Panel Vector Autoregressive Models: A Survey*

Fabio Canova[†]
EUI and CEPR

Matteo Ciccarelli
European Central Bank

April 2013

Abstract

This chapter provides an overview of the panel VAR models used in macroeconomics and finance to study the dynamic relationships between heterogeneous assets, households, firms, sectors, and countries. We discuss what their distinctive features are, what they are used for, and how they can be derived from economic theory. We also describe how they are estimated and how shock identification is performed. We compare panel VAR models to other approaches used in the literature to estimate dynamic models involving heterogeneous units. Finally, we show how structural time variation can be dealt with.

JEL classification: C11, C30, C53.

Key words: Panel VAR, Estimation, Identification, Inference.

*Prepared for Advances in Econometrics, volume 31. We thank Lutz Kilian, Chris Sims and the participants to the 12th Annual workshop on Avances in Econometrics, Dallas for comments and suggestions. The views expressed in this paper are those of the authors and do not necessarily reflect those of the ECB or the Eurosystem.

†Corresponding author: European University Institute, Via Della Piazzuola 43, 50133 Firenze, Italy. Email: fabio.canova@eui.eu.

1 Introduction

Macroeconomic analyses and policy evaluations increasingly require taking into account the interdependencies across sectors, markets and countries. Even national economic issues, while often idiosyncratic, need now to be tackled from a global perspective. Thus, a number of additional channels of transmission need to be considered and additional spillovers accounted for when formulating policies, even for large and developed economies.

Domestic interdependencies have been known to produce domestic business cycle fluctuations from idiosyncratic sectorial shocks at least since Long and Plosser (1983). Spillovers from the financial sector to the real economy are key to understanding the recent global crisis (e.g. Stock and Watson, 2012; Ciccarelli et al., 2012a). Many authors have also argued that the rapidly rising degree of trade and financial market integration has induced closer international interdependencies within the developed world and between developing and developed world (Kose et al., 2003, Canova et al., 2007, Pesaran et al., 2007, Kose and Prasad, 2010, Canova and Ciccarelli, 2012). Thus, a multilateral perspective is crucial, and the failure to recognize this aspect of reality may induce distortions in the evaluation of economic outcomes and erroneous policy decisions.

There are two ways of examining economic issues in interdependent economies. One is to build multi-sector, multi-market, multi-country dynamic stochastic general equilibrium (DSGE) models, where agents are optimizers, and where preferences, technologies, and constraints are fully specified. Structures like these are now extensively used in the policy arena (see e.g. the SIGMA model at the Federal Reserve Board; the global projection model at the IMF; or the EAGLE model at the ECB). Tightly parameterized DSGE models are useful because they offer clear answers to policy questions and provide easy-to-understand welfare prescriptions. However, by construction, they impose a lot of restrictions that are not always consistent with the statistical properties of the data. Thus, the policy prescriptions they provide are hardwired in the assumptions of the model, and must be considered more as a benchmark than as a realistic assessment of the options and constraints faced by policymakers in real world situations.

An alternative approach to dealing with interdependent economies is to build panel VAR models. These models eschew most of the explicit micro structure present in DSGE models and, as their VAR counterparts, attempt to capture the dynamic interdependencies present in the data using a minimal set of restrictions. Shock identification can then transform these reduced form models into structural ones, allowing typical exercises, such as impulse response analyses or policy counterfactuals, to be constructed in a relatively straightforward way. Structural panel VAR models are potentially liable to standard criticisms of structural VAR models (see e.g. Cooley and Le Roy, 1983, Faust and Leeper, 1997, Cooley and Dweyer, 1998, Canova and Pina, 2005, Chari et al., 2008) and thus need to be implemented with care. The information they produce can effectively complement analyses conducted with DSGE models, help to point to the dimensions where these models fail, and provide stylized facts and predictions which can improve the realism of structural models.

The goal of this chapter is to describe what panel VAR models are and what their use is in applied work; how they can capture the heterogeneities present in interdependent economies and how the restricted specifications typically employed in the literature are nested in the general panel VAR framework we consider. We also examine how panel VAR models can be estimated, how shock identification is performed, and how one can conduct inference with such models. We highlight how the evolving nature of the cross unit interdependencies can be accounted for and how alternative frameworks such as factor models, global VAR (GVAR), collections of bilateral VARs, large scale Bayesian VARs, or spatial VARs compare to them. Finally, we discuss the open challenges that researchers face when dealing with dynamic heterogeneous and interdependent panels (of countries, industries, or markets) in applied work.

The rest of the paper is organized as follows. The next section discusses what the distinctive features of panel VAR models are and how they link to DSGE models. Section 3 describes how reduced-form panel VARs are estimated. Section 4 adds time variation in the coefficients. Section 5 deals with shock identification and describes strategies to perform structural analyses. Section 6 compares panel VARs to other approaches that have been used to deal with dynamic models involving interdependent heterogeneous units. The conclusions and some additional considerations are in section 7.

2 What are panel VARs?

VAR models are now well established in applied macroeconomics. In VAR models all variables are treated as endogenous and interdependent, both in a dynamic and in a static sense, although in some relevant cases, exogenous variables could be included (see e.g., the dummy approach pioneered by Ramey and Shapiro, 1998). Let Y_t be a $G \times 1$ vector of endogenous variables. The VAR for Y_t is

$$Y_t = A_0(t) + A(\ell)Y_{t-1} + u_t \quad u_t \sim iid(0, \Sigma_u) \tag{1}$$

where $A(\ell)$ is a polynomial in the lag operator and *iid* means identically and independently distributed. Restrictions are typically imposed on the coefficient matrices A_j to make the variance of Y_t bounded and to make sure that $A(\ell)^{-1}$ exists – for example, one can impose that no roots of $A(e^{-\omega})^{-1}$ are on or inside the unit circle. Sometimes equation (1) is decomposed into its short run and its long run components, following the work of Beveridge and Nelson (1981) or Blanchard and Quah (1989), but for our purposes this distinction is not critical since the available time series dimension will be, at best, of medium length, rendering the long run properties of the model very imprecisely pinned down. For the sake of notation $A_0(t)$ is defined to include all deterministic components of the data. Thus, it should be understood that the representation (1) may include constants, seasonal dummies and deterministic polynomial in time.

A typical variation of (1), used primarily in small open economy analyses, allows the G variables in Y_t to be linear function of W_t, a set of predetermined or exogenous variables. In the latter case, the VAR becomes

$$Y_t = A_0(t) + A(\ell)Y_{1t-1} + F(\ell)W_t + u_t.$$

Such VARX structures have been described in Ocampo and Rodriguez (2011) and used, for example, by Cushman and Zha (1997) in their analysis of the effect of monetary policy shocks in Canada, or in exercises measuring how variables determined in world markets (such as commodity prices or world productivity) affect domestic economies.

Finite order, fixed coefficient VARs like (1) can be derived in many ways. The standard one is to use the Wold theorem (see e.g. Canova, 2007) and assume linearity, time invariance and invertibility of the resulting moving average representation. Under these assumptions, there exists an (infinite lag) VAR representation for any Y_t. To truncate this infinite dimension VAR and use a VAR(p), p finite and small, in empirical analyses we need to assume that the contribution of Y_{t-j} to Y_t, is small when j is large.

Panel VARs have the same structure as VAR models, in the sense that all variables are assumed to be endogenous and interdependent, but a cross sectional dimension is added to the representation. Thus, think of Y_t as the stacked version of y_{it}, the vector of G variables for each unit $i = 1, \ldots, N$, i.e., $Y_t = (y'_{1t}, y'_{2t}, \ldots y'_{Nt})'$. The index i is generic and could indicate countries, sectors, markets or combinations of them. Then a panel VAR is

$$y_{it} = A_{0i}(t) + A_i(\ell)Y_{t-1} + u_{it} \qquad i = 1, \ldots, N \quad t = 1, \ldots, T \tag{2}$$

where u_{it} is a $G \times 1$ vector of random disturbances and, as the notation makes it clear, $A_{0i}(t)$ and A_i may depend on the cross-sectional unit.

When a panel VARX is considered, the representation is

$$y_{it} = A_{0i}(t) + A_i(\ell)Y_{1t-1} + F_i(\ell)W_t + u_{it} \tag{3}$$

where $u_t = [u_{1t}, u_{2t}, \ldots, u_{Nt}]' \sim iid(0, \Sigma)$, $F_{i,j}$ are $G \times M$ matrices for each lag $j = 1, \ldots, q$, and W_t is a $M \times 1$ vector of exogenous variables, common to all units i.

Simple inspection of (2) or (3) suggests that a panel VAR has three characteristic features. First, lags of all endogenous variables of all units enter the model for unit i: we call this feature "dynamic interdependencies". Second, u_{it} are generally correlated across i: we call this feature "static interdependencies". In addition, since the same variables are present in each unit, there are restrictions on the covariance matrix of the the shocks. Third, the intercept, the slope and the variance of the shocks u_{1it} may be unit specific: we call this feature "cross sectional heterogeneity". These features distinguish a panel VAR typically used in macroeconomic and finance from the panel VAR used in micro studies, such as Holtz Eakin et al. (1988) or the more recently Vidangos (2009), where interdependencies are typically disregarded and sectorial homogeneity (allowing for certain time-invariant individual characteristics) is typically assumed. These features also distinguish the panel VAR setup from others used in the macroeconomic literature, where either cross sectional homogeneity is assumed and/or dynamic interdependencies are a-priori excluded (see e.g. Benetrix and Lane, 2010, Beetsma and Giuliadori, 2011). In a way, a panel VAR is similar to large scale VARs where dynamic and static interdependencies are allowed for. It differs because cross sectional heterogeneity imposes a structure on the covariance matrix of the error terms. A detailed comparison with other approaches designed to handle multi-unit dynamics is in section 6.

2.1 An example

For expository purposes, it is useful to consider a simple example. Suppose that $G = 5$ variables, $N = 3$ countries and that there are $M = 2$ weakly exogenous variables. Then, omitting deterministic terms, the panel VARX model is

$$
\begin{aligned}
y_{1t} &= A_{11}(\ell)y_{1t-1} + A_{12}(\ell)y_{2t-1} + A_{13}(\ell)y_{3t-1} + F_1(\ell)W_t + u_{1t} && (4) \\
y_{2t} &= A_{21}(\ell)y_{1t-1} + A_{22}(\ell)y_{2t-1} + A_{23}(\ell)y_{3t-1} + F_2(\ell)W_t + u_{2t} && (5) \\
y_{3t} &= A_{31}(\ell)y_{1t-1} + A_{32}(\ell)y_{2t-1} + A_{33}(\ell)y_{3t-1} + F_3(\ell)W_t + u_{3t} && (6) \\
W_t &= M(\ell)W_{t-1} + w_t && (7)
\end{aligned}
$$

where $A_{ih,j}, i, h = 1, 2, 3$ are 5×5 matrices for each j and $F_{i,j}, i = 1, 2, 3$ are 5×2 matrices for each j. Furthermore, $E(u_t u_t') \equiv \Sigma_u = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$ is a full matrix and there is additional structure on the 5×5 matrices σ_{ij} $i, j = 1, 2, 3$, since the G variables are the same for each unit. In this setup, there are dynamic interdependencies ($A_{ik,j} = 0, k \neq i$ for some j), there are static interdependencies ($\sigma_{i,k} \neq 0, k \neq i$) and there are cross sectional heterogeneities ($A_{i,k} \neq A_{i+1,k}, k \neq i, i+1$).

Clearly, not all three distinguishing features of panel VARs need to be used in all applications. For example, when analyzing the transmission of shocks across the financial markets of different countries, static interdependencies are probably sufficient if the time period of the analysis is a month or a quarter. Similarly, when analyzing countries in a monetary union, it may be more important to allow for slope heterogeneities (e.g. different countries may respond differently to a fiscal shock) than for variance heterogeneities (the shocks hitting different countries have different magnitude). Finally, dynamic cross sectional differences are likely to be important when the panel includes developed and developing countries, or when it lumps together markets with different trading volumes, different transaction costs, etc.

