially examining a series of hypotheses regarding the structure of the model, one may want to take a general view about the uncertainty surrounding the number of indices to be included in (6). In this case, let M_1 be the model with one index and M_h

the model with h indices, $h = 2, \dots H$, and suppose we run a sequences of tests of model h against model 1. Let $\mathcal{B}_{h1}$ be the corresponding Bayes factor. Then the posterior probability for 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. Using such an expression in (18), it is immediate to recognize that model uncertainty is accounted for by weighting the posterior density of $y^{t+\tau}$ by the posterior probability of the model.

6 Dynamic analysis

Impulse response profiles can be computed as posterior revisions of forecasts. Since the model is only conditionally linear, the impulse responses we compute differ from those obtained in standard VARs. In fact, forecasts for $y_{it+\tau}$ may change for two reasons: because of the model's error are different from zero and because the coefficients vary over time. Since coefficients are time varying, impulse responses depend on the history and the point in time in which these revisions are computed (as in Gallant, Rossi and Tauchen (1993) or Koop, Pesaran, Potter (1995)).

We briefly illustrate how these revisions can be computed using the output of the Gibbs sampler. Rewrite the model as:

$$Y_t = X_t \delta_t + \Gamma \tilde{e}_t \quad (19)$$

$$\delta_t = \Xi((I - \mathcal{C})(\mathcal{P}\mu + \epsilon) + \mathcal{C}\theta_{t-1} + \eta_t) + u_t \quad (20)$$

where $\Gamma\Gamma' = \Omega$, $\tilde{e}_t \sim (0, I)$. The companion form version of (19) is

$$\mathbf{Y}_t = \Delta_t \mathbf{Y}_{t-1} + \iota_t \quad (21)$$

where $\delta_t = vec(\Delta_{1t})$ and Δ_{1t} is the first row of Δ_t.

Iterating τ times on (21), using a matrix $J = [I, 0, .., 0]$ such that $J\mathbf{Y}_t = Y_t$, $J'J = I$ and $J\iota_t = \Gamma \tilde{e}_t$, we have

$$Y_{t+\tau} = J(\prod_{s=0}^{\tau-1} \Delta_{t+\tau-s})\mathbf{Y}_t + \sum_{m=0}^{\tau-1} \Phi_{m,t+\tau}\, \tilde{e}_{t+\tau-m} \quad (22)$$

where $\Phi_{m,t+\tau} = J(\prod_{s=0}^{m-1} \Delta_{t+\tau-s})J'\Gamma$ and $\Phi_{0,t+\tau} = I$. Iterating on (20) we have

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

Define impulse responses at step j, given information at t and terminal horizon τ as $IR_{j,\tau} = E_{t+j}Y_{t+\tau} - E_t Y_{t+\tau}$, $\forall \tau \geq j+1$. Since $E_t Y_{t+\tau} = J E_t \left(\prod_{s=0}^{\tau-1} \Delta_{t+\tau-s}\right) \mathbf{Y}_t$, we have

$$\begin{aligned} IR_{j,\tau} &= \sum_{s=0}^{j-1} (E_{t+j}\Phi_{\tau-j+s,t+\tau})\, \tilde{e}_{t+j-s} \\ &\quad + J\left[E_{t+j}\left(\prod_{s=0}^{\tau-j-1} \Delta_{t+\tau-s}\right) \prod_{s=\tau-j}^{\tau-1} \Delta_{t+\tau-s} - E_t \left(\prod_{s=0}^{\tau-1} \Delta_{t+\tau-s}\right)\right] \mathbf{Y}_t \quad (23) \end{aligned}$$

From (23) it is clear that revisions of the forecast at $t+\tau$ can occur because $\tilde{e}_{t+\tau-s} \neq 0$ or because shocks at some stage of the the hierarchy $(u_{t+\tau-s}, \eta_{t+\tau-s}, \epsilon)$ make $\delta_{t+\tau-s}$ change. Equation (23) also indicates that impulse responses depend on the point where they are generated and on the initial conditions. To operatively see this note, for instance, that

$$\begin{aligned} IR_{1,2} &= E_{t+1}Y_{t+2} - E_t Y_{t+2} \\ &= E_{t+1}\left(\Phi_{1,t+2}\right)\tilde{e}_{t+1} + J\left[E_{t+1}\left(\Delta_{t+2}\right)\Delta_{t+1} - E_t\left(\Delta_{t+2}\Delta_{t+1}\right)\right]\mathbf{Y}_t \end{aligned}$$

