16 Structural VAR Analysis in a Data-Rich Environment

Typical VAR models used for policy analysis include only small numbers of variables. One motivation for considering larger VAR models is that policy institutions such as central banks and government organizations consider large panels of time series variables in making policy decisions. If important variables are not included in a VAR model, that model becomes informationally deficient and estimates of the responses to policy shocks will be distorted by omitted-variable bias. Thus, unless a variable is known to be irrelevant, one should ideally include it in the structural VAR model.

Deciding on the relevance of a particular variable for an empirical model is a difficult task because the variables for which data are available may not correspond exactly to the variables used in theoretical models. For example, consider a monetary policy reaction function that includes inflation and the output gap as explanatory variables. It is well known that the output gap is difficult to measure. Hence, one can make the case for including all variables that contain information about the output gap because they could all be important for the analysis of the impact of monetary policy.

Another motivation for considering larger VAR models is that we may wish to examine the impact of monetary policy shocks at a more disaggregate level. For example, one may be interested not only in the response of the overall price level to a monetary policy shock, but also in the response of sub-indices corresponding to specific expenditure components. Such an analysis necessitates the inclusion of disaggregate price level data in the model. Likewise, one may be interested in the output response in specific sectors of the economy. In that case, again, these additional variables have to be included in the analysis.

Conventional unrestricted VAR models do not allow for the situations described above because the inclusion of many additional variables undermines the precision of the model estimates in small samples. Moreover, the extent to which a VAR model can be enlarged is limited by the fact that the

 https://doi.org/10.1017/9781108164818.017

number of regressors cannot exceed the number of observations. This restriction can easily become binding when working with large-dimensional VAR models because the number of parameters in a VAR model increases with the square of the number of variables included.

Thus, the analyst often faces a dilemma in setting up the model. Whereas degrees-of-freedom constraints suggest including only a small number of variables, concerns over omitted-variables bias make it desirable that we include a large number of variables in the model. In response to this problem, several techniques have been developed that make it possible to increase the information content of VAR models. Two such techniques are factor-augmented VAR (FAVAR) models and large-dimensionl Bayesian VAR models. Related methods include panel VAR models, global VAR (GVAR) models, and spatial VAR models.

In Section 16.1, we show how factor models may be used to condense the information in a large panel of variables to a small number of factors. We distinguish between static and dynamic factor models. Section 16.2 illustrates how factors may be incorporated into structural VAR analysis. We consider two classes of models. The class of FAVAR models refers to VAR models that include one or more factors in addition to observables. Identification of the structural shocks within this augmented VAR model may be achieved by conventional methods. An alternative is the class of dynamic factor models. Whereas FAVAR models may include both factors and observed variables, in structural dynamic factor models all observables are expressed as a weighted average of factors. The evolution of these factors depends on the structural shocks. The identification of these shocks is achieved by restricting the implied responses of the observed variables to the structural shocks. These models play an increasingly important role in applied work both in forecasting and in structural analysis (see, e.g., Bernanke, Boivin, and Eliasz 2005; Favero, Marcellino, and Neglia 2005; Del Negro and Otrok 2007).

Imposing Bayesian restrictions on the parameters of a structural VAR model is an alternative approach to dealing with many variables in VAR analysis. Large-scale BVAR models have gained popularity lately, in particular as forecasting models (see, e.g., Carriero, Kapetanios, and Marcellino 2009, 2012; and Koop 2013a). They have also been used for structural analysis, however (see, e.g., Bańbura, Giannone, and Reichlin 2010). In Section 16.3, problems specific to the analysis of large-dimensional structural BVARs are discussed.

Finally, some alternative approaches to fitting large-dimensional VAR models are considered in Section 16.4. In particular, we briefly discuss panel VAR models, GVAR models, and spatial VAR models. Section 16.5 contains a summary of the pros and cons of alternative modeling approaches for large panels of time series.

16.1 Factor Models

This section shows how common factors may be extracted from a large cross-section of time series data. FAVAR models rely on a small number of these factors to summarize the information contained in the original large-dimensional data set. The construction of structural FAVAR models and related models is discussed in Section 16.2.

In Section 16.1.1, we review static factor models. Static factor models were originally designed for cross-sectional analysis. In Section 16.1.2, we present the more general dynamic factor model (DFM) specifically designed for time series analysis. We present alternative versions of the DFM. The problem of determining the appropriate number of factors is reviewed in Section 16.1.3.

There are a number of good surveys on factor analysis for time series data (see, e.g., Stock and Watson 2005, 2016; Breitung and Eickmeier 2006; Bai and Ng 2008; Barhoumi, Darné, and Ferrara 2014). DFMs in particular have been used extensively for forecasting (e.g., Stock and Watson 2002a, 2006, 2011). Some of that literature is also relevant in the present context. Important results on statistical inference for DFMs are available in Forni, Hallin, Lippi, and Reichlin (2000, 2004), Breitung and Tenhofen (2011), Choi (2012), Stock and Watson (2002a, 2005), Bai (2003), and Gonçalves and Perron (2014), among others. A detailed survey of how to conduct inference in DFMs is provided in Bai and Ng (2008).

16.1.1 Static Factor Models

Model Setup. The classical static factor model assumes the form

$$x_t = \Lambda f_t + v_t, \tag{16.1.1}$$

where $x_t \overset{iid}{\sim} (0, \Sigma_x)$ is a vector of N observed variables, f_t is an r-dimensional vector of unobserved common factors, and r is typically much smaller than N, $r << N$. Accordingly Λ is a $N \times r$ matrix of factor loadings. Finally, $v_t \overset{iid}{\sim} (0, \Sigma_v)$ is an N-dimensional vector of uncorrelated idiosyncratic components such that Σ_v is diagonal. Moreover, the common factors and idiosyncratic components are assumed to be orthogonal, i.e., $\mathbb{E}(f_t v_s') = 0$ for all s and t. Hence,

$$\Sigma_x = \Lambda \Sigma_f \Lambda' + \Sigma_v, \tag{16.1.2}$$

where $\Sigma_f = \mathbb{E}(f_t f_t')$ is the covariance matrix of the factors. If the factors are mutually uncorrelated such that Σ_f is diagonal, the factors are said to be orthogonal. Otherwise they are oblique. This basic model has been in use in statistical analysis for many decades. For a detailed review see Anderson

(2003), for example, who traces such models back to Spearman (1904). Notice that in the basic model (16.1.1) the observed variables are assumed to have mean zero. Thus, in practice, the data have to be transformed prior to the analysis.

Obviously, in model (16.1.1) the factors and factor loadings are not separately identified. For any nonsingular $r \times r$ matrix Q, defining $f_t^* = Qf_t$ and $\Lambda^* = \Lambda Q^{-1}$, implies $\Lambda f_t = \Lambda^* f_t^*$. In practice, it is common to choose the factor loading matrix such that it has orthonormal columns, implying that

$$\Lambda'\Lambda = I_r, \tag{16.1.3}$$

or to choose uncorrelated factors with variances normalized to 1,

$$f_t \sim (0, I_r). \tag{16.1.4}$$

In the latter case, the factors are orthogonal and

$$\Sigma_x = \Lambda\Lambda' + \Sigma_v.$$

Such normalizations are useful for developing estimation algorithms. They are not sufficient for uniquely identifying the model, however. For instance, if we normalize the factors as in (16.1.4), Λ is still not unique without further restrictions. This can be seen by choosing an orthogonal matrix Q and defining $\Lambda^* = \Lambda Q$. Thereby we arrive at the decomposition

$$\Sigma_x = \Lambda^*\Lambda^{*\prime} + \Sigma_v.$$

Exact identification of the static factor model can be ensured by choosing Λ such that expression (16.1.3) holds and choosing the factors such that Σ_f is a diagonal matrix with distinct diagonal elements in decreasing order. In other words, the first factor has the largest variance and, hence, explains the largest part of the variance of x_t among the common factors. The second factor, f_{2t}, has the second largest variance, etc. The requirement that the factor variances have to be distinct means that the columns of Λ cannot simply be reordered. Table 16.1.1 summarizes four alternative sets of identification conditions for factors and factor loadings (see Bai and Ng 2013). It should be noted that even when these conditions are satisfied, the Λ matrix is unique only up to the sign of its columns. For a thorough discussion of alternative identification conditions, see Anderson (2003, section 14.2.2). If the model parameters are identified, their estimation is straightforward.

Estimating Static Factor Models. If the factor loadings were known and normalized such that $\Lambda'\Lambda = I_r$, a natural estimator for the factors would be obtained by left-multiplying (16.1.1) with Λ' and dropping the idiosyncratic term,

$$\hat{f}_t = \Lambda' x_t. \tag{16.1.5}$$

Table 16.1. *Identification Conditions for Factors and Factor Loadings*

Restrictions for Λ	Restrictions for Σ_f
$\Lambda'\Lambda = I_r$	Σ_f diagonal with decreasing diagonal elements
$\Lambda'\Lambda$ diagonal with distinct, decreasing diagonal elements	$\Sigma_f = I_r$
$\Lambda = \begin{bmatrix} \lambda_{11} & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{r1} & \lambda_{r2} & \cdots & \lambda_{rr} \\ \vdots & \vdots & & \vdots \\ \lambda_{N1} & \lambda_{N2} & \cdots & \lambda_{Nr} \end{bmatrix}$ $\lambda_{ii} \neq 0, i = 1, \ldots, r$	$\Sigma_f = I_r$
$\Lambda = \begin{bmatrix} I_r \\ \Lambda_2 \end{bmatrix}$	Σ_f unrestricted

In practice, the factor loadings are typically unknown. A possible objective function for estimation in that case is the sum of the squared idiosyncratic errors. Minimizing the variance of the idiosyncratic components amounts to maximizing the part of the variance of the observed variables explained by the common factors. In other words, we may estimate the factor loadings and factors by minimizing the sum of squared errors,

$$\min_{\Lambda, f_1, \ldots, f_T} T^{-1} \sum_{t=1}^{T} (x_t - \Lambda f_t)'(x_t - \Lambda f_t)$$
$$= \min_{\Lambda, f_1, \ldots, f_T} \operatorname{tr}\left(T^{-1} \sum_{t=1}^{T} (x_t - \Lambda f_t)(x_t - \Lambda f_t)' \right). \tag{16.1.6}$$

A solution to this minimization problem is obtained in three steps:

1. Find the r largest eigenvalues $\lambda_1 > \cdots > \lambda_r$ of $S_x = T^{-1} \sum_{t=1}^{T} x_t x_t'$ and the corresponding orthonormal eigenvectors $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r$.
2. Choose $\widehat{\Lambda} = [\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r]$.
3. Express the factor estimate as $\hat{f}_t = \widehat{\Lambda}' x_t$.

The estimator $\widehat{\Lambda}$ is the so-called principal components (PC) estimator of Λ. Given the orthogonality of the eigenvectors, it satisfies $\widehat{\Lambda}'\widehat{\Lambda} = I_r$. The factors are the principal components and $\widehat{\Sigma}_f = T^{-1} \sum_{t=1}^{T} \hat{f}_t \hat{f}_t' = \widehat{\Lambda}' S_x \widehat{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_r)$. The eigenvalues $\lambda_1, \ldots, \lambda_r$ are the empirical variances of the factors with λ_1 representing the variance of the principal component with the largest contribution to the variance of the data, λ_2 represents the variance of the second-most important component, etc.

The asymptotic properties of estimators of static factor models for $T \to \infty$ and fixed N can be found in Anderson (2003, chapter 14). Results for more general factor models under the assumption that both the number of variables N and the sample size T go to infinity are derived in Stock and Watson (2002a) and Bai (2003), among others. In particular, these authors establish the consistency of $\widehat{\Lambda}$ and its asymptotic normality, provided N and T go to infinity at suitable rates and some further regularity conditions hold (see also Bai and Ng 2008).

The PC estimator is the ML estimator if the observations x_t come from a normal distribution and the idiosyncratic components have equal variances such that $\Sigma_v = \sigma^2 I_N$. In that case, the factors and idiosyncratic components are normally distributed according to $f_t \overset{iid}{\sim} \mathcal{N}(0, \Sigma_f)$ and $v_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2 I_N)$ (see Anderson 2003). If the variances of the idiosyncratic components are heterogeneous, i.e., $\Sigma_v = \operatorname{diag}(\sigma_1^2, \dots, \sigma_N^2) \neq \sigma^2 I_K$, the log-likelihood is

$$\log l(\Lambda, f_1, \dots, f_T, \Sigma_v) = \text{constant} - \frac{T}{2} \log(\det(\Sigma_v)) \\ - \frac{1}{2} \operatorname{tr}\left(\sum_{t=1}^{T} (x_t - \Lambda f_t)(x_t - \Lambda f_t)' \Sigma_v^{-1} \right).$$

Anderson (2003, section 14.4) points out that this likelihood function is unbounded in general and, hence, does not have a global maximum. Thus, standard ML estimation cannot be used. The likelihood function has local maxima, however, allowing us to consider local maxima in the neighbourhood of the true parameter vector (e.g., Breitung and Tenhofen 2011; Bai and Li 2012).