Several interesting submodels are nested in the specification allowing us to test some of the model restrictions. For example, one might wish to know whether a model without dynamic interdependencies is sufficient to characterize the available data. This is the typical setup employed when all units are small and do not exercise dynamic effects on the other units, but shocks in different units have a common component. It is also the setup used in certain macro studies which treat units as isolated islands (see e.g. Rebucci, 2010a; De Graeve and Karas, 2012; and Sa et al., 2012, for recent examples).

Another restricted specification nested in the general framework and often used in the literature is one where all interdependencies are eschewed and cross sectional slope homogeneity is assumed. This is the typical setup used in micro studies, but it is potentially problematic in macroeconomic analyses dealing with countries or regions. Even within this restricted setup, micro and macro panel approaches differ in an important respect: the cross sectional dimension is typically large in micro studies and small or moderate for macro panels. Vice versa, micro panels typically feature a very short time series dimension while macro panels have a moderate time series dimension. These differences have important implications for the identification of the dynamics effects of interest.

2.2 What have panel VARs been used for?

Panel VARs have been used to address a variety of issues of interest to applied macroeconomists and policymakers. Within the realm of the business cycle literature, Canova et al. (2007) have employed a panel VAR to study the similarities and convergences among G7 cycles, while Canova and Ciccarelli (2012) employ them to examine the cross-sectional dynamics of Mediterranean business cycles. They can also be used to construct coincident or leading indicators of economic activity (see Canova and Ciccarelli, 2009) or to forecast out-of-sample, for example, output and inflation, taking into account potential cross unit spillovers effects. As we will see, cyclical indicators of both coincident and leading nature can be easily constructed from a panel VAR and (density) forecasts can be constructed with straightforward Monte Carlo methods.

Panel VARs are particularly suited to analyzing the transmission of idiosyncratic shocks across units and time. For example, Canova et al. (2012) have studied how shocks to U.S. interest rates are propagated to ten European economies, seven in the Euro area and three outside of it, and how German shocks, defined as shocks which simultaneously increase output, employment, consumption and investment in Germany, are transmitted to the remaining nine economies. Ciccarelli et al. (2012a) investigate the heterogeneity of macro-financial linkages in a number of developed economies and compare the transmission of real and financial shocks during the 2008-2009 recession. Caivano (2006) examines how disturbances generated in the Euro area are transmitted to the U.S. and vice versa, when these two geographic units are included into a world economy. Beetsma and Giuliadori (2011) and Lane and Benetrix (2011) look at the transmission of government spending shocks and Boubtane et al. (2010) examine how immigration shocks are transmitted in a variety of countries. Finally, Love and Zicchino (2006) measure the effect of shocks to "financial factors" on a cross section of U.S. firms.

Panel VARs have also been frequently used to construct average effects – possibly across heterogeneous groups of units – and to characterize unit specific differences relative to the average. For example, one may want to know if government expenditure is more countercyclical, on average, in countries or states which have fiscal restrictions included in the constitution, or whether the instantaneous fiscal rule depends on the type of fiscal restrictions that are in place (see Canova and Pappa, 2004). One may also be interested in knowing whether inflation dynamics in a monetary union may depend on geographical, political, institutional or cultural characteristics, or on whether fiscal and monetary interactions are relevant (see Canova and Pappa, 2007). Alternatively, one may want to examine whether shocks generated outside a country (or an area) dominate the variability of domestic variables (see Canova, 2005; Rebucci 2010a). Finally, one may want to examine what channels of transmission may make responses to international shocks different across countries and from the average, or how financial fragility may induce a different transmission mechanism of monetary policy shocks in different groups of countries (Ciccarelli et al., 2012b).

Another potential use of panel VARs is in analyzing the importance of interdependencies, and in checking whether feedbacks are generalized or only involve certain pairs of units. Thus, one may want to use a panel VARs to test the small open economy assumption or to evaluate certain exogeneity assumptions, often made in the international economics literature. Finally, panel VARs may be used to examine the extent of dynamic heterogeneity and of convergence clubs (see Canova, 2004), to endogenously group units or to characterize their differences. For instance, De Graeve and Karas (2012) show that the dynamics of deposits and interest rates of "good" and "bad" banks differs in response to bank run shocks. They also show that differences in the health of their balance sheet are of second order importance and what truly matters is whether banks are insured or not by regulators.

2.3 Are panel VAR models consistent with economic theory?

As with standard VARs, one may wonder whether panel VARs can be used to "test" theories or to inform researchers about the relative validity of different economic paradigms. Panel VARs can be easily generated from standard intertemporal optimization problems under constraints, as long as the decision rules are log-linearized around the steady state. For example, Canova (2007, Chapter 8) shows that, if all variables are observed, a small open economy version of the Solow growth model generates a heterogeneous panel VAR with static but without dynamic interdependencies. More importantly, he shows that the panel VAR that the theory generates either has both fixed effects and dynamic heterogeneity or none of them – i.e., either the steady states and the dynamics are heterogeneous or both are homogeneous. Thus, it is difficult to justify the common practice of specifying panel VARs which allow only for intercept heterogeneity but impose dynamic homogeneity.

Panel VARs with dynamic interdependencies can be obtained if the small open economy assumption is dropped and at least one asset is traded in financial markets or if intermediate factors of production are exchanged in open markets (see e.g. Canova and

Marrinan, 1998). In this case, market clearing implies that the excess demands present in a unit is compensated by excess supply in other units, and these spillovers, together with adjustments in the relative prices of goods and/or assets, imply generalized feedbacks from one unit to all the others. Also here, whenever there are heterogeneities in the steady states, there will be heterogeneities in the dynamic responses to domestic and international shocks.

It is well known that if the VAR omits relevant states of the optimization problem, there is no insurance that the innovations obtained in a structural VAR (SVAR) will display the same characteristics as the innovations appearing in theoretical models (see e.g. Fernandez Villaverde et al., 2005). Since the states of the problem are often not observed, this mismatch has caused several researchers to question the use of SVARs. However, it is also well know that this mismatch does not have an either/or consequence and that there are situations where the structural innovations the VAR recovers may have characteristics that are different from the innovations of the theoretical model but the dynamics they induce are similar (see e.g. Sims, 2012). Panel VARs are not different in this respect: if important states are omitted from the list of variables for each unit, standard non-fundamentalness issues may arise. Thus, care must be exercised in choosing the variables for each unit. The presence of a cross section does not help in general to reduce the non-fundamentalness problem, unless it happens that cross sectional data reveals information about the missing states that the data of single unit is not able to provide – an event which is, probabilistically, quite remote.

3 Reduced-form estimation

Depending on the exact specification, different approaches can be used to obtain estimates of the unknowns of the model. Because of the added complexity, we first discuss the case of panel VARs without dynamic interdependencies and then analyze what happens when these dynamic interdependencies are allowed for. Both classical and Bayesian estimators are presented even though, for this particular problem, a Bayesian perspective is preferable, as it gives important insights into the estimation problem.

3.1 Panel VARs without dynamic interdependencies

Suppose there is a domestic VAR for each unit but the reduced form shocks may be correlated across units. For later reference, we call this setup *a collection of unit specific VARs*. Let the cross sectional size N be large. If we are willing to assume that the data generating process features dynamic homogeneity, and to condition on initial values of the endogenous variables, pooled estimation with fixed effect – potentially capturing idiosyncratic but constant heterogeneities across variables and/or units – is the standard classical approach to estimating the parameters of the model. However, when T is fixed, the pooled estimator is biased and one may want to employ the GMM approach of Arellano and Bonds (1991), which is consistent even when T is small. A GMM approach, however, requires differencing the specification, throws away sample information and may make inference less accurate when the information being ignored is important for the parameters of interest. Rather than differencing, one may want to impose a-priori restrictions which ensure consistency in such an environment. Sims (2000) emphasizes that, in a model of this type, lagged dependent variables are not independent of the unit specific intercepts and describes how this information can be used to recover parameter consistency, for both stationary and non-stationary models. Generally speaking, the inconsistency problem arises when the conditional pdf is used as the likelihood, because the number of parameters grows with the cross sectional size. If the unconditional pdf is instead employed, where the density of the initial observations is a function of the unit specific intercept and of the common slope parameters, consistency may be obtained. To be specific, let the model be

$$y_{it} = A_{0i} + A(\ell)y_{it-1} + u_{it} \tag{8}$$

where $u_{it}|y_{it-1} \sim N(0, \Sigma_i)$ and, for the sake of illustration, the roots of $A(e^{-iw})$ are all outside the unit circle. Then, the unconditional distribution of the initial conditions is $N((I - A(\ell))A_{0i}, \Omega_i)$, where Ω_i is implicitly defined by $\Omega_i = \Sigma_i + \sum_j A_j \Omega_i A_j'$. If this density is used together with the standard conditional density to build the likelihood, the maximum likelihood estimator of the parameters will be consistent, even with T short.

If T is large enough, one could also consider estimating the VAR for different units separately and averaging the results across units. Such a mean group estimator is inefficient relative to the pooled estimator under dynamic homogeneity, but gives consistent estimates of the average dynamic effect of shocks if dynamic heterogeneity is present, whereas the pooled estimator does not (see e.g. Pesaran and Smith, 1995, and Rebucci, 2010b). The pooled estimator is inconsistent under dynamic heterogeneity because the regressors are correlated with the error term. If the data generating process features dynamic heterogeneity, both a within and a between estimator will also give inconsistent estimates of the parameters, even when N and T are large, since the error term is also likely to be correlated with the regressors. With dynamic heterogeneity, a GMM strategy may be difficult to employ since it is hard to find instruments which are simultaneously correlated with the regressors and uncorrelated with the error term.

When T and N are moderate and dynamic heterogeneity suspected, some form of "partial pooling" may help to improve the quality of the coefficients' estimate. One standard format leading to partial pooling is a random coefficients model. The setup is the following. The model (without dynamic interdependencies) is $y_{it} = A_{0i} + A_i(\ell)y_{it-1} + u_{it}$, where now the slope parameters are potentially unit specific. If we impose that

$$\alpha_i = \bar{\alpha} + v_i \tag{9}$$

where $\alpha_i = [vec(A_i(\ell)), vec(A_{0i})]'$ and $v_i \sim N(0, \Sigma_v)$, an estimator which partially pools the information present in different units can be constructed. Note that (9) implies that the coefficients of the VAR in different units are different, but are drawn from a distribution whose mean and variance are constant across i.

Given this structure, several estimators are available in the literature (see e.g. Canova, 2007, chapter 8). The two most popular ones are a classical estimator and a Bayesian one. In the classical estimator proposed by Swamy (1970) (9) is substituted into the model and GLS is applied. GLS is required because of the particular error structure that substitution of (9) in the model generates. Importantly, this setup does not allow the estimation of the individual unit coefficients: only the mean $\bar{\alpha}$ is estimated. An estimate of the degree of cross sectional heterogeneity is also difficult to obtain since Σ_v enters in a complicated way in the variance of the composite error term of the model.

The Bayesian alternative treats (9) as an exchangeable prior. This prior is then combined with the likelihood of the data to obtain the posterior distribution of the individual α_i, and of the average value, $\bar{\alpha}$, if that is of interest. Thus, a Bayesian perspective allows us to quantify the heterogeneity present in the dynamics, while a classical approach does not. If e_i and u_i are normally distributed, and $\bar{\alpha}$ and Σ_v known, conditional on the initial observations, the posterior of α_i is normal with mean

$$\tilde{\alpha}_i = (X_i' \Sigma_{i,ols}^{-1} X_i + \Sigma_v^{-1})^{-1} (X_i' \Sigma_{i,ols}^{-1} X_i \alpha_{i,ols} + \Sigma_v^{-1} \bar{\alpha}) \tag{10}$$

where $\alpha_{i,ols}$ is the OLS estimator of α_i, $\Sigma_{i,ols}^{-1}$ is the OLS estimate of Σ_i and X_i is the matrix containing the right hand side variables, and variance

$$(X_i' \Sigma_{i,ols}^{-1} X_i + \Sigma_v^{-1})^{-1} \tag{11}$$

The moments of the posterior distribution of α_i have the usual convenient format: the posterior mean is a linear combination of sample and prior information, with weights given by the relative precision of the two types of information and the posterior variance is a weighted average of the prior and of the sample variance. Note that, if one wants to use the information present in the initial conditions for estimation (which could be very important when both N and T are short), a specification like the one suggested by Sims

(2000), where a joint distribution for (α_i, y_{i0}) is a-priori specified, could be used.