where $\Phi_{1,t+2} = J\Delta_{t+2}J'\Gamma$, and that

$$\begin{aligned} IR_{2,3} &= E_{t+2}Y_{t+3} - E_t Y_{t+3} \\ &= \sum_{s=0}^{1}\left(E_{t+2}\Phi_{1+s,t+3}\right)\tilde{e}_{t+2-s} \\ &\quad + J\left[E_{t+2}\left(\Delta_{t+3}\right)\Delta_{t+2}\Delta_{t+1} - E_t\left(\Delta_{t+3}\Delta_{t+2}\Delta_{t+2}\right)\right]\mathbf{Y}_t \end{aligned} \tag{24}$$

where $\sum_{s=0}^{1}\left(E_{t+2}\Phi_{1+s,t+3}\right)e_{t+2-s} = JE_{t+2}\left(A_{t+3}\right)J'\Gamma e_{t+2} + JE_{t+2}\left(\Delta_{t+3}\right)\Delta_{t+2}J'\Gamma e_{t+1}$.

Hence, forecast revisions in Y_{t+3} due to structural innovations are

$$JE_{t+2}\left(\Delta_{t+3}\right)J'\Gamma\tilde{e}_{t+2} + JE_{t+2}\left(\Delta_{t+3}\right)\Delta_{t+2}J'\Gamma\tilde{e}_{t+1}$$

while movements due to innovations in the coefficients are

$$J\left[E_{t+2}\left(\Delta_{t+3}\right)\Delta_{t+2}\Delta_{t+1} - E_t\left(\Delta_{t+3}\Delta_{t+2}\Delta_{t+1}\right)\right]\mathbf{Y}_t.$$

To further illustrate the point consider a scalar time varying AR(1) model

$$y_t = a_t y_{t-1} + e_t \tag{25}$$

$$a_t = (I-C)\bar{a} + Ca_{t-1} + \epsilon_t \tag{26}$$

Then

$$a_{t+\tau} = (C)^{\tau+1}a_{t-1} + \sum_{s=1}^{\tau}(C^s)(I-C)\bar{a} + \sum_{s=1}^{\tau}(C^s)(\epsilon_{t+\tau-s}) \tag{27}$$

$$y_{t+\tau} = (\prod_{s=0}^{\tau} a_{t-\tau-s})y_{t-1} + \sum_{m=1}^{\tau}(\prod_{s=0}^{m-1} a_{t+\tau-s})e_{t+\tau-m} + e_{t+j} \tag{28}$$

Note that if only $e_{t+1} \neq 0$ the summation term disappears from (28). On the other hand if only ϵ_t is different from zero the summation term in (27) disappears. Forecast revisions of $y_{t+\tau}$ in (28) occur because a shock $\bar{e}$ has realized - this will feed into $y_{t+\tau}$ via changes in a_t - or because a shock $\bar{\epsilon}$ has realized - this will feed into $y_{t+\tau}$ via changes in y_t.

The output of the Gibbs sampling can be used to compute the expressions in (23). For example, consider one period revisions (one-step ahead impulse responses) constructed at t. To construct $IR_{1,2}$ we need the following three steps:

1. Given Γ, draw $\tilde{e}_{t+1}$ and $\Delta_{t+1}, \Delta_{t+2}$ from the posterior distribution $L+1$ times.

2. For each draw $\ell = 2, \ldots, L+1$ compute $d_t^\ell = \Delta_{t+2}^\ell \Delta_{t+1}^\ell$ and the quantities $\hat{d}_{1,t} = \frac{1}{L}\sum_{\ell=2}^{L+1} d_t^\ell$, and $\hat{d}_{2,t} = \frac{1}{L}\sum_{\ell=2}^{L+1} \Delta_{t+2}^\ell$.