If an estimator $\widetilde{\Sigma}_v$ of Σ_v is available, the factor loadings and factors may be estimated by a feasible GLS method based on the minimization problem

$$\min_{\Lambda, f_1, \dots, f_T} T^{-1} \sum_{t=1}^{T} (x_t - \Lambda f_t)' \widetilde{\Sigma}_v^{-1} (x_t - \Lambda f_t).$$

Feasible GLS estimators for factor models and their asymptotic properties are discussed in Choi (2012).

If the normalization in expression (16.1.4) is used for the common factors and if the observations are normally distributed, ML estimation of the factor loadings and idiosyncratic variances involves maximizing the log-likelihood

$$\log l(\Lambda, \Sigma_v) = \text{constant} - \frac{T}{2} \log(\det(\Sigma_x)) - \frac{1}{2} \operatorname{tr}(T S_x \Sigma_x^{-1}) \\ = \text{constant} - \frac{T}{2} \log(\det(\Lambda \Lambda' + \Sigma_v)) - \frac{1}{2} \operatorname{tr}\left[T S_x (\Lambda \Lambda' + \Sigma_v)^{-1} \right]$$

by numerical methods. Suitable algorithms are discussed, for instance, by Magnus and Neudecker (1988, chapter 17).

Approximate Static Factor Models. So far we have considered what might be called an exact static factor model in which the idiosyncratic components are clearly separated from each other and from the factors. For economic data such an assumption may be too strict. An alternative approach is to work with an approximate factor model that allows for the possibility that the common factors specified by the researcher do not include all of the common factors in the data. In other words, some common factors will be contained in what the static factor model labels as the idiosyncratic component. In the approximate factor model

$$\Sigma_x = \Lambda \Sigma_f \Lambda' + \Sigma_v,$$

where Σ_v is not necessarily a diagonal matrix. Assuming that the common factors are normalized to have variance one, Chamberlain and Rothschild (1983) define an approximate factor model by the condition that Σ_x has only r unbounded eigenvalues when $N \to \infty$, where N is the number of variables considered by the researcher. The common factors are defined by the requirement that there exists a sequence of $N \times r$ matrices Λ and positive definite covariances Σ_v such that

$$\Sigma_x = \Lambda \Lambda' + \Sigma_v$$

and the maximum eigenvalue of Σ_v is bounded when $N \to \infty$. Thus, the relative variance share of each idiosyncratic component is small, when the number of variables is large. Chamberlain and Rothschild (1983) apply this model to a financial market with many assets (see also Connor and Korajczyk 1986, 1993).

Obviously, in that case identification of the model becomes more difficult and conditions different from those stated earlier are required. In fact, it is even possible that Σ_v has a factor decomposition that needs to be clearly separated from the common factor part captured by $\Lambda \Lambda'$, at least asymptotically. Choi (2012) considers estimation of models of that type and provides general asymptotic results.

16.1.2 Dynamic Factor Models

Taking into account the serial dependence in the data is essential for forecasting and for structural analysis. As we have seen earlier, static factor models and approximate static factor models ignore the time series dependence of the data in extracting the common factors. Dynamic factor models are obtained by allowing f_t and v_t in (16.1.1) to be serially correlated or more generally dependent stochastic processes. Which model is obtained depends on the assumptions made about the stochastic processes f_t and v_t. A number of special cases may be considered.

For example, if v_t is white noise, x_t inherits all its serial dependence from the common factors. An early example of such a model for time series data is considered by Peña and Box (1987) who postulate that the factors have a VARMA-DGP and that the Σ_v matrix is not necessarily diagonal. Inference may be conducted as in the static factor model (e.g., Choi 2012). This case is therefore not considered here. From a practical point of view such models are typically too restrictive.

If f_t and v_t have parametric VAR representations, model (16.1.1) may be viewed as a dynamic factor model (DFM). Some authors, including Boivin and Ng (2005), refer to this model as a static factor model to distinguish it from a model where lagged factors f_{t-j} appear on the right-hand side of (16.1.1) in addition to the contemporaneous factors. We do not use this terminology here because it can be shown that every dynamic factor models in the sense of Boivin and Ng (2005) can also be expressed in static form.

Finally, when the common component and the idiosyncratic components are general stochastic processes, we follow part of the literature in calling this model a generalized dynamic factor model (GDFM). We caution the reader, however, that it is important to check which assumptions are made and which terminology is used in each case, when reading the original literature.

The remainder of this section is structured as follows. DFMs may be written equivalently in static form or in dynamic form. First, the static form of the DFM is presented and its estimation is discussed. Then the dynamic form of the DFM is considered. Finally the GDFM is presented.

Throughout this section, we presume that the variables have been transformed to be stationary. It should be noted that factor models can be adapted to allow for integrated variables, but the presence of stochastic trends means that the methods described in this chapter would have to be modified. For example, the estimation procedures discussed in this chapter are based on covariance matrix and spectral density estimators that are not meaningful when dealing with integrated variables. Extensions of factor models to allow for integrated variables can be found in Bai (2004), for example. In fact, there is a close relationship between factor models and the cointegrated VAR models discussed in Chapter 3 in that cointegrated variables share a common trend that may be viewed as a common factor. A detailed analysis of the DFM for integrated and possibly cointegrated processes can be found in Barigozzi, Lippi, and Luciani (2016).

Static Form of the DFM. Consider the model (16.1.1),

$$x_t = \Lambda^f f_t + v_t, \tag{16.1.7}$$

with dynamic factors being generated as

$$f_t = \Gamma_1 f_{t-1} + \cdots + \Gamma_s f_{t-s} + \eta_t$$

and idiosyncratic components being generated as

$$v_t = A_1 v_{t-1} + \cdots + A_p v_{t-p} + u_t,$$

where the A_i, $i = 1, \ldots, p$, are diagonal matrices and u_t is white noise with diagonal covariance matrix Σ_u. Using lag operator notation,

$$\Gamma(L) f_t = \eta_t \quad \text{and} \quad A(L) v_t = u_t,$$

where $A(L) = \text{diag}[a_1(L), \ldots, a_N(L)]$. This model is called the static form of the DFM because the relation between the observed x_t and the dynamic factors is contemporaneous. No lagged f_t appear in equation (16.1.7).

Estimation. Estimation of the factors and factor loadings in model (16.1.7) for a given number of factors, r, can be carried out by PC estimation, ignoring any serial dependence in the error terms. Bai (2003) derives the properties of this estimator. PC estimation is generally inefficient because the dependence structure of the errors is ignored. Choi (2012) develops a GLS estimation procedure that can accommodate heteroskedastic idiosyncratic components, and Breitung and Tenhofen (2011) propose a GLS estimation procedure that can deal with a more general dependence structure in the error terms. This procedure applies even if the model is just an approximate factor model. Bai and Li (2012) discuss ML estimation of such models.

Dynamic Form of the DFM. A more general formulation of a DFM is obtained if the factors are also allowed to enter in lagged form such that

$$x_t = \Lambda_0^f f_t + \Lambda_1^f f_{t-1} + \cdots + \Lambda_{q^*}^f f_{t-q^*} + v_t. \tag{16.1.8}$$

Assuming the same DGP for f_t and v_t as in the static form (16.1.7) of the DFM, this model can be written in lag operator notation as

$$x_t = \Lambda^f(L) f_t + v_t, \quad \Gamma(L) f_t = \eta_t, \quad A(L) v_t = u_t,$$

where

$$\Lambda^f(L) = \Lambda_0^f + \Lambda_1^f L + \cdots + \Lambda_{q^*}^f L^{q^*},$$

$$\Gamma(L) = I_r - \Gamma_1 L - \cdots - \Gamma_s L^s,$$

$$A(L) = \text{diag}[a_1(L), \ldots, a_N(L)],$$

$f_t = (f_{1t}, \ldots, f_{rt})'$ are the common factors, $v_t = (v_{1t}, \ldots, v_{Nt})'$ is the vector of idiosyncratic components, and η_t is white noise such that $\mathbb{E}(u_t \eta_s') = 0 \; \forall \, t, s$.

Defining $F_t = (f_t', \dots, f_{t-q^*}')'$ and $\Lambda^F = [\Lambda_0^f, \Lambda_1^f, \dots, \Lambda_{q^*}^f]$, model (16.1.8) can equivalently be written in static form as

$$x_t = \Lambda^F F_t + v_t,$$

the only difference being that the dimension of the factor vector is larger. The vector F_t is often referred to as the vector of static factors, whereas the corresponding shorter vector f_t is called the vector of primitive dynamic factors (or, simply, the vector of dynamic factors).

In order to estimate the DFM (16.1.8), it is useful to left-multiply model (16.1.8) by $A(L)$ which yields

$$A(L)x_t = \Lambda(L)f_t + u_t, \tag{16.1.9}$$

where $\Lambda(L) = A(L)\Lambda^f(L)$ is a matrix polynomial of order $q \leq pq^*$. Assuming without loss of generality that $q \geq s$, the model (16.1.9) can be written in static form as

$$A(L)x_t = \Lambda F_t + u_t, \qquad F_t = \mathbf{\Gamma} F_{t-1} + G\eta_t, \tag{16.1.10}$$

where, using similar notation as before, $F_t = (f_t', \dots, f_{t-q}')'$, $\Lambda = [\Lambda_0, \Lambda_1, \dots, \Lambda_q]$, and

$$\mathbf{\Gamma} = \underbrace{\begin{bmatrix} \Gamma_1 & \Gamma_2 & \cdots & \Gamma_q & \Gamma_{q+1} \\ I_r & 0 & \cdots & 0 & 0 \\ 0 & I_r & & 0 & 0 \\ \vdots & & \ddots & 0 & 0 \\ 0 & 0 & \cdots & I_r & 0 \end{bmatrix}}_{R\times R} \quad \text{and} \quad G = \underbrace{\begin{bmatrix} I_r \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{R\times r}.$$

Here $R = r(q+1)$ and $\Gamma_i = 0$ for $i > s$.

Following Chamberlain and Rothschild (1983), Stock and Watson (2005) refer to model (16.1.8) as an exact DFM if $A(L)$ has a diagonal structure and the error covariance matrix $\mathbb{E}(u_t u_t') = \Sigma_u$ is diagonal, which implies mutually uncorrelated idiosyncratic components. Models of this type were used in the earlier econometrics literature by Sargent and Sims (1977). They are also closely related to the index models considered by Reinsel (1983) and reduced-rank VAR models discussed by Velu, Reinsel, and Wichern (1986), Tso (1981), Ahn and Reinsel (1988), Reinsel (1993), Reinsel and Velu (1998) and Anderson (1999, 2002). Such models differ from the DFM in (16.1.8) by their assumptions about the error term v_t. They assume that v_t is white noise with a general, not necessarily diagonal covariance matrix. In other words, the error term cannot be interpreted easily as a vector of idiosyncratic components. In contrast to exact DFMs, approximate DFMs also allow for dependence between the idiosyncratic components. In the following we treat $A(L)$ and Σ_u as diagonal, unless noted otherwise.

Estimation of DFMs. We defer the question of how to specify the model until later and take as given the number of lags and the number of factors in the model.

Before estimating the DFM, it is recommended that the variables are first standardized such that they have zero mean and variance one. Following Stock and Watson (2005), the dynamic form of the DFM can be estimated as follows:

Step 1. Construct an initial estimate $\tilde{A}(L)$ of $A(L) = \text{diag}[a_1(L), \dots, a_N(L)]$, for example, by regressing the individual variables on their own lags.

Step 2. Compute the PC estimator $\widehat{\Lambda}$ of Λ, as discussed earlier, from the model $\tilde{A}(L)x_t = \Lambda F_t + \tilde{u}_t$, where $\tilde{A}(L)x_t$ assumes the role of x_t in equation (16.1.1), and estimate the factors as $\widehat{F}_t = \widehat{\Lambda}'\tilde{A}(L)x_t$.

Step 3. Estimate $A(L)x_t = \Lambda\widehat{F}_t + \hat{u}_t$ by single-equation LS for each equation to update the estimate of $A(L)$.

Step 4. Iterate Steps 2 and 3 until convergence.