The above posterior distribution does not take into account the fact that the shocks in each unit VAR may be correlated. Thus, efficiency can be improved by stacking the sample information of all the units. Since the resulting model has a SUR structure, the convenient weighted average property of posterior moments is maintained.

Three important points need to be made regarding the posterior moment in (10) and (11). First, the formulas are valid under the assumption that $\bar{\alpha}$ and Σ_v are known. If they are not, one can specify a prior distribution for these unknown and use the Gibbs sampler to construct draws for the marginal of α_i.We will describe how the Gibbs sampler can be used in a more complicated version of this hierarchical model in section 4. Shortcuts designed to decrease the complexity of the computations are available. For example, modal estimates α_i^* can be easily computed plugging in

$$\bar{\alpha}^* = \frac{1}{n}\sum_{i=1}^{n}\alpha_i^* \tag{12}$$

$$(\sigma_i^*)^2 = \frac{1}{T+2}[(y_i - X_i\alpha_i^*)'(y_i - X_i\alpha_i^*)] \tag{13}$$

$$\Sigma_v^* = \frac{1}{n-\dim(\alpha)-1}[\sum_i(\alpha_i^* - \bar{\alpha}^*)(\alpha_i^* - \bar{\alpha}^*) + \Phi \tag{14}$$

where "*" indicates modal estimates and $\Phi = \text{diag}[0.001]$ in the above formulas. In equation (14) an arbitrary diagonal matrix is added since Σ_v^* may not be positive definite. Alternatively, one could estimate $\bar{\alpha}$ and Σ_v from a training sample, if one is available, or from information contained in units left out from the cross-section. In this situations, the formulas in (10) and (11) are still valid with plug-in estimates in place of true values of $\bar{\alpha}$ and Σ_v. Clearly, with plug-in estimates, the uncertainty present in the posterior of α_i will be underestimated – the estimation error present in $\bar{\alpha}$ and Σ_v is disregarded, and one may want to use simple corrections to take these errors into account (see Canova, 2004).

Second, the mean of the posterior $\tilde{\alpha}_i$ collapses to a standard OLS estimator constructed from unit specific information if heterogeneity is large, i.e., $\Sigma_v \to \infty$, and it collapses to the prior mean, if the sample is uninformative. Thus, if one of the two types of information is highly imprecise, it is disregarded in the construction of the posterior. Third, the classical GLS estimator satisfies $\alpha_{GLS} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\alpha}_i$. Thus, the classical estimator for the mean effect is the arithmetic average of the posterior means of the individual units.

Canova (2005), Canova and Pappa (2007) and more recently Calza et al. (2012), apply such a Bayesian random coefficient approach to estimate the dynamic responses to shocks of a potentially heterogeneous collection of unit specific VARs. These papers, however, rather than modeling cross sectional heterogeneities in VAR coefficients, as described in (9), specify directly the nature of the heterogeneities present in the MA representation of the data for each unit. Thus, the model is

$$\tilde{y}_{it} = B_i(\ell)u_{it} \quad i = 1, \dots, N \tag{15}$$

where $\tilde{y}_{it}$ represents the original vector of series in country i in deviation from the deterministic components. Let $\beta_i = vec(B_i(\ell))$. Then, one can assume that the vector of MA coefficients β_i are random around a mean

$$\beta_i = \bar{\beta} + v_i \tag{16}$$

where $v_i \sim N(0, \Sigma_v)$. A Bayesian estimator of β_i still maintains a weighted average structure with weights given by the relative precision of the two types of information. The advantage of (16) is that an economically reasonable prior for the dynamic responses to shocks may be much easier to formulate than a prior for the VAR coefficients.

In many applications, an estimator of the average effect $\bar{\alpha}$ ($\bar{\beta}$) is of interest. If a Bayesian approach is followed, one may obtain the average effect by averaging the posterior mean over i. Thus, the classical and the Bayesian estimator of the mean effect coincide. An alternative is available if, in addition to treating $\alpha_i(\beta_i)$ as random around a mean, $\bar{\alpha}$ ($\bar{\beta}$) and Σ_v are also treated as random. In this case, the posterior distribution for, e.g., $\bar{\beta}$ can be obtained integrating out β_i and Σ_v from the joint posterior of $(\beta_i, \bar{\beta}, \Sigma_v)$ using standard hierarchical methods (see e.g. Canova, 2007; Jarocinski, 2010).

Improved estimates of β_i can also be obtained in other ways. For example, Zellner and Hong (1989) suggested a prior specification for the α_i that results in a posterior distribution for the VAR coefficients combining unit specific and average sample information. In particular, when (9) holds and v_i has a normal distribution with mean zero and variance $\Sigma_v = \phi^{-1}\sigma_u^2 I_k$ (k being the dimension of α_i), and $\Sigma_i = \Sigma = \sigma_u^2 I_G$, the (conditional) posterior distribution for $\alpha = (\alpha_1, \ldots, \alpha_n)'$ will be normal with mean $\tilde{\alpha}$ given by

$$\tilde{\alpha} = (Z'Z + \phi I_{nk})^{-1}(Z'Z\alpha_{ols} + \phi J\bar{\alpha}) \tag{17}$$

where $\alpha_{ols} = [\alpha_{1,ols}, \ldots, \alpha_{n,ols}]$ is the OLS estimator of α, unit by unit, Z is a block diagonal matrix containing in each diagonal block the regressors of each unit, and $J' = (I_k, \ldots, I_k)$. The variance of the posterior is $(Z'Z + \phi I_{nk})^{-1}$. Zellner and Hong (1989) replace $\bar{\alpha}$ with the mean group estimator. Alternatively, an additional prior can be added for $\bar{\alpha}$ and its posterior distribution derived in a fully-fledged hierarchical setup.

Improved classical estimators, which combine unit specific and average information, also exist. For example, a James-Stein estimator for the above model is

$$\alpha_i = \alpha_p + (1 - \frac{\kappa}{F})(\alpha_{i,ols} - \alpha_p) \quad i = 1, \ldots, n \tag{18}$$

$\alpha_{i,ols}$ is the OLS estimator with unit i data, α_p is the pooled estimator, F is the statistic for the null hypothesis $\alpha_i = \alpha, \forall i$, and $\kappa = [(NG-1)dim(\alpha)-2]/[NGT - dim(\alpha)+2]$. Thus, the shrinkage factor κ depends on the dimension of α relative to T. When $dim(\alpha) >> T$, $(1 - \frac{\kappa}{F})$ is smaller, therefore pulling α_i closer to α_p. It is rare to see estimators of this type in the macroeconomic literature primarily because the pre-testing required to construct (18) is not typically performed.

3.2 Panel VARs with dynamic interdependencies

The estimation problem becomes more complicated if dynamic interdependencies are allowed for. The problem is related to the curse of dimensionality: excluding deterministic components, since there are $k = NGp$ coefficients in each equation, the total number of parameters (NGk) to be estimated in the model easily exceeds the sample one has available. One way to solve this problem is to selectively model the dynamic links across units while imposing zero-restrictions on others. Thus, for example, one can assume that only the variables of unit j enter the equations of unit i, as in the spatial VAR model discussed in section 6.2. It is unclear how to do this in a way that avoids data mining. An alternative is to group cross sectional units into clubs and assume random coefficients within each group but no relationship across groups. Such an approach has been used, for example, in Canova (2004) to examine regional convergence rates, but it naturally applies to a situation where dynamic interdependencies are excluded. Thus, some ingenuity is required to extend it to a framework where interdependencies are allowed for.

Canova and Ciccarelli (2004 and 2009) have suggested different cross sectional shrinkage approaches which can deal with the curse of dimensionality and thus allow the estimation of models with dynamic interdependencies. Del Negro and Schorfheide (2010) provide an overview of the approach.

To see what the procedure involves, rewrite (2) in a simultaneous equations format:

$$Y_t = Z_t \alpha + U_t \tag{19}$$

where $Z_t = I_{NG} \otimes X_t'$; $X_t' = (I, Y_{t-1}',\ Y_{t-2}', \ldots, Y_{t-p}')$, $\alpha = (\alpha_1', \ldots, \alpha_N')'$ and α_i are $Gk \times 1$ vectors containing, stacked, the G rows of the matrices ($A_{oi}(t)$,$A_i(\ell)$), while Y_t and U_t are $NG \times 1$ vectors. Since α varies with cross-sectional units, its sheer dimensionality prevents any meaningful unconstrained estimation. Thus assume that α depends on a much lower dimension vector θ and posit the following linear structure:

$$\alpha = \Xi_1 \theta_1 + \Xi_2 \theta_2 + \Xi_3 \theta_3 + \Xi_4 \theta_4 + ... + e_t \tag{20}$$

where $\Xi_1,\ \Xi_2,\ \Xi_3, \Xi_4$ are matrices of dimensions $NGk \times q$, $NGk \times N$, $NGk \times G$, $NGk \times 1$ respectively and $\theta_i, i = 1, 2 \ldots$, are factors, capturing the determinants of α. For example,

θ_1 could capture components in the coefficient vector which are common across units and variables (or groups of them) – its dimension is, say, q; θ_2 could capture components which are common within units, thus its dimension equals N; θ_3 could capture components which are variable specific, thus its dimension is equal to G; θ_4 could capture components in the lagged coefficients and its dimension is equal to $p_1 < p$, and so on. Finally, e_t captures all the unmodelled features of the coefficient vector, which may have to do with time specific or other idiosyncratic effects.

Factoring α as in (20) is advantageous in many respects. Computationally, it reduces the problem of estimating NGk coefficients into the one of estimating $q+N+G+p_1$ factors characterizing them. Practically, the factorization (31) transforms an overparametrized panel VAR into a parsimonious SUR model, where the regressors are averages of certain right-hand side VAR variables. In fact, using (20) in (19) we have

$$Y_t = \sum_{j=1}^{r} \mathcal{Z}_{jt}\theta_j + \gamma_t \tag{21}$$

where $\mathcal{Z}_{jt} = Z_t\Xi_j$ capture respectively, common, unit specific, variable specific, lag specific information present in the lagged dependent variables, and $\gamma_t = U_t + Z_t e_t$. Since, by construction, Z_{it} has a moving average structure, the regressors in (21) will capture low frequency movements present in the VAR and this feature is valuable in medium term out-of-sample forecasting exercises. Economically, the decomposition in (21) conveniently allows us to measure, for example, the relative importance of common and unit specific influences for fluctuations in Y_t. In fact, $WLI_t = \mathcal{Z}_{1t}\theta_1$ plays the role of a (vector) of common indicators, while $CLI_t = \mathcal{Z}_{2t}\theta_2$ plays the role of a vector of unit specific indicators. In general, WLI_t and CLI_t are correlated – a portion of the variables in $\mathcal{Z}_{1t}$ also enter in $\mathcal{Z}_{2t}$ – but the correlation tends to zero as N increases.