3. Given Y_t, the draws for $\tilde{e}_{t+1}$ $\Delta_{t+1}, \Delta_{t+2}$ from step [1], and $\hat{d}_{1,t}$ and $\hat{d}_{2,t}$ from step [2] compute $IR_{1,2}$

Assuming that $\tilde{e}_t \neq 0$ and that all future values of the shocks are integrated out, we can generalize the above algorithm to any j-period revision as follows. Given Γ,

1. Draw $(\tilde{e}_t, \tilde{e}_{t+1}, \tilde{e}_{t+2}, \ \ldots, \tilde{e}_{t+j})$ and $(\Delta_{t+1}, \Delta_{t+2}, \ldots, \Delta_{t+j})$ from the posterior distribution $L+1$ times

2. For each $\ell = 2, \ldots L+1$ compute $d_{t,j}^\ell = (\prod_{s=0}^{j} \Delta_{t+\tau-s}^\ell)$. Calculate $\hat{d}_t = \frac{1}{L}\sum_{\ell=2}^{L+1} d_{t,j}^\ell$.

3. For each draw compute $\hat{e}_{t+\tau} = \sum_{\ell=2}^{L+1} (e_{t+\tau})^\ell$.

4. Given Y_t, the draws $(\tilde{e}_{t+j}, \ \Delta_{t+j})$ from step [1], $\hat{d}_{t,j}$ from step [2] and $\hat{e}_{t+\tau}$ from step [3] compute $IR_{j,\tau}$

Note that using the output of the Gibbs sampler drastically simplifies the calculation of the impulse response profiles relative e.g. Gallant, Rossi and Tauchen (1993). Also the state space nature of the model allows us to characterize what move conditional expectations, while this is impossible in general nonlinear impulse response analyses.

7 Leading Indicators of Euro inflation and output growth

There are many interesting problems to which apply we could apply the framework of analysis we have described in this paper. Here we discuss how to construct leading indicators of economic activity and inflation for the Euro area using information coming from the cross section of G-7 countries. The last twenty years have witnessed an increased globalization of world economies. Given the current high level of integration of G-7 economies, 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 future developments in the continent. Furthermore, the evolutionary nature of the relationship suggests that a time varying specification will be probably most useful in modelling cross country interdependencies. Given that Italy, France and Germany account for about 70% of total activity in the Euro area, and since several European countries have closely related cycles, we approximate area wide aggregates using real and nominal variables of Italy, France and Germany.

For each of the G-7 countries we use 4 endogenous variables (real GDP growth, CPI inflation, employment growth, and rent inflation); three predetermined ones (a commodity price index, the median stock return and the trade weighted US real exchange rate) and a constant. Besides GDP growth and CPI inflation, the other two endogenous variables have been selected because they have the largest in-sample predictive power across countries. While this preliminary step is unnecessary from the point of view of the model specification, we find it is useful to reduce the noise in the estimates caused by the near-collinear informational content of certain variables. Five lags of the endogenous variables and two lags of the predetermined variables are used. Therefore, each equation has k=7*4*5+2*3+1=146 coefficients. The sample covers the period 1980:1-2000:4. We calculate leading indicators four and eight quarters ahead, as these are the most interesting horizons for policymakers, directly from the model, i.e. we set up (1) with $D_{it}(L)$ and $C_{it}(L)$ different from zero either for $L \geq 4$ or $L \geq 8$. By doing so we avoid to have a separate model to forecast future values of the exogenous variables.

We produce 24,000 iterations of the MCMC algorithm starting from arbitrary initial conditions. Runs of 40 elements were drawn 600 times and the last observation of the final 500 runs was used for inference. We checked convergence by calculating the mean of ψ for 200, 300, 400, 500 observations. We found that convergence was achieved using 200 to 300 observations. Convergence was confirmed also when splitting the sample in two parts and testing (in a classical sense) for differences in the means across subsamples.

We conducted the analysis with a partially non-informative specification ($\Psi = 0$, $\mathcal{P} = I$, $\zeta = 0$, $B_t = b_i * I, i = 1,2,3$ and $p(b_i)$ is proportional to a constant. $\bar{\theta}$ is initialized with a sequential OLS on the sample 1975-1980 on the time invariant version of (16) and σ^2 is calculated averaging the estimated variances of NG AR(p) models. The vector δ_t is decomposed into 3 factors: a 2×1 vector of common factors (λ_t) - one for the Euro area and one for the rest of the world; a 7×1 vector of country specific factors (α_t); and a 4×1 vector of variable specific factors (ρ_t). Hence, $\theta_t = (\lambda_t', \alpha_t', \rho_t')'$ is 13×1 vector.

