

1351

Discussion  
Papers

# Structural Vector Autoregressive Analysis in a Data Rich Environment

A Survey

Helmut Lütkepohl

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

#### IMPRESSUM

© DIW Berlin, 2014

DIW Berlin  
German Institute for Economic Research  
Mohrenstr. 58  
10117 Berlin

Tel. +49 (30) 897 89-0  
Fax +49 (30) 897 89-200  
<http://www.diw.de>

ISSN print edition 1433-0210  
ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:  
<http://www.diw.de/discussionpapers>

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:  
<http://ideas.repec.org/s/diw/diwwpp.html>  
<http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html>

# Structural Vector Autoregressive Analysis in a Data Rich Environment: A Survey

Helmut Lütkepohl <sup>1</sup>  
DIW and Freie Universität Berlin  
Mohrenstr. 58  
10117 Berlin, Germany  
email: hluetkepohl@diw.de

January 9, 2014

**Abstract:** Large panels of variables are used by policy makers in deciding on policy actions. Therefore it is desirable to include large information sets in models for economic analysis. In this survey methods are reviewed for accounting for the information in large sets of variables in vector autoregressive (VAR) models. This can be done by aggregating the variables or by reducing the parameter space to a manageable dimension. Factor models reduce the space of variables whereas large Bayesian VAR models and panel VARs reduce the parameter space. Global VARs use a mixed approach. They aggregate the variables and use a parsimonious parametrisation. All these methods are discussed in this survey although the main emphasis is on factor models.

*Key Words:* factor models, structural vector autoregressive model, global vector autoregression, panel data, Bayesian vector autoregression

*JEL classification:* C32

---

<sup>1</sup>Helpful comments by Jörg Breitung are gratefully acknowledged. This paper was written while the author was a Bundesbank Professor. This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 “Economic Risk”.

# Contents

<b>1</b>	<b>Factor Models</b>	<b>5</b>
1.1	Static Factor Models . . . . .	5
1.1.1	Model Setup . . . . .	5
1.1.2	Estimating Static Factor Models . . . . .	6
1.1.3	Approximate Static Factor Models . . . . .	8
1.2	Dynamic Factor Models . . . . .	9
1.2.1	Static Form of a DFM . . . . .	11
1.2.2	Dynamic Form of the Factor Model . . . . .	11
1.2.3	FAVAR Models . . . . .	15
1.2.4	Generalized Dynamic Factor Models . . . . .	17
1.2.5	Comparison of Dynamic Factor Models . . . . .	18
1.3	Selecting the Number of Factors and Specifying the Model . . . . .	19
1.3.1	Specification of DFMs . . . . .	19
1.3.2	GDFMs . . . . .	21
1.4	Structural Identification . . . . .	21
1.4.1	FAVAR Models . . . . .	21
1.4.2	Identification of Shocks in DFMs . . . . .	23
1.5	Applications . . . . .	28
1.6	Critique of Structural Analysis with FAVARs . . . . .	32
<b>2</b>	<b>Large Bayesian VAR Models</b>	<b>33</b>
2.1	Priors for Large Bayesian VARs . . . . .	34
2.2	Structural Identification in Large BVARs . . . . .	37
<b>3</b>	<b>Alternative Models</b>	<b>38</b>
3.1	Panel VARs . . . . .	38
3.2	Global VARs . . . . .	40
3.3	Other Ideas . . . . .	41
<b>4</b>	<b>Model Comparison and Discussion</b>	<b>41</b>

Typical vector autoregressive (VAR) models used for policy analysis include only small numbers of variables. On the other hand, in policy institutions such as central banks and government organisations large panels of variables are processed and used for policy decisions. If important variables are not included in a VAR model, there will be omitted variables bias in impulse responses etc. This suggests that one should include all variables that are potentially important in a structural vector autoregressive (SVAR) model. In other words, if a variable is not known to be irrelevant, one should in principle include it in the SVAR model. It should be understood that deciding on the importance of a particular variable in an empirical model is a difficult task for a number of reasons. For example, the variables for which data are available may not be exactly the ones that an economist has in mind in a theoretical model. As an example consider the Taylor rule. It includes the output gap as an explanatory variable which is not easy to measure. Thus, one would need to include all variables in the model related to or containing information on the output gap. They could all be important in an analysis that wants to investigate the impact of monetary policy. Moreover, it may be of interest to see the impact of monetary policy shocks at a more disaggregate level. For example, one may not only be interested in the response of the general price level to a monetary policy shock but also in the reaction of sub-indices of specific sectors of the economy. Likewise, one may be interested in the output response in specific sectors of the economy. In that case all variables of interest have to be included in the analysis.

On the other hand, the number of parameters in a VAR increases with the square of the number of variables included. Hence, in a conventional frequentist analysis estimation precision suffers from including many variables and degrees of freedom limitations keep the number of variables included in SVAR models low. Thus, the analyst often faces a dilemma in setting up the model. On the one hand, degrees of freedom considerations suggest including only a small number of variables whereas possible omitted variables bias and other considerations make a large number of variables desirable in a model. Therefore techniques have been developed that make it possible to include the information content in a large number of variables in a VAR model. A couple of possibilities, namely factor augmented VAR models and large Bayesian VAR models are discussed in this chapter. We also discuss other related methods such as panel VAR models and generalized VAR models briefly.

Factor augmented VAR models summarize the information contained in a large panel of variables in a small number of factors and include those factors in the SVAR analysis. By summarizing a large set of variables in factors these models impose additional structure on the data that reduces the dimensionality of the estimation problem and, hence, standard frequentist estimation and analysis methods can be used. The idea is to decompose the

observed variables in common factors and idiosyncratic components. The common factors incorporate the relations between the variables that are of central interest for a specific analysis. The factors can be static or dynamic.

In the next section a number of important properties of static factor models are reviewed. They are potentially useful for cross-sectional data and show some specific features of factor models that are important to understand when generalizing them for time series data. In Section 1.2 dynamic factor models (DFMs) for time series variables are presented as a general framework for factor models suitable for time series data. There are different representations of such models that will be discussed and that are also the basis for SVAR analysis. The problem of determining the appropriate number of factors is treated in Section 1.3. Structural identification is considered in Section 1.4. Applications are discussed in Section 1.5 and some critical thoughts about structural analysis with factor models are presented in Section 1.6.

There are a number of good surveys of factors in SVAR models, e.g., Stock and Watson (2005), Breitung and Eickmeier (2006), Barhoumi, Darné and Ferrara (2013). DFMs have been used extensively for forecasting (e.g., Stock and Watson (2002a, 2006, 2011) and many more listed in Breitung and Eickmeier (2006) and Barhoumi et al. (2013)). 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 many others. For a survey see Bai and Ng (2008).

Imposing Bayesian restrictions on the parameters of a VAR model is another alternative for dealing with many variables in a VAR analysis. Large BVAR models have gained popularity lately in particular in the context of forecasting. Examples of related studies are, for instance, Bańbura, Giannone and Reichlin (2010), Gupta, Jurgilas, Kabundi and Miller (2009), Carriero, Kapetanios and Marcellino (2009, 2012), Koop (2013). In Section 2 specific problems that result from including large panels of variables in the present context of structural modelling are discussed. Finally, some alternative approaches to fitting large VAR models are considered in Section 3. In particular, panel VARs and global VARs are treated. Concluding remarks with a critical evaluation of structural VAR modelling in the presence of many variables are presented in Section 4.

# 1 Factor Models

## 1.1 Static Factor Models

### 1.1.1 Model Setup

The classical static factor model is a model for cross-sectional data. It has the form

$$y_t = \Lambda f_t + v_t, \tag{1.1}$$

where  $y_t \sim \text{iid}(0, \Sigma_y)$  is a vector of  $K$  observed variables,  $f_t$  is an  $r$ -dimensional vector of unobserved *common factors* and  $r$  is typically much smaller than  $K$ ,  $r \ll K$ . Accordingly  $\Lambda$  is a  $(K \times r)$  matrix of *factor loadings*. Finally,  $v_t \sim \text{iid}(0, \Sigma_v)$  is a  $K$ -dimensional vector of uncorrelated *idiosyncratic components*, that is,  $\Sigma_v$  is diagonal. Moreover, the common factors and idiosyncratic components are assumed to be orthogonal, that is,  $E(f_t v_s') = 0$  for all  $s$  and  $t$ . Hence,

$$\Sigma_y = \Lambda \Sigma_f \Lambda' + \Sigma_v, \tag{1.2}$$

where  $\Sigma_f = E(f_t f_t')$  is the covariance matrix of the factors. If the factors are mutually uncorrelated, that is, if  $\Sigma_f$  is diagonal, the factors are said to be *orthogonal*. Otherwise they are *oblique*. This basic model has been used for statistical analysis already for many decades. For a detailed treatment see, e.g., Anderson (2003) who traces such models back to Spearman (1904). Notice that in the basic model (1.1) the observed variables are assumed to have mean zero. In practice that may require mean-adjustment prior to an analysis based on the model (1.1).

Obviously, in the model (1.1) the factors and factor loadings are not separately identified. For any nonsingular  $(r \times r)$  matrix  $Q$ , defining  $f_t^* = Q f_t$  and  $\Lambda^* = \Lambda Q^{-1}$  gives  $\Lambda f_t = \Lambda^* f_t^*$ . Thus, we may choose the factor loading matrix such that it has orthonormal columns, that is,

$$\Lambda' \Lambda = I_r, \tag{1.3}$$

or we may choose uncorrelated factors with variances normalized to 1,

$$f_t \sim (0, I_K). \tag{1.4}$$

In the latter case, the factors are orthogonal and

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

Such normalizations are useful for developing estimation algorithms. They are not sufficient for uniquely identifying the model. For instance, if we normalize the factors as in (1.4),  $\Lambda$

Table 1: Identification Conditions for Factors and Factor Loadings

Restrictions for $\Lambda$	Restrictions for $\Sigma_f$
$\Lambda'\Lambda = I_r$	$\Sigma_f$ diagonal with decreasing diagonal elements
$\Lambda'\Lambda$ diagonal with distinct, decreasing diagonal elements	$\Sigma_f = I_r$
$\Lambda = \begin{bmatrix} \lambda_{11} & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{r1} & \lambda_{r2} & \cdots & \lambda_{rr} \\ \vdots & \vdots & & \vdots \\ \lambda_{K1} & \lambda_{K2} & \cdots & \lambda_{Kr} \end{bmatrix}$ $\lambda_{ii} \neq 0, i = 1, \dots, r$	$\Sigma_f = I_r$
$\Lambda = \begin{bmatrix} I_r \\ \Lambda_2 \end{bmatrix}$	$\Sigma_f$ unrestricted

is still not unique without further restrictions. This can be seen by choosing an orthogonal matrix  $Q$  and defining  $\Lambda^* = \Lambda Q$ . Thereby we get a decomposition

$$\Sigma_y = \Lambda^* \Lambda^{*'} + \Sigma_v.$$

Uniqueness (identification) can be ensured by choosing  $\Lambda$  such that (1.3) holds and the factors such that they are mutually uncorrelated, that is,  $\Sigma_f$  is a diagonal matrix and the diagonal elements are distinct and ordered from largest to smallest. In other words, the first factor has the largest variance and, hence, explains the largest part of the variance of  $y_t$  that is explained by common factors. The second factor,  $f_{2t}$ , has the second largest variance etc.. The requirement that the factor variances have to be distinct ensures that the columns of  $\Lambda$  cannot simply be reordered. In Table 1 some sets of identification conditions for factors and factor loadings from Bai and Ng (2013) are presented. It should be noted that even when these conditions are satisfied, the  $\Lambda$  matrix is unique only up to sign changes of its columns. For a thorough discussion of identification conditions see also Anderson (2003, Section 14.2.2). If the model parameters are identified, they can be estimated straightforwardly. Being aware of conditions for uniqueness of the factors is also important for identifying shocks of interest, as will be seen in Section 1.4.