To illustrate the structure of the $\mathcal{Z}_{jt}$'s, suppose there are $G = 2$ variables, $N = 2$ countries, $q = 1$ common component, $p = 1$ lags, and omit deterministic components, for convenience. Then:

$$\begin{bmatrix} y_t^1 \\ x_t^1 \\ y_t^2 \\ x_t^2 \end{bmatrix} = \begin{bmatrix} A_{1,1}^{1,y} & A_{2,1}^{1,y} & A_{1,2}^{1,y} & A_{2,2}^{1,y} \\ A_{1,1}^{1,x} & A_{2,1}^{1,x} & A_{1,2}^{1,x} & A_{2,2}^{1,x} \\ A_{1,1}^{2,y} & A_{2,1}^{2,y} & A_{1,2}^{2,y} & A_{2,2}^{2,y} \\ A_{1,1}^{2,x} & A_{2,1}^{2,x} & A_{1,2}^{2,x} & A_{2,2}^{2,x} \end{bmatrix} \begin{bmatrix} y_{t-1}^1 \\ x_{t-1}^1 \\ y_{t-1}^2 \\ x_{t-1}^2 \end{bmatrix} + U_t \tag{22}$$

$\alpha = [A_{1,1}^{1,y}, A_{2,1}^{1,y}, A_{1,2}^{1,y}, A_{2,2}^{1,y}, A_{1,1}^{1,x}, A_{2,1}^{1,x}, A_{1,2}^{1,x}, A_{2,2}^{1,x}, A_{1,1}^{2,y}, A_{2,1}^{2,y}, A_{1,2}^{2,y}, A_{2,2}^{2,y}, A_{1,1}^{2,x}, A_{2,1}^{2,x}, A_{1,2}^{2,x}, A_{2,2}^{2,x}]'$ is a 16×1 vector and the typical element of α, $\alpha_{l,h}^{i,j}$, is indexed by the unit i, the variable j, the variable in an equation l (independent of the unit), and the unit in an equation h (independent of variable). If we are not interested in modelling all these aspects, one possible factorization of α is

$$\alpha = \Xi_1\theta_1 + \Xi_2\theta_2 + \Xi_3\theta_3 + e_t \tag{23}$$

where e_t captures unaccounted features, and for each t, θ_1 is a scalar, θ_2 is a 2×1 vector, θ_3 is a 2×1 vector, Ξ_1 is a 16×1 vector of ones,

$$\underset{(16\times 2)}{\Xi_2} = \begin{bmatrix} \iota_1 & 0 \\ \iota_1 & 0 \\ 0 & \iota_2 \\ 0 & \iota_2 \end{bmatrix} \qquad \underset{(16\times 2)}{\Xi_3} = \begin{bmatrix} \iota_3 & 0 \\ 0 & \iota_4 \\ \iota_3 & 0 \\ 0 & \iota_4 \end{bmatrix}$$

with $\iota_1 = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix}'$, $\iota_2 = \begin{pmatrix} 0 & 0 & 1 & 1 \end{pmatrix}'$, $\iota_3 = \begin{pmatrix} 1 & 0 & 1 & 0 \end{pmatrix}'$ and $\iota_4 = \begin{pmatrix} 0 & 1 & 0 & 1 \end{pmatrix}'$. Substituting (23) into the model, the panel VAR can be rewritten as

$$\begin{bmatrix} y_t^1 \\ x_t^1 \\ y_t^2 \\ x_t^2 \end{bmatrix} = \begin{bmatrix} \mathcal{Z}_{1t} \\ \mathcal{Z}_{1t} \\ \mathcal{Z}_{1t} \\ \mathcal{Z}_{1t} \end{bmatrix} \theta_1 + \begin{bmatrix} \mathcal{Z}_{2,1,t} & 0 \\ \mathcal{Z}_{2,1,t} & 0 \\ 0 & \mathcal{Z}_{2,2,t} \\ 0 & \mathcal{Z}_{2,2,t} \end{bmatrix} \theta_2 + \begin{bmatrix} \mathcal{Z}_{3,1,t} & 0 \\ 0 & \mathcal{Z}_{3,2,t} \\ \mathcal{Z}_{3,1,t} & 0 \\ 0 & \mathcal{Z}_{3,2,t} \end{bmatrix} \theta_3 + \gamma_t \tag{24}$$

where $\mathcal{Z}_{1t} = y_{t-1}^1 + x_{t-1}^1 + y_{t-1}^2 + x_{t-1}^2 + 1$, $\mathcal{Z}_{2,1,t} = y_{t-1}^1 + x_{t-1}^1$, $\mathcal{Z}_{2,2,t} = y_{t-1}^2 + x_{t-1}^2$, $\mathcal{Z}_{3,1,t} = y_{t-1}^1 + y_{t-1}^2$, $\mathcal{Z}_{3,2,t} = x_{t-1}^1 + x_{t-1}^2$.

The specification in (21) is preferable to a collection of VARs or bilateral VARs for two reasons. First, the parsimonious use of cross sectional information helps to get more accurate estimates of the parameters and to reduce standard errors. Second, if the momentum that the shocks induce across countries is the result of a complicated structure of lagged interdependencies, the specification will be able to capture it. Such a structure would instead emerge as "common shocks" in the two alternative frameworks.

It is easy to estimate a model like (21). Stack the t observations in a vector so that

$$Y = \sum_{j=1}^{r} \mathcal{Z}_j \theta_j + \gamma \tag{25}$$

Thus, the reparametrized model is simply a multivariate regression model. If the factorization in (20) is exact, the error term is uncorrelated with the regressors and classical OLS can be used to estimate the vector θ and thus the vector α. Consistency is ensured as T grows. When the factorization in (20) allows for an error, γ_t has a particular heteroschedastic covariance matrix which needs to be taken into account. If a Bayesian framework is preferred, the posterior for the unknowns is easy to construct. Let $e_t \sim N(0, \Sigma_u \otimes V)$ and further restrict $V = \sigma^2 I$ as in Kadiyala and Karlsson (1997). Then $v_t \sim N(0, \sigma_t \Sigma_u)$ where $\sigma_t = (I + \sigma^2 X_t' X_t)$. Thus, if the prior for $(\theta, \Sigma_u, \sigma^2)$ is, for example, of the semi-conjugate type: $\theta \sim N(\theta_0, \Omega_0)$, $\Sigma_u^{-1} \sim W(z_0, Q_0)$, $\sigma^{-2} \sim G(0.5a_0, 0.5a_0 s^2)$, where $(\theta_0, \Omega_0, z_0, Q_0, a_0, s^2)$ are known quantities, W stands for Wishart distribution, and G for Gamma distribution, one can use the Gibbs sampler to construct sequences for $(\theta, \Sigma_u, \sigma^2)$ from their joint posterior distribution – see next section for details.

4 Adding time variation in the coefficients

The modern macroeconomic literature is taking seriously the idea that the coefficients of a VAR and the variance of the shocks may be varying over time. For example, Cogley and Sargent (2005), Primiceri (2005), pioneered a specification where the VAR coefficients evolve over time like random walks; Sims and Zha (2006) assume that VAR coefficients evolve over time according a Markov switching process, while Auerbach and

Gorodnichenko (2011) specify a smooth transition VAR model, where contemporaneous and lagged coefficients are a function of a pre-specified variable indicator.

Specifications of this type can also be used in a panel VAR framework if time variation in the parameters is suspected. In general, the presence of time variation in the coefficients adds to the curse of dimensionality and some ingenuity is required if one is to obtain meaningful estimates of the parameters and of the responses to the shocks of interest. The approach employed in Canova and Ciccarelli (2004 and 2009), and Canova et al. (2007 and 2012), which extends the shrinkage structure previously described to the case of time varying parameter models, can go a long way in that direction.

Let the time varying panel VAR model be

$$y_{it} = A_{0i}(t) + A_{it}(\ell)Y_{t-1} + F_{it}(\ell)W_t + u_{it} \tag{26}$$

where $A_{it}(\ell)$ are the coefficients on the lag endogenous variables Y_{t-1}, W_t is a $M \times 1$ vector of weakly exogenous variables common to all units and time-varying coefficients are allowed in both $A_{it}(\ell)$ and $F_{it}(\ell)$. Such a specification could be employed, for example, to study time varying business cycle features of a vector of countries, evolutionary patterns in the transmission of structural shocks across variables or units, or the effect of changes in the variance of the shocks on relevant endogenous variables. Time variation in the coefficients add realism to the specification but is costly, since there are $k = (NGp + Mq)$ parameters in each equation and there is only one time period per unit to estimate them.

Two approaches have been proposed to estimate such model. To see what they involve, rewrite (26) in a simultaneous equations format:

$$Y_t = Z_t\alpha_t + U_t \tag{27}$$

where $\alpha_t = (\alpha'_{1t}, \ldots, \alpha'_{Nt})'$ and α_{it} are $Gk \times 1$ vectors containing, stacked, the G rows of the matrices A_{0i}, A_{it} and F_{it}.

4.1 A panel-type hierarchical prior

Here α_t is assumed to be the sum of two independent components: one which is unit specific and constant over time; the other common across units but time-varying, i.e.:

$$\alpha_{it} = \delta_i + \lambda_t$$

An exchangeable prior is assumed for δ_i

$$\delta_i = \bar{\delta} + v_i, \qquad v_i \sim N(0, \Sigma_v) \tag{28}$$

and, a further layer of dimensionality reducing hierarchy can be specified by setting $\bar{\delta} \sim N(\mu, \Psi)$. On the other hand, λ_t is assumed to follow an autoregressive process:

$$\lambda_t = \rho\lambda_{t-1} + (1-\rho)\lambda_0 + e_t \tag{29}$$

Additional assumptions on Σ_v, ρ, λ_0 and e_t complete the specification of the prior.

This setup is convenient and found to be useful in forecasting and turning point analysis (Canova and Ciccarelli, 2004). The fact that the time-varying parameter vector is common across units does not prevent unit-specific structural movements, since α_{it} can be re-written as

$$\alpha_{it} = (1-\rho)(\delta_i + \lambda_{i0}) + (1-\rho)\alpha_{it-1} + e_t$$

where persistent movements in α_{it} are driven by the common coefficient ρ. Note also that the setup provides a general mechanism to account for structural shifts without explicitly modeling their sources.

The assumptions made allow us to recover the posterior of δ_i and of $\bar{\delta}$. Thus, one can distinguish between individual α_{it} and mean effects $\bar{\alpha}_t = \bar{\delta} + \lambda_t$, as in Lindley and Smith (1972). The difference is relevant in a forecasting context, since one may be concerned in predictions with the posterior of the average $\bar{\alpha}_t$ or with the posterior of the distribution of unit specific effects. As in the case of a collection of VARs, the exchangeability assumption on δ_i allows for some degree of cross sectional pooling and, again, this may be useful when there are similarities in the characteristics of the vector of variables across units.

The structure of the model can be summarized with the following hierarchical scheme:

$$\begin{aligned} Y_t &= Z_t\delta + Z_t\mathcal{S}\lambda_t + U_t & U_t \sim N(0, \Sigma_u) \\ \delta &= \mathcal{S}\bar{\delta} + \zeta & \zeta \sim N(0, \Delta) \\ \bar{\delta} &= \mu + \omega & \omega \sim N(0, \Psi) \\ \lambda_t &= \rho\lambda_{t-1} + (1-\rho)\lambda_0 + e_t & e_t \sim N(0, \Sigma_e) \end{aligned} \tag{30}$$

where $\mathcal{S} = \mathbf{e} \otimes I$; $\mathbf{e} = vec(1, 1, ..., 1)$; and $\Delta = I \otimes \Sigma_v$.Canova and Ciccarelli (2004) describe how to construct the posterior distributions for (functions of) the parameters of interest under several prior assumptions on the variance covariance matrices Σ_u, Δ, Ψ and Σ_e, the mean vector μ and the initial λ_0, using the Gibbs sampler.

4.2 A factor structure for the coefficient vector

Another possibility is to allow for time variation in the factorization in (20). Thus, let

$$\alpha_t = \sum_j \Xi_j \theta_{jt} + e_t \tag{31}$$

where Ξ_j are matrices with ones and zeros and θ_{jt} are factors. While in the setup of equation (20) θ's were fixed hyperparameters, now they are stochastic processes and thus a specification of their law of motion is needed to complete the model. Canova and Ciccarelli (2009) study different alternatives for this law of motion. A simple representation, nested in their specification, which illustrates the point is

$$\theta_t = \theta_{t-1} + \eta_t \qquad \eta_t \sim N(0, \Omega_t). \tag{32}$$

where $\theta_t = [\theta_{1t}, \theta_{2t} \ldots]'$, Ω_t is block diagonal and U_t, e_t and η_t be mutually independent.