Using single-equation LS in Step 3 is justified because the idiosyncratic error terms are assumed to be instantaneously uncorrelated such that Σ_u is a diagonal matrix. If that assumption is incorrect, estimation efficiency can be improved by using a feasible GLS procedure because the regressors in the different equations of the system $A(L)x_t = \Lambda\widehat{F}_t + \hat{u}_t$ are not identical.

So far we have only discussed the estimation of the observation equation of the state-space model (16.1.10). Once the estimated factors $\widehat{F}_t$ are available, the $\mathbf{\Gamma}$ coefficient matrix in the transition equation in model (16.1.10) can be estimated as well. It should be noted, however, that estimates of $\mathbf{\Gamma}$ are not required for many applications of dynamic factor models. The estimation of $\mathbf{\Gamma}$ is complicated by the fact that in practice we can only estimate some linear transformation of the static factors $F_t = (f_t', \dots, f_{t-q}')'$. The four-step estimator of the DFM discussed above uses a statistical normalization that may not result in the primitive dynamic factors, f_t. Thus, estimating $\mathbf{\Gamma}$ by regressing $\widehat{F}_t$ on $\widehat{F}_{t-1}$ only produces an estimator $\widehat{\mathbf{\Gamma}}^*$ of some linear transformation of $\mathbf{\Gamma}$.

This concern may be addressed by implementing a second-stage estimator which determines the r linearly independent primitive factors f_t underlying $F_t = (f_t', \dots, f_{t-q}')'$. Let $\widehat{W}$ be the matrix of eigenvectors corresponding to the r largest eigenvalues of the residual covariance matrix $\widehat{\Sigma}_{\mathcal{E}} = T^{-1}\sum_t \widehat{\mathcal{E}}_t\widehat{\mathcal{E}}_t'$, where $\widehat{\mathcal{E}}_t = \widehat{F}_t - \widehat{\mathbf{\Gamma}}^*\widehat{F}_{t-1}$ and $\widehat{\mathbf{\Gamma}}^*$ is the estimator obtained by regressing $\widehat{F}_t$ on $\widehat{F}_{t-1}$. Then $\widehat{\eta}_t = \widehat{W}'\widehat{\mathcal{E}}_t$ and the primitive factors f_t can be estimated as $\hat{f}_t = \widehat{W}'\widehat{F}_t$. If estimates of $\Gamma_1, \dots, \Gamma_{q+1}$ are required, they may be obtained by regressing $\hat{f}_t$ on $\hat{f}_{t-1}, \dots, \hat{f}_{t-q-1}$. Finally, the covariance matrix of η_t can be estimated in the usual way using the residual covariance estimator based on the latter regression. Alternatively, it may be based on $\widehat{\Sigma}_\eta = T^{-1}\sum_t \widehat{\eta}_t\widehat{\eta}_t'$, where $\widehat{\eta}_t = \widehat{W}'\widehat{\mathcal{E}}_t$. Methods for choosing the values of R and r required in this procedure are discussed in Section 16.1.3.

It is also possible to estimate all DFM parameters simultaneously by ML estimation under the assumption that x_t is Gaussian. The Gaussian likelihood may be evaluated with the Kalman filter because model (16.1.10) is in state-space form. Using state-space estimation algorithms is also appealing because it allows one to handle missing observations (see Stock and Watson 2016, section 2.3.4). Missing observations are a common problem in panels of variables, when the beginning of the available time series varies across N, and starting the panel sample at the beginning of the shortest series may result in a substantial loss of sample information for other variables. In addition, the state-space framework can accommodate time series with different sampling frequencies. Computing the full Gaussian ML estimator may be challenging if the panel of variables is large. Doz, Giannone, and Reichlin (2011) propose an alternative two-step estimator based on the Kalman filter that may be helpful in the latter context.

Asymptotic results for the estimators of dynamic factor models can be found in Stock and Watson (2002a), Bai (2003), and Bai and Ng (2008) among others. Despite the fact that the asymptotic properties are derived for T and $N \to \infty$, small-sample simulation evidence in Boivin and Ng (2006) indicates that including more variables in a factor analysis does not necessarily result in more accurate estimates. Likewise, Boivin and Ng show that including more variables need not improve the forecast accuracy of an approximate factor model.

Rather than using frequentist methods, one may also use Bayesian methods for estimating dynamic factor models. We return to Bayesian estimation in the context of large panels of variables in Section 16.3 and therefore do not discuss these methods here. Bayesian methods of estimating DFMs have been used by Otrok and Whiteman (1998), Kose, Otrok, and Whiteman (2003) and Amir Ahmadi and Uhlig (2009), for example.

Generalized Dynamic Factor Models. The generalized dynamic factor model (GDFM) generalizes (16.1.7) by allowing the common and idiosyncratic components to be general stationary processes that may not admit a finite-order VAR representation. We consider the model $x_t = \Lambda f_t + v_t$, where f_t is the process of common factors and v_t is the process of idiosyncratic components. The two processes f_t and v_t can be characterized by their spectral properties (see Forni and Lippi 2001; Forni, Hallin, Lippi, and Reichlin 2000, 2004, 2005). Provided x_t is $I(0)$, a natural generalization of the covariance decomposition (16.1.2) is a decomposition of the spectral density of x_t. Denoting the spectral density functions of x_t, f_t, and v_t by $\Sigma_x(\xi)$, $\Sigma_f(\xi)$, and $\Sigma_v(\xi)$, respectively,

$$\Sigma_x(\xi) = \Lambda \Sigma_f(\xi)\Lambda' + \Sigma_v(\xi), \quad \forall \xi \in [0, 2\pi]. \tag{16.1.11}$$

The matrix $\Sigma_v(\xi)$ is assumed to be diagonal in the exact GDFM, whereas more general structures are allowed for in the approximate GDFM.

Estimation. Estimation of GDFMs based on the decomposition (16.1.11) is considered in Forni and Reichlin (1998) and Forni, Hallin, Lippi, and Reichlin (2000, 2004, 2005). Since they do not assume a parametric model for the DGP of the observables and the common factors, they use a nonparametric frequency-domain estimator developed by Brillinger (1975). Based on the work of Forni et al., Favero, Marcellino, and Neglia (2005) propose the following procedure for estimating the dynamic principal components:

Step 1. For a sample $x_1, \dots, x_T$ of size T, the spectral density matrix of x_t is estimated as

$$\widehat{\Sigma}_x(\xi_j) = \sum_{s=-S}^{S} \omega_s \widehat{\Gamma}_x(s) e^{-is\xi_j}, \quad \xi_j = 2\pi j/(2S+1), \quad j = 0, 1, \dots, 2S,$$

where S is the window width, $\omega_s = 1 - |s|/(S+1)$ are the weights of the Bartlett window and $\widehat{\Gamma}_x(s) = T^{-1}\sum_t (x_t - \bar{x})(x_{t-s} - \bar{x})'$ is the sample covariance matrix of x_t for lag s. The window width has to be chosen such that $S \to \infty$ and $S/T \to 0$ as $T \to \infty$. Forni, Hallin, Lippi, and Reichlin (2000) observe that a choice of $S = 2T^{1/3}/3$ worked well in their simulations.

Step 2. For $j = 0, 1, \dots, 2S$, determine the eigenvectors $\boldsymbol{\lambda}_1(\xi_j), \dots, \boldsymbol{\lambda}_r(\xi_j)$ corresponding to the r largest eigenvalues of $\widehat{\Sigma}_x(\xi_j)$.

Step 3. Defining

$$\boldsymbol{\lambda}_{sk} = \frac{1}{2S+1}\sum_{j=0}^{2S} \boldsymbol{\lambda}_s(\xi_j) e^{ik\xi_j}, \quad k = -S, \dots, S,$$

the dynamic PCs of x_t are obtained as

$$\hat{f}_{st} = \sum_{k=-S}^{S} \boldsymbol{\lambda}'_{sk} x_{t-k}, \quad s = 1, \dots, r,$$

and collected in the vector $\hat{f}_t = (\hat{f}_{1t}, \dots, \hat{f}_{rt})'$.

Step 4. Fit the regression

$$x_t = \Lambda_{-q}\hat{f}_{t+q} + \cdots + \Lambda_p \hat{f}_{t-p} + v_t$$

and estimate the common component as the fitted value

$$\hat{\chi}_t = \widehat{\Lambda}_{-q}\hat{f}_{t+q} + \cdots + \widehat{\Lambda}_p \hat{f}_{t-p},$$

where $\widehat{\Lambda}_j$, $j = -q, \ldots, p$, are the LS estimators. The leads q and lags p used in the regression could be chosen by model selection criteria. In practice, small numbers of leads and lags are used.

Using leads of the estimated factors and, hence, of the observations in Step 4 to reconstruct the common component is not an option in forecasting. An alternative one-sided procedure has been proposed in Forni, Hallin, Lippi, and Reichlin (2005). For impulse response analysis it is not clear that a one-sided procedure is preferred because impulse responses are expectations that are estimated from the full sample. Hence, we presented the two-sided procedure.

Comparison of DFMs and GDFMs. As we have emphasized, the various forms of the DFM are just different representations of the same DGP. The particular form used in empirical analysis is largely a matter of convenience. Of course, factor models may also differ to some extent in their underlying assumptions. For example, GDFMs in principle allow for more general dynamics of the factors and idiosyncratic components than DFMs. In particular, whereas GDFMs allow the factors and idiosyncratic errors to be general stationary stochastic processes, DFMs focus on finite-order VAR and AR approximations to these processes. How important that distinction is in practice, is not clear because linear stationary processes can be approximated by finite-order autoregressive processes. Thus, for all practical purposes, the choice of the model depends on the preferences of the user. Technical considerations are of secondary importance.

16.1.3 Selecting the Number of Factors

The specification of a DFM requires selecting the number of common factors and the various lag orders. Since PC analysis does not require the specification of a lag order, it is possible to specify the number of static factors before the lag orders of the VAR operators are determined. In classical static factor models, criteria such as choosing as many factors as are necessary to explain a prespecified fraction of the overall variance are commonly used. More precisely, we include additional common factors until the sum of the variances of the common factors exceeds a prespecified fraction of the sum of the eigenvalues of the sample covariance matrix. Another criterion in that spirit is the so-called scree test proposed by Cattell (1966), which increases the number of factors until the variance explained by these factors starts to taper off.

Specification of DFMs. For DFMs more formal criteria have been developed that assume the existence of a true number of factors, R_0, and allow for consistent estimation of this quantity when both the cross-section and time dimension become large ($N, T \to \infty$). The most popular criteria are proposed by Bai and

Ng (2002) and take the form

$$IC(R) = \log \mathbb{V}(R) + Rg(N, T), \tag{16.1.12}$$

where $\mathbb{V}(R) = (NT)^{-1} \sum_{t=1}^{T} (x_t - \widehat{\Lambda}^F \hat{F}_t)'(x_t - \widehat{\Lambda}^F \hat{F}_t)$ and $g(N, T)$ is a penalty term. Note the similarity with the information criteria for VAR order selection discussed in Chapter 2. Bai and Ng show that under suitable conditions the estimator $\hat{R} = \operatorname{argmin}_{R=1,\dots,R_{\max}} IC(R)$ is consistent for the true number of factors R_0. Of course, a necessary condition is that $R_{\max} \geq R_0$. Moreover, the penalty term $g(N, T)$ has to go to zero at a suitable rate with growing T and N. According to Breitung and Eickmeier (2006), the most popular criterion from this class chooses $g(N, T) = \left(\frac{N+T}{NT}\right) \log(\min[N, T])$ such that

$$IC_{p2}(R) = \log \mathbb{V}(R) + R \left(\frac{N+T}{NT} \right) \log(\min[N, T]). \tag{16.1.13}$$

Using this criterion we can estimate the number of static factors, which determines the dimension of F_t.

