Dynamic Factor Models

Catherine Doz, Peter Fuleky

WORKING PAPER N° 2019 – 45

Dynamic Factor Models

Catherine Doz
Peter Fuleky

JEL Codes: C32, C38, C53, C55
Keywords : dynamic factor models; big data; two-step estimation; time domain; frequency domain; structural breaks

DYNAMIC FACTOR MODELS *

Catherine Doz[1,2] and Peter Fuleky[3]

[1]Paris School of Economics
[2]Université Paris 1 Panthéon-Sorbonne
[3]University of Hawaii at Manoa

July 2019

*Prepared for *Macroeconomic Forecasting in the Era of Big Data Theory and Practice* (Peter Fuleky ed), Springer.

Chapter 2
Dynamic Factor Models

Catherine Doz and Peter Fuleky

Abstract Dynamic factor models are parsimonious representations of relationships among time series variables. With the surge in data availability, they have proven to be indispensable in macroeconomic forecasting. This chapter surveys the evolution of these models from their pre-big-data origins to the large-scale models of recent years. We review the associated estimation theory, forecasting approaches, and several extensions of the basic framework.

Keywords: dynamic factor models; big data; two-step estimation; time domain; frequency domain; structural breaks;
JEL Codes: C32, C38, C53, C55

2.1 Introduction

Factor analysis is a dimension reduction technique summarizing the sources of variation among variables. The method was introduced in the psychology literature by Spearman (1904), who used an unobserved variable, or factor, to describe the cognitive abilities of an individual. Although originally developed for independently distributed random vectors, the method was extended by Geweke (1977) to capture the co-movements in economic time series. The idea that the co-movement of macroeconomic series can be linked to the business cycle has been put forward by Burns and Mitchell (1946): "a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycle; this sequence of

Catherine Doz
Paris School of Economics and University Paris 1 Panthéon-Sorbonne, 48 Boulevard Jourdan, 75014 Paris, France. e-mail: catherine.doz@univ-paris1.fr

Peter Fuleky
University of Hawaii, 2424 Maile Way, Honolulu, HI 96822, USA. e-mail: fuleky@hawaii.edu

changes is recurrent but not periodic." Early applications of dynamic factor models (DFMs) to macroeconomic data, by Sargent and Sims (1977) and Stock and Watson (1989, 1991, 1993), suggested that a few latent factors can account for much of the dynamic behavior of major economic aggregates. In particular, a dynamic single-factor model can be used to summarize a vector of macroeconomic indicators, and the factor can be seen as an index of economic conditions describing the business cycle. In these studies, the number of time periods in the data set exceeded the number of variables, and identification of the factors required relatively strict assumptions. While increments in the time dimension were limited by the passage of time, the availability of a large number of macroeconomic and financial indicators provided an opportunity to expand the dataset in the cross-sectional dimension and to work with somewhat relaxed assumptions. Chamberlain (1983) and Chamberlain and Rothschild (1983) applied factor models to wide panels of financial data and paved the way for further development of dynamic factor models in macroeconometrics. Indeed, since the 2000's dynamic factor models have been used extensively to analyze large macroeconomic data sets, sometimes containing hundreds of series with hundreds of observations on each. They have proved useful for synthesizing information from variables observed at different frequencies, estimation of the latent business cycle, nowcasting and forecasting, and estimation of recession probabilities and turning points. As Diebold (2003) pointed out, although DFMs "don't analyze *really* Big Data, they certainly represent a movement of macroeconometrics in that direction", and this movement has proved to be very fruitful.

Several very good surveys have been written on dynamic factor models, including Bai and Ng (2008b); Bai and Wang (2016); Barigozzi (2018); Breitung and Eickmeier (2006); Darné et al. (2014); Lütkepohl (2014); Stock and Watson (2006, 2011, 2016). Yet we felt that the readers of this volume would appreciate a chapter covering the evolution of dynamic factor models, their estimation strategies, forecasting approaches, and several extensions to the basic framework. Since both small- and large-dimensional models have advanced over time, we review the progress achieved under both frameworks. In Section 2.2 we describe the distinguishing characteristics of exact and approximate factor models. In Sections 2.3 and 2.4, we review estimating procedures proposed in the time domain and the frequency domain, respectively. Section 2.5 presents approaches for determining the number of factors. Section 2.6 surveys issues associated with forecasting, and Section 2.8 reviews the handling of structural breaks in dynamic factor models.

2.2 From Exact to Approximate Factor Models

In a factor model, the correlations among n variables, $\boldsymbol{x}_1 \ldots \boldsymbol{x}_n$, for which T observations are available, are assumed to be entirely due to a few, $r < n$, latent unobservable variables, called factors. The link between the observable variables and the factors is assumed to be linear. Thus, each observation x_{it} can be decomposed as

$$x_{it} = \mu_i + \boldsymbol{\lambda}_i' \boldsymbol{f}_t + e_{it},$$

where μ_i is the mean of $\boldsymbol{x}_i$, $\boldsymbol{\lambda}_i$ is an $r \times 1$ vector, and e_{it} and $\boldsymbol{f}_t$ are two uncorrelated processes. Thus, for $i = 1 \ldots n$ and $t = 1 \ldots T$, x_{it} is decomposed into the sum of two mutually orthogonal unobserved components: the common component, $\chi_{it} = \boldsymbol{\lambda}_i' \boldsymbol{f}_t$, and the idiosyncratic component, $\xi_{it} = \mu_i + e_{it}$. While the factors drive the correlation between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j, j \neq i$, the idiosyncratic component arises from features that are specific to an individual $\boldsymbol{x}_i$ variable. Further assumptions placed on the two unobserved components result in factor models of different types.

2.2.1 Exact factor models

The exact factor model was introduced by Spearman (1904). The model assumes that the idiosyncratic components are not correlated at any leads and lags so that ξ_{it} and ξ_{ks} are mutually orthogonal for all $k \neq i$ and any s and t, and consequently all correlation among the observable variables is driven by the factors. The model can be written as

$$x_{it} = \mu_i + \sum_{j=1}^{r} \lambda_{ij} f_{jt} + e_{it}, \quad i = 1 \ldots n,\ t = 1 \ldots T, \tag{2.1}$$

where f_{jt} and e_{it} are orthogonal white noises for any i and j; e_{it} and e_{kt} are orthogonal for $k \neq i$; and λ_{ij} is the loading of the j^{th} factor on the i^{th} variable. Equation (2.1) can also be written as:

$$x_{it} = \mu_i + \boldsymbol{\lambda}_i' \boldsymbol{f}_t + e_{it}, \quad i = 1 \ldots n,\ t = 1 \ldots T,$$

with $\boldsymbol{\lambda}_i' = (\lambda_{i1} \ldots \lambda_{ir})$ and $\boldsymbol{f}_t = (f_{1t} \ldots f_{rt})'$. The common component is $\chi_{it} = \sum_{j=1}^{r} \lambda_{ij} f_{jt}$, and the idiosyncratic component is $\xi_{it} = \mu_i + e_{it}$.

In matrix notation, with $\boldsymbol{x}_t = (x_{1t} \ldots x_{nt})'$, $\boldsymbol{f}_t = (f_{1t} \ldots f_{rt})'$ and $\boldsymbol{e}_t = (e_{1t} \ldots e_{nt})'$, the model can be written as

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{e}_t, \tag{2.2}$$

where $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1 \ldots \boldsymbol{\lambda}_n)'$ is a $n \times r$ matrix of full column rank (otherwise fewer factors would suffice), and the covariance matrix of e_t is diagonal since the idiosyncratic terms are assumed to be uncorrelated. Observations available for $t = 1, \ldots T$ can be stacked, and Equation (2.2) can be rewritten as

$$\boldsymbol{X} = \boldsymbol{\iota}_T \otimes \boldsymbol{\mu}' + \boldsymbol{F} \boldsymbol{\Lambda}' + \boldsymbol{E},$$

where $\boldsymbol{X} = (\boldsymbol{x}_1 \ldots \boldsymbol{x}_T)'$ is a $T \times n$ matrix, $\boldsymbol{F} = (\boldsymbol{f}_1 \ldots \boldsymbol{f}_T)'$ a $T \times r$ matrix, $\boldsymbol{E} = (\boldsymbol{e}_1 \ldots \boldsymbol{e}_T)'$ is a $T \times n$ matrix and $\boldsymbol{\iota}_T$ is a $T \times 1$ vector with components equal to 1.

While the core assumption behind factor models is that the two processes $\boldsymbol{f}_t$ and $\boldsymbol{e}_t$ are orthogonal to each other, the exact static factor model further assumes that the idiosyncratic components are also orthogonal to each other, so that any correlation between the observable variables is solely due to the common factors. Both orthogonality assumptions are necessary to ensure the identifiability of the model (see Anderson, 1984; Bartholomew, 1987). Note that $\boldsymbol{f}_t$ is defined only up to a premultiplication by an invertible matrix $\boldsymbol{Q}$ since $\boldsymbol{f}_t$ can be replaced by $\boldsymbol{Q}\boldsymbol{f}_t$ whenever $\boldsymbol{\Lambda}$ is replaced by $\boldsymbol{\Lambda}\boldsymbol{Q}^{-1}$. This means that only the respective spaces spanned by $\boldsymbol{f}_t$ and by $\boldsymbol{\Lambda}$ are uniquely defined. This so-called indeterminacy problem of the factors must be taken into account at the estimation stage.