In (32) the factors driving the coefficients of the panel VAR evolve over time as random walks. This specification is similar to the one employed in the time varying coefficient VAR literature, but it is parsimonious since it concerns θ_t, which is of much smaller dimension than the α_t vector, and allows us to focus on coefficient changes which are permanent. The variance of η_t is, in principle, allowed to be time-varying. Such a specification implies ARCH-M type effects in the representation for Y_t and it is a way of modeling time-varying conditional second moments that provides an alternative to the stochastic volatility specification used, e.g., in Cogley and Sargent (2005) and many others. The main difference is that here volatility changes will be related to coefficient changes. Note that the computational costs involved in using this specification are moderate since the dimension of θ_t is considerably smaller than the dimensionality of Y_t. The block diagonality of Ω_t, on the other hand, guarantees the identifiability of the factors.

To make the specification composed of (27), (31) and (32) estimable, we need assumptions on the error terms of (27) and of (31). If we let $U_t \sim N(0, \Sigma_u)$ and $e_t \sim N(0, \Sigma_u \otimes V)$, where $V = \sigma^2 I_k$ is a $k \times k$ matrix, the reparametrized model has the state space structure:

$$\begin{aligned} Y_t &= (Z_t \Xi)\theta_t + \gamma_t \qquad && \gamma_t = U_t + Z_t e_t \sim N(0, \sigma_t \Sigma_u) \\ \theta_t &= \theta_{t-1} + \eta_t && \eta_t \sim N(0, \Omega_t) \end{aligned} \tag{33}$$

where $\sigma_t = (I + \sigma^2 Z_t' Z_t)$. Bayesian estimation requires prior distributions for $\Sigma_u, \Omega_t, \sigma^2$ and θ_0. Canova and Ciccarelli (2009) show how these joint prior densities can be specified so that the posterior distribution for the quantities of interest can be computed numerically with MCMC methods. Once these distributions are found, location and dispersion measures for any interesting continuous functions of the parameters can be obtained. Similarly, the marginal likelihood and the predictive distributions needed for model checking and model comparisons are easy to construct.

If classical methods are preferred, notice that (33) is a linear state space system, where θ_{it} represents unobservable states. Thus, variations of the Kalman filter algorithm can be used to construct the likelihood function which then can be maximized with respect to the relevant parameters (see e.g. Ljung and Soderstrom, 1983).

4.3 Implementation in small samples

The approaches described in the previous two subsections conditions on the initial p observations. When T is large, the difference conditional and unconditional likelihood is likely to be small. When T is small, the information present in the initial conditions may contain important information for the quantities of interest. Intuitively, with $T = 20$ observations throwing away, say $p = 4$, initial conditions effectively reduces the information by 20 percent, making the likelihood flatter and leaving to the prior the burden of producing enough curvature in the posterior. This problem is relevant for many applications since comparable macro time series across a number of countries or sectors exist only for the last 15-20 years, at best. Thus, one may want to take all the existing information into account when constructing the posterior of the quantities of interest.

In the case of (33), conditional on the first p observations, the likelihood is

$$L(\theta|y,\mathcal{Z}) \propto (\prod_t \sigma_t)|\Sigma_u|^{-0.5T} exp[-0.5 \sum_t (Y_t - \mathcal{Z}_t\theta_t)'(\sigma_t\Sigma_u)^{-1}(Y_t - \mathcal{Z}_t\theta_t)] \tag{34}$$

Since $\mathcal{Z}_t = Z_t\Xi_j$, the likelihood of $\mathcal{Z}_{jt}$ is proportional to the likelihood of Z_t. The likelihood of the initial conditions can be written as

$$L(\mathcal{Z}|\psi) \propto exp[-0.5 \sum_i (Z_t - \bar{Z})'(\Sigma_Z)^{-1}(Z_t - \bar{Z})] \tag{35}$$

where $\bar{Z}$ is a vector of mean parameters. The full likelihood of the sample is then simply the product of (34) and (35) and can be combined with the prior to yield a posterior kernel or a conditional posterior for the unknowns which can be used in the Gibbs sampler.

5 Impulse responses and shock identification

Shock identification can be performed with standard methods. To decrease the number of identification restrictions, it is typical to assume that Σ_u is block diagonal, with blocks corresponding to each unit, employ symmetric identification restrictions across units (these could be zero, long run, sign restrictions or a combination of the them) and require the structural shocks to be orthogonal. Block diagonality implies differences in the responses within and across units: within a unit, variables are allowed to move instantaneously; across units, variables may react but only with a lag. Symmetric identification restrictions imply that the nature of the disturbances (i.e., being demand or supply shocks) is independent of the unit.

Restrictions of this type have been used to obtain the (cross sectional) average responses or the average responses of particular groups of units which are homogeneous in their dynamics. Ciccarelli and Rebucci (2006), for instance, study the transmission of a German monetary policy shock, identified as an innovation to the reaction function of the Bundesbank, in the four largest EMU countries using pre-EMU data. Jarocinski (2010) compares responses to monetary policy shocks in the Euro area countries before the EMU to those in the new member states from Eastern Central Europe. A monetary policy shock is identified with the same restrictions in each group of countries. Rebucci (2010a) is interested in assessing the role of external and policy shocks for domestic growth variability. External and policy shocks are identified by imposing the same Choleski decomposition in all countries. Finally, Ciccarelli et al. (2012b) analyze whether financial fragility has altered the transmission mechanism of monetary policy in the Euro area. A panel VAR is estimated for core countries and countries under financial stress, allowing the slopes and the contemporaneous impact matrix to be group dependent, but restricting them to be common within groups. A monetary policy shock is identified using the same restrictions in the two groups and average responses for each group are constructed.

Shock identification is somewhat more complicated when static interdependencies are allowed for and cross unit symmetry in shock identification cannot be assumed. A convenient tool to be used in this situation is described in Canova and Ciccarelli (2009). Researchers using panel VARs with static and dynamic interdependencies and, possibly, time variation in the coefficients may be interested in computing the responses of the endogenous variables to shocks in the variables or to shocks to the coefficients (via shocks to the common λ_t or shocks to the factors θ_t) and in describing their evolution over time. In this situation, responses can be obtained as the difference between two conditional forecasts: one where a particular variable (coefficient) is shocked and one where the disturbance is set to zero.

Formally, let y^t be the history for y_t, θ^t the trajectory for the coefficients up to t, Ω^t the trajectory for the variance of the coefficients up to t; $y_{t+1}^{t+\tau} = [y'_{t+1}, ...y'_{t+\tau}]'$ be a collection of future observations and $\theta_{t+1}^{t+\tau} = [\theta'_{t+1}, ...\theta'_{t+\tau}]'$ a collection of future trajectories for θ_t. Let $\mathcal{W} = (\Sigma_u, \sigma^2)$; set $\xi'_t = [u'_{1t}, u'_{2t}, \eta'_t]$, where u_{1t} are shocks to the endogenous variables and u_{2t} shocks to the exogenous variables (if there are any). Let $\xi_{j,t}^{\delta}$ be a realization of $\xi_{j,t}$ of size δ, and $\mathcal{F}_t^1 = \{y^t, \theta^t, \Omega^t, \mathcal{W}, J_t, \xi_{j,t}^{\delta}, \xi_{-j,t}, \xi_{t+1}^{t+\tau}\}$ and $\mathcal{F}_t^2 = \{y^t, \theta^t, \Omega^t, \mathcal{W}, J_t, \xi_t, \xi_{t+1}^{t+\tau}\}$ two conditioning sets, where $\xi_{-j,t}$ indicates all shocks, excluding the one in the j-th component, and $J_t J'_t = \Sigma_u$. Then, responses at horizon τ to a δ impulse in $\xi_{j,t}$, $j = 1, \ldots$ are

$$IR_y^j(t, \tau) = E(y_{t+\tau}|\mathcal{F}_t^1) - E(y_{t+\tau}|\mathcal{F}_t^2) \qquad \tau = 1, 2, \ldots \tag{36}$$

Notice that in (36), the history of the coefficient and of their variance is taken as given at each t and that the size of the impulse, can be positive or negative, but it is also taken as given. This is because one may be interested in comparing responses over time for a given trajectory of the coefficients and their variance (rather than their average values) and because the relevant size of the impulse is generally determined by policy considerations. When the coefficients are constant, $\xi'_t = [u'_{1t}, u'_{2t}]$ and (36) produces the traditional impulse response function to structural shocks.

A proper identification strategy, i.e., the selection of the (large) matrix J_t and its time evolution, is an open area for research since its sheer dimensionality makes it hard to find enough constraints to achieve identification of all shocks. Shortcuts, such as a block structure, may not be very appealing – one can envision situations where shocks are transmitted across unit within a time period. Alternative shortcuts, such as the one imposed in Canova et al. (2012), where shocks occurring in one unit (Germany) are allowed to feed contemporaneously on all other units (European countries) but not viceversa, may be acceptable if there are good economic reasons to justify them. In both cases, dynamics interdependencies are left unrestricted.

Which kind of restrictions should be used for identification is an open question. Zero restrictions and Choleski ordering for $J_t = J$, all t, are the most common ones and just identification is typically sought, even though overidentification requires only a simple extension of the tools described in Canova and Perez Forero (2012) for standard VARs. As in single unit VARs, one could also employ external information to identify shocks in different units. For example, one could measure the elasticity of tax revenues and government expenditure to output shocks in each cross sectional unit and use this information to identify government spending and tax revenue shocks in all units, adding the restriction that domestic government expenditure and domestic revenues do not instantaneously respond to shocks generated in other units. Long run restrictions a la Blanchard and Quah (1989) as well as heteroschedasticity restrictions a la Lanne and Lütkepohl (2010) are also possible. In both cases, one has to clearly state what happens to the variables of other units and often the restrictions needed to achieve identification are economically difficult to justify. For example, when long run restrictions are used, one has to impose that foreign supply shocks have no long run effect on domestic real variables or that they have the same effect as domestic supply shocks, both of which are not very palatable.

Sign restrictions can also be used (see e.g. Calza et al., 2012, and Sa et al., 2012). In this case, it is typical to use the same type of restrictions on each cross sectional unit. Recently, De Graeve and Karas (2011) have suggested using cross sectional heterogeneity to identify certain structural shocks. Their approach involves imposing sign and inequality restrictions on $\frac{\partial Y^{\Lambda}_{m,t+s}}{\partial u_{k,t}}$, the response of variable Y_m at horizon $t+s, s = 0, 1, 2, \ldots,$ to shock k at time t, $u_{k,t}$, for a subset of the units $\Lambda = \{1, \ldots, M\}$.

The approach is best understood with an example. Suppose the cross sectional dimension of the panel can be stratified according to an observable indicator, for example, in a sample of banks, whether bank deposits are insured or not. Suppose the endogenous variables are deposits D_t and the average interest rate they earn R_t. Then, a bank run is associated with a fall in deposits and an increase in the interest rate offered by uninsured banks. The insured banks may also respond, because of contagion effects, but the responses will be smaller because deposit insurance makes them less liable to the run. Thus, together with the sign restrictions, De Graeve and Karas impose that $|D_{t,u}| > |D_{t,I}|$ and $|R_{t,u}| > |R_{t,I}|$, where u stands for uninsured and I for insured banks. Hence, following a bank run, the deposit outflow in insured banks cannot be larger than in uninsured banks, and the corresponding increase in deposit rates must be smaller for insured than for uninsured banks.

Cross sectional identification restrictions of this type seem useful if one can sharply stratify the data with some exogenous indicator. For example, one could impose such restrictions to identify shocks originating in less developed countries (LDC), when the sample includes LDC and developed countries, once it is recognized that shocks originating in LDCs are unlikely to generate the same amount of volatility in the two groups. Alternatively, one could identify shocks primarily affecting small open economies integrated in the world economy. If the impact of particular shocks is strong in open economies but weak in relatively more closed economy or if more closed economies respond to the shocks only via second round effects, sign and differential magnitude restrictions can help to isolate them. Finally, an approach that combines sign and inequality restrictions can be used also to distinguish shocks taking place in units with slow versus fast adjustments or in markets affected differently by the presence of frictions.