For structural analysis, the number of primitive dynamic factors, which corresponds to the dimension r of f_t in the dynamic form (16.1.9), is of primary interest. Bai and Ng (2007) propose a procedure for determining r. They utilize the fact that the error term in the transition equation in the static form (16.1.10), $G\eta_t$, has covariance matrix $G\Sigma_\eta G'$ of rank r and devise a procedure for determining that rank. Starting from estimates $\widehat{F}_t$ of the static factors, they propose to fit a VAR model to the $\widehat{F}_t$. In our current framework that VAR model is of order one because we have assumed that $q \geq s$ (see model (16.1.10)). Thus, fitting

$$\widehat{F}_t = \mathbf{\Gamma} \widehat{F}_{t-1} + \mathcal{E}_t$$

generates residuals $\widehat{\mathcal{E}}_t, t = 1, \dots, T$ (see Section 16.1.2). Let $\rho_1 \geq \rho_2 \geq \dots \geq \rho_R$ be the eigenvalues obtained from a PC analysis of the estimated residual covariance matrix $T^{-1} \sum_{t=1}^{T} \widehat{\mathcal{E}}_t \widehat{\mathcal{E}}_t'$ and define

$$\hat{D}_1(r) = \left(\frac{\rho_{r+1}^2}{\sum_{i=1}^{R} \rho_i^2} \right)^{1/2} \tag{16.1.14}$$

and

$$\hat{D}_2(r) = \left(\frac{\sum_{i=r+1}^{R} \rho_i^2}{\sum_{i=1}^{R} \rho_i^2} \right)^{1/2}. \tag{16.1.15}$$

Based on these quantities Bai and Ng (2007) propose to estimate the number of primitive dynamic factors as

$$\hat{r} = \min r \in \left\{ r \,\middle|\, \hat{D}_1(r) < \frac{1}{\min(N^{1/2-\delta}, T^{1/2-\delta})} \right\} \tag{16.1.16}$$

or as

$$\hat{r} = \min r \in \left\{ r \,\middle|\, \hat{D}_2(r) < \frac{1}{\min(N^{1/2-\delta}, T^{1/2-\delta})} \right\}, \qquad (16.1.17)$$

where δ is a small number between 0 and 1/2. In a simulation study they choose $\delta = 1/4$ which appears to give reasonable results.

Because the lag length of $\Lambda(L)$ in the DFM (16.1.9) cannot be larger than R/r, the number of static factors, R, bounds the lag length of $\Lambda(L)$ for given r. More precisely, we can choose $q + 1 = R/r$, if the latter ratio is an integer. If R/r is not an integer, a plausible value for q is the smallest integer larger than $(R/r) - 1$. Thus, choosing the number of factors is related to selecting the lag length of at least one of the lag polynomials in the DFM. The other lag lengths can be chosen by standard model selection criteria. Jacobs and Otter (2008) propose a procedure for determining the number of dynamic factors and their lags simultaneously.

A number of studies address the problem of estimating the number of factors in DFMs. Further important contributions include Amengual and Watson (2007) and Breitung and Pigorsch (2013). For a thorough review see Bai and Ng (2008). A more recent contribution is Caner and Han (2014) who propose to choose the number of factors based on a bridge estimator.

Specification of GDFMs. There are also methods specifically designed to estimate the number of factors in GDFMs. For example, Onatski (2009) presents a testing procedure for the number of static factors that may be viewed as a formalization of the scree test mentioned earlier. Hallin and Liška (2007) develop information criteria for estimating the number of dynamic factors based on the eigenvalues $\lambda_1(\xi_j), \dots, \lambda_S(\xi_j)$ of the estimated spectral density matrix $\widehat{\Sigma}_x(\xi_j)$, $\xi_j = 2\pi j/(2S + 1)$, $j = 0, 1, \dots, 2S$. These criteria are defined as

$$PCP(k) = \frac{1}{N} \sum_{j=k+1}^{N} \frac{1}{2S+1} \sum_{s=-S}^{S} \lambda_j(\xi_s) + k\varphi(N, T)$$

and

$$IC(k) = \log \left(\frac{1}{N} \sum_{j=k+1}^{N} \frac{1}{2S+1} \sum_{s=-S}^{S} \lambda_j(\xi_s) \right) + k\varphi(N, T).$$

The bandwidth S has to be chosen such that $S \to \infty$ and $S/T \to 0$ for $T \to \infty$, and the penalty term $\varphi(N, T)$ has to satisfy

$$\varphi(N, T) \to 0 \quad \text{and} \quad \min\{N, S^{-2}, (ST)^{1/2}\}\varphi(N, T) \to \infty.$$

The search is done over $k \in \{0, \dots, r_{\max}\}$ with $r_{\max}$ greater than or at least equal to the true rank in order to obtain consistent estimates.

16.1.4 Structural Change

As in other models, there may be parameter instability in dynamic factor models. A number of authors have considered that possibility and have developed statistical tools for diagnosing structural change in factor models (e.g., Stock and Watson 2009; Breitung and Eickmeier 2011; Chen, Dolado, and Gonzalo 2014; Han and Inoue 2015). A survey of related issues and procedures is provided in Stock and Watson (2016, section 2.5).

16.2 Factor-Augmented Structural VAR Models and Related Techniques

Section 16.1 focused on the question of how to extract the common factors from a large panel of time series. The statistical models used to estimate these factors by themselves have no structural economic interpretation. In this section we show how factor structures may be incorporated into structural VAR models and how these models may be used to construct structural impulse responses. Section 16.2.1 examines structural FAVAR models. In Section 16.2.2, we discuss structural analysis with dynamic factor models. Section 16.2.3 reviews empirical examples.

16.2.1 Structural FAVAR Models

One way of obtaining a FAVAR model is to augment the set of observed variables, z_t, in a given VAR model by latent factors, F_t, extracted from a set of observed variables x_t that does not include z_t. This approach is often intended to address the informational deficiencies of conventional small-scale structural VAR models (see Chapter 17). It also allows the researcher to approximate the responses of many more variables to a given structural shock than is possible in conventional VAR models.

Another motivation for this approach is that often the observed economic variables are only proxies for the economic concepts used in economic theory or that these variables may be subject to measurement error. In these situations a more accurate measure of the variable of economic interest may be obtained as a factor extracted from alternative measures of this variable. For example, Aruoba, Diebold, Nalewaik, Schorfheide, and Song (2016) propose viewing ‘true GDP’ as a latent variable for which we have several indicators, the two most obvious being an expenditure-based measure of GDP and an income-based measure of GDP. They propose to extract this latent ‘true GDP’ measure using a DFM. The resulting factors may be used instead of any one of the observed imperfect measures of the variables of interest, possibly in combination with observed variables that are either accurately measured or that are known to matter to economic agents. For example, the European Central Bank

(ECB) is required by law to maintain price stability as measured by the harmonized index of consumer prices in the European Monetary System. Thus, this variable has to enter directly the ECB's policy reaction function when specifying a VAR model of European monetary policy.

An example of such a model is Bernanke, Boivin, and Eliasz (2005). This study proposes a structural FAVAR model for $y_t = (F_t', z_t')'$ as

$$B(L)\begin{bmatrix} F_t \\ z_t \end{bmatrix} = w_t, \tag{16.2.1}$$

where the vector of structural shocks w_t is $(R+M)$-dimensional white noise, $B(L) = B_0 + B_1 L + \cdots + B_p L^p$ is a $(R+M)\times(R+M)$ matrix operator, F_t is a vector of R unobserved common factors that are related to the $N \times 1$ vector of informational variables x_t by the observation equation

$$x_t = \Lambda^F F_t + \Lambda^z z_t + e_t, \tag{16.2.2}$$

where Λ^F is the $N \times R$ matrix of factor loadings and Λ^z is an $N \times M$ matrix of coefficients. Bernanke, Boivin, and Eliasz (2005) assume that the upper $R \times R$ block of Λ^F is an identity matrix and the upper $R \times M$ block of Λ^z is a zero matrix, noting that these conditions are sufficient for the identification of the factors. Alternative conditions for identifying FAVAR models are discussed in Bai, Li, and Lu (2016).

The M observed variables z_t are usually a small set of variables of interest from an economic point of view that drive the dynamics of the system together with the unobserved factors F_t. The z_t variables must not be included in x_t because otherwise some of the idiosyncratic components would be zero and, hence, the covariance matrix of the idiosyncratic components would be singular. It should be noted that the factors F_t in model (16.2.1) need not be ordered first. They may also be placed below z_t or in between the elements of z_t.

Estimation. Estimation of model (16.2.1) requires knowledge of F_t. One approach taken in many studies is to first extract F_t by PC analysis from a large set of informational variables, x_t, that do not include the observed variables of interest, z_t, before estimating the FAVAR model (16.2.1) with the estimated factors replacing the true factors. This procedure is inspired by the work of Stock and Watson (2002b). It is also used by Favero, Marcellino, and Neglia (2005), Bernanke and Boivin (2003), and, in modified form, by Bernanke, Boivin, and Eliasz (2005).[1]

[1] In a detailed analysis of the Bernanke, Boivin, and Eliasz (2005) procedure, Ouliaris, Pagan, and Restrepo (2014) show that this modified procedure implies a VARMA-DGP with dependent errors for the factors and the key observed variables, necessitating a sieve approach to estimating the FAVAR model.

The unknown factors F_t may be constructed from model (16.1.10) using the four-step procedure for the estimation of DFMs discussed in Section 16.1.2. Bai and Ng (2006) show that the uncertainty in the estimates of F_t is negligible when N is large relative to T, allowing one to condition on the estimated factors as though they were known.

For example, Kilian and Lewis (2011) replace the real output variable in an otherwise standard VAR model of U.S. monetary policy by the Chicago Fed National Activity Index (CFNAI), which is constructed as the leading principal component of a wide range of monthly indicators of U.S. real activity, including five categories of data: output and income (21 series); employment, unemployment, and hours (24 series); personal consumption, housing starts, and sales (13 series); manufacturing and trade sales (11 series); and inventories and orders (16 series). They specify a VAR(12) with intercept for $y_t = (\Delta rpcom_t, \Delta rpoil_t, F_t, \pi_t, i_t)'$, where $\Delta rpcom_t$ is the percentage change in the real price of imported commodities (excluding crude oil), $\Delta rpoil_t$ is the percentage change in the real price of imported crude oil, F_t is the CFNAI real activity index, π_t is the CPI inflation rate, and i_t is the federal funds rate. After fitting the reduced-form VAR model $A(L)y_t = \nu + u_t$, the structural shocks are recovered by applying a lower-triangular Cholesky decomposition to Σ_u.

This approach does not impose the additional restrictions implied by the observation equation (16.2.2). Boivin and Giannoni (2009) discuss an alternative estimation procedure that is intended to account for all restrictions implied by the structural FAVAR model. They construct the initial value of the factors, $F_t^{(0)}$, by performing a PC analysis on x_t. They then estimate Λ^F and Λ^z by applying equation-by-equation LS to

$$x_t = \Lambda^F F_t^{(i-1)} + \Lambda^z z_t + e_t. \tag{16.2.3}$$

Denoting the estimate of Λ^z by $\widehat{\Lambda}^z$, they then perform a PC analysis of $x_t - \widehat{\Lambda}^z z_t$, resulting in factors $F_t^{(i)}$. The last step is repeated for $i = 1, 2, \ldots,$ until convergence.

This alternative procedure has its own limitations, however. As Ouliaris, Pagan, and Restrepo (2014) point out, the effects of z_t on the unobserved factors $F_t^{(i)}$ will not tend to zero as i increases. Hence, there is no clear separation in this estimation procedure between the unobserved factors and the observed variables z_t. Thus, this alternative method is not necessarily superior to the simpler approach that ignores the observation equation (16.2.2) in estimating the state-space model.

Bai, Li, and Lu (2016) discuss an alternative two-step estimation procedure for FAVAR models. In the initial stage of the procedure the observed variables z_t are concentrated out of equation (16.2.2) and the unobserved factors F_t are estimated by a quasi-ML procedure rather than by PC estimation. Then a reduced-form version of model (16.2.1) with $B_0 = I_{R+M}$ is estimated by LS and the estimates are used to update the initial estimates of the factors and parameters of the FAVAR model (16.2.1). Bai et al. prove the consistency and asymptotic normality of this estimator under alternative sets of identifying conditions, provided that suitable moment conditions hold and that the data are stationary. They also show that their quasi-ML procedure is asymptotically more efficient than PC estimation.

Extensions. There are several extensions of this framework. First, Dufour and Stevanović (2013) generalize the FAVAR model to the factor-augmented vector autoregressive moving average (FAVARMA) model. They note that, if the factors are driven by a finite-order VAR process, the informational variable x_t follows a VARMA process. They propose a suitable estimation procedure for such models and provide evidence that the FAVARMA model may have higher forecast accuracy than the FAVAR model.

Second, Banerjee and Marcellino (2008) and Banerjee, Marcellino, and Masten (2014a, 2014b) consider factor-augmented cointegrated VAR models in error correction form. They discuss estimation, forecasting, and structural analysis based on such factor-augmented VECMs (FECMs). The advantage of FECMs is that they explicitly allow for integrated variables, whereas standard dynamic factor models are designed for stationary variables. An obvious advantage of including integrated variables in levels is that the models can also capture cointegration relations.

Third, Balabanova and Brüggemann (2017) consider FAVAR models with factors entering as exogenous variables

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \Pi_0 F_t + \cdots + \Pi_q F_{t-q} + u_t,$$

resulting in a FAVARX model.