Using the 500 draws for θ_t we examined posterior support for a number of hypotheses. First, we checked whether a model with three indices can be reduced or not. We find that the predictive Bayes factor for the 1996:1-2000:4 period always prefers a model with three indices to a model with any combination of two of the three indices. The least favorable specification obtains in comparison with a model which excludes country specific factors (Bayes factor is 1.08). In all other cases the Bayes factor exceeds 1.25.

Second, we examined the support for the exactness of the three factor decomposition. That is, we examine if the posterior for σ^2 is concentrated around 0. Since when $\sigma^2 = 0$ the prior for σ_t^{-1} is centered around 1 and since figure 1 suggests that the posterior time series for σ_t^{-1} is on average around 0.8, the posterior and the prior distributions are concentrated around different values. Hence, the posterior support for $\sigma^2 = 0$ is small.

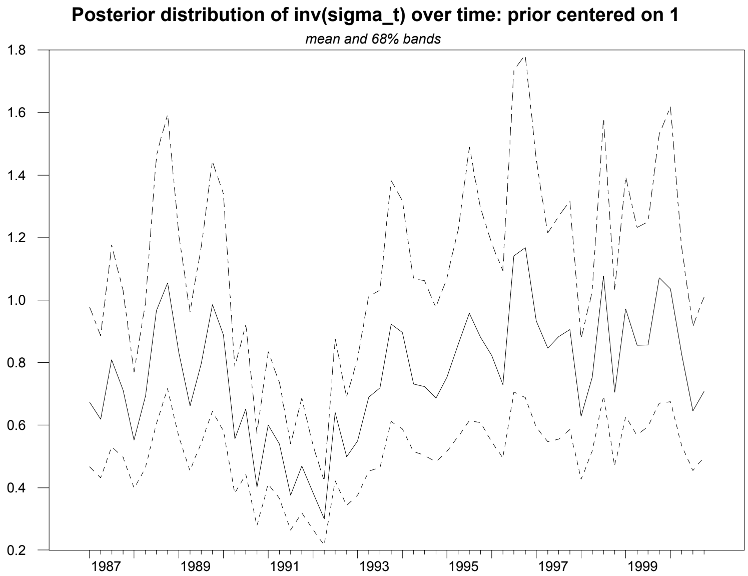


Figure 1

Third, we checked whether interdependencies are an important ingredient to characterize the dynamics of the data. Comparing a model without interdependencies (7 seemingly unrelated VARs) and a model with interdependencies for the period 1996:1-2000:4 we obtain a Bayes factor of 1.14, suggesting that interdependencies play some role in the data. We have also compared the MSE of the forecasts of the two models. The out-of-sample performance of a specification with interdependencies appears to be superior to the one of a model without interdependencies: the relative four (eight) steps ahead MSE for the sample 1996:1-2000:4 of a model with interdependencies is 0.90 (0.82) on average across variables. Diebold and Mariano (1995) test for predictive accuracy also reject the hypothesis that the two specifications are equivalent from a forecasting point of view.

Fourth, we examined whether time variations in the coefficients are important. Using the approach described in section 5, we find that for $\epsilon = 0.008$ the statistic $\mathcal{S}$ for the three indices is 1.085, $>>$ 20, $>>$ 20 respectively. Hence, time variation appears to be significant only in the vector of common factors. To assess the economic importance of these time variations, figure 2 plots the profile response of EU output growth and inflation to one standard deviation change in the non-EU part of λ_t. This picture is constructed using $t = 2000:4$, $\tau = 8$, $j = 0, \ldots, 7$. The initial impact appears to be large but there is little persistence in the responses. Note that the bands shrink with the horizon because the difference between the two expressions appearing in the bracket term in (24) is getting smaller as j increases. Interestingly, inflation and output growth react in the opposite direction over the adjustment path as it would be typical if a supply shock was perturbing the economy.