The combination of sign and relative magnitude restrictions appears to be a very powerful identification device if the stratification employed is relevant. De Graeve and Karas show that in their sample of banks these restrictions allow them to identify a bank run shock which has characteristics that are similar to those obtained with narrative approaches and with ex-post insight. Note that the set of constraints one can impose is quite large, making the combination of sign and relative magnitude restrictions potentially usable in many situations. For example, one could impose relative magnitude restrictions on a particular variables across subsets Λ_1, Λ_2 of the units $\frac{\partial Y^{\Lambda_1}_{m,t+s}}{\partial u_{k,t}} \geq \frac{\partial Y^{\Lambda_2}_{m,t+s}}{\partial u_{k,t}}$, across variables within a particular subset of units, $\frac{\partial Y^{\Lambda}_{m_1,t+s}}{\partial u_{k,t}} \geq \frac{\partial Y^{\Lambda}_{m_2,t+s}}{\partial u_{k,t}}$, or across variables and across subset of units $\frac{\partial Y^{\Lambda_1}_{m_1,t+s}}{\partial u_{k,t}} \geq \frac{\partial Y^{\Lambda_2}_{m_1,t+s}}{\partial u_{k,t}}$. Clearly, which one is used depends on the question and the available data. Theory driven restrictions are clearly preferable, but restrictions obtained from reliable stylized facts characterizing different groups of units can also be used.

While the identification restrictions of De Graeve and Karas involve only the contemporaneous effect of shocks, one could also consider dynamic restrictions involving relative shape and/or relative magnitudes to identify shocks in a panel VAR. Future investigations need to clarify what kind of dynamic restrictions are consistent with economic theory and could be meaningfully employed to recover interesting shocks.

6 A comparison with alternative approaches

As we have emphasized, panel VARs are unique in their ability to model dynamic interdependencies, cross sectional heterogeneities and, at the same time, account for evolving pattern of transmission. However, to estimate them restrictions need to be imposed. Thus, it is natural to ask how panel VAR models compare to other models, which still allow us to study interdependencies and the transmission of shocks across units but impose alternative restrictions on the nature of the interdependencies present in the data.

This section sketches the main features of large scale Bayesian VARs (e.g. Banbura et al, 2010), spatial econometric models (see Anselin, 2010), factor models (see e.g. Stock and Watson, 1989, 2003), global VARs (see Dees et al., 2007 and Pesaran, et al., 2004) and bilateral VARs (see e.g. Kim, 2001, or Peersman, 2004), and highlights their differences with panel VARs.

6.1 Large Scale VARs

A close cousin of panel Bayesian VARs is the large scale Bayesian VAR model suggested by Mol et al. (2008) and recently employed by Banbura et al. (2010). As in panel VARs, both static and dynamic interdependencies are allowed for, but the researcher gives no consideration to the existence of a panel dimension in the data. Thus, all variables are treated symmetrically, regardless of whether they belong to a unit or not, and of whether they measure the same quantity in different units or not. Given the large scale of the model, classical estimation methods are infeasible, especially if time varying features are allowed for. The lack of a panel perspective is reflected in the type of priors imposed in Bayesian estimation, which are typically of Litterman-Minnesota type (see Doan, et al., 1984), and do not exploit any cross sectional information present in the data.

Failure to recognize that there is a cross sectional dimension to the available data set may not be too damaging in terms of forecasting, since it is well known that dimensionality shrinkage is more important than the exact details on how it is implemented – the Litterman-Minnesota prior is indeed a shrinkage prior. However, the choice of priors may limit the type of analyses one can perform, since the covariance matrix of the error has a particular structure which is generally disregarded in the policy exercises.

As in panel VARs, time variation in the coefficients of a large scale VARs are relatively easy to allow for (see e.g. Koop and Korobilis, 2011). However, ingenuity needs to be used since unrestricted time variation on all coefficients is impossible to estimate. Thus a factor structure, like the one described in equation (31), may be necessary to make the estimation problem manageable, and simple processes for time variation need to be imposed for computational ease.

6.2 Spatial VARs

Large scale VARs are also popular in classical econometric frameworks. Here, dimensionality restrictions are directly imposed to make estimation feasible. Restrictions of this type are commonly employed in the spatial econometric literature (see Anselin, 2010). Assuming, for simplicity only one lag of the dependent variables and no deterministic components, a spatial VAR has the form

$$Y_t = \rho S_1 Y_{t-1} + u_t \tag{37}$$

$$u_t = S_2 e_t \tag{38}$$

where S_1 and S_2 are fixed matrix of weights. For example, a typical structure for S_1 is

$$S_1 = \begin{bmatrix} s_{11} & s_{12} & 0 & 0 & \dots & 0 & 0 & \\ s_{21} & s_{22} & s_{23} & 0 & \dots & 0 & 0 & \\ 0 & s_{32} & \sigma_{33} & s_{34} & \dots & 0 & 0 & \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \\ 0 & 0 & 0 & 0 & \dots & s_{N-1,N-2} & s_{N-1,N-1} & s_{N-1,N} \\ 0 & 0 & 0 & 0 & \dots & 0 & s_{N,N-1} & s_{N,N} \end{bmatrix}$$

Here only the neighbors will have dynamic repercussion on unit i within one period while the rest is assumed to have a negligible effects. This structure implies that a shock originating in unit i can be transmitted after one period to unit j if j is a neighbor of i. However, if j is not a neighbor of i, delayed effects are longer and will depends on how many neighbors are between unit j and unit i. The idea of restricting dynamic effects to neighbors has been implemented using, for example, regions which share a border or stores which are located in the same city. There are three main disadvantages to this procedure. First, the weights have to be chosen prior to the estimation and, depending on the focus of the analysis, different weights may be used by different researchers on the same data set. Second, the setup is difficult to entertain in analyzing, for example, countries in a monetary union since generalized feedbacks are possible, or situations where borders (national, regional, etc.) do not reflect the economic separation present across units. Third, while the approach is relatively easy to implement when y_{it} is a scalar, it is much more complicated when y_{it} is vector since different elements of y_{it} may have different relationships across units. For example, when y_{it} includes output and consumption and the units are countries, the neighbor scheme may be appropriate for output, if the units are grouped using their natural resources, but may be highly inappropriate for consumption if migrations to all units are important.

Recently, Chudik and Pesaran (2011) have extended this framework to allow S_1 and S_2 to be full matrices and have derived classical estimators under the assumption that there is weak or strong cross sectional dependence across units. The basic idea of their approach is still to distinguish between neighbors and non-neighbors, where the former have non-negligible static and dynamic effect on unit i and the latter have negligible effects. However, even if the non-neighbor effects may be individually small, the sum of their absolute values may not be small, making aggregate feedback effects potentially large. To account for this possibility, the authors assume that the dynamic feedbacks produced by neighbors are important and independent of N, while the effects of the non-neighbors depend on N. As N increases, if there is weak cross correlation between a unit and the non-neighbors, the model approaches a spatial VAR where the non-neighbor effects are neither interesting nor estimable while the neighboring effects can be consistently estimated simply ignoring the non-neighbor feedbacks. On the other hand, when there is strong cross correlation between a unit and some non-neighbor, the structure approaches a factor model where, controlling for neighboring effects, one unit drives fluctuations in all the others. Here ignoring the feedbacks due to the factor may lead to inconsistent estimates of the parameters.

The main value of this setup is to provide a link between classical parameter shrinkage, as implied by spatial VAR models, and classical data shrinkage, as implied by factor models, both of which attempt to mitigate the curse of dimensionality present in large scale VARs. The approach also provides a justification for using the Global VAR approach described later. The main disadvantages of the procedure are similar to those of spatial VARs, namely that (i) neighbors and non-neighbors need to be chosen a priori; (ii) the approach is difficult to implement if y_{it} is a vector; and (iii) it is hard a-priori to know whether weak or strong cross sectional dependence characterizes the units under consideration.

6.3 Dynamic Factor models

Unobservable factor models are popular in the applied macroeconometric literature because they capture the idea that the comovements present in a large set of series may be driven by a small number of latent variables, and because they are relatively easy to specify and estimate (Forni et al., 2000 and 2005; Stock and Watson, 2002).

Factor models have another appealing feature: their format is consistent with economic theory. To see why, consider a prototypical factor model:

$$\begin{aligned} Y_t &= \chi f_t + u_t \\ A(\ell) f_t &= e_t \\ U(\ell) u_t &= v_t \end{aligned} \tag{39}$$

where Y_t is a $nG \times 1$ vector, f_t is a $m < nG \times 1$ vector of factors. The log-linearized solution of a DSGE model is

$$\begin{aligned} X_{2t} &= AX_{1t} + Bv_t \\ X_{1t} &= CX_{1t-1} + Du_t \end{aligned} \tag{40}$$

For example, in the case of an international RBC model, $X_{1t} = [k_{it}, \zeta_{it}]$, k_{it} is the capital stock, ζ_{it} the technology shock in country i and X_{2t} a vector including consumption, investment, output, hours, etc. in each country. Here $u_t = [u_{1t}, ...u_{nt}]$ is the vector of innovations in the technology process, and $U(\ell) = I$. Simple inspection indicates that the decision rules in (40) have a factor structure and this similarity allows statistical and economic analyses to be better linked.

When nG is small, one can use an EM algorithm or the Gibbs sampler to estimate the factors and free parameters of the model, if v_t and e_t are normally distributed. If the nG instead is large, averaging will insure that the idiosyncratic component u_t will cancel out. In this situation, one can use $Y_t = \chi f_t + u_t \approx \chi \bar{Y}_t + u_t$, where $\bar{Y}_t$ is a $m \times 1$ vector estimated averaging the variables in Y_t (see e.g. Forni and Reichlin, 1998).

Multi-unit (large scale) dynamic factor models and panel VARs differ in a number of dimensions. In terms of specification, the complex structure of dynamic interdependencies is not modelled in the former and is instead captured with a set of unobservable factors. Furthermore, the presence of cross sectional information is generally ignored.

The interpretation of impulse responses is usually much easier in a panel VAR than in factor models. In fact, in factor models dynamic analyses are typically performed by shocking the factors and seeing how impulses in these factors are transmitted to the endogenous variables. Thus, the structure does not allow us to study, say, how a shock generated in one unit is propagated to other units or identify the interdependencies that make an effect large or small. On the other hand, by appropriately selecting a combination of shocks, a panel VAR can mimic the exercises typically performed in factor models.

Finally note, that once the parameter dimensionality reduction described in sections 3 and 4 is performed, a panel VAR with interdependencies can be written in a factor format. Still, two important differences arise. First, the regressors in (21) are combinations of the lags of the right hand side variables of the panel VAR and thus observable. In factor models, factors are unobservable and typically estimated using averages of (subsets) of the current values of endogenous variables. Hence, they are likely to have different characteristics and span a different informational space. Whether lags or current values of the endogenous variables provide superior information for the states of a theoretical model is an open question which deserves further investigation. Second, the regressors of (21) equally weight the information present in the subset of the variables used to construct them. The equal weighting scheme comes directly from (20) and the fact that all variables are measured in the same units (all variables will be demeaned and standardized). In factor models, instead, estimates of the factors reflect the relative variability of the variables used to construct them.

6.4 Global VARs

Global VARs (GVARs) are similar in spirit to factor models and Dees et al. (2007) showed how they can be obtained when the DGP is a factor model with observable and unobservable factors. They are appealing to users because they intuitively capture important features of the cross sectional relationships while maintaining a simple structure that makes estimation easy.

For our purposes, a GVAR can be thought as a collection of unit specific VARs to which one tags on an unobservable common factor. This idea is explicitly put forward in Gilhooly et al. (2012), who estimate with Bayesian techniques a collection of unit specific VARs and assume a common factor structure in the error term. Consider the structure in (26) where now the coefficients are time invariant, only lags of the variables for that particular unit appear and add a new vector of unobservable variables x_t:

$$y_{it} = A_i(\ell)y_{it-1} + F_i(\ell)W_t + H_i(\ell)x_t + e_{it} \tag{41}$$

x_t is, potentially, a vector of autoregressive processes with finite variance and $H_i(\ell)$ is a polynomial in the lag operator. Set for simplicity $W_t = 0, \forall t$. Then (41) is a collection of unit specific VARs linked together by the presence of the unobservable vector of factors. Their presence complicates estimation since Kalman filter techniques need to be used, and unless the cross sectional dimension is small, computations may be demanding.