Structural Identification. Structural shocks in the FAVAR model (16.2.1) can be recovered as in the standard VAR model by a linear transformation of the reduced-form errors from the model

$$A(L)y_t = u_t,$$

where $y_t = (F_t', z_t')'$. As before, the structural shocks are obtained as $w_t = B_0 u_t$, given suitable identifying restrictions on B_0^{-1} or B_0. Of course, the number of variables and, hence, structural shocks to be identified may be larger in FAVAR models, rendering the identification potentially more difficult. In many applications the FAVAR model is only partially identified.

For example, Favero, Marcellino, and Neglia (2005) specify a FAVAR model for

$$(F_t', z_t^{*\prime}, r_t)',$$

where z_t^* contains all directly observed variables apart from the interest rate r_t. They order the monetary policy instrument last in their FAVAR model and impose a recursive ordering because they are only interested in the effects of monetary policy shocks. This ordering implies that a monetary policy shock has no instantaneous impact on any of the observed variables or on the factors.

Of course, the assumption that none of the factors and observed variables reacts to a monetary policy shock within the same period may be questioned, in particular, when fast-moving financial variables are included in the model. Alternatively, one could split up the variables into fast-moving and slow-moving variables, as proposed in Bernanke, Boivin, and Eliasz (2005), and extract factors separately from the two groups of variables. Then one could order the slow-moving factors (f_t^s) above the interest rate and the fast-moving factors (f_t^f) below it, so that they can be instantaneously affected in a lower-triangular recursive identification scheme. In other words, the variables in the FAVAR could be arranged as follows:

$$(f_t^{s\prime}, z_t^{s\prime}, r_t, f_t^{f\prime}, z_t^{f\prime})',$$

where z_t^s and z_t^f contain the slow- and fast-moving observed variables, respectively, in the system. Of course, this identification prevents the central bank from responding to fast-moving variables such as stock returns or changes in commodity prices and, hence, is not particularly credible either.

Suitable restrictions separating slow- and fast-moving factors can also be implemented by imposing zero restrictions on the factor loadings. In other words, one may specify that some factors load only on fast-moving variables and others load only on slow-moving variables. Imposing such restrictions requires suitable estimation algorithms that allow for a restricted loading matrix such as ML or Bayesian methods.

The responses of the factors to structural shocks in the FAVAR setup (16.2.1) in general differ from the responses of the informational variables, x_t. The latter may be approximated using the observation equation (16.2.2) as

$$x_t \approx [\Lambda^F, \Lambda^z] A(L)^{-1} B_0^{-1} w_t. \tag{16.2.4}$$

The approximation error relates to the omission of the idiosyncratic error term, e_t, in the observation equation. The ability of FAVAR models to generate (approximate) structural responses for the large number of variables contained in x_t makes these models particularly useful for practitioners. For example, we may trace the effect of a monetary policy shock on a large number of disaggregate measures of real output or inflation if the latter are contained in x_t.

Of course, in many applications we may not be interested in the responses of x_t, but are satisfied with reporting the responses of z_t and F_t. One approach in the literature has been to treat F_t like an observed variable. In that case, the identification of the structural shocks and the construction of the structural impulse responses and related statistics of economic interest is no different than in earlier chapters. For example, Mumtaz and Surico (2009) and Eickmeier and Hofmann (2013) use a mixture of exclusion and sign restrictions in a FAVAR setting. Mumtaz and Surico (2009) also consider recursive and nonrecursive models identified by short-run exclusion restrictions.

It should be noted, however, that the structural impulse responses implied by the system of equations (16.2.1) and (16.2.2) may not coincide with the structural impulse responses $B(L)^{-1}$ implied by model (16.2.1) alone because the DGP is not a conventional VAR model, but a state-space model with latent structure. Although the responses of z_t and F_t to structural shocks in the system of equations may in principle be computed by iteration, this approach does not appear to have been used in applied work. Nor is it clear how to achieve identification of the structural shocks in the latter framework.

Inference on Structural Impulse Responses. If we treat the estimated factors as though they were known, inference in structural FAVAR models may proceed exactly as in Chapter 12. In practice, the use of bootstrap methods is standard. If the factors are subject to estimation uncertainty, in contrast, the bootstrap procedure has to account for the full DGP described by the state-space model (16.2.1) and (16.2.2). Several studies including Bernanke, Boivin, and Eliasz (2005) have used the bootstrap to approximate the distribution of structural impulse responses in FAVAR models, but without stating explicitly their algorithm or establishing its asymptotic validity.

Building on related ideas in Yamamoto (2016), such an algorithm may look as follows. Consider the reduced-form state-space model with transition equation

$$A(L)\begin{bmatrix} F_t \\ z_t \end{bmatrix} = u_t,$$

and observation equation

$$x_t = \Lambda^F F_t + \Lambda^z z_t + e_t,$$

where u_t and e_t are mutually uncorrelated. Then a bootstrap procedure involves the following steps:

Step 1. Estimate the factors and parameters of the state-space model by one of the procedures discussed earlier, denote the estimates of parameters of the observation equation by $\widehat{\Lambda}^F$ and $\widehat{\Lambda}^z$, and denote the corresponding (mean-adjusted) residuals by $\hat{u}_t$ and $\hat{e}_t$.

Step 2. Draw bootstrap samples u_t^* and e_t^*, $t = 1, \dots, T$, with replacement from the sets of residuals $\hat{u}_t$ and $\hat{e}_t$, respectively, and compute a bootstrap sample of factors, $F_1^*, \dots, F_T^*$, and observed variables, $z_1^*, \dots, z_T^*$, based on the transition equation in the usual way. Then compute the bootstrap sample of the informational variables based on the observation equation as $x_t^* = \widehat{\Lambda}^F F_t^* + \widehat{\Lambda}^z z_t^* + e_t^*$ for $t = 1, \ldots, T$.

Step 3. Use the bootstrap realizations x_t^*, z_t^* for $t = 1, \ldots, T$ to reestimate the factors and the parameters of the state-space model and to recompute the structural impulse responses.

These steps are repeated a large number of times to obtain the bootstrap distribution of the quantities of interest. There are no formal results about the validity of this bootstrap.

16.2.2 Structural Analysis with DFMs

Structural impulse response analysis can alternatively be performed within the framework of a DFM or GDFM. There is one important difference. Whereas in FAVAR models we are concerned with structural shocks to either F_t or z_t, in the current setting all observed variables are in x_t and the structural innovations are driving the primitive factors f_t.

Identification. Recall that the DFM (16.1.7) can be written in compact form as

$$x_t = \Lambda^f f_t + v_t,$$

with the dynamic factors being generated as

$$\Gamma(L) f_t = \eta_t \quad \text{and} \quad A(L) v_t = u_t,$$

such that $f_t = \Gamma(L)^{-1}\eta_t$. Substituting the latter expression into equation (16.1.7) yields

$$x_t = \Phi(L)\eta_t + v_t, \tag{16.2.5}$$

where $\Phi(L) = \Lambda^f \Gamma(L)^{-1}$. For our discussion of structural forms and of the identification of the structural shocks we treat the reduced-form parameters $\Phi(L)$ and Σ_η as known. If necessary, they can be estimated as discussed in Section 16.1.2. It should be noted that the representation (16.2.5) remains valid when working with GDFMs. The only difference is how the parameters of representation (16.2.5) are derived from the underlying model.

Assuming as usual that the $r \times 1$ vector of reduced-form residuals η_t is related to the $r \times 1$ vector of structural shocks w_t by a linear transformation $B_0\eta_t = w_t$, the structural form corresponding to model (16.2.5) is

$$x_t = \Phi(L) B_0^{-1} w_t + v_t. \tag{16.2.6}$$

If the structural shocks are instantaneously uncorrelated and the variances are normalized to 1, we know that $w_t \sim (0, I_r)$. In that case, B_0 has to satisfy

$B_0^{-1}B_0^{-1\prime} = \Sigma_\eta$ and we need at least $r(r-1)/2$ more restrictions for identifying the $r \times r$ matrix B_0. In other words, identifying the structural shocks requires putting enough restrictions on B_0 or its inverse to obtain uniqueness. Because the primitive factors f_t have no natural economic interpretation, economic identifying restrictions cannot be imposed on the effects on f_t, but only on the effects of the structural shocks on x_t. These restrictions may come in the form of exclusion restrictions on the impact effects or on the long-run effects of the structural shocks on x_t. They may also be available in the form of sign restrictions. Some specific identification strategies are discussed next.

Restrictions on the Impact Effects of Shocks. Note that the impact matrix Φ_0 in $\Phi(L) = \sum_{i=0}^{\infty} \Phi_i L^i$ will in general not be an identity matrix. In fact, $\Phi(L)$ is $N \times r$ and is typically not a square matrix. Therefore the impact effects of the structural shocks, w_t, on x_t are given by $\Phi_0 B_0^{-1}$, and exclusion restrictions on the impact effects are zero restrictions on the elements of the matrix product $\Phi_0 B_0^{-1}$. For example, one may want to impose a recursive identification scheme on the impact effects, as is often done in conventional structural VAR analysis. This amounts to choosing a suitable nonsingular $r \times r$ submatrix of $\Phi_0 B_0^{-1}$ to be lower triangular. Such restrictions would suffice for identifying B_0 and, hence, the structural shocks. Denoting the $r \times r$ submatrix of Φ_0 that is of interest in the present context by $\bar{\Phi}_0$, the corresponding B_0^{-1} matrix can be obtained by noting that $\bar{\Phi}_0 B_0^{-1} B_0^{-1\prime} \bar{\Phi}_0' = \bar{\Phi}_0 \Sigma_\eta \bar{\Phi}_0'$. Thus, computing a Cholesky decomposition of this matrix and left-multiplying by $\bar{\Phi}_0^{-1}$ yields a suitable B_0^{-1} matrix.

An illustration of this approach is provided in Forni and Gambetti (2010) who specify a GDFM for a panel of 112 monthly series. In their benchmark model the number of primitive dynamic factors is four. They use industrial production, a consumer price index, the federal funds rate and the Swiss/U.S. real exchange rate as the first four variables in their panel in that order. The structural shocks are identified recursively with the monetary policy shock specified to be the third shock. This policy shock is identified as a shock that does not have an instantaneous effect on industrial production and on the price level, but may induce immediate responses of the real exchange rate. All four shocks are allowed to have impact effects on the other variables.

More generally, exclusion restrictions can be imposed on the impact effects by choosing a suitable $\frac{1}{2}r(r-1) \times r^2$ selection matrix J such that

$$J\text{vec}(\bar{\Phi}_0 B_0^{-1}) = 0$$

which implies restrictions

$$J(I_r \otimes \bar{\Phi}_0)\text{vec}(B_0^{-1}) = 0$$

for B_0^{-1}.

As in standard structural VAR models identified by exclusion restrictions on the impact effects, we may get away with imposing fewer than $r(r-1)/2$ structural restrictions if fewer than r shocks are of interest in a particular application. The other shocks can then be identified arbitrarily. For example, in the Forni and Gambetti (2010) study, because only the monetary policy shock is of interest, the restrictions used for making the other shocks unique are not important.

It is also possible in the current framework to identify the shocks by placing restrictions directly on the factor loadings Λ^f in a structural version of model (16.1.7), exploiting the fact that $\Phi_0 = \Lambda^f$ and normalizing Σ_η to be diagonal. This approach may be more natural if the common factors have a direct economic interpretation. For example, a study of the international business cycle by Kose, Otrok, and Whiteman (2003), based on a panel of macroeconomic aggregates for 60 countries from 7 different regions in the world, postulates the existence of one world factor, a factor for each region, and a country-specific factor for each country. Only the world factor is allowed to have a direct impact on all variables in x_t. The effects of the other factors are restricted by imposing suitable zero constraints on the corresponding loadings. Of course, this approach is not structural in the conventional sense.

Restrictions on the Long-Run Effects of Shocks. Long-run restrictions as in Blanchard and Quah (1989) may also be used in the current context, provided x_t includes at least some variables that are expressed in growth rates. The approach of restricting the long-run cumulative effects of the structural shocks on x_t is more natural here than specifying a vector error correction model because x_t typically does not include $I(1)$ variables. This method involves restricting an $r \times r$ submatrix of $\Phi(L)$, say $\bar{\Phi}(L)$, such that $\bar{\Phi}(1)$ is nonsingular and $\bar{\Phi}(1)B_0^{-1}$ is lower-triangular. In this case the corresponding B_0^{-1} is obtained by computing a Cholesky decomposition of $\bar{\Phi}(1)B_0^{-1}B_0^{-1\prime}\bar{\Phi}(1)' = \bar{\Phi}(1)\Sigma_\eta\bar{\Phi}(1)'$ and left-multiplying by $\bar{\Phi}(1)^{-1}$.