Factor models have been introduced in economics by Engle and Watson (1981); Geweke (1977); Sargent and Sims (1977). These authors generalized the model above to capture dynamics in the data. Using the same notation as in Equation (2.2), the dynamic exact factor model can be written as

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{\Lambda}_0\boldsymbol{f}_t + \boldsymbol{\Lambda}_1\boldsymbol{f}_{t-1} + \cdots + \boldsymbol{\Lambda}_s\boldsymbol{f}_{t-s} + \boldsymbol{e}_t,$$

or more compactly

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{\Lambda}(L)\boldsymbol{f}_t + \boldsymbol{e}_t, \tag{2.3}$$

where L is the lag operator.

In this model, $\boldsymbol{f}_t$ and $\boldsymbol{e}_t$ are no longer assumed to be white noises, but are instead allowed to be autocorrelated dynamic processes evolving according to $\boldsymbol{f}_t = \boldsymbol{\Theta}(L)\boldsymbol{u}_t$ and $\boldsymbol{e}_t = \boldsymbol{\rho}(L)\boldsymbol{\varepsilon}_t$, where the q and n dimensional vectors $\boldsymbol{u}_t$ and $\boldsymbol{\varepsilon}_t$, respectively, contain iid errors. The dimension of $\boldsymbol{f}_t$ is also q, which is therefore referred to as the number of *dynamic factors*. In most of the dynamic factor models literature, $\boldsymbol{f}_t$ and $\boldsymbol{e}_t$ are generally assumed to be stationary processes, so if necessary, the observable variables are pre-processed to be stationary. (In this chapter, we only consider stationary variables; non-stationary models will be discussed in Chapter 17.)

The model admits a *static representation*

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{F}_t + \boldsymbol{e}_t, \tag{2.4}$$

with $\boldsymbol{F}_t = (\boldsymbol{f}_t', \boldsymbol{f}_{t-1}', \ldots, \boldsymbol{f}_{t-s}')'$, an $r = q(1+s)$ dimensional vector of *static factors*, and $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_0, \boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_s)$, a $n \times r$ matrix of loading coefficients. The dynamic representation in (2.2.1) and (2.3) captures the dependence of the observed variables on the lags of the factors explicitly, while the static representation in (2.4) embeds those dynamics implicitly. The two forms lead to different estimation methods to be discussed below.

Exact dynamic factor models assume that the cross-sectional dimension of the data set, $n < T$, is finite, and they are usually used with small, $n \ll T$, samples. There are two reasons for these models to be passed over in the $n \to \infty$ case. First, maximum likelihood estimation, the typical method of choice, requires specifying a full parametric model and imposes a practical limitation on the number of parameters that can be estimated as $n \to \infty$. Second, as $n \to \infty$, some of the unrealistic

assumptions imposed on exact factor models can be relaxed, and the approximate factor model framework, discussed in the next section, can be used instead.

2.2.2 Approximate factor models

As noted above, exact factor models rely on a very strict assumption of no cross-correlation between the idiosyncratic components. In two seminal papers Chamberlain (1983) and Chamberlain and Rothschild (1983) introduced approximate factor models by relaxing this assumption. They allowed the idiosyncratic components to be mildly cross-correlated and provided a set of conditions ensuring that approximate factor models were asymptotically identified as $n \to \infty$.

Let $\boldsymbol{x}_t^n = (x_{1t}, \ldots, x_{nt})'$ denote the vector containing the t^{th} observation of the first n variables as $n \to \infty$, and let $\boldsymbol{\Sigma}_n = \text{cov}(\boldsymbol{x}_t^n)$ be the covariance matrix of $\boldsymbol{x}_t$. Denoting by $\lambda_1(\boldsymbol{A}) \geq \lambda_2(\boldsymbol{A}) \geq \cdots \geq \lambda_n(\boldsymbol{A})$ the ordered eigenvalues of any symmetric matrix $\boldsymbol{A}$ with size $(n \times n)$, the assumptions underlying approximate factor models are the following:

(CR1) $\text{Sup}_n \lambda_r(\boldsymbol{\Sigma}_n) = \infty$ (the r largest eigenvalues of $\boldsymbol{\Sigma}_n$ are diverging)
(CR2) $\text{Sup}_n \lambda_{r+1}(\boldsymbol{\Sigma}_n) < \infty$ (the remaining eigenvalues of $\boldsymbol{\Sigma}_n$ are bounded)
(CR3) $\text{Inf}_n \lambda_n(\boldsymbol{\Sigma}_n) > 0$ ($\boldsymbol{\Sigma}_n$ does not approach singularity)

The authors show that under assumptions (CR1)-(CR3), there exists a unique decomposition $\boldsymbol{\Sigma}_n = \boldsymbol{\Lambda}_n \boldsymbol{\Lambda}_n' + \boldsymbol{\Psi}_n$, where $\boldsymbol{\Lambda}_n$ is a sequence of nested $n \times r$ matrices with rank r and $\lambda_i(\boldsymbol{\Lambda}_n \boldsymbol{\Lambda}_n') \to \infty$, $\forall i = 1 \ldots r$, and $\lambda_1(\boldsymbol{\Psi}_n) < \infty$. Alternatively, $\boldsymbol{x}_t$ can be decomposed using a pair of mutually orthogonal random vector processes $\boldsymbol{f}_t$ and $\boldsymbol{e}_t^n$

$$\boldsymbol{x}_t^n = \boldsymbol{\mu}_n + \boldsymbol{\Lambda}_n \boldsymbol{f}_t + \boldsymbol{e}_t^n,$$

with $\text{cov}(\boldsymbol{f}_t) = \boldsymbol{I}_r$ and $\text{cov}(\boldsymbol{e}_t^n) = \boldsymbol{\Psi}_n$, where the r common factors are pervasive, in the sense that the number of variables affected by each factor grows with n, and the idiosyncratic terms may be mildly correlated with bounded covariances.

Although originally developed in the finance literature (see also Connor and Korajczyk, 1986, 1988, 1993), the approximate static factor model made its way into macroeconometrics in the early 2000's (see for example Bai, 2003; Bai and Ng, 2002; Stock and Watson, 2002a,b). These papers use assumptions that are analogous to (CR1)-(CR3) for the covariance matrices of the factor loadings and the idiosyncratic terms, but they add complementary assumptions (which vary across authors for technical reasons) to accommodate the fact that the data under study are autocorrelated time series. The models are mainly used with data that is stationary or preprocessed to be stationary, but they have also been used in a non-stationary framework (see Bai, 2004; Bai and Ng, 2004). The analysis of non-stationary data will be addressed in detail in Chapter 17.

Similarly to exact dynamic factor models, approximate dynamic factor models also rely on an equation linking the observable series to the factors and their lags,

but here the idiosyncratic terms can be mildly cross-correlated, and the number of series is assumed to tend to infinity. The model has a dynamic

$$\boldsymbol{x}_t^n = \boldsymbol{\mu}_n + \boldsymbol{\Lambda}_n(L)\boldsymbol{f}_t + \boldsymbol{e}_t^n$$

and a static representation

$$\boldsymbol{x}_t^n = \boldsymbol{\mu}_n + \boldsymbol{\Lambda}_n\boldsymbol{F}_t + \boldsymbol{e}_t^n$$

equivalent to (2.3) and (2.4), respecively.

By centering the observed series, $\boldsymbol{\mu}_n$ can be set equal to zero. For the sake of simplicity, in the rest of the chapter we will assume that the variables are centered, and we also drop the sub- and superscript n in $\boldsymbol{\Lambda}_n$, $\boldsymbol{x}_t^n$ and $\boldsymbol{e}_t^n$. We will always assume that in exact factor models the number of series under study is finite, while in approximate factor models $n \to \infty$, and a set of assumptions that is analogous to (CR1)-(CR3)—but suited to the general framework of stationary autocorrelated processes—is satisfied.

2.3 Estimation in the Time Domain

2.3.1 Maximum likelihood estimation of small factor models

The static exact factor model has generally been estimated by maximum likelihood under the assumption that $(\boldsymbol{f}_t)_{t\in\mathbb{Z}}$ and $(\boldsymbol{e}_t)_{t\in\mathbb{Z}}$ are two orthogonal iid gaussian processes. Unique identification of the model requires that we impose some restrictions on the model. One of these originates from the definition of the exact factor model: the idiosyncratic components are set to be mutually orthogonal processes with a diagonal variance matrix. A second restriction sets the variance of the factors to be the identity matrix, $Var(\boldsymbol{f}_t) = \boldsymbol{I}_r$. While the estimator does not have a closed form analytical solution, for small n the number of parameters is small, and estimates can be obtained through any numerical optimization procedure. Two specific methods have been proposed for this problem: the so-called zig-zag routine, an algorithm which solves the first order conditions (see for instance Anderson, 1984; Lawley and Maxwell, 1971; Magnus and Neudecker, 2019) and the Jöreskog (1967) approach, which relies on the maximization of the concentrated likelihood using a Fletcher-Powell algorithm. Both approaches impose an additional identifying restriction on $\boldsymbol{\Lambda}$. The maximum likelihood estimators $\hat{\boldsymbol{\Lambda}}$, $\hat{\boldsymbol{\Psi}}$ of the model parameters $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}$ are $\sqrt{T}$ consistent and asymptotically gaussian (see Anderson and Rubin, 1956). Under standard stationarity assumptions, these asymptotic results are still valid even if the true distribution of $\boldsymbol{f}_t$ or $\boldsymbol{e}_t$ is not gaussian: in this case $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\Psi}}$ are QML estimators of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$.