### 1.1.2 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 (1.1) with  $\Lambda'$  and dropping the idiosyncratic

term,

$$\hat{f}_t = \Lambda' y_t. \quad (1.5)$$

In practice the factor loadings are typically unknown. A possible objective function for estimation in that case is the sum of 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 so as to minimize the sum of squared errors,

$$\min_{\Lambda, f_1, \dots, f_T} T^{-1} \sum_{t=1}^T (y_t - \Lambda f_t)' (y_t - \Lambda f_t) = \min_{\Lambda, f_1, \dots, f_T} \text{tr} T^{-1} \sum_{t=1}^T (y_t - \Lambda f_t)(y_t - \Lambda f_t)'. \quad (1.6)$$

A solution to this minimization problem is obtained by considering the  $r$  largest eigenvalues  $\lambda_1 > \dots > \lambda_r$  of  $S_y = T^{-1} \sum_{t=1}^T y_t y_t'$  with corresponding orthonormal eigenvectors  $\lambda_1, \dots, \lambda_r$ , choosing  $\hat{\Lambda} = [\lambda_1, \dots, \lambda_r]$  and using  $\hat{f}_t = \hat{\Lambda}' y_t$ . Notice that  $\hat{\Lambda}$  is the so-called *principal components* (PC) estimator of  $\Lambda$ . Given the orthogonality of the eigenvectors, it satisfies  $\hat{\Lambda}' \hat{\Lambda} = I_r$ . The factors are the principal components and  $\hat{\Sigma}_f = T^{-1} \sum_{t=1}^T f_t f_t' = \hat{\Lambda}' S_y \hat{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_r)$ , that is, the eigenvalues  $\lambda_1, \dots, \lambda_r$  are the empirical variances of the factors so that their variances are ordered from largest to smallest.

Asymptotic properties of estimators of factor models can be found in Anderson (2003, Chapter 14) for  $T \rightarrow \infty$  and fixed  $K$  and results for more general factor models under the assumption that both the number of components  $K$  and the sample size  $T$  go to infinity are derived by Stock and Watson (2002a), Bai (2003) and many others. In particular, these authors show consistency of the estimators and asymptotic normality if  $K$  and  $T$  go to infinity at suitable rates and suitable normalisations are made. In addition some further regularity conditions are necessary for these results to hold (see Bai and Ng (2008) for a review of conditions and results).

The PC estimator is the ML estimator if the observations  $y_t$  come from a normal distribution and the idiosyncratic components have equal variances,  $\Sigma_v = \sigma^2 I_K$ . In other words, it is assumed that the factors and idiosyncratic components are normally distributed,  $f_t \sim \text{iid} \mathcal{N}(0, \Sigma_f)$  and  $v_t \sim \text{iid} \mathcal{N}(0, \sigma^2 I_K)$  (see Anderson (2003)). If the variances of the idiosyncratic components are heterogeneous,  $\Sigma_v = \text{diag}(\sigma_1^2, \dots, \sigma_K^2) \neq \sigma^2 I_K$ , the log-likelihood becomes

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

Anderson (2003, Section 14.4) points out that the likelihood function is unbounded in general and, hence, it does not have a global maximum. Thus, standard ML estimation cannot be

used. Instead a local maximum in the neighbourhood of the true parameter vector has to be considered (e.g., Breitung and Tenhofen (2011) and Bai and Li (2012)).

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

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

See Choi (2012) for FGLS procedures for factor models and related asymptotic properties of the estimators.

If the normalization in (1.4) is used for the common factors and the observations are normally distributed, ML estimation of the factor loadings and idiosyncratic variances is achieved by maximizing the log-likelihood

$$\begin{aligned} \log l(\Lambda, \Sigma_v) &= \text{constant} - \frac{T}{2} \log |\Sigma_y| - \frac{1}{2} \text{tr}(T S_y \Sigma_y^{-1}) \\ &= \text{constant} - \frac{T}{2} \log |\Lambda \Lambda' + \Sigma_v| - \frac{1}{2} \text{tr} [T S_y (\Lambda \Lambda' + \Sigma_v)^{-1}]. \end{aligned}$$

Again this maximization problem calls for numerical methods. Suitable algorithms are discussed, for instance, by Magnus and Neudecker (1988, Chapter 17).

### 1.1.3 Approximate Static Factor Models

So far we have considered what might be called an *exact static factor model* where the idiosyncratic components are clearly separated from each other and the factors. For economic data such an assumption may be too strict, in particular, if large sets of variables are considered. In that case, one may want to assume that there are infinitely many potentially interesting variables and a model could be specified under the assumption that  $K \rightarrow \infty$ . Such a model was, for instance, considered by Chamberlain and Rothschild (1983) for investigating a market with many assets (see also Connor and Korajczyk (1986, 1993)). In that case it is of interest to look at *approximate factor models* that allow for some correlation within the idiosyncratic components or, in other words, models where the common factors do not fully capture all relations between the observed variables, that is,

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

where  $\Sigma_v$  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 to exist if  $\Sigma_y$  has only  $r$  unbounded eigenvalues when  $K \rightarrow \infty$ . The common factors are defined by the requirement that there exists a sequence of  $(K \times r)$  matrices  $\Lambda$

and positive definite covariances  $\Sigma_v$  such that

$$\Sigma_y = \Lambda\Lambda' + \Sigma_v$$

and the maximum eigenvalue of  $\Sigma_v$  is bounded when  $K \rightarrow \infty$ . Thus, the relative variance share of each idiosyncratic components is small when the number of variables is large.

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

Approximate factor models are particularly relevant if time series data are considered. We turn to that case next.

## 1.2 Dynamic Factor Models

If time series data are under consideration, taking into account the serial dependence is essential for forecasting and structural analysis. Hence, dealing with models that capture dynamic relations is important. In other words, in the context of factor models the classical static model has to be generalized to allow for dynamic structures. Of course, serial dependence may well be represented by a model of the form (1.1) if  $f_t$  and  $v_t$  are not serially uncorrelated or independent. Thus, dynamic factor models are obtained by allowing  $f_t$  and  $v_t$  to be general stochastic processes. If we remain in the stationary world, a natural extension of the covariance decomposition in (1.2) would be a decomposition of the spectral density of  $y_t$ . Denoting the spectral density functions of  $y_t$ ,  $f_t$  and  $v_t$  by  $\Sigma_y(\xi)$ ,  $\Sigma_f(\xi)$  and  $\Sigma_v(\xi)$ , respectively,

$$\Sigma_y(\xi) = \Lambda\Sigma_f(\xi)\Lambda' + \Sigma_v(\xi), \tag{1.7}$$

where  $\Sigma_v(\xi)$  is assumed to be a diagonal matrix in the exact dynamic factor model while more general assumptions are made in an approximate dynamic factor model.

Different dynamic factor models for time series data that decompose the spectral density of the observations in the way shown in (1.7) have been proposed in the literature. Clearly, it depends on the assumptions made for the stochastic processes considered for  $f_t$  and  $v_t$  which model is obtained. A number of special cases are considered in the following.

If  $v_t$  is white noise, that is,  $\Sigma_v(\xi) = \Sigma_v$ , then  $y_t$  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 assume that the factors have a vector ARMA generation process

and the  $\Sigma_v$  matrix is not necessarily diagonal. For inference purposes there is no difference to the static factor model (e.g., Choi (2012)). This case is therefore not specifically considered here. From a practical point of view such models are typically too restrictive.

If the model can be written in the form (1.1) and  $f_t$  and  $v_t$  have parametric VAR representations, then the model is a dynamic factor model in static form. In fact, Boivin and Ng (2005), for example, simply call this model a static factor model to distinguish it from a model where lagged factors  $f_{t-j}$  appear on the right-hand side of (1.1) in addition to the contemporaneous factors. We do not use this terminology here because, as we will see, dynamic factor models in the sense of Boivin and Ng (2005) can always be written in static form. Instead we call any dynamic factor model with parametric VAR representation of the factors and idiosyncratic components a dynamic factor model (DFM). A more general model where the common component and the idiosyncratic components are general stochastic processes is called a generalized dynamic factor model (GDFM). This terminology is in line with some of the related literature. Generally, when reading that literature, it is worth checking which assumptions precisely are made and which terminology is used.

The remainder of this section is structured as follows. In Subsection 1.2.1 the static form of a DFM is presented and its estimation is discussed. In Subsection 1.2.2 the dynamic form of a DFM is considered and it is shown how it can be written in static form. A specific variant of the DFM contains observed variables in addition to dynamic factors. It is the model that is commonly known as factor augmented VAR (FAVAR) model and is considered in Section 1.2.3 because it is of particular importance for structural VAR analysis. Finally the GDFM is presented in Section 1.2.4. Estimation methods for all the models are also presented.

In this section we have in mind models for stationary variables. Factor models can also be considered for integrated variables although in that case the inference and analysis methods have to be modified. In fact, a model with cointegrated variables captures the common trends in a system of integrated variables. The common trends can be viewed as common factors. Given the differences in inference procedures relative to stationary models, it is perhaps not surprising that adjustments are necessary if the variables have stochastic trends. In fact, the estimation procedures presented in the following are based on covariance matrix and spectral density estimates that are not meaningful for integrated variables. Extensions of factor models to allow for integrated variables can be found in the cointegration literature or Bai (2004), for example.

### 1.2.1 Static Form of a DFM

Consider the model (1.1),

$$y_t = \Lambda^f f_t + v_t, \quad (1.8)$$

with dynamic factors being generated as

$$f_t = \Gamma_1 f_{t-1} + \dots + \Gamma_s f_{t-s} + \eta_t \quad \text{and} \quad v_t = A_1 v_{t-1} + \dots + A_p v_{t-p} + u_t,$$

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

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

where  $A(L) = \text{diag}[\alpha_1(L), \dots, \alpha_K(L)]$ . This model is called a static form of a DFM because the relation between the observed  $y_t$  and the dynamic factors can be described as instantaneous, that is, no lagged  $f_t$  appears in (1.8).

**Estimation** Estimation of the factors and factor loadings in (1.8) for a given number of factors,  $r$ , can be done by PC, ignoring all serial dependence in the error terms. Bai (2003) derives properties of the estimators. 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. In fact, it works even if the model is just an approximate factor model with a more general dependence structure of the error terms. Bai and Li (2012) discuss maximum likelihood estimation of such models.

### 1.2.2 Dynamic Form of the Factor Model

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

$$y_t = \Lambda_0^f f_t + \Lambda_1^f f_{t-1} + \dots + \Lambda_{q^*}^f f_{t-q^*} + v_t. \quad (1.9)$$

Assuming the same generation mechanisms for  $f_t$  and  $v_t$  as in the static form (1.8), the model can be written in lag operator notation as

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

where

$$A(L) = \text{diag}[\alpha_1(L), \dots, \alpha_K(L)],$$

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

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

$f_t = (f_{1t}, \dots, f_{rt})'$  are the common factors as before,  $v_t = (v_{1t}, \dots, v_{Kt})'$  is the vector of idiosyncratic components and  $\eta_t$  is white noise such that  $E(u_t \eta_s') = 0 \forall t, s$ .

Defining  $F_t = (f_t', \dots, f_{t-q^*}')'$  and  $\Lambda^F = [\Lambda_0^f, \Lambda_1^f, \dots, \Lambda_{q^*}^f]$  the model (1.9) can be written in static form,

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

where just the dimension of the factor vector is larger. It 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*.

Left-multiplying (1.9) by  $A(L)$  gives

$$A(L)y_t = \Lambda(L)f_t + u_t, \tag{1.10}$$

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 (1.10) can be written in static form as

$$A(L)y_t = \Lambda F_t + u_t, \quad F_t = \mathbf{\Gamma} F_{t-1} + G\eta_t, \tag{1.11}$$

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

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

$(R \times R)$   $(R \times r)$

Here  $R = r(q+1)$  and  $\Gamma_i = 0$  for  $i > s$ . The overall model in *VAR form* can be written as

$$\begin{bmatrix} I_R - \mathbf{\Gamma}L & 0 \\ -\Lambda\mathbf{\Gamma}L & A(L) \end{bmatrix} \begin{bmatrix} F_t \\ y_t \end{bmatrix} = \begin{bmatrix} G\eta_t \\ \Lambda G\eta_t + u_t \end{bmatrix}. \tag{1.12}$$

This DFM is a restricted version of the factor-augmented VAR (FAVAR) model considered in the next subsection. In particular, the VAR coefficient matrices contain a block of zeros and the residuals have a specific structure.

Following Chamberlain and Rothschild (1983), Stock and Watson (2005) call the model (1.9) an *exact DFM* if  $A(L)$  has a diagonal structure and the error covariance matrix

$E(u_t u_t') = \Sigma_u$  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 *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 (1.9) by their assumptions for 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 more dependence between the idiosyncratic components. In the following we treat  $A(L)$  and  $\Sigma_u$  as diagonal, unless otherwise specified.

Before we discuss the unrestricted FAVAR model we consider estimation of the restricted model (1.12).

**Estimation** We now discuss estimation of DFMs for a given number of lags and a given number of factors. Of course, these quantities have to be decided first. It is still useful to consider estimation for given numbers of lags and factors because determining these quantities requires estimation of the models. We discuss model specification later.