Identification Through Instruments. Identification of the $r \times r$ matrix B_0^{-1} and thereby of the structural shocks can also be achieved by using instruments with suitable properties, as discussed in Chapter 15. Suppose that a variable ζ_t is available that is correlated with the k^{th} structural shock, but uncorrelated with all other shocks such that

$$\mathbb{E}(w_{it}\zeta_t) = \begin{cases} \rho \neq 0 & \text{for} \quad i = k, \\ 0 & \text{for} \quad i \neq k. \end{cases}$$

Then, using $\eta_t = B_0^{-1}w_t$ and denoting the columns of B_0^{-1} by b_i, $i = 1, \ldots, r$, we obtain

$$\mathbb{E}(\eta_t\zeta_t) = B_0^{-1}\mathbb{E}(w_t\zeta_t) = b_k\rho. \tag{16.2.7}$$

Thus, a multiple of the k^{th} column of B_0^{-1} can be obtained as the covariance of the reduced form error η_t and the instrument ζ_t. If a suitable instrumental variable is available that is only correlated with the k^{th} structural shock and is uncorrelated with all other structural shocks, a natural estimator of $b_k \rho$ is

$$T^{-1} \sum_{t=1}^{T} \eta_t \zeta_t,$$

where η_t can be replaced by the estimator $\widehat{\eta}_t = \widehat{W}' \widehat{\mathcal{E}}_t$, as discussed in Section 16.1.2. Stock and Watson (2012) call such instruments external instruments if they are not part of the database x_t used in the factor analysis (see Chapter 15).

Note that for identification it is sufficient to know a multiple of the k^{th} column of B_0^{-1} because the responses of the variables to the k^{th} shock are obtained as $\Phi(L)b_k$ and multiplying the shocks by some constant only changes the magnitude of the shock, but does not change the shape of the response function. Hence, the magnitude of the shock can be chosen freely. Stock and Watson (2012) propose to normalize the initial response of one of the variables to one. For example, a monetary policy shock may be chosen such that the policy instrument changes by one unit on impact.

So far we have just explained how to identify the columns of B_0^{-1} and, hence, the structural impulse responses. Stock and Watson (2012) suggest that the corresponding structural shocks can be determined by regressing ζ_t on η_t and using the predicted values of that regression as estimates of the structural shocks (see Section III.a of Stock and Watson (2012) for details). Of course, different sets of instruments will imply different estimates of the structural shocks. As observed by Sims (1998), there is no limit to how different the estimated structural shock time series might appear, when using alternative equally valid instruments.

In principle, identification using external instruments is straightforward if suitable instruments can be found. Stock and Watson (2012) use as instruments external estimates of exogenous shocks of the type discussed in Chapter 7. For example, as an instrument for the productivity shock in the factor model they consider an estimate of the productivity shock series from the DSGE model of Smets and Wouters (2007). As instruments for the monetary policy shocks they use the corresponding shocks identified by Sims and Zha (2006b) and the shock to the monetary policy reaction function of the Smets and Wouters (2007) model.

Other Identification Strategies. Earlier we discussed the identification of structural shocks based on exclusion restrictions. In principle it is also possible to use sign restrictions in DFMs (see Yamamoto 2016). This approach does not appear to have been used in applied work so far.

It is also possible to identify structural shocks based on a penalty function. For example, Giannone, Reichlin, and Sala (2004) postulate the existence of two factors in a system of U.S. macroeconomic variables and, hence, of two structural shocks. They identify the first shock as a real shock that maximizes the share of the variance of the real variables explained by the factor. The other shock is taken to be the nominal shock.

Another proposal is to link the factors to economic variables in a DSGE model (see Bäurle 2013). Other authors have considered so-called multi-level or hierarchical factor models. In these models, the variables are partitioned into blocks and there are block-specific and global common factors (e.g., Moench and Ng 2011; Hallin and Liška 2011). For example, Hallin and Liška investigate industrial production in a multiple-country study in which the blocks refer to different countries. Such models allow the separate identification of block-specific and global shocks.

Inference on Structural Impulse Responses. Structural impulse responses are functions of the structural parameters of the factor model and are obtained from the parameter estimates. Thus, asymptotic properties can be derived for the impulse responses in the usual way from the asymptotic properties of the parameter estimators. Since in frequentist impulse response analysis bootstrap methods are typically used for constructing confidence intervals around impulse responses, the question arises of how to bootstrap a DFM or GDFM. Based on earlier work of Gonçalves and Perron (2014), Yamamoto (2016) proposes and compares two procedures for a DFM in static form.

For expository purposes, consider the simplified model

$$x_t = \Lambda^f f_t + v_t, \quad f_t = \Gamma_1 f_{t-1} + \cdots + \Gamma_s f_{t-s} + \eta_t, \tag{16.2.8}$$

where v_t and η_t are mutually uncorrelated white noise processes. Moreover, it is assumed that the reduced-form parameters are identified by suitable restrictions, and the structural impulse responses are just-identified. Then a residual-based bootstrap for the structural impulse responses can be executed in the following steps.

Step 1. Estimate the factors and parameters of model (16.2.8) by one of the procedures presented in Section 16.1.2 and denote the parameter estimates by $\widehat{\Lambda}^f$ and $\widehat{\Gamma}_i$, $i = 1, \ldots, s$. Let the corresponding (mean-adjusted) residuals be $\hat{v}_t$ and $\hat{\eta}_t$.

Step 2. Draw a bootstrap sample v_t^* and η_t^*, $t = 1, \ldots, T$, with replacement from the sets of residuals $\hat{v}_t$ and $\hat{\eta}_t$, respectively, and compute the bootstrap sample of factors and observed variables in the usual way as $f_t^* = \widehat{\Gamma}_1 f_{t-1}^* + \cdots + \widehat{\Gamma}_s f_{t-s}^* + \eta_t^*$ and $x_t^* = \widehat{\Lambda}^f f_t^* + v_t^*$.

Step 3. Use the bootstrap realizations x_t^* for $t = 1, \dots, T$ to reestimate the parameters and the factors of model (16.2.8). Then recompute the structural impulse responses.

These steps are repeated a large number of times to obtain the bootstrap distribution of the quantities of interest. Yamamoto (2016) makes the case that this procedure is asymptotically valid. He also suggests that other bootstrap procedures could be used to accommodate more general assumptions about the error term. Stock and Watson (2016, section 5.1.3) propose a related bootstrap procedure that allows for autoregressive dynamics in the idiosyncratic errors, v_t, and involves drawing the bootstrap innovations η_t^* as well as the bootstrap innovations v_t^* of the autoregressive process from normal distributions.

In Step 3 of the procedure the parameters and factors of the model are reestimated from the bootstrap sample x_t^*. Yamamoto (2016) also considers an alternative procedure in which the factors are not reestimated, but instead the bootstrap factors are used in estimating the model parameters. In other words, Step 3 of the procedure is replaced by

Step 3*. Use the bootstrap sample x_t^* and factors f_t^* to estimate the parameters of model (16.2.8) by regression methods and then compute the quantities of interest from these estimates.

He argues that this procedure is also asymptotically valid. Based on a simulation study he finds that the procedure that reestimates the factors as in Step 3, has higher coverage accuracy in some situations than the procedure based on Step 3*. There do not appear to be results in the literature on bootstrapping structural impulse responses from GDFMs.

A Critique of Structural Analysis with DFMs. The identifying assumptions discussed earlier are, of course, critical for the structural analysis considered in this chapter. One could question the identification of structural shocks as a linear transformation of the errors η_t driving the factors in model (16.2.5). It is difficult to see why this assumption makes sense in approximate DFMs that allow more structure in the part of the model not explained by the common factors. Such approximate DFMs make explicit that the relationship between the variables is only partly captured by the common factors. Hence, the transmission of the shocks to x_t is only partly captured by $\Phi(L)B_0^{-1}$. Even if the DGP is an exact DFM, the Wold MA representation of the observed variables x_t is different from representation (16.2.5). Hence, building on the MA representation of x_t results in a different transmission of the shocks.

Stock and Watson (2005) rightly point out that if the DFM is taken as the DGP, the factors contain all the dynamic interaction between the variables. Conditioning on the factors, none of the variables is Granger-causal for any other variable. Still this does not mean that the shocks are best extracted from the model driving the factors. It may well be that important shocks enter through the idiosyncratic component of x_t rather than through the η_t errors in the representation (16.2.6). For example, a discretionary change in monetary policy can be viewed as an idiosyncratic shock to the interest rate that affects both the factors and x_t.

Although dynamic factor models help address the missing information problem in low-dimensional VAR models by allowing us to include a large set of variables in the analysis, extending the data set also creates new problems. For example, not much is known about the sensitivity of structural impulse responses with respect to changes in the information set and/or the model structure. In particular, the choice of the number of factors and the question of which factors to include play an important role. A sensitivity analysis with respect to the set of variables, the sampling period, the model specification, and the type and number of factors is therefore recommended.

16.2.3 Empirical Examples of FAVAR Models and DFMs

As already discussed, one of the reasons why researchers have been attracted to structural FAVAR and dynamic factor models is the ability to mitigate informational deficiencies in conventional structural VAR models. Another reason is that these models may be used to characterize the response of many economic variables to a given shock. For example, we may describe the responses of the economy by sector or of the inflation rate by expenditure item. This feature helps alleviate concerns that conventional structural VAR models are silent on the response of many of the variables that users of macroeconomic models are interested in and that traditional DSEMs were able to capture (see Chapter 6).

There are many examples of structural analysis with factor models in the literature. Many of these studies deal with one of two applications: monetary policy analysis and international business cycle analysis. Next, a small number of examples is reviewed to convey the flavor of these studies.

The Effects of Monetary Policy. Structural FAVAR analysis was first proposed by Bernanke, Boivin, and Eliasz (2005). For their analysis of U.S. monetary policy they extracted several factors from a panel of 120 monthly time series. Their baseline FAVAR model includes industrial production, the consumer price index (CPI), the federal funds rate, and one of the factors. In another specification they add three factors to the federal funds rate. Identification of the monetary policy shock is achieved by assuming that none of the other variables or factors respond instantaneously to a monetary policy shock. This assumption is consistent with a recursive identification scheme in which the federal funds rate is ordered last. Bernanke et al. find that taking into account the additional information summarized in the factors makes a substantial difference for the structural impulse responses and paint a different picture of the transmission of monetary policy shocks than a conventional small-scale VAR model.

Del Negro and Otrok (2007) use a FAVAR model for quarterly U.S. variables to investigate the impact of monetary policy shocks on house prices. They include information on state-level house prices in a VAR model of monetary policy. They find that part of the house price changes in each state can be attributed to a national factor constructed from these price series. They include this factor in a FAVAR model consisting of six variables: the factor, total reserves, CPI inflation, the GDP growth rate, a 30-year mortgage rate, and the federal funds rate. They use sign restrictions for identifying the monetary policy shock (see Chapter 13). Specifically, they assume that the federal funds rate increases and the growth rate of total reserves, changes in CPI inflation, and changes in GDP growth do not increase for several quarters after a monetary policy shock. They find that monetary policy shocks have an effect on house prices and hence contribute to housing price booms, although the impact is small.

Favero, Marcellino, and Neglia (2005) use a FAVAR model to investigate the impact of monetary policy shocks in the United States and four large European economies (Germany, France, Italy, Spain). They compare the Stock-Watson principal-components approach with estimates of the dynamic factors constructed as in the work of Forni, Hallin, Lippi and Reichlin. The common factors are extracted from large monthly panels of variables from the U.S. and the four European countries. The country-specific FAVAR models include small sets of macroeconomic variables such as output, inflation, commodity price inflation, an exchange rate, and, in addition, the interest rate in the case of the United States and selected foreign variables for European countries. For example, U.S. inflation is included in the model for Germany. The study presents estimates for models augmented with alternative sets of factors. Identification is achieved by a recursive scheme with the interest rate ordered last. The authors conclude that including common factors can make a difference for impulse response analysis. For example, it may remove the ‘price puzzle’, which refers to an increase in inflation after a contractionary monetary policy shock. This phenomenon is often attributed to omitted variables bias (see Chapter 8). Including further information in the form of factors can be thought of as mitigating this omitted variables bias.