Various formulas have been proposed to estimate the factors, given the parameter estimates $\hat{\boldsymbol{\Lambda}}$, $\hat{\boldsymbol{\Psi}}$. Commonly used ones include

- $\widehat{\boldsymbol{f}}_t = \widehat{\boldsymbol{\Lambda}}'(\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Lambda}}'+\widehat{\boldsymbol{\Psi}})^{-1}\boldsymbol{x}_t$, the conditional expectation of $\boldsymbol{f}_t$ given $\boldsymbol{x}_t$ and the estimated values of the parameters, which, after elementary calculations, can also be written as $\widehat{\boldsymbol{f}}_t = (\boldsymbol{I}_r + \widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Psi}}^{-1}\widehat{\boldsymbol{\Lambda}})^{-1}\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Psi}}^{-1}\boldsymbol{x}_t$, and
- $\widehat{\boldsymbol{f}}_t = (\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Psi}}^{-1}\widehat{\boldsymbol{\Lambda}})^{-1}\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Psi}}^{-1}\boldsymbol{x}_t$ which is the FGLS estimator of $\boldsymbol{f}_t$, given the estimated loadings.

Additional details about these two estimators are provided by Anderson (1984). Since the formulas are equivalent up to an invertible matrix, the spaces spanned by the estimated factors are identical across these methods.

A dynamic exact factor model can also be estimated by maximum likelihood under the assumption of gaussian $(\boldsymbol{f}_t', \boldsymbol{e}_t')_{t\in\mathbb{Z}}$[1]. In this case, the factors are assumed to follow vector autoregressive processes, and the model can be cast in state-space form. To make things more precise, let us consider a factor model where the vector of factors follows a VAR(p) process and enters the equation for $\boldsymbol{x}_t$ with s lags

$$\boldsymbol{x}_t = \boldsymbol{\Lambda}_0\boldsymbol{f}_t + \boldsymbol{\Lambda}_1\boldsymbol{f}_{t-1} + \cdots + \boldsymbol{\Lambda}_s\boldsymbol{f}_{t-s} + \boldsymbol{e}_t, \tag{2.5}$$

$$\boldsymbol{f}_t = \boldsymbol{\Phi}_1\boldsymbol{f}_t + \cdots + \boldsymbol{\Phi}_p\boldsymbol{f}_{t-p} + \boldsymbol{u}_t, \tag{2.6}$$

where the VAR(p) coefficient matrices, $\boldsymbol{\Phi}$, capture the dynamics of the factors. A commonly used identification restriction sets the variance of the innovations to the identity matrix, $\operatorname{cov}(\boldsymbol{u}_t) = \boldsymbol{I}_r$, and additional identifying restrictions are imposed on the factor loadings. The state-space representation is very flexible and can accommodate different cases as shown below:

- If $s \geq p-1$ and if $\boldsymbol{e}_t$ is a white noise, the measurement equation is $\boldsymbol{x}_t = \boldsymbol{\Lambda}\boldsymbol{F}_t + \boldsymbol{e}_t$ with $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_0\ \boldsymbol{\Lambda}_1 \ldots \boldsymbol{\Lambda}_s)$ and $\boldsymbol{F}_t = (\boldsymbol{f}_t'\boldsymbol{f}_{t-1}' \ldots \boldsymbol{f}_{t-s}')'$ (static representation of the dynamic model), and the state equation is

$$\begin{bmatrix} \boldsymbol{f}_t \\ \boldsymbol{f}_{t-1} \\ \vdots \\ \boldsymbol{f}_{t-p+1} \\ \vdots \\ \boldsymbol{f}_{t-s} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \ldots & \boldsymbol{\Phi}_p & \boldsymbol{O} & \ldots & \boldsymbol{O} \\ \boldsymbol{I}_q & \boldsymbol{O} & \ldots & \ldots & \ldots & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{I}_q & \boldsymbol{O} & \ldots & \ldots & \boldsymbol{O} \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \boldsymbol{O} & \ldots & \ldots & \boldsymbol{O} & \boldsymbol{I}_q & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_{t-1} \\ \boldsymbol{f}_{t-2} \\ \vdots \\ \boldsymbol{f}_{t-p} \\ \vdots \\ \boldsymbol{f}_{t-s-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{I}_q \\ \boldsymbol{O} \\ \vdots \\ \boldsymbol{O} \end{bmatrix} \boldsymbol{u}_t$$

- If $s < p-1$ and if $\boldsymbol{e}_t$ is a white noise, the measurement equation is $\boldsymbol{x}_t = \boldsymbol{\Lambda}\boldsymbol{F}_t + \boldsymbol{e}_t$ with $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_0\ \boldsymbol{\Lambda}_1 \ldots \boldsymbol{\Lambda}_s\ \boldsymbol{O}\ \ldots\ \boldsymbol{O})$ and $\boldsymbol{F}_t = (\boldsymbol{f}_t'\boldsymbol{f}_{t-1}' \ldots \boldsymbol{f}_{t-p+1}')'$, and the state equation is

$$\begin{bmatrix} \boldsymbol{f}_t \\ \boldsymbol{f}_{t-1} \\ \vdots \\ \boldsymbol{f}_{t-p+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \ldots & \boldsymbol{\Phi}_p \\ \boldsymbol{I}_q & \boldsymbol{O} & \ldots & \boldsymbol{O} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{O} & \ldots & \boldsymbol{I}_q & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_{t-1} \\ \boldsymbol{f}_{t-2} \\ \vdots \\ \boldsymbol{f}_{t-p} \end{bmatrix} + \begin{bmatrix} \boldsymbol{I}_q \\ \boldsymbol{O} \\ \vdots \\ \boldsymbol{O} \end{bmatrix} \boldsymbol{u}_t \tag{2.7}$$

[1] When $(\boldsymbol{f}_t)$ and $(\boldsymbol{e}_t)$ are not gaussian, the procedure gives QML estimators.

- The state-space representation can also accommodate the case where each of the idiosyncratic components is itself an autoregressive process. For instance, if $s = p = 2$ and e_{it} follows a second order autoregressive process with $e_{it} = d_{i1}e_{it-1} + d_{i2}e_{it-2} + \varepsilon_{it}$ for $i = 1 \dots n$, then the measurement equation can be written as $\boldsymbol{x}_t = \boldsymbol{\Lambda}\boldsymbol{\alpha}_t$ with $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_0\ \boldsymbol{\Lambda}_1\ \boldsymbol{\Lambda}_2\ \boldsymbol{I}_n\ \mathbf{O})$ and $\boldsymbol{\alpha}_t = (\boldsymbol{f}_t'\ \boldsymbol{f}_{t-1}'\ \boldsymbol{f}_{t-2}'\ \boldsymbol{e}_t'\ \boldsymbol{e}_{t-1}')'$, and the state equation is

$$\begin{bmatrix} \boldsymbol{f}_t \\ \boldsymbol{f}_{t-1} \\ \boldsymbol{f}_{t-2} \\ \boldsymbol{e}_t \\ \boldsymbol{e}_{t-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \boldsymbol{I}_q & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{I}_q & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \boldsymbol{D}_1 & \boldsymbol{D}_2 \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \boldsymbol{I}_n & \mathbf{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_{t-1} \\ \boldsymbol{f}_{t-2} \\ \boldsymbol{f}_{t-3} \\ \boldsymbol{e}_{t-1} \\ \boldsymbol{e}_{t-2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{I}_q & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{I}_n \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{\varepsilon}_t \end{bmatrix}, \tag{2.8}$$

  with $\boldsymbol{D}_j = \operatorname{diag}\left(d_{1j} \dots d_{nj}\right)$ for $j = 1,\ 2$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t} \dots \varepsilon_{nt})'$. This model, with $n = 4$ and $r = 1$, was used by Stock and Watson (1991) to build a coincident index for the US economy.