Before estimating a DFM it may be a good idea to scale the variables such that they have zero mean and variance one, that is, one may want to mean-adjust the variables and scale them by the sample standard deviation. Of course, the static form of the DFM can be estimated easily as described earlier. Following Stock and Watson (2005), the dynamic form of the DFM can be estimated as follows:

**Step 1** Get an initial estimate  $\tilde{A}(L)$  of  $A(L) = \text{diag}(\alpha_{11}(L), \dots, \alpha_{KK}(L))$ , for example, by regressing the individual variables on their own lags.

**Step 2** Compute the PC estimator  $\hat{\Lambda}$  of  $\Lambda$  from the model  $\tilde{A}(L)y_t = \Lambda F_t + \tilde{u}_t$  and estimate the factors as  $\hat{F}_t = \hat{\Lambda}' \tilde{A}(L)y_t$ .

**Step 3** Estimate  $A(L)y_t = \Lambda \hat{F}_t + \hat{u}_t$  by single equation OLS for each equation separately to get new estimates of  $A(L)$  and  $\Lambda$  and choose  $\hat{F}_t = \hat{\Lambda}' \hat{A}(L)y_t$ .

**Step 4** Iterate Step 3 until convergence.

Using single equation OLS in Step 3 is justified because the idiosyncratic error terms are assumed to be instantaneously uncorrelated, that is,  $\Sigma_u$  is a diagonal matrix. If that assumption is false estimation efficiency can be improved by using a feasible GLS procedure

because the regressors in the different equations of the system  $A(L)y_t = \Lambda \widehat{F}_t + \hat{u}_t$  are not identical if the diagonal structure of  $A(L)$  is taken into account.

Once the estimated factors  $\widehat{F}_t$  are available, the  $\mathbf{\Gamma}$  coefficient matrix of the transition equation in (1.11) can then be estimated by regressing  $\widehat{F}_t$  on  $\widehat{F}_{t-1}$ . Thereby we have estimates of all the parameters in the model (1.12).

Unfortunately, the procedure will only deliver a linear transformation of the true factors because the PC estimator uses just some statistical normalisation or identification that may not result in the primitive dynamic factors,  $f_t$ , of which the static factors,  $F_t$ , are composed. Therefore we need to do another PC estimation to determine the  $r$  linearly independent factors  $f_t$  underlying  $F_t$ . Let  $\widehat{W}$  be the matrix of eigenvectors corresponding to the  $r$  largest eigenvalues of the residual covariance matrix  $\widehat{\Sigma}_U = T^{-1} \sum_t \widehat{U}_t \widehat{U}_t'$ , where  $\widehat{U}_t = \widehat{F}_t - \widehat{\mathbf{\Gamma}}^* \widehat{F}_{t-1}$  and  $\widehat{\mathbf{\Gamma}}^*$  is an estimator obtained by regressing  $\widehat{F}_t$  on  $\widehat{F}_{t-1}$ . Then  $\widehat{\eta}_t = \widehat{W}' \widehat{U}_t$  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  $\eta_t$  can be estimated in the usual way using the covariance estimator of the latter regression. Alternatively, it may be based on the  $\widehat{\eta}_t$ , that is,  $\widehat{\Sigma}_\eta = T^{-1} \sum_t \widehat{\eta}_t \widehat{\eta}_t'$  may be used. Methods for choosing the values of  $R$  and  $r$  required in this procedure are discussed in Section 1.3.

It may be worth pointing out that there does not appear to be a standard procedure in the literature for estimating the  $\Gamma_1, \dots, \Gamma_{q+1}$ . The one presented here may not be generally appealing. Perhaps this is one reason for the growing popularity of the FAVAR approach presented in the next section.

It is also possible to use ML estimation under normality assumptions for all parameters simultaneously, that is, one may set up the log-likelihood and maximize that by some nonlinear optimization algorithm. The actual evaluation of the log-likelihood can be done with the Kalman filter because (1.11) is in state space form. The computations may still be challenging if a large panel of variables is considered. Doz, Giannone and Reichlin (2011) propose an alternative two-step estimator based on the Kalman filter that may be helpful for large panels of variables.

Asymptotic results for 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 asymptotic properties are obtained for  $T$  and  $K \rightarrow \infty$ , small sample results by Boivin and Ng (2006) indicate that including more variables in a factor analysis does not necessarily result in better estimates. In particular, they find that including more variables may not improve forecasts of an approximate factor model.

Rather than using frequentist estimation 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 2 and therefore do not discuss these methods here but just mention that they have been used by Otrok and Whiteman (1998), Kose, Otrok and Whiteman (2003) and Amir Ahmadi and Uhlig (2009), for example, in the context of estimating dynamic factor models.

### 1.2.3 FAVAR Models

Bernanke, Boivin and Elias (2005) consider a more general, unrestricted version of a FAVAR model,

$$\mathbf{A}(L) \begin{bmatrix} F_t \\ y_t \end{bmatrix} = w_t, \quad (1.13)$$

where  $w_t$  is  $(R + K)$ -dimensional white noise,  $\mathbf{A}(L) = \mathbf{A}_0 + \mathbf{A}_1L + \dots + \mathbf{A}_pL^p$  is a  $((R + K) \times (R + K))$  matrix operator,  $F_t$  is a vector of  $R$  unobserved common factors that are related to a large number of  $N$  *informational* variables  $x_t$  by the observation equation

$$x_t = \Lambda^F F_t + \Lambda^y y_t + e_t. \quad (1.14)$$

The  $K$  observed variables  $y_t$  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  $y_t$  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. Hence, although the DFM (1.12) may be viewed as a restricted FAVAR model, alternatively the observation equation in (1.14) can be interpreted as a specific DFM where some of the factors are observed variables. Thus, it is not clear which of the models should be viewed as more general.

For identifying the factors, Bernanke et al. (2005) assume that the upper  $(R \times R)$  block of  $\Lambda^F$  is an identity matrix and the upper  $(R \times K)$  block of  $\Lambda^y$  is a zero matrix. They mention that these conditions are sufficient for identification. Clearly, they are not necessary and they may well be over-identifying and, hence, may imply unwanted distortions.

As an example Bernanke et al. (2005) consider the following variables: output,  $q_t$ , potential output,  $q_t^p$ , a cost push shock,  $s_t$ , an inflation rate,  $\pi_t$ , and a nominal interest rate,  $R_t$ , and they mention that variables such as the output gap and a cost push shock are not observable and, hence, should be replaced by a set of informational variables. In fact, one could even go further and argue that policy makers view variables such as output and inflation as latent variables that are measured as factors.

**Estimation** Bernanke and Boivin (2003) augment a VAR model by factors for forecasting purposes. They estimate the factors in a first step by PC analysis and then include them

in a VAR model together with observed variables. This procedure is inspired by Stock and Watson (2002b). It is also used by Favero, Marcellino and Neglia (2005) and, in modified form, by Bernanke et al. (2005) for estimating the factors and their FAVAR model. The factors are estimated in the first step by a PC analysis of the large set of informational variables,  $x_t$ , that do not include the observed variables of interest,  $y_t$ . Then the FAVAR model (1.13) is estimated with the estimated factors replacing the true factors. Bernanke et al. (2005) also use another method that leads to similar conclusions, suggesting that the method works reasonably well.

Alternatively, estimation of the factors may be based directly on the observation equation (1.14). We define  $X = [x_1, \dots, x_T]$ ,  $F = [F_1, \dots, F_T]$ ,  $Y = [y_1, \dots, y_T]$ , and  $E = [e_1, \dots, e_T]$  and write the relation in matrix form as

$$X = \Lambda^F F + \Lambda^y Y + E. \tag{1.15}$$

This form suggests the following iterative estimation procedure inspired by the approach used in Boivin and Giannoni (2009).

**Step 1. Initial estimate of  $\Lambda^y$**

$$\hat{\Lambda}_{(0)}^y = XY'(YY')^{-1}$$

**Step 2. Iteration for  $i = 1, 2, \dots$**

Let  $\hat{\Lambda}_{(i)}^F$  be the  $R$  largest PCs of  $T^{-1}(X - \hat{\Lambda}_{(i-1)}^y Y)(X - \hat{\Lambda}_{(i-1)}^y Y)'$ . Compute  $\hat{F}_{(i)} = \hat{\Lambda}_{(i)}^{F'}(X - \hat{\Lambda}_{(i-1)}^y Y)$  and  $\hat{\Lambda}_{(i)}^y = (X - \hat{\Lambda}_{(i)}^F \hat{F}_{(i)})Y'(YY')^{-1}$ .

Iterating the second step until convergence avoids the identification restrictions for  $\Lambda^y$  that are imposed by Bernanke et al. (2005).

**Extensions** Dufour and Stevanović (2013) extend the FAVAR model to a factor augmented vector autoregressive moving average (FAVARMA) model. They argue that if the factors are driven by a finite order VAR process this implies a mixed VARMA generation process for the  $y_t$  variables in  $y_t = \Lambda F_t + v_t$ . Hence it is natural to consider VARMA rather than VAR models. They also propose a suitable estimation procedure for such models and find improved forecast performance of a FAVARMA model for U.S. macro data relative to a FAVAR model.

Banerjee and Marcellino (2008) and Banerjee, Marcellino and Masten (2013, 2014) consider factor augmented cointegrated VAR models and set them up in error correction form. They abbreviate the factor augmented VECM as FECM and discuss estimation, forecasting

and structural analysis based on such models. The advantage of FECMs is that they explicitly allow for integrated variables whereas standard dynamic factor models are typically designed for stationary variables. An obvious advantage of including integrated variables in levels is that the models can also capture cointegration relations.

#### 1.2.4 Generalized Dynamic Factor Models

The *generalized dynamic factor model* (GDFM) generalizes (1.8) by allowing the common and idiosyncratic components to be general stationary processes that may not admit a finite order VAR representation, that is, we consider a model  $y_t = \Lambda f_t + v_t$ , 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 via their frequency domain or spectral properties as in (1.7) (see Forni and Lippi (2001) and Forni et al. (2000, 2004, 2005)). The decomposition in (1.7) also suggests estimation methods which are presented now.

**Estimation** Estimation of GDFMs is considered by Forni and Reichlin (1998) and Forni et al. (2000, 2004, 2005). Since they do not assume a parametric model for the generation process of the observables and the common factors, they use a nonparametric frequency domain PC analysis as developed by Brillinger (1975). Based on the work of Forni et al., Favero et al. (2005) propose the following procedure for estimating the dynamic PCs and common components:

**Step 1** For a sample  $y_1, \dots, y_T$  of size  $T$ , the spectral density matrix of  $y_t$  is estimated as

$$\widehat{\Sigma}_y(\xi_j) = \sum_{m=-M}^M w_m \widehat{\Gamma}_y(m) e^{-im\xi_j}, \quad \xi_j = 2\pi j / (2M + 1), \quad j = 0, 1, \dots, 2M,$$

where  $M$  is the window width,  $w_m = 1 - |m| / (M + 1)$  are the weights of the Bartlett window and  $\widehat{\Gamma}_y(m) = T^{-1} \sum_t (y_t - \bar{y})(y_{t-m} - \bar{y})'$  is the sample covariance matrix of  $y_t$  for lag  $m$ . The window width has to be chosen such that  $M \rightarrow \infty$  and  $M/T \rightarrow 0$  as  $T \rightarrow \infty$ . Forni et al. (2000) remark that a choice of  $M = 2T^{1/3}/3$  worked well in simulations.

**Step 2** For  $j = 0, 1, \dots, 2M$ , determine the eigenvectors  $\lambda_1(\xi_j), \dots, \lambda_r(\xi_j)$  corresponding to the  $r$  largest eigenvalues of  $\widehat{\Sigma}_y(\xi_j)$ .

**Step 3** Defining

$$\lambda_{mk} = \frac{1}{2M + 1} \sum_{j=0}^{2M} \lambda_m(\xi_j) e^{ik\xi_j}, \quad k = -M, \dots, M,$$

the dynamic PCs of  $y_t$  are obtained as

$$\hat{f}_{mt} = \sum_{k=-M}^M \lambda'_{mk} y_{t-k}, \quad m = 1, \dots, r,$$

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

**Step 4** Run a regression

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

and estimate the common component as

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

where  $\hat{\Lambda}_j$ ,  $j = -q, \dots, p$ , are the OLS estimators, that is, the common component  $\chi_t$  is estimated as the fitted value of the regression. 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 seem to be used.

Using leads of the estimated factors and, hence, of the observations in Step 4 to reconstruct the common component may be a disadvantage in forecasting and impulse response analysis. Therefore it may be worth knowing that a one-sided procedure has been proposed (see Forni, Hallin, Lippi and Reichlin (2005)). For impulse response analysis it is in fact not clear that a one-sided procedure has advantages. Although impulse responses can be seen as forecasts, estimating them from the full sample is also common in standard VAR analysis. Hence, we have presented the two-sided procedure here.