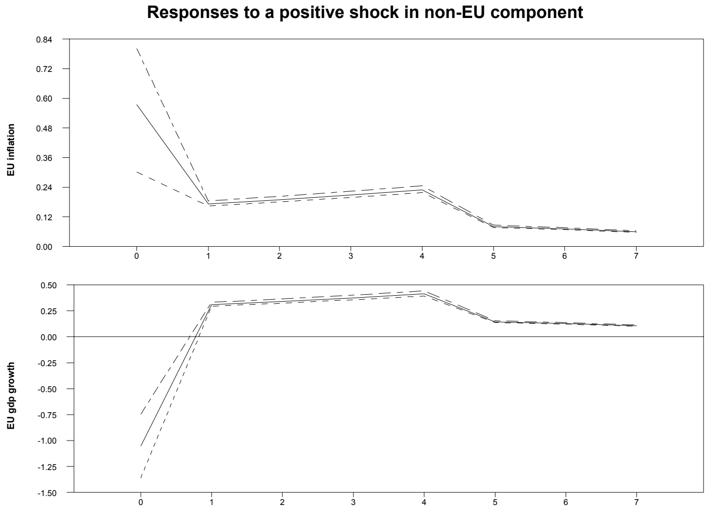


Figure 2

Fifth, we examined how important are cross sectional differences. We have already seen that the country specific factor appears to be the least important of the three factors, at least according to predictive Bayes analysis. In figure 3 we plot the time series for the posterior mean and the 68% band for α_t (constructed using information one year ahead).

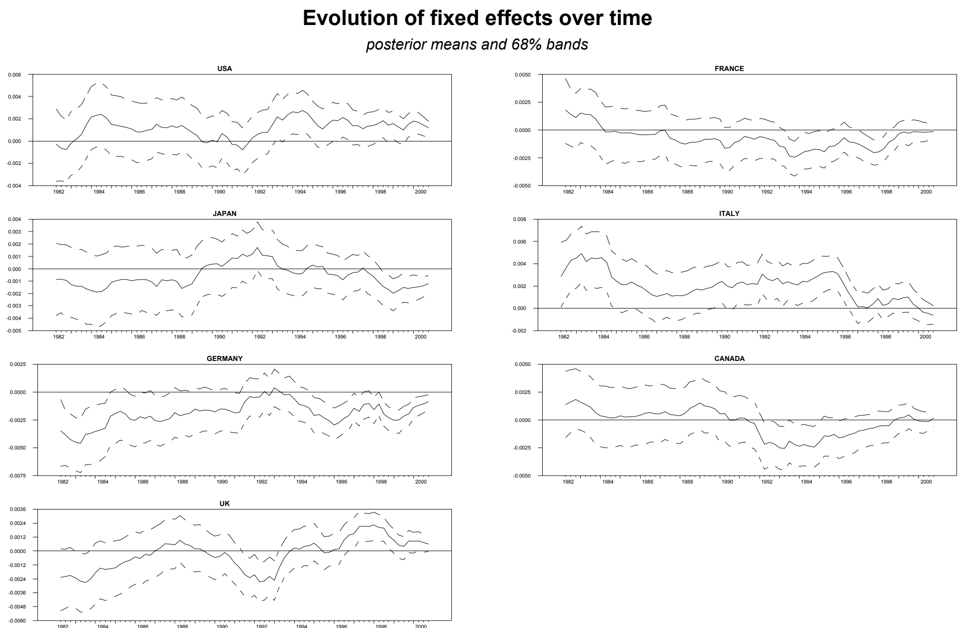


Figure 3

Visual inspection indicates that there is a modest amount of time variation, that the seven components are small in size and not different from zero at most of the dates. Concentrating on the last five years of the sample, we see that Germany and Japan's α's deviate (negatively) from the time path determined by the world factors, while Italy and again Germany display significant country specific variations in the dynamics in the early part of the 1990's.