One computational challenge is keeping the dimension of the state vector small when n is large. The inclusion of the idiosyncratic component in $\boldsymbol{\alpha}_t$ implies that the dimension of system matrices in the state equation (2.8) grows with n. Indeed, this formulation requires an element for each lag of each idiosyncratic component. To limit the computational cost, an alternative approach applies the filter $\boldsymbol{I}_n - \boldsymbol{D}_1 L - \boldsymbol{D}_2 L^2$ to the measurement equation. This pre-whitening is intended to control for serial correlation in the idiosyncratic terms, so that they don't need to be included in the state equation. The transformed measurement equation takes the form $\boldsymbol{x}_t = \boldsymbol{c}_t + \widetilde{\boldsymbol{\Lambda}}\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t$, for $t \geq 2$, with $\boldsymbol{c}_t = \boldsymbol{D}_1\boldsymbol{x}_{t-1} + \boldsymbol{D}_2\boldsymbol{x}_{t-2}$, $\widetilde{\boldsymbol{\Lambda}} = \left(\widetilde{\boldsymbol{\Lambda}}_0\ \dots\ \widetilde{\boldsymbol{\Lambda}}_4\right)$ and $\boldsymbol{\alpha}_t = (\boldsymbol{f}_t\ \boldsymbol{f}_{t-1}\ \dots\ \boldsymbol{f}_{t-4})'$, since

$$\begin{aligned} (\boldsymbol{I}_n - \boldsymbol{D}_1 L - \boldsymbol{D}_2 L^2)\boldsymbol{x}_t = &\boldsymbol{\Lambda}_0\boldsymbol{f}_t + (\boldsymbol{\Lambda}_1 - \boldsymbol{D}_1\boldsymbol{\Lambda}_0)\boldsymbol{f}_{t-1} + \\ &(\boldsymbol{\Lambda}_2 - \boldsymbol{D}_1\boldsymbol{\Lambda}_1 - \boldsymbol{D}_2\boldsymbol{\Lambda}_0)\boldsymbol{f}_{t-2} - \\ &(\boldsymbol{D}_1\boldsymbol{\Lambda}_2 + \boldsymbol{D}_2\boldsymbol{\Lambda}_1)\boldsymbol{f}_{t-3} - (\boldsymbol{D}_2\boldsymbol{\Lambda}_2)\boldsymbol{f}_{t-4} + \boldsymbol{\varepsilon}_t \end{aligned} \tag{2.9}$$

The introduction of lags of $\boldsymbol{x}_t$ in the measurement equation does not cause further complications; they can be incorporated in the Kalman filter since they are known at time t. The associated state equation is straightforward and the dimension of $\boldsymbol{\alpha}_t$ is smaller than in (2.8).

Once the model is written in state-space form, the gaussian likelihood can be computed using the Kalman filter for any value of the parameters (see for instance Harvey, 1989), and the likelihood can be maximized by any numerical optimization procedure over the parameter space. Watson and Engle (1983) proposed to use a score algorithm, or the EM algorithm, or a combination of both, but any other numerical procedure can be used when n is small. With the parameter estimates, $\widehat{\boldsymbol{\theta}}$, in hand, the Kalman smoother provides an approximation of $\boldsymbol{f}_t$ using information from all observations $\widehat{\boldsymbol{f}}_{t|T} = E(\boldsymbol{f}_t|\boldsymbol{x}_1, \dots, \boldsymbol{x}_T, \widehat{\boldsymbol{\theta}})$. Asymptotic consistency and normality of

the parameter estimators and the factors follow from general results concerning maximum likelihood estimation with the Kalman filter.

To further improve the computational efficiency of estimation, Jungbacker and Koopman (2014) reduced a high-dimensional dynamic factor model to a low-dimensional state space model. Using a suitable transformation of the measurement equation, they partitioned the observation vector into two mutually uncorrelated subvectors, with only one of them depending on the unobserved state. The transformation can be summarized by

$$\boldsymbol{x}_t^* = \boldsymbol{A}\boldsymbol{x}_t \quad \text{with} \quad \boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}^L \\ \boldsymbol{A}^H \end{bmatrix} \quad \text{and} \quad \boldsymbol{x}_t^* = \begin{bmatrix} \boldsymbol{x}_t^L \\ \boldsymbol{x}_t^H \end{bmatrix},$$

where the model for $\boldsymbol{x}_t^*$ can be written as

$$\boldsymbol{x}_t^L = \boldsymbol{A}^L\boldsymbol{\Lambda}\boldsymbol{F}_t + \boldsymbol{e}_t^L \quad \text{and} \quad \boldsymbol{x}_t^H = \boldsymbol{e}_t^H,$$

with $\boldsymbol{e}_t^L = \boldsymbol{A}^L\boldsymbol{e}_t$ and $\boldsymbol{e}_t^H = \boldsymbol{A}^H\boldsymbol{e}_t$. Consequently, the $\boldsymbol{x}_t^H$ subvector is not required for signal extraction and the Kalman filter can be applied to a lower dimensional collapsed model, leading to substantial computational savings. This approach can be combined with controlling for idiosyncratic dynamics in the measurement equation, as described above in (2.9).

2.3.2 Principal component analysis of large approximate factor models

Chamberlain and Rothschild (1983) suggested to use principal component analysis (PCA) to estimate the approximate static factor model, and Stock and Watson (2002a,b) and Bai and Ng (2002) popularized this approach in macro-econometrics. PCA will be explored in greater detail in Chapter 8, but we state the main results here.

Considering centered data and assuming that the number of factors, r, is known, PCA allows to simultaneously estimate the factors and their loadings by solving the least squares problem

$$\min_{\boldsymbol{\Lambda},\boldsymbol{F}} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it} - \lambda_i'\boldsymbol{f}_t\right)^2 = \min_{\boldsymbol{\Lambda},\boldsymbol{F}} \frac{1}{nT}\sum_{t=1}^{T}(\boldsymbol{x}_t - \boldsymbol{\Lambda}\boldsymbol{f}_t)'(\boldsymbol{x}_t - \boldsymbol{\Lambda}\boldsymbol{f}_t). \tag{2.10}$$

Due to the aforementioned rotational indeterminacy of the factors and their loadings, the parameter estimates have to be constrained to get a unique solution. Generally, one of the following two normalization conditions is imposed

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{f}}_t\widehat{\boldsymbol{f}}_t' = \boldsymbol{I}_r \quad \text{or} \quad \frac{\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Lambda}}}{n} = \boldsymbol{I}_r.$$

Using the first normalization and concentrating out $\mathbf{\Lambda}$ gives an estimated factor matrix, $\widehat{\boldsymbol{F}}$, which is T times the eigenvectors corresponding to the r largest eigenvalues of the $T \times T$ matrix $\boldsymbol{X}\boldsymbol{X}'$. Given $\widehat{\boldsymbol{F}}$, $\widehat{\mathbf{\Lambda}} = (\widehat{\boldsymbol{F}}'\widehat{\boldsymbol{F}})^{-1}\widehat{\boldsymbol{F}}'\boldsymbol{X} = \widehat{\boldsymbol{F}}'\boldsymbol{X}/T$ is the corresponding matrix of factor loadings. The solution to the minimization problem above is not unique, even though the sum of squared residuals is unique. Another solution is given by $\widetilde{\mathbf{\Lambda}}$ constructed as n times the eigenvectors corresponding to the r largest eigenvalues of the $n \times n$ matrix $\boldsymbol{X}'\boldsymbol{X}$. Using the second normalization here implies $\widetilde{\boldsymbol{F}} = \boldsymbol{X}\widetilde{\mathbf{\Lambda}}/n$. Bai and Ng (2002) indicated that the latter approach is computationally less costly when $T > n$, while the former is less demanding when $T < n$. In both cases, the idiosyncratic components are estimated by $\widehat{\boldsymbol{e}}_t = \boldsymbol{x}_t - \widehat{\mathbf{\Lambda}}\widehat{\boldsymbol{f}}_t$, and their covariance is estimated by the empirical covariance matrix of $\widehat{\boldsymbol{e}}_t$. Since PCA is not scale invariant, many authors (for example Stock and Watson, 2002a,b) center and standardize the series, generally measured in different units, and as a result PCA is applied to the sample correlation matrix in this case.

Stock and Watson (2002a,b) and Bai and Ng (2002) proved the consistency of these estimators, and Bai (2003) obtained their asymptotic distribution under a stronger set of assumptions. We refer the reader to those papers for the details, but let us note that these authors replace assumption (CR1) by the following stronger assumption[2]

$$\lim_{n\to\infty} \frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{n} = \mathbf{\Sigma}_{\mathbf{\Lambda}}$$

and replace assumptions (CR2) and (CR3), which were designed for white noise data, by analogous assumptions taking autocorrelation in the factors and idiosyncratic terms into account. Under these assumptions, the factors and the loadings are proved to be consistent, up to an invertible matrix $\boldsymbol{H}$, converging at rate $\boldsymbol{\delta}_{nT} = 1/\min(\sqrt{n}, \sqrt{T})\boldsymbol{\iota}_r$, so that

$$\widehat{\boldsymbol{f}}_t - \boldsymbol{H}\boldsymbol{f}_t = O_P(\boldsymbol{\delta}_{nT}) \text{ and } \widehat{\boldsymbol{\lambda}}_i - \boldsymbol{H}^{-1}\boldsymbol{\lambda}_i = O_P(\boldsymbol{\delta}_{nT}),\ \forall\, i = 1 \ldots n.$$

Under a more stringent set of assumptions, Bai (2003) also obtains the following asymptotic distribution results:

- If $\sqrt{n}/T \to 0$ then, for each t: $\sqrt{n}(\widehat{\boldsymbol{f}}_t - \boldsymbol{H}'\boldsymbol{f}_t) \underset{d}{\to} \mathcal{N}(\boldsymbol{0}, \mathbf{\Omega}_t)$ where $\mathbf{\Omega}_t$ is known.
- If $\sqrt{T}/n \to 0$ then, for each i: $\sqrt{T}(\widehat{\boldsymbol{\lambda}}_i - \boldsymbol{H}^{-1}\boldsymbol{\lambda}_i) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{W}_i)$ where $\boldsymbol{W}_i$ is known.