### 1.2.5 Comparison of Dynamic Factor Models

As we have seen in the foregoing, the various forms of the dynamic factor models are to some extent just different representations of the same data generation process. The particular form used in a specific analysis is a matter of convenience in setting up inference procedures or using them in a specific analysis. Of course, the models also differ to some extent in the underlying assumptions. This is in particular true for DFM and GDFM models. The latter allow in principle more general underlying dynamics of the factors and idiosyncratic components. As noted earlier, GDFMs allow the factors and idiosyncratic errors to be general stochastic processes while DFMs focus on finite order VAR and AR processes for these quantities. But even that distinction is not substantial in practice because stationary processes can be approximated arbitrarily well by finite order AR or VAR processes if the lag order is not restricted. Thus, for practical purposes the choice of model may just be made by

convenience and personal preference. This issue is to some extent important when it comes to structural analysis. The model or model form that is most suitable for identifying the shocks of interest is then the model of choice. Technical considerations may be of secondary importance.

Before we discuss structural analysis in more detail, we first consider choosing the number of factors.

### 1.3 Selecting the Number of Factors and Specifying the Model

A full specification of a DFM requires selecting the number of common factors and the various lag orders. Since a PC analysis does not require that the lag order is specified, it is in fact possible to specify the number of static factors before lag orders of the VAR operators are determined. In classical static factor models, subjective criteria such as choosing as many factors as are necessary to explain a prespecified fraction of the overall variance have been used traditionally. More precisely, if PC analysis is used, the sum of the variances of the PCs considered as common factors has to exceed a prespecified fraction of the sum of the eigenvalues of the sample covariance matrix. Another criterion of that kind is the so-called *scree test* proposed by Cattell (1966). It is based on assessing for which number of factors the variance explained by the factors starts to taper off.

#### 1.3.1 Specification of DFMs

For DFMs more formal criteria have been developed that assume 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 ( $K, T \rightarrow \infty$ ). The most popular criteria are proposed by Bai and Ng (2002) and take the form

$$IC(R) = \log V(R) + Rg(K, T) \tag{1.16}$$

where  $V(R) = (KT)^{-1} \sum_{t=1}^T (y_t - \hat{\Lambda}^F \hat{F}_t)' (y_t - \hat{\Lambda}^F \hat{F}_t)$ . Notice the similarity with information criteria for VAR order selection. These authors 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 minimum condition is that  $R_{max} \geq R_0$ . Moreover, the penalty term  $g(K, T)$  has to go to zero at a suitable rate with growing  $T$  and  $K$ . According to Breitung and Eickmeier (2006), the most popular criterion from this class chooses  $g(K, T) = \left(\frac{K+T}{KT}\right) \log(\min[K, T])$ , that is, the criterion becomes

$$IC_{p2}(R) = \log V(R) + R \left(\frac{K+T}{KT}\right) \log(\min[K, T]). \tag{1.17}$$

Using this criterion we can estimate the number of static factors, that is, the dimension of  $F_t$ .

For structural analysis the number of primitive dynamic factors, that is, the dimension  $r$  of  $f_t$  in the dynamic model form (1.10) is of prime interest, however. Bai and Ng (2007) propose a procedure for determining them or the number of related primitive shocks, as they call them. They utilize that the error term in the transition equation in the static form (1.11),  $G\eta_t$ , 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 present framework that VAR model is of order one because we have assumed that  $q \geq s$ . Thus, fitting

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

gives estimated residuals  $\widehat{U}_t$ ,  $t = 1, \dots, T$  (see Section 1.2.2). 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  $T^{-1} \sum_{t=1}^T \widehat{U}_t \widehat{U}_t'$  and define

$$\widehat{D}_1(r) = \left( \frac{\rho_{r+1}^2}{\sum_{i=1}^R \rho_i^2} \right)^{1/2} \quad (1.18)$$

and

$$\widehat{D}_2(r) = \left( \frac{\sum_{i=r+1}^R \rho_i^2}{\sum_{i=1}^R \rho_i^2} \right)^{1/2}. \quad (1.19)$$

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

$$\widehat{r} = \min r \in \left\{ r : \widehat{D}_1(r) < \frac{1}{\min(K^{1/2-\delta}, T^{1/2-\delta})} \right\} \quad (1.20)$$

or as

$$\widehat{r} = \min r \in \left\{ r : \widehat{D}_2(r) < \frac{1}{\min(K^{1/2-\delta}, T^{1/2-\delta})} \right\}, \quad (1.21)$$

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

Notice that by choosing the number of static factors and the underlying dynamic primitive factors suggests that the lag length of  $\Lambda(L)$  in the DFM (1.10) cannot be longer than  $R/r$ . 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 publications 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).

### 1.3.2 GDFMs

There are also methods specifically designed to estimate the number of factors in GDMFs. 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  $\lambda_1(\xi_j), \dots, \lambda_K(\xi_j)$  of the estimated spectral density matrix  $\widehat{\Sigma}_y(\xi_j)$ ,  $\xi_j = 2\pi j/(2M + 1)$ ,  $j = 0, 1, \dots, 2M$ . They are defined as

$$PCP(k) = \frac{1}{K} \sum_{j=k+1}^K \frac{1}{2M+1} \sum_{m=-M}^M \lambda_j(\xi_m) + k\varphi(K, T)$$

and

$$IC(k) = \log \left( \frac{1}{K} \sum_{j=k+1}^K \frac{1}{2M+1} \sum_{m=-M}^M \lambda_j(\xi_m) \right) + k\varphi(K, T).$$

The bandwidth  $M$  has to be such that  $M \rightarrow \infty$  and  $M/T \rightarrow 0$  for  $T \rightarrow \infty$ , the penalty term has to be such that

$$\varphi(K, T) \rightarrow 0 \quad \text{and} \quad \min\{K, M^{-2}, (MT)^{1/2}\}\varphi(K, T) \rightarrow \infty$$

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

## 1.4 Structural Identification

### 1.4.1 FAVAR Models

If the FAVAR model (1.13) is considered, structural shocks can be recovered as in a standard VAR model by a linear transformation of the reduced form residuals, that is, the structural shocks are obtained as  $\varepsilon_t = B^{-1}w_t$ . In that setup identification of the shocks can be done as in a conventional VAR model. Of course, the number of variables and, hence, potential shocks may be larger which may make identification more difficult. However, if only some of the shocks are of interest, only those shocks have to be identified.

For example, Favero et al. (2005) first extract factors from large data sets and then use these as additional variables in a FAVAR model. They order the policy interest rate last and as they are only interested in the effects of monetary policy shocks, they use a recursive identification scheme, that is, the impact effects matrix is lower triangular. In other words, the variables in their FAVAR are

$$(F_t', y_t^{*'}, r_t)',$$

where  $y_t^*$  contains all observed key variables apart from the interest rate  $r_t$ . They assume that a monetary policy shock has no instantaneous impact on any of the observed variables and the factors. Since the other shocks are not of interest in their analysis they can be identified arbitrarily.

Of course, the assumption that none of the factors and observed variables reacts instantaneously (or, more precisely, within the period of sampling frequency) to a monetary policy shock may be regarded as restrictive, in particular, if fast-moving financial variables are included in the model. If one wants to avoid such an assumption, one could split up the variables in fast-moving and slow-moving variables, as in Bernanke et al. (2005) and extract factors separately from the two groups of variables. Then one could order the slow-moving factors ( $f_t^s$ ) before the interest rate and the fast-moving factors ( $f_t^f$ ) behind it so that they can be instantaneously affected in a lower-triangular recursive identification scheme. In other words, the variables in the FAVAR are arranged as follows:

$$(f_t^{s'}, y_t^{s'}, r_t, f_t^{f'}, y_t^{f'})',$$

where  $y_t^s$  and  $y_t^f$  contain the slow- and fast-moving observed variables, respectively, in the system. 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. Maximum likelihood and Bayesian methods can in principle be used although they may be more difficult to implement than methods that impose only identifying (uniqueness) restrictions on the loading matrix.

In this approach no overall model for the DGP is considered and the factors are treated as actual variables measured without errors. Clearly this is a strong assumption, although, as we will see shortly, also an impulse response analysis based on DFMs requires strong assumptions. In any case, if the FAVAR setup (1.13) is used for impulse response analysis, the responses of the informational variables are not obtained. They may also be of interest and can be obtained via (1.14) by considering

$$x_t = [\Lambda^F : \Lambda^y]A(L)^{-1}w_t + e_t. \tag{1.22}$$

Alternatively, impulse response analysis can be performed in the framework of DFMs or GDFMs.

### 1.4.2 Identification of Shocks in DFMs

In a DFM the dynamics of the system and in particular the relationships between the variables are determined by the factors. Hence, the shocks are also assumed to be transmitted through the factors and we replace the factors in the DFM by their MA representation and get a *reduced form*

$$y_t = \Phi(L)\eta_t + v_t, \quad (1.23)$$

where  $\Phi(L) = \Lambda^f(L)\Gamma(L)^{-1} = A(L)^{-1}\Lambda(L)\Gamma(L)^{-1} = A(L)^{-1}\Lambda(I_R - \Gamma L)^{-1}G$  if we start from (1.9). For our discussion of structural forms and identifying structural shocks we assume that the reduced form parameters  $\Phi(L)$  and  $\Sigma_\eta$  are known. They can be estimated from the data as discussed in Section 1.2.

Assuming as usual that the  $(r \times 1)$  vector of reduced form residuals  $\eta_t$  is related to the  $(r \times 1)$  vector of structural shocks  $\varepsilon_t$  by a linear transformation  $\eta_t = B\varepsilon_t$ , the structural form corresponding to (1.23) is

$$y_t = \Phi(L)B\varepsilon_t + u_t. \quad (1.24)$$

If the structural shocks are instantaneously uncorrelated and the variances are normalized to 1 we get  $\varepsilon_t \sim (0, I_r)$ . Hence,  $B$  has to satisfy  $BB' = \Sigma_\eta$  and, as in the standard case, we need at least  $r(r-1)/2$  more restrictions for identification of the  $(r \times r)$  matrix  $B$ . In other words, identifying the structural shocks requires putting enough restrictions on  $B$  to obtain uniqueness. These restrictions may come in the form of exclusion restrictions on the impact effects or the long-run effects of the shocks. They may also be available in the form of sign restrictions. Some specific restrictions are discussed in the following.

**Restrictions on the Impact Effects of Shocks** Notice that the impact matrix  $\Phi_0$  in  $\Phi(L) = \sum_{i=0}^{\infty} \Phi_i L^i$  will in general not be an identity matrix. In fact,  $\Phi(L)$  is  $(K \times r)$  and is typically not a square matrix. Therefore the impact effects of the shocks are given by  $\Phi_0 B$  and exclusion restrictions on the impact effects are zero restrictions on the elements of the matrix product  $\Phi_0 B$ . For example, one may want to impose a recursive identification scheme on the impact effects, as is often done in a standard SVAR analysis. This amounts to choosing a suitable  $(r \times r)$  submatrix of  $\Phi_0 B$  to be lower triangular. Such restrictions would suffice for identifying  $B$  and, hence, the structural shocks. Denoting the  $(r \times r)$  submatrix of  $\Phi_0$  that is of interest in the present context by  $\Phi_{0b}$ , the corresponding  $B$  matrix can be

obtained by noting that  $\Phi_{0b}BB'\Phi'_{0b} = \Phi_{0b}\Sigma_{\eta}\Phi'_{0b}$ . Thus, computing a Cholesky decomposition of this matrix and left-multiplying by  $\Phi_{0b}^{-1}$  gives a suitable  $B$  matrix.

As an example consider a study by Forni and Gambetti (2010) who use a GDFM for analysing the effects of U.S. monetary policy. They use a panel of 112 monthly series for a period 1973 - 2007 and work with different numbers of factors. In their benchmark model the number of dynamic primitive factors is four. They use industrial production, a consumer price index, the federal funds rate and the Swiss/US real exchange rate as the first four variables in their panel in that order. The structural shocks are identified recursively and the monetary policy shock is specified to be the third one. Hence, it is identified as a shock that does not have an instantaneous effect on industrial production and the price index but may induce immediate reactions of the exchange rate. Of course, in this setup all four shocks can have impact effects on many 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

$$J\text{vec}(\Phi_{0b}B) = 0$$

which implies restrictions

$$J(I_r \otimes \Phi_{0b})\text{vec}(B) = 0$$

for  $B$ .