Boivin, Giannoni, and Mihov (2009) investigate the impact of macroeconomic factors and monetary policy shocks on sectorally disaggregated consumer and producer prices. They construct a FAVAR model based on a large number of monthly U.S. time series for the period 1976m1-2005m6. The number of factors is five. The federal funds rate is viewed as the policy instrument. Identification of the monetary shocks is achieved by the assumption that none of the common factors reacts instantaneously to surprise changes in the policy rate, which amounts to a lower-triangular recursive scheme where the interest rate is ordered last. Boivin et al. find that the reaction of sector specific prices to macroeconomic shocks and sector-specific shocks is very different. The response of disaggregated prices to a monetary shock is delayed and little evidence is found for a price puzzle.

International Business Cycle Analysis. The objective of business cycle analysis with dynamic factor models is typically to find a factor that describes the business cycle fluctuations globally or in a large region. For example, Kose, Otrok, and Whiteman (2003) use a dynamic factor model and Bayesian estimation techniques to investigate the business cycle fluctuations in a set of 60 countries that covers seven regions of the world. They consider aggregate output, consumption, and investment variables and find a dynamic factor that explains some of the fluctuations in the aggregates in most countries and can thus be viewed as a world business cycle factor. They decompose the variance into components that can be attributed to the different factors and thereby determine how much of the variance in specific variables is determined by the business cycle factor and how much is accounted for by other factors. They find that a large part of the fluctuations in many aggregate variables can be attributed to the global business cycle factor, whereas region-specific factors are less important in determining fluctuations in economic activity. In a related study Kose, Otrok, and Whiteman (2008) investigate possible differences in the business cycle dynamics over specific historic periods. Other FAVAR studies of the international business cycle include Eickmeier (2007), Mansour (2003), Helbling and Bayoumi (2003), and Bordo and Helbling (2010).

Transmission of Oil Supply Shocks. In a comprehensive study, Stock and Watson (2016) investigate the interaction between the global oil market and the U.S. economy using structural DFM and FAVAR models, building on the analysis in Kilian (2009). They find that global oil supply shocks explain only a fraction of the variation in the price of oil and a very small fraction of the variation in major U.S. macroeconomic aggregates.

16.3 Large Bayesian VAR Models

As mentioned earlier, instead of frequentist estimation methods, one may use Bayesian methods for estimating dynamic factor models or FAVAR models (see, e.g., Otrok and Whiteman 1998; Kose, Otrok, and Whiteman 2003; Amir Ahmadi and Uhlig 2009). If Bayesian methods are used, it is not obvious, however, that one would want to focus on factor models. Recall that the motivation for using factor models is that they allow us to investigate large panels of variables in a structural VAR analysis. In the context of Bayesian estimation, suitable priors on VAR models may serve the same purpose (see, e.g., De Mol, Giannone, and Reichlin 2008), even when not imposing a factor structure. In fact, as pointed out by Bańbura, Giannone, and Reichlin (2010), using Bayesian shrinkage methods to overcome the degrees-of-freedom problem in structural VAR analysis has several advantages. For example, having no limits on the number of observed variables included in the VAR model allows one to include all variables considered by macroeconomists. Even sectoral information can be included and the impact of specific shocks such as monetary policy shocks on the disaggregated variables can be traced. Thereby large-scale international comparisons become possible without imposing ad hoc restrictions on the structure of the model.

Another advantage of starting from an unrestricted VAR model rather than summarizing some of the information in factors is that levels variables can be included easily. Recall that standard factor analysis tends to be based on stationary variables without stochastic trends. Thereby they may miss out on common trend structures except to the extent that known cointegration constraints are imposed by including cointegration relations as stationary variables. Although factor analysis can in principle also be applied to trending variables, the additional assumptions regarding the stochastic trends required for proper inference can be restrictive. From a practical point of view it may therefore be advantageous to work with the variables in levels, thereby potentially accommodating unit roots, long-range dependence, near unit root behaviour and the like.

A crucial problem with using large-scale BVAR models is the choice of the prior. This issue is reviewed in Section 16.3.1. Structural identification is discussed in Section 16.3.2.

16.3.1 Priors for Large Bayesian VARs

Bańbura, Giannone, and Reichlin (2010) use the so-called Minnesota or Litterman prior as their point of departure, which postulates a reduced-form Gaussian VAR model,

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

and imposes a normal prior with a random walk prior mean,

$$A^* = [0, I_K, 0, \ldots, 0],$$

where $A = [\nu, A_1, A_2, \ldots, A_p]$ and hence, the prior mean of A_1 is the identity matrix (see Chapter 5). This prior has to be modified if there are variables that are not persistent. For the latter variables the prior mean of the corresponding diagonal element of A_1 is set to zero instead of 1. The prior variance of the ij^{th} element of A_l is

$$v_{ij,l} = \begin{cases} (\lambda/l)^2 & \text{if } i = j, \\ (\lambda\theta\sigma_i/l\sigma_j)^2 & \text{if } i \neq j, \end{cases}$$

where λ is the prior standard deviation of $a_{ii,1}$, $0 < \theta < 1$, and σ_i^2 is the i^{th} diagonal element of the reduced form residual covariance matrix Σ_u. If Σ_u is known, the posterior is also normal and quite easy to deal with. Thus, if one is prepared to replace the covariance matrix by some known quantity such as a plausible estimator, the Bayesian estimation problem is basically solved. Of course, we cannot just replace Σ_u by its unrestricted LS estimator because this estimator is typically infeasible, given the degrees-of-freedom limitations. One alternative is to estimate the variances by fitting univariate AR models by LS and to assume that Σ_u is diagonal. This solution is sometimes used in practice (see, e.g., Koop 2013a).

The original Minnesota prior is regarded as unattractive by Bańbura, Giannone, and Reichlin (2010) because of the restrictive requirements for the reduced form residual covariance matrix. Instead they propose using a Gaussian-inverse Wishart prior, which is a natural conjugate prior if $\theta = 1$. Using that prior, the posterior mean is

$$\bar{A} = \left(A^* V(\lambda)^{-1} + YZ'\right)\left(V(\lambda)^{-1} + ZZ'\right)^{-1},$$

where $Y = [y_1, \dots, y_T]$, Z is the corresponding matrix of regressors, and the prior covariance matrix is $V(\lambda) \otimes \Sigma_u$. For given Σ_u and assuming $\theta = 1$, the prior covariance matrix depends only on the tightness parameter λ, which has to be provided by the user.

The posterior mean may be interpreted as a shrinkage estimator where the shrinkage is completely determined by λ. For large models the matrix ZZ' will not even be invertible and the posterior mean can only be determined by adding another matrix $(V(\lambda)^{-1})$ that makes the sum invertible and, hence, effectively determines the outcome of the estimation. In other words, the prior determines the estimation outcome to a large extent. Thus, the question is how to choose the shrinkage or tightness parameter. For example, if forecasting is the objective, one could choose λ such that the model tends to forecast accurately (see, e.g., Carriero, Kapetanios, and Marcellino 2009). Alternatively, λ may be chosen so as to maximize the marginal likelihood in a hierarchical modeling framework, as proposed by Giannone, Lenza, and Primiceri (2015) (see also Carriero, Kapetanios, and Marcellino 2012 for a similar approach). For models with hundreds of variables the latter procedure poses computational challenges, however.

In related work, De Mol, Giannone, and Reichlin (2008) and Bańbura, Giannone, and Reichlin (2010) propose decreasing the parameter λ, as the model grows larger, such that the estimated model has the same in-sample fit as a small VAR model estimated by LS that only includes the key economic variables of interest. This procedure works as follows.

Denote the posterior means of the parameters obtained from a model with tightness parameter λ and K variables by $\nu^{(\lambda,K)}$ and $A_i^{(\lambda,K)}$, $i = 1, \dots, p$, and the corresponding 1-step ahead predictions as

$$y_{t|t-1}^{(\lambda,K)} = \nu^{(\lambda,K)} + A_1^{(\lambda,K)} y_{t-1} + \dots + A_p^{(\lambda,K)} y_{t-p}.$$

Moreover, let $y_{k,t|t-1}^{(\lambda,K)}$ be the k^{th} component of $y_{t|t-1}^{(\lambda,K)}$, i.e., $y_{k,t|t-1}^{(\lambda,K)}$ is the 1-step ahead prediction of the k^{th} variable of a system with K variables and prior tightness parameter λ. The corresponding full-sample mean squared prediction error (MSPE) is

$$\text{MSPE}_k^{(\lambda,K)} = \frac{1}{T-p} \sum_{t=p+1}^{T} (y_{k,t|t-1}^{(\lambda,K)} - y_{k,t})^2.$$

Suppose that there is a small number, K^*, of variables of central interest with index set $\mathcal{K}$. Then the tightness parameter for the large model including all K variables, λ_K, is chosen such that

$$\lambda_K = \arg\min_{\lambda} \left| \text{Fit} - \frac{1}{K^*} \sum_{k \in \mathcal{K}} \frac{\text{MSPE}_k^{(\lambda,K)}}{\text{MSPE}_k^{(0,K)}} \right|, \tag{16.3.1}$$

where $\text{MSPE}_k^{(0,K)}$ corresponds to the MSPE of the large model evaluated at its prior mean and $\text{MSPE}_k^{(\lambda,K)}$ corresponds to the MSPE of the large model for a given choice of λ.

The benchmark fit is defined analogously as

$$\text{Fit} = \frac{1}{K^*} \sum_{k \in \mathcal{K}} \frac{\text{MSPE}_k^{(\infty,K^*)}}{\text{MSPE}_k^{(0,K^*)}},$$

where the small model only includes $K^* < K$ variables, $\text{MSPE}_k^{(\infty,K^*)}$ corresponds to the MSPE of the small model estimated by LS, and $\text{MSPE}_k^{(0,K^*)}$ corresponds to the MSPE of the small model evaluated at its prior mean. The actual minimization with respect to λ in (16.3.1) can be done by a grid search over λ because only one parameter is involved (see Koop 2013a).

In practice, after estimating the small model, λ is chosen for the large-scale model such that the in-sample fit for the equations corresponding to the central variables remains constant. This procedure worked well in a forecasting experiment reported in Bańbura, Giannone, and Reichlin (2010). Note that their choice of the tightness parameter amounts to specifying a tighter prior for larger models with more variables and lags.

Koop (2013a) explores an analogous procedure for the original Minnesota prior. He expresses concern about the very restrictive nature of the Minnesota prior, however, which uses only one or two parameters to determine the degree of shrinkage. Koop therefore considers another prior based on a proposal by George, Sun, and Ni (2008) that allows different parameters to be shrunk differently. In Koop's forecast comparison the effectiveness of this prior tends to deteriorate, when the number of variables in the VAR increases. Given that we

 https://doi.org/10.1017/9781108164818.017

are primarily interested in dealing with large BVAR models, we do not consider this prior.

In related work, Bańbura, Giannone, and Reichlin (2010) report that a sum-of-coefficients prior which is a variant of the Minnesota prior worked best in their forecasting application. They recommend that prior also for structural analysis. As discussed in Chapter 5, the sum-of-coefficients prior accommodates long-run relations more easily and may therefore improve the accuracy of impulse response estimates when cointegration relations exist in the data.

Finally, Korobilis (2013) proposes to combine BVAR models with Bayesian variable selection. In his approach an indicator variable is specified for each parameter that tells us whether the parameter is included or its coefficient is set to zero. The prior for the indicator variables can be combined easily with the Minnesota prior, for example. Korobilis (2013) presents a modification of this approach intended to make it feasible to deal with large panels of variables. However, the largest model he considers contains only 13 variables, which is far from the dimensions we have in mind in this section.

16.3.2 Structural Identification in Large BVARs

In large-scale BVAR models, the identification of the structural shocks is most easily achieved by linking the properties of the shocks to their impact effects. For example, exclusion restrictions can be specified for the impact effects. Bańbura, Giannone, and Reichlin (2010) are interested in the effects of monetary policy shocks. As in Bernanke, Boivin, and Eliasz (2005), they split their set of variables into those that move slowly in response to the monetary policy shock and those that may respond to this shock within the impact period such that

$$y_t = (y_t^{s\prime}, r_t, y_t^{f\prime})',$$

where y_t^f contains the fast moving variables such as financial variables, y_t^s is the vector of slow moving variables such as prices and real variables, and r_t is the policy interest rate. They identify the monetary policy shock by a lower-triangular Cholesky decomposition of the reduced-form covariance matrix Σ_u. Thereby the fast-moving variables are allowed to be affected instantaneously while the slow-moving variables are assumed to be known to the policymaker at the time when a decision is made.

In principle, one could also identify the shock of interest by sign restrictions. Although plausible sign restrictions for a large set of variables may be available, as argued in Amir Ahmadi and Uhlig (2009), such an approach is complicated by the large dimension of the structural-form covariance matrix and the corresponding dimension of the possible rotation matrices that have to be considered in computing admissible shocks. Thus, such an approach may be computationally infeasible with current technology. A possible solution may be to reduce these computational problems by combining sign restrictions with exclusion restrictions.