2.3.3 Generalized principal component analysis of large approximate factor models

Generalized principal components estimation mimics generalized least squares to deal with a nonspherical variance matrix of the idiosyncratic components. Although

[2] A weaker assumption can also be used: $0 < \underline{c} \leq \liminf_{n\to\infty} \lambda_r(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{n}) < \limsup_{n\to\infty} \lambda_1(\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{n}) \leq \overline{c} < \infty$ (see Doz et al., 2011).

the method had been used earlier, Choi (2012) was the one who proved that efficiency gains can be achieved by including a weighting matrix in the minimization

$$\min_{\boldsymbol{\Lambda},\boldsymbol{F}} \frac{1}{nT}\sum_{t=1}^{T}(\boldsymbol{x}_t - \boldsymbol{\Lambda}\boldsymbol{f}_t)'\boldsymbol{\Psi}^{-1}(\boldsymbol{x}_t - \boldsymbol{\Lambda}\boldsymbol{f}_t).$$

The feasible generalized principal component estimator replaces the unknown $\boldsymbol{\Psi}$ by an estimator $\widehat{\boldsymbol{\Psi}}$. The diagonal elements of $\boldsymbol{\Psi}$ can be estimated by the variances of the individual idiosyncratic terms (see Boivin and Ng, 2006; Jones, 2001). Bai and Liao (2013) derive the conditions under which the estimated $\widehat{\boldsymbol{\Psi}}$ can be treated as known. The weighted minimization problem above relies on the assumption of independent idiosyncratic shocks, which may be too restrictive in practice. Stock and Watson (2005) applied a diagonal filter $\boldsymbol{D}(L)$ to the idiosyncratic terms to deal with serial correlation, so the problem becomes

$$\min_{\boldsymbol{D}(L),\boldsymbol{\Lambda},\widetilde{\boldsymbol{F}}} \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{D}(L)\boldsymbol{x}_t - \boldsymbol{\Lambda}\widetilde{\boldsymbol{f}}_t)'\widetilde{\boldsymbol{\Psi}}^{-1}(\boldsymbol{D}(L)\boldsymbol{x}_t - \boldsymbol{\Lambda}\widetilde{\boldsymbol{f}}_t),$$

where $\widetilde{\boldsymbol{F}} = (\widetilde{\boldsymbol{f}}_1 \dots \widetilde{\boldsymbol{f}}_T)'$ with $\widetilde{\boldsymbol{f}}_t = \boldsymbol{D}(L)\boldsymbol{f}_t$ and $\widetilde{\boldsymbol{\Psi}} = \mathrm{E}[\widetilde{\boldsymbol{e}}_t\widetilde{\boldsymbol{e}}_t']$ with $\widetilde{\boldsymbol{e}}_t = \boldsymbol{D}(L)\boldsymbol{e}_t$. Estimation of $\boldsymbol{D}(L)$ and $\widetilde{\boldsymbol{F}}$ can be done sequentially, iterating to convergence. Breitung and Tenhofen (2011) propose a similar two-step estimation procedure that allows for heteroskedastic and serially correlated idiosyncratic terms.

Even though PCA was first used to estimate approximate factor models where the factors and the idiosyncratic terms were *iid* white noise, most asymptotic results carried through to the case when the factors and the idiosyncratic terms were stationary autocorrelated processes. Consequently, the method has been widely used as a building block in approximate dynamic factor model estimation, but it required extensions since PCA on its own does not capture dynamics. Next we review several approaches that attempt to incorporate dynamic behavior in large scale approximate factor models.

2.3.4 Two-step and quasi-maximum likelihood estimation of large approximate factor models

Doz et al. (2011) proposed a two-step estimator that takes into account the dynamics of the factors. They assume that the factors in the approximate dynamic factor model $\boldsymbol{x}_t = \boldsymbol{\Lambda}\boldsymbol{f}_t + \boldsymbol{e}_t$, follow a vector autoregressive process, $\boldsymbol{f}_t = \boldsymbol{\Phi}_1\boldsymbol{f}_{t-1} + \cdots + \boldsymbol{\Phi}_p\boldsymbol{f}_{t-p} + \boldsymbol{u}_t$, and they allow the idiosyncratic terms to be autocorrelated but do not specify their dynamics. As illustrated in Section 2.3.1, this model can easily be cast in state-space form. The estimation procedure is the following:

Step 1 Preliminary estimators of the loadings, $\widehat{\boldsymbol{\Lambda}}$, and factors, $\widehat{\boldsymbol{f}}_t$, are obtained by principal component analysis. The idiosyncratic terms are estimated by $\widehat{e}_{it} =$

$x_{it} - \widehat{\boldsymbol{\lambda}}_i' \widehat{\boldsymbol{f}}_t$, and their variance is estimated by the associated empirical variance $\widehat{\psi}_{ii}$. The estimated factors, $\widehat{\boldsymbol{f}}_t$, are used in a vector-autoregressive model to obtain the estimates $\widehat{\boldsymbol{\Phi}}_j, j = 1 \dots p$.

Step 2 The model is cast in state-space form as in (2.7), with the variance of the common shocks set to the identity matrix, $\mathrm{cov}(\boldsymbol{u}_t) = \boldsymbol{I}_r$, and $\mathrm{cov}(\boldsymbol{e}_t)$ defined as $\boldsymbol{\Psi} = \mathrm{diag}(\psi_{11} \dots \psi_{nn})$. Using the parameter estimates $\widehat{\boldsymbol{\Lambda}}, \widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Phi}}_j, j = 1 \dots p$ obtained in the first step, one run of the Kalman smoother is then applied to the data. It produces a new estimate of the factor, $\widehat{\boldsymbol{f}}_{t|T} = \mathrm{E}(\boldsymbol{f}_t | \boldsymbol{x}_1, \dots, \boldsymbol{x}_T, \widehat{\boldsymbol{\theta}})$, where $\widehat{\boldsymbol{\theta}}$ is a vector containing the first step estimates of all parameters. It is important to notice that, in this second step, the idiosyncratic terms are misspecified, since they are taken as mutually orthogonal white noises.

Doz et al. (2011) prove that, under their assumptions, the parameter estimates, $\widehat{\boldsymbol{\theta}}$, are consistent, converging at rate $\left(\min(\sqrt{n}, \sqrt{T})\right)^{-1}$. They also prove that, when the Kalman smoother is run with those parameter estimates instead of the true parameters, the resulting two step-estimator of $\boldsymbol{f}_t$ is also $\min(\sqrt{n}, \sqrt{T})$ consistent.

Doz et al. (2012) proposed to estimate a large scale approximate dynamic factor model by quasi-maximum likelihood (QML). In line with Doz et al. (2011), the quasi-likelihood is based on the assumption of mutually orthogonal iid gaussian idiosyncratic terms (so that the model is treated as if it were an exact factor model, even though it is not), and a gaussian VAR model for the factors. The corresponding log-likelihood can be obtained from the Kalman filter for given values of the parameters, and they use an EM algorithm to compute the maximum likelihood estimator. The EM algorithm, proposed by Dempster et al. (1977) and first implemented for dynamic factor models by Watson and Engle (1983), alternates an expectation step relying on a pass of the Kalman smoother for the current parameter values and a maximization step relying on multivariate regressions. The application of the algorithm is tantamount to successive applications of the two-step approach. The calculations are feasible even when n is large, since in each iteration of the algorithm, the current estimate of the i^{th} loading, $\boldsymbol{\lambda}_i$, is obtained by ordinary least squares regression of x_{it} on the current estimate of the factors. The authors prove that, under a standard set of assumptions, the estimated factors are mean square consistent. Their results remain valid even if the processes are not gaussian, or if the idiosyncratic terms are not iid, or not mutually orthogonal, as long as they are only weakly cross-correlated. Reis and Watson (2010) apply this approach to a model with serially correlated idiosyncratic terms. Jungbacker and Koopman (2014) also use the EM algorithm to estimate a dynamic factor model with autocorrelated idiosyncratic terms, but instead of extending the state vector as in Equation (2.8), they transform the measurement equation as described in Section 2.3.5.

Bai and Li (2012) study QML estimation in the more restricted case where the quasi-likelihood is associated with the static exact factor model. They obtain consistency and asymptotic normality of the estimated loadings and factors under a set of appropriate assumptions. Bai and Li (2016) incorporate these estimators into the two-step approach put forward by Doz et al. (2011) and obtain similar asymptotic

results. They also follow Jungbacker and Koopman (2014) to handle the case where the idiosyncratic terms are autoregressive processes. Bai and Liao (2016) extend the approach of Bai and Li (2012) to the case where the idiosyncratic covariance matrix is sparse, instead of being diagonal, and propose a penalized maximum likelihood estimator.