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

In a standard SVAR analysis the identifying restrictions are typically placed on the impact effects of the shocks ( $B$ -model) or on the instantaneous relations between the variables ( $A$ -model). The previously discussed restrictions correspond to a  $B$ -model setup. It is also possible, of course, to identify the shocks by placing restrictions directly on the factor loadings in a structural version of the model (1.9) or (1.10). Such an approach may be more natural if the common factors have a direct interpretation. As an example consider a study of the international business cycle by Kose et al. (2003). These authors investigate a panel of 60 countries from 7 different regions in the world and consider output, consumption and investment macroeconomic aggregates from each country. They identify one world factor, a factor for each region and a country specific factor for each country. The world factor

is the only one that is allowed to have a direct impact on all variables and the effects of the other factors are restricted by imposing suitable zero constraints on the corresponding loadings. The shocks are identified by imposing that the VAR process driving the factors has instantaneously uncorrelated residuals, that is, in our notation,  $\eta_t$  has a diagonal covariance matrix.

**Restrictions on the Long-run Effects of Shocks** Long-run restrictions à la Blanchard-Quah are also easy to use in the present context. They amount to choosing an  $(r \times r)$  submatrix of  $\Phi(L)$ , say  $\Phi_b(L)$ , such that  $\Phi_b(1)B$  is lower-triangular. In this case the corresponding  $B$  is obtained by computing a Cholesky decomposition of  $\Phi_b(1)BB'\Phi_b(1)' = \Phi_b(1)\Sigma_\eta\Phi_b(1)'$  and left-multiplying by  $\Phi_b(1)^{-1}$ .

In principle other long-run restrictions can also be used, e.g., those proposed by King, Plosser, Stock and Watson (1991). They are less natural in the present context, however, because factor models are more commonly based on  $I(0)$  variables without stochastic trends.

**Identification Through Instruments** Identification of  $B$  and thereby the structural shocks can also be achieved by using instruments with suitable properties. Suppose a variable  $z_t$  is available that is correlated with the  $k$ th structural shock but uncorrelated with all other shocks, that is, it has the property

$$E(\varepsilon_{it}z_t) = \begin{cases} \rho \neq 0 & \text{for } i = k, \\ 0 & \text{for } i \neq k. \end{cases}$$

Then, using  $\eta_t = B\varepsilon_t$  and denoting the columns of  $B$  by  $b_i$ ,  $i = 1, \dots, K$ , we get

$$E(\eta_t z_t) = BE(\varepsilon_t z_t) = b_k \rho. \tag{1.25}$$

Thus, a multiple of the  $k$ th column of  $B$  can be obtained as the covariance of the reduced form error  $\eta_t$  and the instrument  $z_t$ . A natural estimator of  $b_k \rho$  is

$$T^{-1} \sum_{t=1}^T \eta_t z_t.$$

Of course, in practice  $\eta_t$  has to be replaced by an estimator. More precisely it will be replaced by the estimated quantities from the reduced form model (1.23). This is an easy way to estimate a multiple of the  $k$ th column of  $B$  if a suitable instrument variable is available that is only correlated with the  $k$ th structural shock and uncorrelated with all other structural shocks. Stock and Watson (2012) call such instruments *external instruments* if they are not part of the database  $y_t$  used in the factor analysis but use outside information.

Note that for identification it is sufficient to know a multiple of the  $k$ th column of  $B$  because the responses of the variables to the  $k$ th shock are obtained as  $\Phi(L)b_k$  and, hence, considering a multiple just changes the scaling. In other words, multiplying the shock by some constant just changes the size of the shock but not the shape of the response. Hence, the size of the shock can be chosen freely. For example, Stock and Watson (2012) propose to choose it such that the initial response of a specific variable is one. For example, a monetary shock may be chosen such that the policy interest rate changes by one unit on impact. It is important to note that changing the scaling does not change the shape of the response function. Similarly, changing the sign of the shock just reflects the response function at the zero line but leaves the shape otherwise invariant because we are currently dealing with a linear model.

So far we have just explained how to identify the columns of  $B$  and, hence, the impulse responses. Stock and Watson (2012) mention that the associated shocks can be determined by regressing  $z_t$  on  $\eta_t$  and using the predicted values of that regression as estimated structural shocks (see Section III.a of Stock and Watson (2012) for details).

So technically identification via instruments is easy. The important practical question of interest in this context is, of course, from where to get suitable instruments. Stock and Watson (2012) use for that purpose shocks that have been constructed in other studies based on other models and assumptions. For example, as an instrument for a productivity shock they consider the productivity shock series from a DSGE model constructed by Smets and Wouters (2007). As monetary shocks they use the corresponding shocks identified by Sims and Zha (2006) and the shock to the monetary policy reaction function of the Smets and Wouters (2007) DSGE model.

If there are two different instruments identifying the same shocks, e.g., the monetary policy shocks in the above example, then the resulting shocks should in principle be perfectly correlated. Of course, in practice this is not the case and getting very different shocks and impulse responses with different instruments can be a reason of concern regarding the suitability of the instruments. Similarly, two different structural shocks should be uncorrelated in theory. In practice, that may not hold precisely because the instruments are not necessarily chosen such that uncorrelated structural shocks are guaranteed. In that case, the empirical correlation between different structural shocks can give an indication of the suitability of the instruments. If, for example, a technology shock and a monetary policy shock turn out to be highly correlated, then this suggests that the instruments used for their identification are not suitable. Of course, it has to be kept in mind that in practice we do not have the true shocks and impulse responses but only estimates. Hence, small empirical correlation may be acceptable.

**Sign Restrictions** So far we have just discussed identification of the shocks via exact restrictions, for instance, exclusion restrictions for the impact effects or long-run responses. In principle it is, of course, also possible to use sign restrictions. In the context of FAVAR models they have been considered by Mumtaz and Surico (2009), Amir Ahmadi and Uhlig (2009) and Amir Ahmadi and Ritschel (2009), for example. The way the sign restrictions are imposed is quite different in these studies. Mumtaz and Surico (2009) use a mixture of exclusion and sign restrictions in a FAVAR setting, that is, they place sign restrictions on the effects of shocks on the factors. In contrast, Amir Ahmadi and Uhlig (2009) and Amir Ahmadi and Ritschel (2009) impose sign restrictions on the effects of the disaggregated, informational variables. In other words, a shock is identified by its impact in the equations (1.22). It is admissible if the responses of the  $x_t$  variables have the correct sign.

For example, Amir Ahmadi and Uhlig (2009) consider a panel of 120 monthly U.S. macroeconomic time series. Their objective is to investigate the impact of U.S. monetary policy on the economy. In one of their scenarios they define a contractionary monetary policy shock as a shock that rises the federal funds rate and lowers inflation measured by several consumer and producer price indices, the M1 monetary aggregate and nonborrowed reserves.

**Other Restrictions** It is also possible to link the identifying assumptions for shocks directly to the properties of factors. For example, Giannone, Reichlin and Sala (2004) determine two main factors in a system of U.S. macroeconomic variables and, hence, two shocks are of central importance in driving the system. They identify the real shock such that the corresponding factor maximizes the share of the variance of the real variables explained by the factor. The other shock is taken to be the nominal shock.

Bäurle (2013) assumes that the factors correspond to economic quantities that are related according to a DSGE model. He uses that model to identify the structural shocks and develops Bayesian methods based on ideas of Del Negro and Schorfheide (2004) for estimation. Thus, he also links identification to the factors.

Some authors consider so-called multi-level or hierarchical factor models where the variables are partitioned in 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 where the blocks refer to the different countries. Such models open up the possibility for identifying block-specific and global shocks separately.

## 1.5 Applications

There are many examples of structural analysis with factor models in the literature. We have mentioned already some of them previously. Many of them fall in two categories, monetary policy analysis and international business cycle analysis. In the following a small number of these studies is reviewed to provide a flavour of the variety of models and approaches used in applied work. Those interested in more detailed examples may find them useful further reading.

**The Effects of Monetary Policy** As mentioned earlier, structural FAVAR analysis was proposed by Bernanke et al. (2005). They perform an analysis of U.S. monetary policy. The factors are determined from a panel of 120 monthly time series and are used to augment small macro models. For example, they consider a VAR model for industrial production, CPI, the federal funds rate and one factor. In another specification they add three factors to the federal funds rate. Identification of the monetary policy shocks is achieved by assuming that none of the other variables or factors responds instantaneously to a monetary monetary shock. This can be imposed by using a recursive identification scheme in which the federal funds rate is ordered last. They find that taking into account the additional information summarized in the factors makes a substantial difference for the impulse responses. Hence, the additional information presents a different picture of the transmission of monetary policy shocks than a standard small VAR model.

Del Negro and Otrok (2007) use a FAVAR model for quarterly U.S. variables to investigate the impact of monetary policy on house prices. They are explicitly interested in including information on regional house prices in the analysis. Using state level house prices, they find that there is a period in which the house price increase can be attributed mainly to a national factor constructed from the regional price series. Therefore they include that factor in a FAVAR model consisting of six variables: the house price 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 shocks. 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 can have an effect on house prices and may hence contribute to housing booms although the impact of monetary policy may be small.

Favero et al. (2005) use FAVAR models to investigate the impact of monetary policy shocks in the U.S. and four large European economies (Germany, France, Italy, Spain). They are mainly interested in comparing different methods for constructing factors. They use the Stock-Watson principle components approach and contrast that with the dynamic

factors constructed à la Forni, Hallin, Lippi, Reichlin. The common factors are extracted from large monthly panels of variables from the U.S. and the four European countries. The country FAVAR models include small sets of economic variables such as output, inflation, commodity price inflation, an exchange rate and the policy rate in the case of the U.S. and in addition foreign variables for the European countries. For example, U.S. inflation is included in the model for Germany. Moreover, results are compared for models augmented with different sets of factors. Identification is achieved by a recursive scheme where the policy interest rate is ordered last. The authors conclude that including common factors can make a difference for impulse response analysis. For example, it can remove the ‘price puzzle’, that is an increase in inflation after a contractionary monetary policy shock. This phenomenon is often attributed to omitted variables bias and, hence, removing it by including further information in the form of factors is plausible. Although the type of factors added makes a difference, the effect is often not great. There is no clear recommendation in favor of which type of factors to include.

We have already mentioned earlier that Amir Ahmadi and Uhlig (2009) consider a panel of 120 monthly U.S. macroeconomic time series to investigate the impact of U.S. monetary policy on the economy. They use sign restrictions for identification. A contractionary monetary policy shock is characterized as a shock that rises the federal funds rate and lowers inflation measured by several consumer and producer price indices, the M1 monetary aggregate and nonborrowed reserves. They specifically account for the disaggregate effects and do not simply confine the impulse response analysis to a small dimensional FAVAR model but use the dynamic factor model setup for investigating the responses of a large panel of variables to monetary policy shocks. They find reasonable responses to monetary policy shocks even for samples that include the recent financial crisis period. In particular, the response of output to a contractionary monetary policy shock is negative but of modest size. Amir Ahmadi and Ritschel (2009) use a similar approach to investigate the role of monetary policy for a historic period of the inter war Great Depression. They find that monetary policy may have had only a very modest impact in that period.

Boivin and Giannoni (2009) consider the impact of global forces on the U.S. economy and in particular on the transmission of monetary shocks. They extract domestic factors from a panel of 671 quarterly macroeconomic and financial series for the period 1984Q1 - 2005Q2 and foreign factors from 49 series from other countries. Then they set up a FAVAR model with 10 domestic factors, four foreign factors and the federal funds rate in their preferred model. Identification of the monetary shocks is done by assuming that surprise changes of the federal funds rate impact on the factors only with a delay of at least one quarter. In other words, a recursive identification scheme with federal funds rate ordered last specifies

the monetary policy shocks. The authors are interested in the effects of monetary shocks on some key economic variables and the corresponding impulse responses are obtained via (1.22). They do not find strong evidence for significant changes in the transmission of monetary shocks due to international factors and conclude that global forces may have a stronger impact in the last part of their sample period at best.

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. series for the period 1976M1 - 2005M6 as informational variables. The number of extracted factors is five and in addition the policy interest rate taken to be the federal funds rate is included. 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. They 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.

Eickmeier and Hofmann (2013) consider the contribution of monetary policy to the housing boom and financial imbalances in the U.S. and find that it was considerable in the 2001 - 2006 period. They obtain their conclusions from a FAVAR analysis based on a quarterly model for real GDP growth, inflation based on the GDP deflator, the federal funds rate and a panel of 232 financial variables as informational variables. Identification of monetary shocks is based on a combination of zero restrictions on the impact effects and sign restrictions. The restrictions are such that instantaneous interaction between the policy rate and the financial factors is ensured.