As in large scale factor models, the basic idea of GVARs is that, if N is sufficiently large, one can proxy x_t with cross unit averages of y_{it} (and W_t, when they are present). Thus, rather than estimating (41) one estimates

$$y_{it} = A_i(\ell)y_{it-1} + H_i(\ell)y_{it}^* + e_{it} \tag{42}$$

where $y_{it}^* = \sum_{j=1}^N \pi_{ij}y_{jt}$ with $\pi_{ii} = 0$ and π_{ij} is a set of country specific weights which reflect the relative importance of the unit in the aggregate. For example, if the units are countries, one does not expect them to be equally important in the world economy and may want to weight country specific variables by their share in world trade. Alternatively, they could capture relative variability if, for example, the cross section contains units featuring cyclical fluctuations with different amplitudes.

A model like (42) can be estimated in two steps. First, country specific VARs are estimated and all endogenous variables of the model collected. Second, the vector $Y_t = [Y_{1t}, \ldots, Y_{NT}]$, where each $Y_{it} = [y_{it}, y_{1t}^*]$ is simultaneously solved from the model. Pesaran et al (2007) show that this is equivalent to specifying a large scale VAR of the form

$$Y_t = D(\ell)Y_{t-1} + u_t \tag{43}$$

where $I - D(\ell)$ is $NG \times 1$ matrix whose typical i element is $1 - D_i(\ell)\pi_i$.

Hence, a Global VAR is a restricted large scale VAR, where variables of different units in an equation are weighted according to π_i. Since the weights are country specific and a-priori determined by the investigator, a GVAR imposes a particular structure on the interdependencies present in the data. In particular, it selectively chooses what feedbacks may be a-priori important based, for example, on trade or financial considerations, and forces the same dynamics on the variables belonging to all units, apart from a scale factor. In this sense, it resembles an extreme version of a Minnesota prior in that variables of the units different from the one appearing on the left hand side of the relationship have weights which are smaller than their own.

To state the concept differently, a GVAR becomes estimable imposing the restriction that the dynamics produced by variables of different units on the variables of unit i are proportional to the weights. This effectively collapses the number of estimated coefficients to a more manageable number, comparable to those one would estimate using a collection of single unit VARs.

6.5 Bilateral VARs

It is common to run bilateral VARs with units representing countries, sectors or disaggregated components of important macro variables, even if the DGP is suspected to be much more complicated (see, e.g., Kim, 2001, or Peersman, 2004). One reason for doing so is ease of interpretation; another is to reduce the size of the parameter vector to be estimated. However, such an approach may distort both the properties of the estimated structural shocks and the dynamics of their transmission.

To see how this can happen, consider the following three unit structural panel VAR(1)

$$y_{1t} = A_{11}y_{1t-1} + A_{12}y_{2t-1} + A_{13}y_{3t-1} + J_{11}u_{1t} \quad (44)$$

$$y_{2t} = A_{21}y_{1t-1} + A_{22}y_{2t-1} + A_{23}y_{3t-1} + J_{21}u_{1t} + J_{22}u_{2t} + J_{23}u_{3t} \quad (45)$$

$$y_{3t} = A_{31}y_{1t-1} + A_{32}y_{2t-1} + A_{33}y_{3t-1} + J_{31}u_{1t} + J_{32}u_{2t} + J_{33}u_{3t} \quad (46)$$

where y_{it} is of dimension $G \times 1$ and all the roots of the A matrix are outside the unit circle. In this system the reduced form innovations in y_{1t} are proportional to u_{1t} while the reduced form innovations in y_{2t} and y_{3t} are linear combinations of the three structural shocks (u_{1t}, u_{2t}, u_{3t}).

Suppose that a researcher decides to use data from units 1 and 2 only to form a bilateral

VAR. Then the estimated model would be

$$y_{1t} = (A_{11} + A_{13}(I - A_{33}\ell)^{-1}A_{31})y_{1t-1} + (A_{12} + A_{13}(I - A_{33}\ell)^{-1}A_{32})y_{2t-1} + e_{1t} \quad (47)$$

$$y_{2t} = (A_{21} + A_{23}(I - A_{33}\ell)^{-1}A_{31})y_{1t-1} + (A_{22} + A_{23}(I - A_{33}\ell)^{-1}A_{32})y_{2t-1} + e_{2t} \quad (48)$$

where

$$e_{1t} = J_{11}u_{1t} + (I - A_{33}\ell)^{-1}(J_{31}u_{1t} + J_{32}u_{2t} + J_{33}u_{3t}) \quad (49)$$

$$e_{2t} = J_{21}u_{1t} + J_{22}u_{2t} + J_{23}u_{3t} + (I - A_{33}\ell)^{-1}(J_{31}u_{1t} + J_{32}u_{2t} + J_{33}u_{3t}) \quad (50)$$

The dynamic responses induced by reduced form shocks will be different in the two systems. In particular, in the estimated system, the true dynamics will be contaminated by the responses of the variables of unit 3 to the structural shocks. For example, the reduced form dynamics of unit 1 will be correctly captured only when A_{13} is zero – which requires unit 1 to be exogenous with respect to the system. However, even in this special case, structural dynamics will be incorrectly measured in the estimated system for three reasons. First, while in the true model the reduced form innovations to unit 1 are a scaled version of the true structural shocks, in the estimated system e_{1t} mixes structural shocks from different units and this is true even when the original system has only lagged interdependencies but no static interdependencies, i.e. $J = I$. In addition, while in the original system unit 1 was predetermined, in the estimated one it is not. Hence, while a particular Choleski block ordering would be able to recover the innovations to the first unit in the original system, the imposition of such a structure would induce important identification errors in the estimated system.

Second, even if the reduced form innovations in the original system were serially uncorrelated, serial correlation would appear in the estimated ones, since the marginalization due the elimination of unit 3 data creates moving average components in the reduced form errors of the estimated system. Thus, either the lag length of the estimated panel VAR is appropriately increased or the reduced form errors will be serially correlated. In other words, to approximate the original panel VAR(1) with a smaller number of units, we need either a panel VARMA or a VAR(p) with p generally large.

Third, idiosyncratic shocks in the original system may show up as common shocks in the estimated system. For example, when $J = I$ structural shocks to the variables of the third unit will show up in the estimated system as common shocks to units 1 and 2. Hence, one should be cautious in interpreting the empirical evidence in favour of common shocks in such systems.

It general, it seems a bad practice to circumvent the curse of dimensionality problem using bilateral systems when the data generating process may be more complex. Omitted variables create important distortions in the estimated structural shocks and hamper the ability of researchers to interpret the estimated dynamic responses.

6.6 Summary

All available approaches impose restrictions. The large scale Bayesian VARs and the Bayesian panel VARs leave the model unrestricted but employ a shrinkage prior to effectively reduce the dimensionality of the coefficient vector. The spatial econometric model, the factor model, the global VARs, and the bilateral VARs on the other hand, impose that the interdependencies can be captured either with a set of factors, or that there is only a very limited number of neighbor effects, or that the off-diagonal elements of the matrix $D(\ell)$ are proportional to the diagonal elements, or that the estimated system is of lower dimension than the DGP. All restrictions may be violated in practice and it is unclear which ones are preferable. In theory, prior restrictions are superior to dogmatic restrictions. An interesting question for future research is whether and in what way different sets of restrictions affect our ability to capture and interpret interdependencies economies with heterogenous features.

7 Conclusions

Over the last fifteen years, there has been considerable improvement and unification in the standards of data collection and substantial efforts to create detailed and comparable data (on banks, firms, industries) in various countries and regions of the world. This means that while empirical analyses were previously limited to a handful of developed countries, interdependencies were hardly explored, and cross-country comparisons very scant, now an important panel dimension is added to the exercises and studies analyzing differences between, say, emerging markets and developed economies, or open and relatively closed economies are more frequent.

Together with improvements in the data collection, there has been a gradual but steady increase in the interdependencies among regions, countries and sectors. This phenomenon is not only the object of academic studies. Terms like "global economies", "global interdependencies", "global transmission" have become part of everyday discussions in the popular press. Economies, regions or sectors can no longer be treated in isolation and spillovers are now prevalent. In this new global order, shocks are quickly propagated and contagion effects are important, but heterogeneities across sectors and countries remain. Asymmetries both in the pace and in the magnitude of the recovery from the 2008 recession and the increasing North-South divide caused by the prolonged European debt crisis are simple examples of these heterogeneities. Their origins is probably complex, but factors such as income, initial conditions, geography, trade and financial developments, institutions and culture are good suspects to consider.

Since the growth path, the dynamic responses to shocks and the transmission across sectors, markets or countries may differ substantially, it is inadvisable from an economic and from an econometric point of view to treat all units symmetrically or to disregard country specific features even within the EU or the Euro zone. The presence of dynamic heterogeneities suggests that there is ample room to study how shocks are transmitted across units; to characterize not only average effects but also cross sectional differences that help understand the potential sources of heterogeneities; to analyze how past tendencies have created the current status quo and how one should expect the current situation to evolve in the future; and to provide policymakers with evidence that can help to build alternative scenarios and formulate policy decisions.

Panel VARs seem particularly suited to addressing issues that are currently at the center stage of policy discussions as they are able to (i) capture both static and dynamic interdependencies, (ii) treat the links across units in an unrestricted fashion, (iii) easily incorporate time variation in the coefficients and in the variance of the shocks, and (iv) account for cross sectional dynamic heterogeneities. The recent boom in empirical analyses using panel VARs in macroeconomics, banking and finance, and international economics attests to this fact.

Panel VARs are built on the same logic as standard VARs but add a cross sectional dimension. The purpose of this chapter was to point out their distinctive features and their potential applications, describe how they are estimated and how shocks are typically identified, how one deals with structural time variation; what differences are there between the panel VAR models used in microeconomic and macroeconomic studies; how panel VARs and a collection of VARs compare and how panel VARs relate to other popular alternatives such as large scale VARs, Factor models or GVARs.

The large dimension of panel VARs typically makes the curse of dimensionality an issue especially when researchers are interested in examining the input-output links of a region, such as Latin America, or an area, such as the Euro area, where the time series dimension of the panel is short. We present a shrinkage approach which goes a long way toward dealing with dimensionality issues without compromising too much on the structure and on the ability to address interesting economic questions.

Many challenges remain and future work can be expected to improve on existing approaches, both in terms of estimation and inference. For example, Koop and Korobilis (2012) have suggested fast algorithms to estimate large scale time varying coefficients VARs but it is unclear whether these will work also in time varying coefficients panel VAR, where cross sectional shrinkage becomes important.

Similarly, it is still unclear how to expand the Markov switching methods of Sims and Zha (2006) to a panel framework, especially when transition probabilities differ across heterogeneous units. The properties of estimators used have not been evaluated in relevant economic situations and it is unclear whether tests for model selection or model validation are powerful or not.

When it comes to identification, except for De Graeve and Karas (2012), the techniques are the traditional ones used in VARs and no effort has been made to exploit the richness of the cross sectional information. Nor have there been efforts to directly link panel VARs to global DSGE models and to study whether they can be used as testing ground for different theories of transmission.

All in all, panel VARs have the potential to become as important as VARs to answer relevant economic questions that do not require the detailed specification of the structure of the economy.

References

[1] Anselin, L., 2010. Thirty Years of Spatial Econometrics. Papers in Regional Science, 89, 3-25.

[2] Arellano, M. and Bond, S., 1991. Some tests of Specification for Panel data: Monte Carlo evidence and an Application to Employment Equations, Review of Economic Studies, 58, 277-297.

[3] Auerbach, A. and Gorodnichenko, Y., 2011. Measuring Output responses to fiscal shocks. American Economic Journal: Public Policy, 4,1-27.

[4] Beetsma, R. and Giuliadori, M., 2011. The effects of government purchase shocks: review and estimates for the EU, Economic Journal, 121, F4-F32.

[5] Beveridge, S. and Nelson, C.1981. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. Journal of Monetary Economics, 7, 151-174

[6] Blanchard, O, J. and Quah, D., 1989. The dynamic effect of Aggregate demand and supply disturbances. American Economic Review, 79, 655-673.