In summary, the large-scale BVAR model has some advantages, but also some drawbacks. On the positive side, it dispenses with the additional structure imposed by FAVAR and dynamic factor models and it avoids some of the ambiguities of structural models based on factors. On the negative side, the priors required for the estimation of large-scale VAR models may distort the structural impulse response estimates without the user being able to assess the extent of this bias. Thus, the prior induces an element of arbitrariness.

16.4 Alternative Large-Dimensional VAR Models

So far we have discussed how to deal with large panels of time series data either by imposing a factor structure or by applying Bayesian shrinkage methods in estimation. Other proposals for dealing with this problem include panel VAR models, global VAR models, and spatial VAR models (see Canova and Ciccarelli 2013; Pesaran, Schuermann, and Weiner 2004; Chudik and Pesaran 2011; or Pesaran 2015). Each of these models imposes specific types of restrictions on the VAR parameters to allow larger VAR models to be estimated. These restrictions in some cases are strong and unrealistic, and they may distort the structural impulse responses in unknown ways. Of course, if some restrictions on the model can be defended on economic grounds, this approach will be preferable to imposing a factor structure or to fitting BVAR models, so it is important to think carefully about the types of restrictions that are best suited for a specific application.

16.4.1 Panel VARs

Large panels of variables often arise in studies of different countries or regions, but also when dealing with sectors, firms, plants, or households. Such a situation makes it convenient to assign an additional subscript to a variable. For example, we may denote the t^{th} observation for the i^{th} variable of country n by y_{int}, where $i = 1, \ldots, M$ and $n = 1, \ldots, N$. Thus, using our earlier notation, $K = M \cdot N$. Let $y_{nt} = (y_{1nt}, \ldots, y_{Mnt})'$ be an M-dimensional vector and denote by $Y_{n,t-1}$ and Y_{t-1} vectors of lags of y_{nt} and all variables in the panel, respectively. Then the model for y_{nt} has the general form

$$y_{nt} = \nu_n + \mathbf{A}_n Y_{t-1} + u_{nt}, \tag{16.4.1}$$

with fully general error covariance matrix Σ_u for the system of all N units, where Σ_u is the covariance matrix of $u_t = (u'_{1t}, \ldots, u'_{Nt})'$.

This panel structure has three characteristics. First, lags of all endogenous variables of all panel units are allowed to affect y_{nt}. Second, u_{nt} is in general correlated across n. Third, the model parameters may be specific to each n, allowing for cross-sectional heterogeneity. In this general form, the model is not different from large-dimensional BVAR models.

Clearly, it is difficult to estimate the panel VAR model in its general form (16.4.1) without imposing additional prior structure. Even when Bayesian estimation methods are used, it is common to impose additional restrictions in estimation. For example, we may impose the restriction that in small open economies global variables are exogenously given. Often researchers further restrict the panel VAR model such that every panel unit corresponds to a separate VAR model

$$y_{nt} = \nu_n + A_n Y_{n,t-1} + u_{nt},$$

where the coefficient matrix A_n is much smaller than $\mathbf{A}_n$ in (16.4.1), with the interaction across units determined by the covariances $\mathbb{E}(u_{nt}u'_{ms})$. Unlike model (16.4.1), this restricted specification does not explicitly model the dynamic interactions across the panel units.

Even more restrictions can be imposed if the assumption of dynamic homogeneity can be justified. In that case, all units are assumed to have the same VAR coefficients such that

$$y_{nt} = \nu_n + AY_{n,t-1} + u_{nt}.$$

Suitable estimation methods for panel VAR models are discussed in Canova and Ciccarelli (2013). In typical economic applications, of course, one would not expect the dynamics to be homogeneous across countries, sectors, or firms.

Panel VAR techniques based on (16.4.1) have become increasingly popular recently. As noted by Canova and Ciccarelli (2013), panel VAR models are particularly suited for analyzing the transmission of unit-specific shocks across panel units and time. For example, Canova, Ciccarelli, and Ortega (2012) study how shocks to U.S. interest rates are propagated to ten European countries, some in the Euro area and some not. Panel VAR models have also been used to construct the average effects across heterogeneous groups of countries. They may also be used to assess the small open economy assumption or to assess the existence of convergence clubs.

The estimation of panel VAR models in practice can be computationally challenging. An additional challenge in structural panel VAR models is the identification of the structural shocks. Typically such models are only partially identified. For these and other issues the reader is referred to the surveys by Breitung (2015) and Canova and Ciccarelli (2013).

16.4.2 Global VARs

Global VARs (GVARs) represent another class of models designed specifically to handle panels of time series data for many countries, which explains its name (e.g., Dees, Di Mauro, Pesaran, and Smith 2007; Chudik and Pesaran 2016; Pesaran 2015). The idea is to augment a VAR for each unit by global variables capturing the other countries. In that respect GVARs have some similarity with FAVAR models. More specifically, the model for the n^{th} country is set up as

$$y_{nt} = \nu_n + A_n Y_{n,t-1} + W(L) y^*_{nt} + u_{nt},$$

where $W(L)$ is a matrix polynomial in the lag operator and y^*_{nt} summarizes global variables such as the price of oil and possibly aggregated variables from other countries. Typically the latter variables are weighted averages of the variables from other countries corresponding to y_{nt} of the form

$$y^*_{nt} = \sum_{\substack{j=1 \\ j\neq n}}^{N} \omega_{nj} y_{jt},$$

where ω_{nj} is the weight attached to country j in the model for the n^{th} country. For example, Dees, Di Mauro, Pesaran, and Smith (2007) link the weights to the share of country j in world trade. These weights may also be matrices, if different variables are weighted differently. Generally y^*_{nt} can be thought of as common factors consisting of global variables and foreign variables computed from data for the other countries in the panel. Thus, the common factors are not necessarily determined by a statistical procedure, but typically are based on economic considerations. Apart from this important difference, the models can be viewed as factor-augmented VAR models.

The advantage of this approach is that several countries or regions can be modeled jointly as a global model (or GVAR model), which is easily recognized as a large restricted VAR model that may also include some additional unmodeled variables. The parameter restrictions are partly due to the choice of weights in aggregating the foreign variables and partly they are just exogeneity restrictions.

Since the global model has a VAR structure it is also possible to use standard tools such as impulse responses, provided shocks of interest can be identified. The GVAR literature has typically used non-structural generalized impulse responses to study the dynamics of the system, but occasionally structural identifying assumptions have been invoked. For example, Dees, Di Mauro, Pesaran, and Smith (2007) identify a U.S. monetary policy shock by ordering the U.S. block first and imposing a recursive ordering on the U.S. variables. The effects on the other countries are left open and are allowed to be instantaneous. Such an approach may be justified if a dominant country such as the United States is considered. The identification of shocks becomes more difficult, however, if the effects of monetary policy shocks in other countries on the United States are of interest or, more generally, the effects of many shocks. It has also been common to trace out responses to variation in variables that are assumed exogenous, not unlike what one would do in traditional DSEMs (see Chapter 6).

The difficulties in identifying structural shocks is one drawback of GVAR analysis. Another drawback is that the transmission of the shocks is affected by the assumptions made when aggregating the variables from other countries. A review of the GVAR approach which addresses many related technical issues and includes a large number of further references is provided in Chudik and Pesaran (2016). This study also summarizes several applications of the GVAR approach.

16.4.3 Spatial Models

There are also other ideas about how to restrict the number of variables or the parameter space in large-dimensional VAR models. For example, spatial models assume that a region depends more strongly on its close neighbours than on more distant regions (see Anselin 2006; Chudik and Pesaran 2011; Canova and Ciccarelli 2013). In other words, the distance between units is used to impose restrictions on the dependence across units, thereby reducing the parameter space. Such assumptions can be problematic. In many applications it is unsatisfactory to link the relationship between panel units to their physical distance alone. For example, U.S. shocks may affect many other countries in the world more than would be expected based on their physical distance alone. Although there are more general norms of distance accounting for multiple attributes, the implicit weights used in constructing a measure of distance tend to be inherently ad hoc. Thus, spatial VAR models have not played an important role in applied work to date.

16.5 Discussion

The proposals for dealing with large panels of time series variables considered in this chapter all amount to imposing restrictions on the VAR model. This can be accomplished either by reducing the dimension of the set of model variables or by shrinking the parameter space. Both approaches deal with the curse of dimensionality. At one extreme, factor models rely entirely on a reduction in the number of model variables; at the other extreme, large-dimensional BVAR models only shrink the parameter space, while preserving the model dimension. Other approaches are somewhat in-between. For example, GVAR models to some extent reduce the space of variables by restricting the parameter space, but they also impose prior structure by choosing the aggregation weights for the model variables.

All the models considered in this chapter have their pros and cons. For example, factor models extract information from the variables first and aggregate it in factors. These factors are then used in a model together with selected observed variables to construct moderately-sized models that can be analyzed with standard frequentist methods. Aggregation of the data, however, causes distortions in the structural responses, the quantitative importance of which is usually unknown. Moreover, standard factor analysis is tailored to stationary variables with time-invariant moments. Estimates may be sensitive to the assumptions regarding the order of integration or to the long-range dependence properties of the model variables.

In contrast, large-scale BVAR models completely avoid this aggregation bias. Moreover, variables may be included in levels regardless of their unit root and trending properties. Recall that degrees-of-freedom deficiencies make frequentist estimation impossible, when VAR models become large. The Bayesian solution is to impose a prior on the parameter space. The priors typically imposed in large-scale BVAR analysis shrink the VAR parameters to zero or to values corresponding to multivariate random walks. The drawback of large-scale BVAR models is that the priors required for estimation do not account for the actual economic structures underlying a panel of time series variables. Even if they are not meant to be restrictive, they may lead to substantial distortions in the estimates of the structural impulse responses and related statistics.

In a forecast comparison based on large panels of variables, Bańbura, Giannone, and Reichlin (2010) and Koop (2013a) find that large BVARs forecast more accurately overall than factor models. In fact, De Mol, Giannone, and Reichlin (2008) provide conditions based on asymptotic theory which ensure that Bayesian shrinkage for large panels of time series that are driven by a limited number of factors, results in optimal forecasts asymptotically, if both the number of variables and the time series dimension go to infinity. Although such results can be used to make a case for BVARs, it is not clear that out-of-sample forecast accuracy is the best criterion for evaluating the transmission mechanism of shocks.

A potential alternative to Bayesian shrinkage is the use of LASSO-type estimators, as proposed in Song and Bickel (2011).[2] A key difference between Bayesian shrinkage and the LASSO approach is that the latter not only shrinks the parameter estimates towards zero but may also eliminate parameters from the model, resulting in a potentially sparse large-dimensional VAR structure. Like the BVAR approach, the estimator of Song and Bickel facilitates structural impulse response analysis directly on the model variables. However, the use of a penalty term involves an additional auxiliary parameter that

[2] LASSO stands for Least Absolute Shrinkage and Selection Operator. LASSO minimizes the residual sum of squares subject to a penalty term involving the absolute value of the regression coefficients (see Tibshirani 1996).

determines the degree of shrinkage. This additional structure may distort the structural impulse responses much like the prior in high-dimensional BVAR models. In addition, it is not clear how robust the LASSO approach is to the presence of unit roots and cointegration among the model variables.

The alternative of imposing a panel structure also involves imposing restrictions on the parameter space. Unlike the large-scale BVAR model, the panel VAR model often is restricted further by imposing additional exclusion restrictions. To the extent that the possibly restricted structural panel VAR model tends to be estimated by Bayesian methods, there is also an important element of shrinkage, however.

Global VAR models in turn may be viewed as a combination of imposing direct restrictions on the parameter space and imposing a specific factor structure. The factors are not chosen based on purely statistical criteria, but on the basis of economic considerations. For example, the use of trade weights in aggregating macroeconomic time series from different countries is common. Alternatively, exclusion restrictions on model parameters may be suggested by the fact that some countries are small open economies. The fact that the factor structure is not selected based on purely statistical criteria, of course, does not mean that the aggregation cannot distort estimates of the structural impulse responses. Another problem in the GVAR literature is the lack of convincing strategies for the identification of structural economic shocks. As mentioned in Section 16.4.2, the so-called generalized impulse responses often reported in the GVAR literature are based on shocks that are only statistically identified and lack a clear economic interpretation.

For applied work the distinction between reducing the set of model variables and shrinking the parameter space is not of prime importance. What is important is the question of whether the restrictions required for estimation can be defended in a particular application. Thus, knowing the pros and cons of these alternative models is important.