Finally, we mention the study by Bäurle (2013) again that was already referred to earlier because of its specific way to identify the shocks. Recall that he uses a Bayesian setup and that his factors correspond to economic quantities that are related according to a DSGE model that is used to identify the structural shocks. He considers a dynamic factor model and assumes that there are as many shocks as factors. In other words, the shocks are driving the observed variables via their impact on the factors. He performs an analysis based on a large panel of quarterly U.S. macro series for the period 1985 - 2007. He compares the responses to monetary shocks identified by his DSGE model with those obtained from sign restrictions and a recursive identification scheme. The sign restrictions in this approach are imposed on the responses of the factors. In particular, one factor is viewed as a price factor and another one as an interest rate factor. A contractionary monetary policy shock is then characterized as a shock that does not increase the price factor and does not decrease the interest rate

factor. He finds that the DSGE and sign identification do not lead to conflicting results but the error bands around the impulse responses are smaller for the DSGE identification scheme. In contrast to the latter two identification schemes, the Cholesky identification leads to 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 et al. (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 in 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 while region-specific factors are only 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.

Eickmeier (2007) uses structural factor models to study international business cycle transmission between the U.S. and Germany. She uses quarterly data for the period 1975 - 2002 for a large set of U.S. and German series. The factors are assumed to be generated by a VAR model and the shocks driving the factors are identified by extracting the two shocks that explain as much as possible of the forecast error variance of the common component of U.S. GDP over a six year horizon and then characterizing them as supply and demand shocks by using sign restrictions. She then investigates how much the U.S. shocks affected Germany and in particular their role in German business cycle fluctuations.

Mansour (2003) uses generalized dynamic factor models with orthogonal shocks driving the common components. He considers annual growth rates of GDP for a panel of 113 countries over the period 1961 - 1989. He interprets the shocks driving the common factors as global shocks and investigates their effects on the characteristics of cyclical fluctuations in the countries under investigation. Then he studies how much the individual countries are affected by the global shocks and he analyses business cycle synchronization in different regions of the world.

Helbling and Bayoumi (2003) also use dynamic factor models for analyzing cyclical fluctuations in the G7 countries. They identify the global business cycle with two global factors and find that these two factors contribute an important part of the output gaps in the countries under consideration. There are a number of other studies of international business cycle fluctuations based on dynamic factor models including Bordo and Helbling (2010) who consider business cycle synchronization in 16 industrial countries over the last century.

As mentioned earlier, factor models have become increasingly popular tools for structural economic analysis. Therefore there are now many empirical studies in the literature and more are likely to appear in the future. Hence, there is a rich set of examples interested readers may refer to.

## 1.6 Critique of Structural Analysis with FAVARs

The identifying assumptions discussed in Section 1.4 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 residuals  $\eta_t$  driving the factors in (1.23). This assumption may be justified if there is a latent structure in the background of the DGP that describes the economic structure of interest. Moreover, it is difficult to see that it makes sense in approximate DFMs that allow more structure in the part of the model not explained by the common factors. Such models admit effectively that the relation in the variables is only partly captured by the common factors and, hence, it is conceivable that the transmission of shocks is only partly captured by  $\Phi(L)B$ . Clearly, in practice the assumption that the transmission of the shocks of interest is via  $\Phi(L)B$  may be questioned and has to be justified carefully in any particular analysis. In any case, even if the true DGP is an exact DFM, the Wold MA representation of the observed variables  $y_t$  is different from (1.24). Hence, building on the MA representation of  $y_t$  results in a different transmission of the shocks. The shocks of the DFM based on (1.24) instead are aggregated shocks extracting information from a large panel of variables.

Stock and Watson (2005) rightly point out that if the DFM is taken as true model, 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 come in through individual variables, that is, they may enter through the idiosyncratic components. Moreover, the idiosyncratic components may well play an important role in the propagation of the shocks that is ignored by extracting them from the  $\eta_t$  errors in the representation (1.24). For example, if a monetary policy shock is implemented by the central bank through a change in the policy interest rate, this is a

change in an individual variable and, hence, may be better viewed as an idiosyncratic shock that may, of course, have an impact on the factors and other variables. Thus, even if factor models are used it is important to think carefully about the informational variables to be included in the dataset. It can have a crucial impact on the results, in particular, if only a small number of factors is considered that is likely to represent just an approximate factor model.

Although factor models solve the missing information problem by allowing to include a large set of variables in an analysis, including such rich data sets can also cause problems. For example, not much is known about the sensitivity of structural results with respect to changes in the information set and/or the model structure. How many factors are included may, for instance, make a difference for some results and that in turn may depend on the variables considered for analysis. Having to deal with a large number of variables may require a substantial updating effort when revisions and new data become available. That FAVAR analysis may well be affected substantially by such factors was also mentioned by Amir Ahmadi and Uhlig (2009) who note that a Cholesky recursive identification scheme for monetary policy shocks results in implausible responses of some of the variables in a FAVAR analysis. In particular, they find an increase in inflation due to a contractionary monetary policy shock while Favero et al. (2005) find a more plausible reaction for some of their scenarios. Thus, the results appear to depend crucially on such things as the model setup, the number of factors and the variables and data used. An analysis of the sensitivity of the results with respect to the set of variables, sampling period, model setup as well as type and number of factors is therefore highly recommended for every study.

## 2 Large Bayesian VAR Models

As mentioned earlier, instead of frequentist estimation methods, one may use Bayesian methods for estimating dynamic factor models (see, e.g., Otrok and Whiteman (1998), Kose et al. (2003), Amir Ahmadi and Uhlig (2009)). If Bayesian methods are used, it is not obvious, however, that one wants to focus on factor models. Recall that a motivation for using factor models is that they allow to integrate large panels of variables in a SVAR analysis. In the context of Bayesian estimation, suitable priors serve the same purpose (see, e.g., De Mol, Giannone and Reichlin (2008)). In fact, as pointed out by Bańbura et al. (2010), using Bayesian shrinkage methods to overcome the degrees of freedom (curse of dimensionality) problem in a SVAR analysis has several advantages. First of all, having no limits on the number of observed variables included in a VAR model for macroeconomic analysis allows to include all the variables desired by macroeconomists. Second, sectoral information can

be included in disaggregated form 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 restrictions as in global VARs or panel VARs that are used just to account for the degrees of freedom limitations otherwise encountered and not because economic reasoning would suggest them. Similarly, in particular in monetary policy analysis considering the effects on disaggregated price series may be desirable and can be studied if a large number of disaggregate price series is included in 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. As explained earlier, factor analysis can in principle also be done for trending variables. In that case some assumptions regarding the stochastic trends are necessary for deriving proper inference procedures. From a practical point of view it may be advantageous to do without such assumptions and include the variables in levels and thereby potentially accommodate unit roots, long-range dependence, near unit root behaviour and the like.

A crucial problem for using large scale VAR models is the choice of prior that makes estimation feasible. We discuss that issue in the next subsection and consider specific issues related to structural identification in Section 2.2.

## 2.1 Priors for Large Bayesian VARs

Bańbura et al. (2010) use the so-called Minnesota or Litterman prior as their point of departure. Recall that it assumes a reduced form Gaussian VAR model,

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

and imposes a normal prior with a random walk mean, that is, the prior mean is

$$B^* = [0, I_K, 0, \dots, 0],$$

where  $B = [\nu, A_1, A_2, \dots, A_p]$  and hence, the prior mean of  $A_1$  is the identity matrix. This prior should be modified if there are  $I(0)$  variables. For those 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  $\lambda$  is the prior standard deviation of  $\alpha_{ii,1}$ ,  $0 < \theta < 1$ , and  $\sigma_i^2$  is the  $i$ th diagonal element of the reduced form residual covariance matrix  $\Sigma_u$ . If  $\Sigma_u$  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. Note, however, that we cannot just replace  $\Sigma_u$  by its unrestricted OLS estimator because this estimator is typically not available given the degrees of freedom limitations. An alternative would be to estimate the variances by fitting univariate AR models by OLS and assuming that  $\Sigma_u$  is diagonal. This solution is sometimes used in practice. For example, Koop (2013) uses it to evaluate the forecast implications of a number of different priors for large Bayesian VARs (BVARs).

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

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

where  $Y = [y_1, \dots, y_T]$ ,  $Z$  is the corresponding matrix of regressors and  $V(\lambda)$  is such that the prior covariance matrix is  $V(\lambda) \otimes \Sigma_u$ . For given  $\Sigma_u$  and assuming  $\theta = 1$ , the prior covariance matrix depends only on  $\lambda$  and therefore the tightness parameter  $\lambda$  is explicitly indicated.

The posterior mean may be interpreted as a shrinkage estimator where the shrinkage is completely determined by  $\lambda$ . 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. Of course, if forecasting is the objective, one could choose it such that the model forecasts well as in Carriero et al. (2009). Alternatively, it may be chosen so as to maximize the marginal likelihood in a hierarchical modelling framework as proposed by Giannone, Lenza and Primiceri (2010) (see also Carriero, Kapetanios and Marcellino (2012) for a similar approach). For models with hundreds of variables the latter procedure poses computational challenges, however.

Based on an investigation of the issue of selecting the shrinkage parameter by De Mol et al. (2008), Bańbura et al. (2010) propose choosing this parameter tighter when the model gets larger. More precisely, they propose choosing  $\lambda$  such that the estimated model has the same in-sample fit as a small VAR model estimated by OLS. The procedure works as follows.

Denote the posterior means of the parameters obtained from a model with tightness parameter  $\lambda$  and  $K$  variables by  $\nu^{(\lambda, K)}$  and  $A_i^{(\lambda, K)}$ ,  $i = 1, \dots, p$ , and the corresponding 1-step ahead forecasts 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  $y_{t|t-1}^{(\lambda,K)}$ , that is,  $y_{k,t|t-1}^{(\lambda,K)}$  is the 1-step ahead forecast of the  $k$ th variable of a system with  $K$  variables and prior tightness parameter  $\lambda$ . The corresponding in-sample mean squared forecast error is

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

Suppose there is a small number,  $K^*$ , of variables of central interest and their index set is  $\mathcal{K}$ , then the tightness parameter for a large panel of a total of  $K$  variables,  $\lambda_K$ , is chosen such that

$$\lambda_K = \arg \min_{\lambda} \left| \text{FIT} - \frac{1}{K^*} \sum_{k \in \mathcal{K}} \frac{\text{msfe}_k^{(\lambda,K)}}{\text{msfe}_k^{(0,K)}} \right|. \quad (2.1)$$

In other words, the MSEs are evaluated relative to the forecast MSEs obtained for a forecast based on the prior mean. Here the benchmark fit is obtained from a small model for the variables of central interest. Using that set of  $K^*$  variables and fitting a VAR model by OLS, the fit is defined as

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

The actual minimization of  $\lambda_K$  in (2.1) can be done by a grid search over  $\lambda$  because only one parameter is involved (see Koop (2013)).

Thus, a small model with the central variables for a particular application of interest is set up and estimated by OLS first. Then  $\lambda$  is chosen for large models 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 by Bańbura et al. (2010). Notice that this choice of tightness parameter amounts to specifying a tighter prior for larger models with more variables and lags. Koop (2013) also uses an analogous procedure for the original Minnesota prior where he chooses both  $\lambda$  and  $\theta$  so as to minimize the fit of the central variables. One could also use the prior that optimizes the forecasts for a given set of variables as in Carriero et al. (2009) or one could consider choosing the shrinkage parameters by maximizing the marginal likelihood in a hierarchical Bayesian setting as in Giannone et al. (2010).

Bańbura et al. (2010) find in their forecasting experiment that a sum-of-coefficients prior which is a variant of the Minnesota prior worked better in their application. Hence, they recommend that also for structural analysis. It can accommodate long-run relations more easily and may therefore give more realistic impulse responses if, for instance, cointegration relations exist. Koop (2013) points out the very restrictive nature of the Minnesota prior

which uses only one or two parameters to determine the degree of shrinkage. He also considers another prior based on a proposal by George, Sun and Ni (2008). They obtain flexibility by allowing different parameters to be shrunk differently. In Koop’s forecast comparison this prior tends to deteriorate when the number of variables in the VAR increases. Given that we are primarily interested in dealing with large BVARs, we therefore do not further discuss this prior here.

Korobilis (2013) proposes to combine BVARs with Bayesian variable selection. In his approach an indicator variable is specified for each parameter that indicates whether the parameter is included or set to zero. A prior is specified for the indicator variables that can be combined easily, for example, with the Minnesota prior. Korobilis (2013) also presents a modification that makes it feasible to deal with large panels of variables. However, the largest model he uses in a forecast assessment exercise contains 13 variables and is far from the dimensions we have in mind when we talk about large BVARs. Given the limited evidence on the performance of these procedures we do not elaborate on them here.