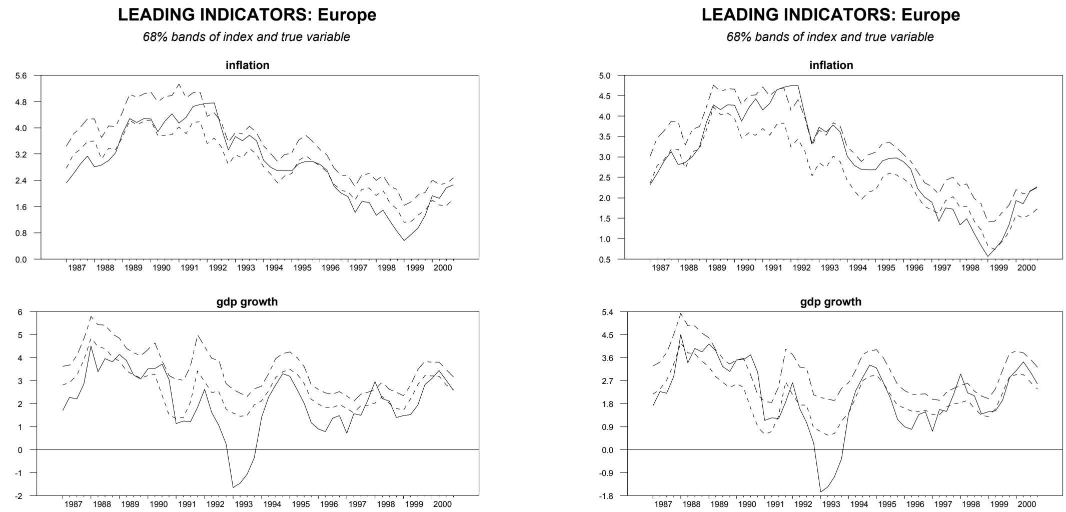
To conclude our specification searches, we examined whether the predictive ability of the model depends on the prior specification and compared it with the one of three other forecasting specifications: a univariate AR(4), a BVAR(4) with a simple Minnesota prior and an observable factor model like (16) with constant loadings and no prior (Benchmark). Table 1 reports the Theil-U statistics for GDP growth and inflation in the Euro area, four and eight steps ahead, the total number of turning points correctly recognized for seven countries and the total number of existing turning points. Turning points are identified using a standard two-quarters rule. These statistics are constructed in real time and recursively over the 1996:1-2000:4 period. The hyperparameters of the informative prior are estimated on the sample 1975-1980 with a rough grid search.

Table 1: Forecasting statistics

	Step	Inflation	GDP growth	Downturns Recorded/Actual	Upturns Recorded/Actual
Non-informative	4	0.59	0.47	35/57	22/36
	8	0.39	0.48	27/39	19/30
Informative	4	0.45	0.97	36/57	18/36
	8	0.35	0.90	26/39	14/30
Benchmark	4	1.34	0.81	31/57	21/36
	8	1.02	0.85	25/39	18/30
AR(4)	4	0.94	1.07	28/57	23/36
	8	0.79	1.30	21/39	20/30
BVAR(4)	4	0.93	0.89	32/57	23/36
	8	1.09	0.92	24/39	19/30

Several features of the table deserve comments. First, our basic specification is superior in MSE sense to all competitors in forecasting GDP growth and comparable to a specification with informative priors for inflation two years ahead. Gains are large and exceed 50% in almost all cases. Second, turning point recognition is also superior but the improvements over other specifications are more limited. Third, it matters which prior specification one uses: for inflation an informative prior produces smaller MSE while for GDP growth and upturns a non-informative one is preferrable. These results are robust, for example, to the substitution of the MAD statistic to the MSE or to alternative turning point dating rules. Apparently, the information contained in the 1975-1980 sample is important to understand the developments of inflation in the latest part of the 1990's but not those of GDP growth.

Using a model with three indices, time variation in the vector of common factors and a partially uninformative prior we constructed leading indicators for the two variables of interest recursively in real time (i.e. draws for θ_t are from posterior estimates consistent with the information available up to t). Figures 4 plots posterior 68% bands for the two leading indicators, constructed using information available one and two years ahead.



4 quarters ahead 8 quarters ahead

Figure 4

Both indicators appear to be remarkably good in tracking the ups and downs of inflation over the sample. For the period between 1992 and 1996 the actual value of inflation is at the lower edge of the posterior 68% bands suggesting a slight overstatement of expected inflation, but also in this case, the direction of changes and the shape of the resulting dynamics are appropriate. The one year ahead indicator of GDP growth is also remarkably good in capturing the ups and downs of the variable over the sample. In fact, using a simple two quarters decline/increase rule (and one quarter tolerance on each side) we find that our indicator misses only one turning point for the 15 years sample. In levels the one year ahead indicator is reasonably good except for the period 1992-93, a sharp and strongly recessionary period in Europe. For example, the probability that our indicator is equal to -1.8 in 1992:2 (the actual GDP growth for that quarter) is less than 1.0% even though the probability that a recession is recorder at that time is 54%. For the two years ahead indicator, the actual value appears to be often around the lower hedge of the 68% posterior band, but the probability that GDP growth falls by 1.8 in 1992:2 is now around 10%. The figure also shows that, thanks to time variations in the parameters, the model is able to quickly adjust when mistakes are made without the need of any exogenous correction. Furthermore, by exploiting time variations and cross sectional information, the model captures changing local trends in GDP growth and inflation which are common across countries. Therefore, it produces leading indicators which are stable and reliable over time, contrary to many existing specifications, which track either variables only over short stretches of time.