[7] Boubtane, E., Coulibaly, D., Rault, C., 2010. Immigration, growth and unemployment: Panel VAR evidence from OECD countries, manuscript.

[8] Banbura, M., Giannone, D., and Reichlin, L., 2010. Large scale Bayesian VARs, Journal of Applied Econometrics, 25, 71-92.

[9] Caivano, M. , 2006. The transmission of shocks between the U.S. and the Euro area, Bank of Italy, manuscript.

[10] Calza, A., Monacelli, T. and Stracca, L., 2012. Housing finance and Monetary policy, forthcoming Journal of the European Economic Association.

[11] Canova, F., 2004. Testing for Convergence Clubs in Income per Capita: A Predictive Density Approach. International Economic Review, 45, 49-77.

[12] Canova, F., 2005. The Transmission of U.S. Shocks to Latin America, Journal of Applied Econometrics, 20, 2005, 229-251.

[13] Canova, F., 2007. Methods for Applied Macroeconomic Research, Princeton University Press, Princeton, NJ.

[14] Canova, F. and Ciccarelli, M., 2004. Forecasting and Turning Point Predictions in a Bayesian Panel VAR Model, Journal of Econometrics, 120, 2004, 327-359.

[15] Canova, F., and Ciccarelli, M., 2009. Estimating Multi-country VAR models. International Economic Review, 50, 929-961.

[16] Canova, F. and Ciccarelli, M., 2012. ClubMed? Cyclical fluctuations in the Mediterranean basin, Journal of International Economics, 88, 162-175.

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

[18] Canova, F., Ciccarelli, M., Ortega, E., 2012. Do institutional changes affect business cycles?, Journal of Economic Dynamics and Control, 36, 1520-1533.

[19] Canova, F., Gambetti, L., 2009. Structural Changes in the U.S. economy, Is there a role for Monetary Policy. Journal of Economic Dynamics and Control, 33, 477-490.

[20] Canova, F. and Marrinan, J., 1998. Sources and Propagation of International Output Cycles: Common Shocks or Transmission?. Journal of International Economics, 46, 133-166.

[21] Canova, F. and Pappa, P., 2004. Does it Cost to be Virtuous? The Macroeconomic Effects of Fiscal Constraints, in Clarida, R., Frankel, J., Giavazzi, F. and K. West (eds.)NBER International Seminar on Macroeconomics, National Bureau of Economic Research.

[22] Canova, F. and Pappa, P., 2007. Price Differentials in Monetary Unions: The Role of Fiscal Shocks (with E. Pappa). Economic Journal, 117, 2007, 713-737.

[23] Canova, F. and Pina, J., 2005. Monetary Policy Misspecification in VAR models, in Diebolt, C. and Kyrtsou, C. (eds.), *New Trends In Macroeconomic*, Springer Verlag.

[24] Canova, F. and Perez Forero, F., 2012. Estimating overidentified, non-recursive, time varying coefficients, structural VARs, UPF manuscript.

[25] Chari, V.V., Kehoe, P. and McGrattan, E., 2008. Are structural VAR with Long Run restrictions useful for developing Business Cycle Theory, Journal of Monetary Economics, 55,1337-1352.

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

[27] Chudik, A. and Pesaran, H., 2011. Infinite-dimensional VARs and factor models. Journal of Econometrics, 163, 4-22.

[28] Ciccarelli, M., E. Ortega and M.T. Valderrama, 2012a. Heterogeneity and cross-country spillovers in macroeconomic-financial linkages. ECB working paper.

[29] Ciccarelli, M., A. Maddaloni and J.L. Peydró, 2012b. Heterogeneous transmission mechanism: Monetary policy and financial fragility in the Euro area. Paper prepared for the 56th Economic Policy Panel.

[30] Ciccarelli, M. and A. Rebucci, 2006. Has the transmission mechanism of European monetary policy changed in the run-up to EMU? European Economic Review, 50, 737-776.

[31] Cogley, T., Sargent, T., 2005. Drifts and Volatilities: Monetary Policy and Outcomes in the post WWII U.S.. Review of Economic Dynamics, 8, 262-302.

[32] Cooley, T. and LeRoy, S., 1985. Atheoretical macroeconomics: A critique. Journal of Monetary Economics, 16, 283-308.

[33] Cooley, T. and Dwyer, M, 1998. Business Cycle Analysis without much Theory: A Look at Structural VARs. Journal of Econometrics, 83, 57-88.

[34] Cushman, D. and Zha, T. , 1997. Identifying monetary policy in a small open economy under flexible exchange rates. Journal of Monetary Economics, 39, 433-448.

[35] De Graeve, F and Karas, A., 2012. Identifying VARs through Heterogeneity: An Application to Bank runs, Riskbank manuscript.

[36] Dees, S, Di Mauro, F., Pesaran, H. and Smith, V., 2007. Exploring the International Linkages of the Euro area: A Global VAR analysis. Journal of Applied Econometrics, 22, 1-38.

[37] Del Negro, M., Otrok, C., 2003. Dynamic Factor Models with Time-Varying Parameters: Measuring Changes in International Business Cycles, Federal Reserve Bank of New York, manuscript.

[38] Del Negro, M. and Schorfheide, F., 2010. Bayesian Macroeconometrics, in J. Geweke, G. Koop and H. Van Dijk (eds.) The Oxford Handbook of Bayesian Econometrics.

[39] Doan, T., Litterman R., Sims, C., 1984. Forecasting and Conditional Projections Using Realist Prior Distributions. Econometric Review, 3, 1-100.

[40] Eun, C. and Shin, S., 1989. International transmission of stock market movements. Journal of Financial and Quantitative Analysis, 24, 241-256.

Faust, J. and Leeper, E., 1997. Do Long Run Restrictions Really Identify Anything? Journal of Business and Economic Statistics, 15, 345-353.

[41] Fernandez Villaverde, J., Rubio Ramirez, J, Sargent , T.and M. Watson, 2007. The A,B,C and D for understanding VARs, American Economic Review, 97, 1021-1026.

[42] Forni, M., Reichlin, L., 1998. Let's get Real: A factor Analytic Approach to Disaggregated Business Cycle Dynamics. Review of Economic Studies, 65, 453-473.

[43] Gertler, M., Comin, D., 2006. Medium Term Cycles. American Economic Review, 96, 523–551.

[44] Geweke, J., 2000. Simulation Based Bayesian Inference for Economic Time Series in Mariano, R., Schuermann, T. and Weeks, M. (eds.). Simulation Based Inference in Econometrics: Methods and Applications, Cambridge, UK: Cambridge University Press.

[45] Gilhooly, R., Weale M. and Wieladek T., 2012. Estimation of short dynamic panels in the presence of cross-sectional dependence and dynamic heterogeneity. Bank of England, External MPC Unit, Discussion Paper No. 38.

[46] Holtz-Eakin,D., Newey, W. and Rosen, H., 1998. Estimating Vector autoregressions with panel data. Econometrica, 56, 1371-1395.

[47] Jarocinski, M., 2010. Responses to monetary policy shocks in the east and the west of Europe: a comparison. Journal of Applied Econometrics, 25, 833-868.

[48] Kadiyala, K.R., Karlsson, S., 1997. Numerical Methods for Estimation and Inference in Bayesian-VAR Models. Journal of Applied Econometrics 12, 99–132.

[49] Kilian, L., 2012. Structural Vector Autoregressions. Forthcoming in: N. Haschimzade and M. Thornton (eds.), Handbook or Research Methods and Applications in Empirical Macroeconomics, Camberley, UK: Edward Elgar.

[50] Kim, S., 2001. International transmission of US monetary policy shocks. Evidence from VARs. Journal of Monetary Economics, 48, 339-372.

[51] Koop, G. and Korobilis, D., 2012. Large scale time varying VARs. University of Strathclyde, manuscript.

[52] Kose, A. Otrok, C. and Whiteman, C., 2003. International Business Cycles: World, Region, and Country-Specific Factors. American Economic Review, American Economic Association, 93, 1216-1239.

[53] Kose, A. and Prasad, E., 2010. Emerging markets: resilience and growth amid global turmoil. Brooking Institution Press, Washington, DC.

[54] Long J. and Plosser, C., 1983. Real Business cycles, Journal of Political Economy, 91, 39-69.

[55] Lane, P. and Benetrix, A., 2010. International differences in fiscal policy during the global crisis. CEPR working paper 8009.

[56] Lanne M., and Lütkepohl, H., 2008. Identifying Monetary Policy Shocks via Changes in Volatility, Journal of Money, Credit and Banking, 40, 1131-1149.

[57] Love, I., and Zicchino, L., 2006. Financial development and dynamic investment behavior: Evidence from a Panel VAR. The Quarterly Review of Economic and Finance, 46, 190-210.

[58] L. Ljung, and Soderstrom, T., 1984. Theory and Practice of Recursive Identification, MIT Press. Cambridge, Mass.

[59] Mol, C. D., Giannone, D., and Reichlin, L., 2008. Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? Journal of Econometrics, 146, 318-328.

[60] Ocampo, S and Rodriguez, N., 2011. An introductory review of a structural VARX estimation and applications. Borradores de Economia 686, Banco de la Republica, Colombia.

[61] Peersman, G., 2004. The transmission of monetary policy shocks in the Euro area. Are the effects different across countries? Oxford Bulletin of Economics and Statistics, 66, 285-308.

[62] Pesaran, H., Schuerman, T. , Wiener, S., 2004. Modelling regional interdependencies using a global error-correcting macroeconomic model. Journal of Business and Economic Statistics, 22, 129-162.

[63] Pesaran, H. and Smith, R., 1995. Estimating Long Run Relationship from Dynamic Heterogeneous Panels. Journal of Econometrics, 68, 79-113.

[64] Primiceri, G., 2005. Time Varying Structural Vector Autoregressions and Monetary Policy. The Review of Economic Studies, 72, 821-852.

[65] Ramey, V. and Shapiro, M., 1998. Costly capital reallocation and the effects of government spending. Carnegie-Rochester Conference Series on Public Policy, 48, 145-194

[66] Rebucci, A. 2010a. Is growth exogenous? Evidence from the 1970s and the 1980s. Applied Economics, 42, 535-543.

[67] Rebucci, A. 2010b. On the heterogeneity bias of pooled estimators in stationary panel VARs. Economic Modeling, 27, 1183–1198.

[68] Sa, F., P. Towbin, and T. Wieladek, 2012. Capital Inflows, Financial Structure and Housing booms, forthcoming, Journal of the European Economic Association.

[69] Sargent, T., 1989. Two models of Measurement and the Investment Accelerator. Journal of Political Economy, 97, 251-287.

[70] Sims, C., 2000. Using a likelihood perspective to sharpen econometrics discourse: Three examples. Journal of Econometrics, 95, 443-462.

[71] Sims, C. and Zha, T., 1998. Bayesian Methods for Multivariate Dynamic Models. International Economic Review, 39, 949-968.

[72] Sims, C. and Zha, T., 2006. Were There Regime Switches in U.S. Monetary Policy?. American Economic Review, 96, 54-81.

[73] Sims, E., 2012. News, Non-Invertibility, and Structural VARs. Forthcoming, Advanced in Econometrics volume 28.

[74] Stock, J., H. Watson, M. W., 1989. New Indices of Coincident and Leading Indicators, in Blanchard, O. and Fisher, S.(eds.), NBER Macroeconomic Annual, 4.

[75] Stock, J. H. and Watson, M. W., 2002a. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association, 97, 1167-1179.

[76] Stock, J. Watson, M., 2003. Understanding Changes in International Business Cycle Dynamics. Journal of European Economic Association, 3, 968-1006

[77] Stock, J., Watson, M., 2012. Disentangling the Channels of the 2007-2009 Recession. NBER working paper 18094.

[78] Swamy, V.,1970. Efficient Inference in random coefficient regression models, Econometrica, 38, 311-323.

[79] Zellner , A. and Hong, C., 1989. Forecasting international growth rates using Bayesian Shrinkage and other procedures. Journal of Econometrics, 40, 183-202.