Of course, structural analysis requires identifying assumptions for the shocks. For the large-scale BVAR context they are discussed next.

## 2.2 Structural Identification in Large BVARs

In large BVAR models identification of structural shocks is most easily achieved by linking the properties of the shocks to their impact effects. For example, if exclusion restrictions can be specified for the impact effects, that is useful identifying information. Bańbura et al. (2010) are interested in the effects of monetary policy shocks and, in line with Bernanke et al. (2005), they split up their variables in those that move slowly after such shocks and those that move fast and may change instantaneously in the same period when the shock hits the system. Thus, they order their variables such that

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

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. Then they identify the monetary policy shock by a lower-triangular Cholesky decomposition of the reduced form covariance matrix  $\Sigma_u$ . Thereby the fast-moving variables are allowed to be affected instantaneously while the slow-moving variables are assumed to be known to the policy makers at the time of their decisions. For another study of monetary policy shocks based on a large BVAR model see Gupta et al. (2009).

In principle one could also identify a 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 combine sign restrictions with exclusion restrictions and thereby reduce and solve the computational problems.

In summary, large-scale BVARs have some advantages but also drawbacks. On the positive side, they allow for inclusion of unlimited numbers of observed variables and, hence, sectorally disaggregated variables can be included and regionally disaggregated analysis becomes feasible. Their drawback is that the prior is based on practical considerations and not on subject matter knowledge. Unfortunately, such diffuse priors distort some results in typically unknown directions. Thereby the results induced by the prior may have an element of arbitrariness. Notice that, without the prior, estimation of large-scale models is usually not possible. So in some sense the estimates are determined by the prior.

### 3 Alternative Models

In this chapter we have discussed possibilities to deal with large panels of data either by summarizing them in factors or by applying Bayesian shrinkage methods for estimation. There are also other proposals for dealing with large panels of variables. For example, panel VARs, global VARs or spatial VARs have been considered for this purpose (see Canova and Ciccarelli (2013), Pesaran, Schuermann and Weiner (2004) or Chudik and Pesaran (2011)). These models impose specific restrictions on the VAR parameters to ensure feasible estimation. Clearly, if many variables are to be included this means that large numbers of restrictions have to be imposed. These restrictions are often viewed as quite strong and perhaps not very realistic. Of course, if specific restrictions can be defended in a particular analysis, they may be preferable to using factors or BVAR methods. In any case, it is important to understand that large panels of variables can be included in a VAR analysis, but only at the cost of imposing some kind of restrictions that may distort the results to some extent. Hence, it is important to think carefully about the types of restrictions that are best suited for a specific analysis.

#### 3.1 Panel VARs

Large panels of variables often come up when studies for different countries or regions or more generally units are considered and a set of variables for each unit is of interest. Such a data 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  $i = 1, \dots, M$  and  $n = 1, \dots, N$ . Thus, using our earlier notation,  $K = M \times N$ . Let  $y_{nt} = (y_{1nt}, \dots, y_{Mnt})'$  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 a full dynamic VAR model for  $y_{nt}$  has the form

$$y_{nt} = \nu_n + \mathbf{A}_n Y_{t-1} + u_{nt}, \quad (3.1)$$

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

Clearly in this general form it may be impossible to estimate the model in unrestricted form due to degrees of freedom limitations. However, the panel structure suggests possible restrictions. For example, there may be no dynamic interdependencies between the units, that is, every unit is represented by 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 (3.1) and may be easy to estimate from the data for the  $n$ th unit. Note that there may still be some dependence between the units via the residual covariance matrix.

Of course, further restrictions can be imposed if an assumption of dynamic homogeneity can be justified. For example, if a panel of countries with similar economic systems is considered such an assumption may be justified, provided the countries are also not dynamically related. In that case, all units are assumed to have the same VAR coefficients, that is,

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

In that case special estimation methods are available for estimating the parameters (see, e.g., Canova and Ciccarelli (2013)).

Of course, such assumptions are too restrictive in many situations and at least some dynamic interdependency may be required given the data properties for a particular application. The panel setup may suggest restrictions that make estimation still feasible in that case. Alternatively one may of course ignore the panel structure and estimate the model with Bayesian methods as proposed in Section 2.

Panel VAR techniques have become increasingly popular recently. There are many issues we have not touched upon in this section. A review of the literature is given by Breitung (2013).

## 3.2 Global VARs

Global VARs (GVARs) are also typically considered if large country panels are under investigation (e.g., Dees, Di Mauro, Pesaran and Smith (2007)). The idea is, however, to augment a VAR for each unit by factors capturing the other countries. In that respect GVARs have some similarity to 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 the  $y_{nt}$ , that is,

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

where  $w_{nj}$  is the weight attached to country  $j$  in the model for the  $n$ th country. For example, Dees et al. (2007) link the weights to the share of country  $j$  in world trade. These weights may also be matrices, of course, if different variables are weighted differently. Generally  $y_{nt}^*$  can be thought of as common factors consisting of global variables and variables computed from the variables for the other countries. Thus, the common factors are not necessarily determined by some statistical procedure but may be based on economic considerations. Apart from that the models can be viewed as factor-augmented VAR models.

The models for many countries or regions can be linked together to a global model, the GVAR, which is easily recognized to be a large restricted VAR model with potentially some additional unmodelled variables, if there are no equations explaining the globally important variables appearing on the right-hand side of all the individual country models. 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 analysis tools such as impulse responses, provided shocks of interest can be identified. The GVAR literature has used non-structural generalized impulse responses to study the dynamics of the system and it has also used limited structural assumptions. For instance, Dees et al. (2007) specify a U.S. monetary shock by ordering the U.S. system first and considering a Cholesky decomposition of the reduced form covariance matrix associated with the U.S. system. Thereby they specify the monetary policy shock similarly to other authors who have studied U.S. monetary policy based on shocks that are identified by restrictions on

the impact effects. The effects on other countries are left open and they are allowed to be instantaneous. Such an approach may be justified if a dominant country like the U.S. is considered. The identification of shocks becomes more difficult, however, if also the effects of monetary shocks in other countries for the U.S. are of interest or, more generally, if the effects of many shocks are to be investigated.

The difficulties in structural shock identification can be seen as a drawback of GVAR analysis. Another drawback may be that the implications of the specific aggregation of variables from other countries for the overall dynamics and the transmission of shocks are difficult to disentangle.

### **3.3 Other Ideas**

There are also other ideas how to restrict the number of variables or the parameter space in VARs (see also Canova and Ciccarelli (2013)). For example, spacial models have been considered where it is assumed that a region depends more strongly on its close neighbours than on more distant regions (see Anselin (2006), Chudik and Pesaran (2011), or Canova and Ciccarelli (2013)). In other words, the distance between units is used to impose restrictions and thereby reduce the parameter space. Again such assumptions are problematic in many applications because it may be unsatisfactory to link the relation between units to their physical distance. For example, if monetary policy is studied the U.S. policy may affect many other countries in the world that are quite far away in distance. In other words, there are global effects that are quite important irrespective of the physical distance of units.

Another idea is to consider smaller submodels involving only a subset of regions or countries. For example, one may just consider two models at a time even though a much larger panel of countries is of interest. Such submodel comparisons can give useful insights regarding differences in the structures. It has to be kept in mind, however, that the impulse responses in a submodel can be quite different from those in the full model. Thus, extracting information on the actual dynamic interactions from them is problematic and requires strong assumptions.

## **4 Model Comparison and Discussion**

The proposals for dealing with large panels of variables in VAR models considered in this chapter all amount to imposing restrictions. This can be accomplished by shrinking either the number of variables or the parameter space. Of course, these two approaches cannot be distinguished perfectly in some approaches. They all result in smaller dimensional parameter spaces and, hence, deal with the curse of dimensionality. Factor models clearly reduce the

number of variables and large BVARs shrink the parameter space. Some other approaches are somewhat in-between. For example, GVAR models to some extent reduce the space of variables by restricting the parameter space. All the models considered in this chapter have their pros and cons. They all enable the researcher to deal with large panels of variables.

Factor models extract important information from the variables first and aggregate it in factors. These factors are then used in a model together with observed key variables of central interest. Thereby quite manageable models may be obtained that can be analysed with standard frequentist methods. On the other hand, aggregation generally leads to distortions in the transmission of shocks. The importance of such distortions is usually unknown in any particular application. Moreover, factor analysis is tailored for stationary variables with time-invariant moments, although these methods have been extended to nonstationary variables with stochastic trends. Proper factor analysis requires assumptions regarding the stochastic trends, however, that may be problematic when many variables are involved. They may have quite different trending properties and assumptions regarding their order of integration or long-range dependence properties may be problematic.

Large-scale BVARs avoid prior aggregation of variables and they may include variables in untransformed form in the VAR model regardless of their unit root and trending properties. Thereby the models may become so large that degrees of freedom deficiencies make frequentist, classical estimation impossible. The Bayesian solution in this case is to impose a prior on the parameter space. The priors typically imposed in large-scale BVAR analysis are rather standardized and shrink the VAR parameters to zero or values corresponding to unit roots. Such priors do not account for the actual economic structures underlying a panel of variables. Even if they are not meant to be restrictive, they may lead to substantial distortions in the standard tools for structural VAR analysis. For example, they may distort impulse responses and, hence, they may provide an incorrect impression of the actual transmission mechanism of shocks.

In a forecast comparison based on large panels of variables, Bańbura et al. (2010) and Koop (2013) find that large BVARs forecast overall better than factor models. In fact, De Mol et al. (2008) give conditions based on asymptotic theory that ensure that Bayesian shrinkage for large panels of variables 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 could be used to make a case for BVARs, it is not clear that out-of-sample forecast performance is the best criterion for evaluating the transmission mechanism of shocks. Also, forecast evaluations are to some extent data dependent and case specific. Hence, for structural analysis it may be more important to think about the underlying economic structure.

Panel VAR models account, for instance, for the regional structure of the data and use that information to impose restrictions on the parameters. They may, for example, impose homogeneity restrictions for a set of countries with similar economic structure. Such restrictions may be justified if a set of similar units is considered. On the other hand, it is not uncommon in practice that, despite some similarities, there are also substantial differences in the units that may be sufficiently important to result in distortions when ignored. In fact, even the unit root and trending properties of the same variables from different countries are often different. Such findings hint already at the presence of substantial differences between the countries. There is no guarantee that by trend-adjusting the data such differences are properly eliminated. In any case, trend-adjustments and other data revisions may lead to changes in the transmission of shocks which is undesirable in a structural analysis.

Global VAR models may be viewed as a combination of imposing direct restrictions on the parameter space and factor models. The factors are not chosen by a purely statistical criterion and procedure but may be chosen on the basis of economic considerations. Also the parameter restrictions may be suggested by subject matter considerations and may be susceptible to statistical testing. Although this means that they are not as removed from the underlying economic structure at the variable selection stage as in a statistical factor analysis, they have, of course, all the problems of data aggregation mentioned earlier. The fact that the factors are not picked with purely statistical procedures does not mean that the aggregation cannot lead to distortions of the transmission of shocks. Another problem in the GVAR literature is the lack of convincing strategies for the identification of proper structural shocks. As mentioned in Section 3.2, in these models shocks are often identified based on mathematical/statistical criteria. The resulting impulse responses are called generalized impulse responses to distinguish them from structural impulse responses.

For applied work the distinction between shrinkage of the variables or the parameter space may not be of prime importance. What is important is the question whether the necessary restrictions can be defended for a specific economic analysis. It seems plausible that purely statistical reductions may not be optimal from an economic point of view. Hence, if economic arguments are available for a specific form of restrictions they may be preferable to purely statistical restrictions. On the other hand, it is also important to account adequately for the statistical data properties. Thus, in the end, mixtures of economic and statistical restrictions are probably the most common result in practice. In any case, ignoring one type of information, either the statistical properties of the data or the information from an economic model is not likely to give satisfactory, widely acceptable results. Therefore a careful consideration of the model to be used for a particular analysis is of prime importance. In that process knowing the alternative models and their pros and cons is important.