2.3.5 Estimation of large approximate factor models with missing data

Observations may be missing from the analyzed data set for several reasons. At the beginning of the sample, certain time series might be available from an earlier start date than others. At the end of the sample, the dates of final observations may differ depending on the release lag of each data series. Finally, observations may be missing within the sample since different series in the data set may be sampled at different frequencies, for example monthly and quarterly. DFM estimation techniques assume that the observations are missing at random, so there is no endogenous sample selection. Missing data are handled differently in principal components and state space applications.

The least squares estimator of principal components in a balanced panel given in Equation (2.10) needs to be modified when some of the nT observations are missing. Stock and Watson (2002b) showed that estimates of $\boldsymbol{F}$ and $\boldsymbol{\Lambda}$ can be obtained numerically by

$$\min_{\boldsymbol{\Lambda},\boldsymbol{F}} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} S_{it} \left(x_{it} - \boldsymbol{\lambda}_i' \boldsymbol{f}_t\right)^2$$

where $S_{it} = 1$ if an observation on x_{it} is available and $S_{it} = 0$ otherwise. The objective function can be minimized by iterations alternating the optimization with respect to $\boldsymbol{\Lambda}$ given $\boldsymbol{F}$ and then $\boldsymbol{F}$ given $\boldsymbol{\Lambda}$; each step in the minimization has a closed form expression. Starting values can be obtained, for example, by principal component estimation using a subset of the series for which there are no missing observations. Alternatively, Stock and Watson (2002b) provide an EM algorithm for handling missing observations.

Step 1 Fill in initial guesses for missing values to obtain a balanced dataset. Estimate the factors and loadings in this balanced dataset by principal components analysis.

Step 2 The values in the place of a missing observations for each variable are updated by the expectation of x_{it} conditional on the observations, and the factors and loadings from the previous iteration.

Step 3 With the updated balanced dataset in hand, reestimate the factors and loadings by principal component analysis. Iterate step 2 and 3 until convergence.

The algorithm provides both, estimates of the factors and estimates of the missing values in the time series.

The state space framework has been adapted to missing data by either allowing the measurement equation to vary depending on what data are available at a given time (see Harvey, 1989, section 6.3.7) or by including a proxy value for the missing observation while adjusting the model so that the Kalman filter places no weight on the missing observation (see Giannone et al., 2008).

When the dataset contains missing values, the formulation in Equation (2.9) is not feasible since the lagged values on the right hand side of the measurement equation are not available in some periods. Jungbacker et al. (2011) addressed the problem by keeping track of periods with missing observations and augmenting the state vector with the idiosyncratic shocks in those periods. This implies that the system matrices and the dimension of the state vector are time-varying. Yet, the model can still be collapsed by transforming the measurement equation and partitioning the observation vector as described above and removing from the model the subvector that does not depend on the state. Under several simplifying assumptions, Pinheiro et al. (2013) developed an analytically and computationally less demanding algorithm for the special case of jagged edges, or observations missing at the end of the sample due to varying publication lags across series.

Since the Kalman filter and smoother can easily accommodate missing data, the two step method of Doz et al. (2011) is also well-suited to handle unbalanced panel datasets. In particular, it also allows to overcome the jagged edge data problem. This feature of the two-step method has been exploited in predicting low frequency macroeconomic releases for the current period, also known as nowcasting (see Section 2.6). For instance, Giannone et al. (2008) used this method to incorporate the real-time informational content of monthly macroeconomic data releases into current-quarter GDP forecasts. Similarly, Bańbura and Rünstler (2011) used the two-step framework to compute the impact of monthly predictors on quarterly GDP forecasts. They extended the method by first computing the weights associated with individual monthly observations in the estimates of the state vector using an algorithm by Koopman and Harvey (2003), which then allowed them to compute the contribution of each variable to the GDP forecast.

In the maximization step of the EM algorithm, the calculation of moments involving data was not feasible when some observations were missing, and therefore the original algorithm required a modification to handle an incomplete data set. Shumway and Stoffer (1982) allowed for missing data but assumed that the factor loading coefficients were known. More recently, Banbura and Modugno (2014) adapted the EM algorithm to a general pattern of missing data by using a selection matrix to carry out the maximization step with available data points. The basic idea behind their approach is to write the likelihood as if the data were complete and to adapt the Kalman filter and smoother to the pattern of missing data in the E-step of the EM algorithm, where the missing data are replaced by their best linear predictions given the information set. They also extend their approach to the case where the idiosyncratic terms are univariate AR processes. Finally they provide a statistical decomposition, which allows one to inspect how the arrival of new data affects the forecast of the variable of interest.

Modeling mixed frequencies via the state space approach makes it possible to associate the missing observations with particular dates and to differentiate between stock variables and flow variables. The state space model typically contains an aggregator that averages the high-frequency observations over one low-frequency period for stocks, and sums them for flows (see Aruoba et al., 2009). However, summing flows is only appropriate when the variables are in levels; for growth rates it is a mere approximation, weakening the forecasting performance of the model (see Fuleky and Bonham, 2015). As pointed out by Mariano and Murasawa (2003) and Proietti and Moauro (2006), appropriate aggregation of flow variables that enter the model in log-differences requires a non-linear state space model.

2.4 Estimation in the Frequency Domain

Classical principal component analysis, described in Section 2.3.2, estimates the space spanned by the factors non-parametrically only from the cross-sectional variation in the data. The two-step approach, discussed in Section 2.3.4, augments principal components estimation with a parametric state space model to capture the dynamics of the factors. Frequency-domain estimation combines some features of the previous two approaches: it relies on non-parametric methods that exploit variation both over time and over the cross-section of variables.

Traditional static principal component analysis focuses on contemporaneous cross-sectional correlation and overlooks serial dependence. It approximates the (contemporaneous) covariance matrix of $\boldsymbol{x}_t$ by a reduced rank covariance matrix. While the correlation of two processes may be negligible contemporaneously, it could be high at leads or lags. Discarding this information could result in loss of predictive capacity. Dynamic principal component analysis overcomes this shortcoming by relying on spectral densities. The $n \times n$ spectral density matrix of a second order stationary process, $\boldsymbol{x}_t$, for frequency $\omega \in [-\pi, \pi]$ is defined as

$$\boldsymbol{\Sigma}_{\boldsymbol{x}}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\omega} \boldsymbol{\Gamma}_{\boldsymbol{x}}(k),$$

where $\boldsymbol{\Gamma}_{\boldsymbol{x}}(k) = \mathrm{E}[\boldsymbol{x}_t \boldsymbol{x}'_{t-k}]$. In analogy to their static counterpart, dynamic principal components approximate the spectrum of $\boldsymbol{x}_t$ by a reduced rank spectral density matrix. Static principal component analysis was generalized to the frequency domain by Brillinger (1981). His algorithm relied on a consistent estimate of $\boldsymbol{x}_t$'s spectral density, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}(\omega)$, at frequency ω. The eigenvectors corresponding to the r largest eigenvalues of $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}(\omega)$, a Hermitian matrix, are then transformed by the inverse Fourier transform to obtain the dynamic principal components.

The method was popularized in econometrics by Forni et al. (2000). Their generalized dynamic factor model encompasses as a special case the approximate factor model of Chamberlain (1983) and Chamberlain and Rothschild (1983), which allows

for correlated idiosyncratic components but is static. And it generalizes the factor model of Sargent and Sims (1977) and Geweke (1977), which is dynamic but assumes orthogonal idiosyncratic components. The method relies on the assumption of an infinite cross-section to identify and consistently estimate the common and idiosyncratic components. The common component is a projection of the data on the space spanned by all leads and lags of the first r dynamic principal components, and the orthogonal residuals from this projection are the idiosyncratic components.

Forni et al. (2000) and Favero et al. (2005) propose the following procedure for estimating the dynamic principal components and common components.

Step 1 For a sample $\boldsymbol{x}_1 \dots \boldsymbol{x}_T$ of size T, estimate the spectral density matrix of $\boldsymbol{x}_t$ by

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}(\omega_h) = \frac{1}{2\pi} \sum_{k=-M}^{M} w_k e^{-ik\omega_h} \widehat{\boldsymbol{\Gamma}}_{\boldsymbol{x}}(k), \quad \omega_h = 2\pi h/(2M+1), \quad h = -M \dots M,$$

where $w_k = 1 - |k|/(M+1)$ are the weights of the Bartlett window of width M and $\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{x}}(k) = (T-k)^{-1} \sum_{t=k+1}^{T} (\boldsymbol{x}_t - \overline{\boldsymbol{x}})(\boldsymbol{x}_{t-k} - \overline{\boldsymbol{x}})'$ is the sample covariance matrix of $\boldsymbol{x}_t$ for lag k. For consistent estimation of $\boldsymbol{\Sigma}_{\boldsymbol{x}}(\omega)$ the window width has to be chosen such that $M \to \infty$ and $M/T \to 0$ as $T \to \infty$. Forni et al. (2000) note that a choice of $M = \frac{2}{3}T^{1/3}$ worked well in simulations.