The output of our model can be used to construct a variety of measures which are of interest for policymakers. In figure 5 we present the time series for the posterior 68% band for a recursive measure of potential world output growth, constructed as $CVLI_t^{GDP} = \mathcal{W}_t\lambda_t + \mathcal{Z}_{1t}\rho_t$ using information two years ahead. Two features are worth emphasizing. First, the cyclical movements of potential output roughly correspond to those of actual output. Second, there is a small trend increase in the level of potential output growth in the last 6-7 years of the sample. The increase is not extraordinary (3.07% as compared to 2.36% of the previous 10 years) but is significant. Note also that our measure of potential output begins to decline already at the beginning of 1999.

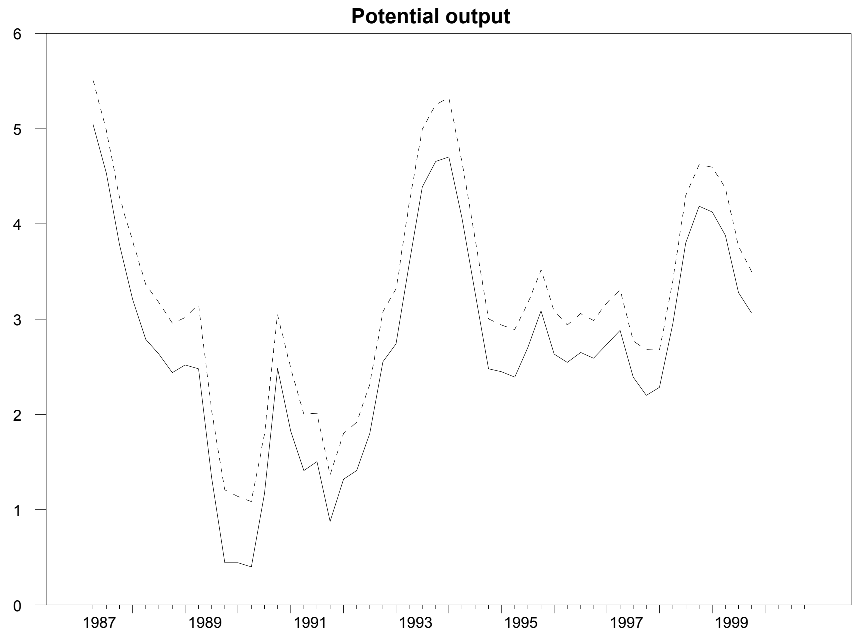


Figure 5

To conclude, one should mention that the computation time needed to obtain posterior estimates for the 28 variable Panel-VAR model used in this section is relatively short. One complete run (drawing 24,000 sequences from the posterior, calculating the predictive density, computing impulse responses and constructing the distribution of leading indicators and potential output growth) took about 45 minutes of CPU time on a Dell Inspiron 4000 with a Pentium IV processor and 256Mb of RAM memory.

8 Conclusions

This paper develops a methodology for conducting inference in time varying coefficient Panel VAR models with cross unit interdependencies. We take a Bayesian viewpoint 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 across units, time variations and heteroskedasticity in the innovations.

The factor structure on the coefficients allow us to transform a potentially overparametrized panel VAR into a parsimonious SUR model where the regressors are a set of observable indices, constructed using certain linear combinations of the right-hand-side variables, and the loadings are the time varying factors. We derive posterior estimates of the vector of loadings using Markov Chain Monte Carlo methods. Posterior distributions for interesting function of these loadings can be obtained as a by-product of the Monte Carlo routine. We show how to compute forecast revisions in response to unexpected perturbations in either the innovations of the panel VAR or in the loadings of one of the observable indices.