## References

- Ahn, S. K. and Reinsel, G. C. (1988). Nested reduced-rank autoregressive models for multiple time series, *Journal of the American Statistical Association* **83**: 849–856.
- Amengual, D. and Watson, M. W. (2007). Consistent estimation of the number of dynamic factors in a large N and T panel, *Journal of Business & Economic Statistics* **25**: 91–96.
- Amir Ahmadi, P. and Ritschel, A. (2009). Depression econometrics: A FAVAR model of monetary policy during the great depression, *SFB 649 Discussion Paper 2009-054*, Humboldt Universität zu Berlin.
- Amir Ahmadi, P. and Uhlig, H. (2009). Measuring the dynamic effects of monetary policy shocks: A Bayesian FAVAR approach with sign restriction, mimeo, Humboldt Universität zu Berlin.
- Anderson, T. W. (1999). Asymptotic distribution of the reduced rank regression estimator under general conditions, *Annals of Statistics* **27**: 1141–1154.
- Anderson, T. W. (2002). Canonical correlation analysis and reduced rank regression in autoregressive models, *Annals of Statistics* **30**: 1134–1154.
- Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, third edn, John Wiley, New York.
- Anselin, L. (2006). Spatial econometrics, in T. C. Mills and K. Patterson (eds), *Palgrave Handbook of Econometrics, Econometric Theory*, Vol. 1, Palgrave Macmillan, Houndmills, Basingstoke, pp. 901–969.
- Bai, J. (2003). Inferential theory for factor models of large dimensions, *Econometrica* **71**: 135–171.
- Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data, *Journal of Econometrics* **122**: 137–183.
- Bai, J. and Li, K. (2012). Maximum likelihood estimation and inference for approximate factor models of high dimension, *Technical Report 42099*, MPRA.
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica* **70**: 191–221.
- Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models, *Journal of Business & Economic Statistics* **25**: 52–60.
- Bai, J. and Ng, S. (2008). Large dimensional factor analysis, *Foundations and Trends in Econometrics* **3**: 89–163.
- Bai, J. and Ng, S. (2013). Principle components estimation and identification of static factors, *Journal of Econometrics* **176**: 18–29.
- Bañbura, M., Giannone, D. and Reichlin, L. (2010). Large Bayesian vector autoregressions, *Journal of Applied Econometrics* **25**: 71–92.
- Banerjee, A. and Marcellino, M. (2008). Factor-augmented error correction models, *Working Paper ECO 2008/15*, European University Institute.
- Banerjee, A., Marcellino, M. and Masten, I. (2013). Structural VECM: Cointegration in large-scale structural FAVAR models, *Technical report*.

- Banerjee, A., Marcellino, M. and Masten, I. (2014). Forecasting with factor-augmented error correction models, *International Journal of Forecasting* .
- Barhoumi, K., Darné, O. and Ferrara, L. (2013). Dynamic factor models: A review of the literature, Discussion Paper, Banque de France.
- Bäurle, G. (2013). Structural dynamic factor analysis using prior information from macroeconomic theory, *Journal of Business & Economic Statistics* **31**: 136–150.
- Bernanke, B. S. and Boivin, J. (2003). Monetary policy in a data-rich environment, *Journal of Monetary Economics* **50**: 525–546.
- Bernanke, B. S., Boivin, J. and Elias, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach, *Quarterly Journal of Economics* **120**: 387–422.
- Boivin, J. and Giannoni, M. P. (2009). Global forces and monetary policy effectiveness, in J. Galí and M. Gertler (eds), *International Dimensions of Monetary Policy*, University of Chicago Press, Chicago, chapter 8, pp. 429–478.
- Boivin, J., Giannoni, M. P. and Mihov, I. (2009). Sticky prices and monetary policy: Evidence from disaggregated US data, *American Economic Review* **99**: 350–384.
- Boivin, J. and Ng, S. (2005). Understanding and comparing factor-based forecasts, *International Journal of Central Banking* **1**: 117–151.
- Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis?, *Journal of Econometrics* **132**: 169–194.
- Bordo, M. D. and Helbling, T. F. (2010). International business cycle synchronization in historical perspective, *Working Paper 16103*, NBER.
- Breitung, J. (2013). The analysis of macroeconomic panel data, mimeo.
- Breitung, J. and Eickmeier, S. (2006). Dynamic factor models, *Allgemeines Statistisches Archiv* **90**: 27–42.
- Breitung, J. and Pigorsch, U. (2013). A canonical correlation approach for selecting the number of dynamic factors, *Oxford Bulletin of Economics and Statistics* **75**: 23–36.
- Breitung, J. and Tenhofen, J. (2011). GLS estimation of dynamic factor models, *Journal of the American Statistical Association* **106**: 1150–1166.
- Brillinger, D. R. (1975). *Time Series Data Analysis and Theory*, Holt, Rinehart and Winston, New York.
- Canova, F. and Ciccarelli, M. (2013). Panel vector autoregressive models: A survey, *Advances in Econometrics* .
- Carriero, A., Kapetanios, G. and Marcellino, M. (2009). Forecasting exchange rates with a large Bayesian VAR, *International Journal of Forecasting* **25**: 400–417.
- Carriero, A., Kapetanios, G. and Marcellino, M. (2012). Forecasting government bond yields with large Bayesian vector autoregressions, *Journal of Banking and Finance* **36**: 2026–2047.
- Cattell, R. B. (1966). A srcee test for the number of factors, *Multivariate Behavioral Research* **1**: 245–276.

- Chamberlain, G. and Rothschild, M. (1983). Arbitrage factor structure, a mean-variance analysis of large asset markets, *Econometrica* **51**: 1281–1304.
- Choi, I. (2012). Efficient estimation of factor models, *Econometric Theory* **28**: 274–308.
- Chudik, A. and Pesaran, M. H. (2011). Infinite-dimensional VARs and factor models, *Journal of Econometrics* **163**: 4–22.
- Connor, G. and Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory - a new framework for analysis, *Journal of Financial Economics* **15**: 373–394.
- Connor, G. and Korajczyk, R. A. (1993). A test for the number of factors in an approximate factor model, *Journal of Finance* **48**: 1263–1291.
- De Mol, C., Giannone, D. and Reichlin, L. (2008). Forecasting using a large number of predictors - is Bayesian regression a valid alternative to principal components?, *Journal of Econometrics* **146**: 318–328.
- Dees, S., Di Mauro, F., Pesaran, M. H. and Smith, V. (2007). Exploring the international linkages of the euro area: A global VAR analysis, *Journal of Applied Econometrics* **22**: 1–38.
- Del Negro, M. and Otrok, C. (2007). 99 Luftballons: Monetary policy and the house price boom across U.S. states, *Journal of Monetary Economics* **54**: 1962–1985.
- Del Negro, M. and Schorfheide, F. (2004). Priors from general equilibrium models for VARs, *International Economic Review* **45**: 643–673.
- Doz, C., Giannone, D. and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering, *Journal of Econometrics* **164**: 188–205.
- Dufour, J.-M. and Stevanović, D. (2013). Factor-augmented VARMA models with macroeconomic applications, *Journal of Business and Economic Statistics* **31**: 491–506.
- Eickmeier, S. (2007). Business cycle transmission from the US to Germany — A structural factor approach, *European Economic Review* **51**: 521–551.
- Eickmeier, S. and Hofmann, B. (2013). Monetary policy, housing booms, and financial (im)balances, *Macroeconomic Dynamics* **17**: 830–860.
- Favero, C. A., Marcellino, M. and Neglia, F. (2005). Principal components at work: The empirical analysis of monetary policy with large data sets, *Journal of Applied Econometrics* **20**: 603–620.
- Forni, M. and Gambetti, L. (2010). The dynamic effects of monetary policy: A structural factor model approach, *Journal of Monetary Economics* **57**(2): 203–216.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation, *Review of Economics and Statistics* **82**: 540–554.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2004). Generalized dynamic factor model consistency and rates, *Journal of Econometrics* **119**: 231–255.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting, *Journal of the American Statistical Association* **100**: 830–840.
- Forni, M. and Lippi, M. (2001). The generalized dynamic factor model: Representation theory, *Econometric Theory* **17**: 1113–1141.

- Forni, M. and Reichlin, L. (1998). Let's get real: A factor analytical approach to disaggregated business cycle dynamics, *Review of Economic Studies* **65**: 453 – 473.
- George, E. I., Sun, D. and Ni, S. (2008). Bayesian stochastic search for VAR model restrictions, *Journal of Econometrics* **142**: 553–580.
- Giannone, D., Lenza, M. and Primiceri, G. (2010). Prior selection for vector autoregressions, *mimeo*, Department of Economics, Free University of Brussels.
- Giannone, D., Reichlin, L. and Sala, L. (2004). Monetary policy in real time, *NBER Macroeconomics Annual* **19**: 161–200.
- Gupta, R., Jurgilas, M., Kabundi, A. and Miller, S. M. (2009). Monetary policy and housing sector dynamics in a large-scale Bayesian vector autoregressive model, *Working Paper 2009-13*, University of Pretoria.
- Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model, *Journal of the American Statistical Association* **102**: 603–617.
- Hallin, M. and Liška, R. (2011). Dynamic factors in the presence of blocks, *Journal of Econometrics* **163**: 29–41.
- Helbling, T. and Bayoumi, T. (2003). Are they all in the same boat? The 2000-2001 growth slowdown and the G7-business cycle linkages, *Working Paper WP/03/46*, International Monetary Fund.
- Jacobs, J. P. A. M. and Otter, P. W. (2008). Determining the number of factors and lag order in dynamic factor models: A minimum entropy approach, *Econometric Reviews* **27**: 385–397.
- King, R. G., Plosser, C. I., Stock, J. H. and Watson, M. W. (1991). Stochastic trends and economic fluctuations, *American Economic Review* **81**: 819–840.
- Koop, G. M. (2013). Forecasting with medium and large Bayesian VARs, *Journal of Applied Econometrics* **28**: 177–203.
- Korobilis, D. (2013). VAR forecasting using Bayesian variable selection, *Journal of Applied Econometrics* **28**: 204–230.
- Kose, M. A., Otrok, C. and Whiteman, C. H. (2003). International business cycles: World, region, and country-specific factors, *American Economic Review* **93**: 1216–1239.
- Kose, M. A., Otrok, C. and Whiteman, C. H. (2008). Understanding the evolution of world business cycles, *Journal of International Economics* **75**: 110–130.
- Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley, Chichester.
- Mansour, J. M. (2003). Do national business cycles have an international origin?, *Empirical Economics* **28**: 223–247.
- Moench, E. and Ng, S. (2011). A hierarchical factor analysis of U.S. housing market dynamics, *Econometrics Journal* **14**: C1–C24.
- Mumtaz, H. and Surico, P. (2009). The transmission of international shocks: A factor-augmented VAR approach, *Journal of Money, Credit and Banking* **41**: 71–100.

- Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models, *Econometrica* **77**: 1447–1479.
- Otrok, C. and Whiteman, C. H. (1998). Bayesian leading indicators: Measuring and predicting economic conditions in Iowa, *International Economic Review* **39**: 997–1014.
- Peña, D. and Box, G. E. P. (1987). Identifying a simplifying structure in time series, *Journal of the American Statistical Association* **82**: 836–843.
- Pesaran, M. H., Schuermann, T. and Weiner, S. M. (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model, *Journal of Business & Economic Statistics* **22**: 129–162.
- Reinsel, G. (1983). Some results on multivariate autoregressive index models, *Biometrika* **70**: 145–156.
- Reinsel, G. C. (1993). *Elements of Multivariate Time Series Analysis*, Springer-Verlag, New York.
- Reinsel, G. C. and Velu, R. P. (1998). *Multivariate Reduced-Rank Regression, Theory and Applications*, Springer-Verlag, New York.
- Sargent, T. J. and Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory, in C. A. Sims (ed.), *New Methods in Business Cycle Research: Proceedings from a Conference*, Federal Reserve Bank of Minneapolis, Minneapolis, pp. 45–109.
- Sims, C. A. and Zha, T. (2006). Were there regime switches in U.S. monetary policy?, *American Economic Review* **96**: 54–81.
- Smets, F. and Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach, *American Economic Review* **97**: 586–606.
- Spearman, C. (1904). “General-intelligence,” objectively determined and measured, *American Journal of Psychology* **15**: 201–293.
- Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association* **97**: 1167–1179.
- Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes, *Journal of Business & Economic Statistics* **20**: 147–162.
- Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factor models for VAR analysis, *manuscript*, Princeton University.
- Stock, J. H. and Watson, M. W. (2006). Forecasting with many predictors, in G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Elsevier, Amsterdam, pp. 515–554.
- Stock, J. H. and Watson, M. W. (2011). Dynamic factor models, in M. P. Clements and D. F. Hendry (eds), *Oxford Handbook of Economic Forecasting*, Oxford University Press, pp. 35–60.
- Stock, J. H. and Watson, M. W. (2012). Disentangling the channels of the 2007-09 recession, *Brookings Papers on Economic Activity* pp. 81–135.
- Tso, M. K.-S. (1981). Reduced-rank regression and canonical analysis, *Journal of the Royal Statistical Society* **B43**: 183–189.
- Velu, R. P., Reinsel, G. C. and Wichern, D. W. (1986). Reduced rank models for multiple time series, *Biometrika* **73**: 105–118.