Step 2 For $h = -M \dots M$, compute the eigenvectors $\boldsymbol{\lambda}_1(\omega_h) \dots \boldsymbol{\lambda}_r(\omega_h)$ corresponding to the r largest eigenvalues of $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}(\omega_h)$. The choice of r is guided by a heuristic inspection of eigenvalues. Note that, for $M = 0$, $\boldsymbol{\lambda}_j(\omega_h)$ is simply the j^{th} eigenvector of the (estimated) variance-covariance matrix of $\boldsymbol{x}_t$: the dynamic principal components then reduce to the static principal components.

Step 3 Define $\boldsymbol{\lambda}_j(L)$ as the two-sided filter

$$\boldsymbol{\lambda}_j(L) = \sum_{k=-M}^{M} \boldsymbol{\lambda}_{jk} L^k, \quad k = -M \dots M,$$

where

$$\boldsymbol{\lambda}_{jk} = \frac{1}{2M+1} \sum_{h=-M}^{M} \boldsymbol{\lambda}_j(\omega_h) e^{-ik\omega_h},$$

The first r dynamic principal components of $\boldsymbol{x}_t$ are $\widehat{\boldsymbol{f}}_{jt} = \boldsymbol{\lambda}_j(L)'\boldsymbol{x}_t, j = 1 \dots r$, which can be collected in the vector $\widehat{\boldsymbol{f}}_t = (\widehat{\boldsymbol{f}}_{1t} \dots \widehat{\boldsymbol{f}}_{rt})'$.

Step 4 Run an OLS regression of $\boldsymbol{x}_t$ on present, past and future dynamic principal components

$$\boldsymbol{x}_t = \boldsymbol{\Lambda}_{-q}\widehat{\boldsymbol{f}}_{t+q} + \dots + \boldsymbol{\Lambda}_p \widehat{\boldsymbol{f}}_{t-p},$$

and estimate the common component as the fitted value

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

where $\widehat{\mathbf{\Lambda}}_l, l = -q \dots p$ are the OLS estimators, and the leads q and lags p used in the regression can be chosen by model selection criteria. The idiosyncratic component is the residual, $\widehat{\xi}_{it} = x_{it} - \widehat{\chi}_{it}$.

Although this method efficiently estimates the common component, its reliance on two-sided filtering rendered it unsuitable for forecasting. Forni et al. (2005) extended the frequency domain approach developed in their earlier paper to forecasting in two steps.

Step 1 In the first step, they estimated the covariances of the common and idiosyncratic components using the inverse Fourier transforms

$$\widehat{\mathbf{\Gamma}}_{\chi}(k) = \frac{1}{2M+1} \sum_{k=-M}^{M} e^{ik\omega_h} \widehat{\mathbf{\Sigma}}_{\chi}(\omega_h) \quad \text{and}$$

$$\widehat{\mathbf{\Gamma}}_{\xi}(k) = \frac{1}{2M+1} \sum_{k=-M}^{M} e^{ik\omega_h} \widehat{\mathbf{\Sigma}}_{\xi}(\omega_h)$$

where $\widehat{\mathbf{\Sigma}}_{\chi}(\omega_h)$ and $\widehat{\mathbf{\Sigma}}_{\xi}(\omega_h) = \widehat{\mathbf{\Sigma}}_{\boldsymbol{x}}(\omega_h) - \widehat{\mathbf{\Sigma}}_{\chi}(\omega_h)$ are the spectral density matrices corresponding to the common and idiosyncratic components, respectively.

Step 2 In the second step, they estimated the factor space using a linear combination of the $\boldsymbol{x}$'s, $\boldsymbol{\lambda}'\boldsymbol{x}_t = \boldsymbol{\lambda}'\boldsymbol{\chi}_t + \boldsymbol{\lambda}'\boldsymbol{\xi}_t$. Specifically, they compute r independent linear combinations $\widehat{\boldsymbol{f}}_{jt} = \widehat{\boldsymbol{\lambda}}_j'\boldsymbol{x}_t$, where the weights $\widehat{\boldsymbol{\lambda}}_j$ maximize the variance of $\boldsymbol{\lambda}'\boldsymbol{\chi}_t$ and are defined recursively

$$\widehat{\boldsymbol{\lambda}}_j = \arg\max_{\boldsymbol{\lambda}\in\mathbb{R}^n} \boldsymbol{\lambda}'\widehat{\mathbf{\Gamma}}_{\chi}(0)\boldsymbol{\lambda} \quad \text{s.t.} \quad \boldsymbol{\lambda}'\widehat{\mathbf{\Gamma}}_{\xi}(0)\boldsymbol{\lambda} = 1 \quad \text{and} \quad \boldsymbol{\lambda}'\widehat{\mathbf{\Gamma}}_{\xi}(0)\widehat{\boldsymbol{\lambda}}_m = 0,$$

for $j = 1 \dots r$ and $1 \le m \le j-1$ (only the first constraint applies for $j = 1$). The solutions of this problem, $\widehat{\boldsymbol{\lambda}}_j$, are the generalized eigenvectors associated with the generalized eigenvalues $\widehat{\nu}_j$ of the matrices, $\widehat{\mathbf{\Gamma}}_{\chi}(0)$ and $\widehat{\mathbf{\Gamma}}_{\xi}(0)$,

$$\widehat{\boldsymbol{\lambda}}_j'\widehat{\mathbf{\Gamma}}_{\chi}(0) = \widehat{\nu}_j\widehat{\boldsymbol{\lambda}}_j'\widehat{\mathbf{\Gamma}}_{\xi}(0), \quad j = 1 \dots n,$$

with the normalization constraint $\widehat{\boldsymbol{\lambda}}_j'\widehat{\mathbf{\Gamma}}_{\xi}(0)\widehat{\boldsymbol{\lambda}}_j = 1$ and $\widehat{\boldsymbol{\lambda}}_i'\widehat{\mathbf{\Gamma}}_{\xi}(0)\widehat{\boldsymbol{\lambda}}_j = 0$ for $i \neq j$. The linear combinations $\widehat{\boldsymbol{f}}_{jt} = \widehat{\boldsymbol{\lambda}}_j'\boldsymbol{x}_t, j = 1 \dots n$ are the generalized principal components of $\boldsymbol{x}_t$ relative to the couple $(\widehat{\mathbf{\Gamma}}_{\chi}(0), \widehat{\mathbf{\Gamma}}_{\xi}(0))$. Defining $\widehat{\mathbf{\Lambda}} = (\widehat{\boldsymbol{\lambda}}_1 \dots \widehat{\boldsymbol{\lambda}}_r)$, the space spanned by the common factors is estimated by the first r generalized principal components of the $\boldsymbol{x}$'s: $\widehat{\mathbf{\Lambda}}'\boldsymbol{x}_t = \widehat{\boldsymbol{\lambda}}_1'\boldsymbol{x}_t \dots \widehat{\boldsymbol{\lambda}}_r'\boldsymbol{x}_t$. The forecast of the common component depends on the covariance between $\boldsymbol{\chi}_{iT+h}$ and $\widehat{\mathbf{\Lambda}}'\boldsymbol{x}_T$. Observing that this covariance equals the covariance between $\boldsymbol{\chi}_{T+h}$ and $\widehat{\mathbf{\Lambda}}'\boldsymbol{\chi}_T$, the forecast of the common component can be obtained from the projection

$$\widehat{\boldsymbol{\chi}}_{T+h|T} = \widehat{\mathbf{\Gamma}}_{\chi}(h)\widehat{\mathbf{\Lambda}}(\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Gamma}}_{\chi}(0)\widehat{\mathbf{\Lambda}})^{-1}\widehat{\mathbf{\Lambda}}'\boldsymbol{x}_T.$$

Since this two-step forecasting method relied on lags but not leads, it avoided the end-of-sample problems caused by two-sided filtering in the authors' earlier study. A simulation study by the authors suggested that this procedure improves upon the forecasting performance of Stock and Watson's (2002) static principal components method for data generating processes with heterogeneous dynamics and heterogeneous variance ratio of the common and idiosyncratic components. In line with most of the earlier literature, Forni et al. (2009) continued to assume that the space spanned by the common components at any time t has a finite-dimension r as n tends to infinity, allowing a static representation of the model. By identifying and estimating cross-sectionally pervasive shocks and their dynamic effect on macroeconomic variables, they showed that dynamic factor models are suitable for structural modeling.

The finite-dimension assumption for the common component rules out certain factor-loading patterns. In the model $x_{it} = a_i(1 - b_i L)^{-1} u_t + \xi_{it}$, where u_t is a scalar white noise and the coefficients b_i are drawn from a uniform distribution (-0.9, 0.9), the space spanned by the common components χ_{it}, $i \in \mathbb{N}$ is infinite dimensional. Forni and Lippi (2011) relaxed the finite-dimensional assumption and proposed a one-sided estimator for the general dynamic factor model of Forni et al. (2000). Forni et al. (2015, 2017) continued to allow the common components—driven by a finite number of common shocks—to span an infinite-dimensional space and investigated the model's one-sided representations and asymptotic properties. Forni et al. (2018) evaluated the model's pseudo real-time forecasting performance for US macroeconomic variables and found that it compares favorably to finite dimensional methods during the Great Moderation. The dynamic relationship between the variables and the factors in this model is more general than in models assuming a finite common component space, but, as pointed out by the authors, its estimation is rather complex.