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 for each equation into the problem of estimating a small number of loadings on certain combination of the right hand side variables of the VAR. The computational advantage inherent in our setup is large: we are able to estimate large multicountry and multisector models with interdependencies and time variations in less than an hour of computer time. Second, since indices are constructed and estimated recursively in real time, they can be employed for a variety of policy and forecasting purposes. In section 7 we have shown how to construct leading indicators of economic activity in the Euro area and a measure of potential world output using the information available at each t. Third, since indices are observable and predetermined with respect to the endogenous variables, it is easy to design a predictive measure to select the number of indices to be used. We propose a simple statistic, based on predictive Bayes factors, which can be used for this purpose. We also suggest how to deal with general forms of model uncertainty using a simple variant of Leamer's measure of posterior uncertainty.

As mentioned, there is a number of applications to which the tools developed in this paper can be applied. For example, Canova, Ciccarelli and Ortega (2003) use a specification similar to the one of section 7 to extract world and national business cycles. 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, or the construction of portfolios of assets in different geographical regions can all be fruitfully studied within the general framework suggested in this paper.

Appendix

This appendix reports the expressions of the parameters of the posterior distributions derived in section 3. They are:

$$\hat{z}_0 = z_0 + 1; \;\; \hat{z}_1 = z_1 + T; \;\; \hat{z}_{2f} = T * \dim\left(\theta_t^f\right) + z_{2f}, \; f = 1, .., F;$$

$$\hat{Q}_o = \left[Q_o^{-1} + \left(\bar{\theta} - \mathcal{P}\mu\right)\left(\bar{\theta} - \mathcal{P}\mu\right)'\right]^{-1}$$

$$\hat{Q}_1 = \left[Q_1^{-1} + \textstyle\sum_t \left(Y_t - X_t \Xi \theta_t\right) \sigma_t^{-1} \left(Y_t - X_t \Xi \theta_t\right)'\right]^{-1}$$

$$\hat{Q}_{2f} = \left[Q_{2f}^{-1} + \textstyle\sum_t \left(\theta_t^f - \theta_{t-1}^{*f}\right)\left(\theta_t^f - \theta_{t-1}^{*f}\right)' / \xi_t\right]^{-1}$$

$$\widehat{\bar{\theta}} = \hat{\Psi}\left[\Psi^{-1}\mathcal{P}\mu + \left(I - \mathcal{C}\right)' \bar{B}^{-1} \textstyle\sum_t \tilde{\theta}_t / \xi_t\right]$$

$$\hat{\Psi} = \left[\Psi^{-1} + \left(I - \mathcal{C}\right)' \bar{B}^{-1} \left(I - \mathcal{C}\right) \textstyle\sum_t 1/\xi_t\right]^{-1}$$

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

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

$$\hat{\zeta} = \zeta + NG$$

$$\mathcal{R} = \zeta s_t + \left(Y_t - X_t \Xi \theta_t\right)' \Upsilon^{-1} \left(Y_t - X_t \Xi \theta_t\right)$$

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

References

[1] 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.

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

[3] Canova, F. and M. Ciccarelli (1999), Forecasting and Turning Point Prediction in a Bayesian Panel VAR Model, forthcoming, *Journal of Econometrics.*

[4] Canova, F., M. Ciccarelli and E. Ortega (2003), Similarities and Convergence of G-7 cycles, manuscript.

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

[6] 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.

[7] Cogley, T. and T. Sargent (2002), Drift and Volatilities: Monetary Policy and Output in Post WWII US, Stanford University, working paper.

[8] Camba Mendez, G., G. Kapetanios, R. Smith and M. Weale (2001), An automatic leading indicator of economic activity in forecasting GDP growth for European countries, *Econometric Journal* 4(1), 37-57.

[9] Forni, M. and L. Reichlin (1998), Let's get real, A Factor Analitical Approach to disaggregate Business Cycle Dynamics, *Review of Economic Studies*, 65, 453-473.

[10] 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.

[11] Diebold, F. and B. Mariano (1995), Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, 253-265.

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

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

[14] 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 University Press.

[15] Granger, C. (2001) Evaluation of forecasts in Hendry, D. and N. Ericsson (eds.) Understanding Economic Forecasts, MIT Press.

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

[17] 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.

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

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

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

[21] Marcellino, M., J. Stock and M. Watson (2003), Macroeconomic Forecasting in the Euro Area: Country Specific versus Area-wide Information, *European Economic Review*, 47, 1- 18

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

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

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

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

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