Hallin and Lippi (2013) give a general presentation of the methodological foundations of dynamic factor models. Fiorentini et al. (2018) introduced a frequency domain version of the EM algorithm for dynamic factor models with latent autoregressive and moving average processes. In this paper the authors focused on an exact factor model with a single common factor, and left approximate factor models with multiple factors for future research. But they extended the basic EM algorithm with an iterated indirect inference procedure based on a sequence of simple auxiliary OLS regressions to speed up computation for models with moving average components.

Although carried out in the time domain, Peña and Yohai's (2016) generalized dynamic principal components model mimics two important features of Forni et al.'s (2000) generalized dynamic factor model: it allows for both a dynamic representation of the common component and nonorthogonal idiosyncratic components. Their procedure chooses the number of common components to achieve a desired degree of accuracy in a mean squared error sense in the reconstruction of the original series. The estimation iterates two steps: the first is a least squares estimator of the loading coefficients assuming the factors are known, and the second is updating the factor estimate based on the estimated coefficients. Since the authors do not place restrictions on the principal components, their method can be applied to potentially nonstationary time series data. Although the proposed method does well for data

reconstruction, it is not suited for forecasting, because—as in Forni et al. (2000)—it uses both leads and lags to reconstruct the series.

2.5 Estimating the Number of Factors

Full specification of the DFM requires selecting the number of common factors. The number of static factors can be determined by information criteria that use penalized objective functions or by analyzing the distribution of eigenvalues. Bai and Ng (2002) used information criteria of the form $IC(r) = \ln V_r(\widehat{\boldsymbol{\Lambda}}, \widehat{\boldsymbol{F}}) + rg(n,T)$, where $V_r(\widehat{\boldsymbol{\Lambda}}, \widehat{\boldsymbol{F}})$ is the least squares objective function (2.10) evaluated with the principal components estimators when r factors are considered, and $g(n,T)$ is a penalty function that satisfies two conditions: $g(n,T) \to 0$ and $\min[n^{1/2}, T^{1/2}] \cdot g(n,T) \to \infty$, as $n,T \to \infty$. The estimator for the number of factors is $\widehat{r}_{IC} = \min_{0 \le r \le rmax} IC(r)$, where $rmax$ is the upper bound of the true number of factors. The authors show that the estimator is consistent without restrictions between n and T, and the results hold under heteroskedasticity in both the time and cross-sectional dimension, as well as under weak serial and cross-sectional correlation. Li et al. (2017) develop a method to estimate the number of factors when the number of factors is allowed to increase as the cross-section and time dimensions increase. This is useful since new factors may emerge as changes in the economic environment trigger structural breaks.

Ahn and Horenstein (2013) and Onatski (2009, 2010) take a different approach by comparing adjacent eigenvalues of the spectral density matrix at a given frequency or of the covariance matrix of the data. The basic idea behind this approach is that the first r eigenvalues will be unbounded, while the remaining values will be bounded. Therefore the ratio of subsequent eigenvalues is maximized at the location of the largest relative cliff in a scree plot (a plot of the ordered eigenvalues against the rank of those eigenvalues). These authors also present alternative statistics using the difference, the ratio of changes, and the growth rates of subsequent eigenvalues.

Estimation of the number of dynamic factors usually requires several steps. As illustrated in Section 2.2.1, the number of dynamic factors, q, will in general be lower than the number of static factors $r = q(1+s)$, and therefore the spectrum of the common component will have a reduced rank with only q nonzero eigenvalues. Based on this result, Hallin and Liska (2007) propose a frequency-domain procedure which uses an information criterion to estimate the rank of the spectral density matrix of the data. Bai and Ng (2007) take a different approach by first estimating the number of static factors and then applying a VAR(p) model to the estimated factors to obtain the residuals. They use the eigenvalues of the residual covariance matrix to estimate the rank q of the covariance of the dynamic (or primitive) shocks. In a similar spirit, Amengual and Watson (2007) first project the observed variables $\boldsymbol{x}_t$ onto p lags of consistently estimated r static principal components $\boldsymbol{f}_t \ldots \boldsymbol{f}_{t-p}$ to obtain serially uncorrelated residuals $\widehat{\boldsymbol{u}}_t = \boldsymbol{x}_t - \sum_{i=1}^{p} \widehat{\boldsymbol{\Pi}}_i \widehat{\boldsymbol{f}}_{t-i}$. These residuals then have a static factor representation with q factors. Applying the Bai and Ng (2002) information

criterion to the sample variance matrix of these residuals yields a consistent estimate of the number of dynamic factors, q.

2.6 Forecasting with Large Dynamic Factor Models

One of the most important uses of dynamic factor models is forecasting. Both small scale (*i.e.* n small) and large scale dynamic factor models have been used to this end. Very often, the forecasted variable is published with a delay: for instance, euro-area GDP is published six weeks after the end of the corresponding quarter. The long delay in data release implies that different types of forecasts can be considered. GDP predictions for quarter Q, made prior to that quarter, in $Q-1, Q-2, \ldots$ for instance, are considered to be "true" out of sample forecasts. But estimates for quarter Q can also be made during that same quarter using high frequency data released within quarter Q. These are called "nowcasts". Finally, since GDP for quarter Q is not known until a few weeks into quarter $Q+1$, forecasters keep estimating it during this time interval. These estimates are considered to be "backcasts". As we have seen, dynamic factor models are very flexible and can easily handle all these predictions.

Forecasts using large dimensional factor models were first introduced in the literature by Stock and Watson (2002a). The method, also denoted diffusion index forecasts, consists of estimating the factors $\boldsymbol{f}_t$ by principal component analysis, and then using those estimates in a regression estimated by ordinary least squares

$$y_{t+h} = \boldsymbol{\beta}_f' \widehat{\boldsymbol{f}}_t + \boldsymbol{\beta}_w' \boldsymbol{w}_t + \varepsilon_{t+h},$$

where y_t is the variable of interest, $\widehat{f}_t$ is the vector of the estimated factors, and $\boldsymbol{w}_t$ is a vector of observable predictors (typically lags of y_t). The direct forecast for time $T+h$ is then computed as

$$y_{T+h|T} = \widehat{\boldsymbol{\beta}}_f' \widehat{\boldsymbol{f}}_T + \widehat{\boldsymbol{\beta}}_w' \boldsymbol{w}_T.$$

The authors prove that, under the assumptions they used to ensure the consistency of the principal component estimates,

$$\operatorname{plim}_{n\to\infty} \left[(\widehat{\boldsymbol{\beta}}_f' \widehat{\boldsymbol{f}}_T + \widehat{\boldsymbol{\beta}}_w' \boldsymbol{w}_T) - (\boldsymbol{\beta}_f' \boldsymbol{f}_T + \boldsymbol{\beta}_w' \boldsymbol{w}_T) \right] = 0,$$

so that the forecast is asymptotically equivalent to what it would have been if the factors had been observed. Furthermore, under a stronger set of assumptions Bai and Ng (2006) show that the forecast error is asymptotically gaussian, with known variance, so that forecast intervals can be computed. Stock and Watson (2002b) considered a more general model to capture the dynamic relationship between the variables and the factors

$$y_{t+h} = \alpha_h + \boldsymbol{\beta}_h(L) \widehat{\boldsymbol{f}}_t + \gamma_h(L) y_t + \varepsilon_{t+h},$$

with forecast equation

$$\widehat{y}_{T+h|T} = \widehat{\alpha}_h + \widehat{\boldsymbol{\beta}}_h(L)\widehat{\boldsymbol{f}}_T + \widehat{\gamma}_h(L)y_T. \tag{2.11}$$

They find that this model produces better forecasts for some variables and forecast horizons, but the improvements are not systematic.

While the diffusion index methodology relied on a distributed lag model, the approach taken by Doz et al. (2011) and Doz et al. (2012) captured factor dynamics explicitly. The estimates of a vector-autoregressive model for $\widehat{\boldsymbol{f}}_t$ can be used to recursively forecast $\widehat{\boldsymbol{f}}_{T+h|T}$ at time T. A mean square consistent forecast for period $T+h$ can then be obtained by $y_{T+h|T} = \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{f}}_{T+h|T}$. This approach has been used by Giannone et al. (2008) and by many others. However Stock and Watson (2002b) point out that the diffusion index and two step approaches are equivalent since the recursive forecast of the factor in the latter implies that $\widehat{\boldsymbol{f}}_{T+h|T}$ is a function of $\widehat{\boldsymbol{f}}_T$ and its lags, as in (2.11).

2.6.1 Targeting predictors and other forecasting refinements

In the diffusion index and two step forecasting methodology, the factors are first estimated from a large number of predictors, $(x_{1t} \dots x_{nt})$, by the method of principal components, and then used in a linear forecasting equation for y_{t+h}. Although the method can parsimoniously summarize information from a large number of predictors and incorporate it into the forecast, the estimated factors do not take into account the predictive power of x_{it} for y_{t+h}. Boivin and Ng (2006) pointed out that expanding the sample size simply by adding data without regard to its quality or usefulness does not necessarily improve forecasts. Bai and Ng (2008a) suggested to target predictors based on their information content about y. They used hard and soft thresholding to determine which variables the factors are to be extracted from and thereby reduce the influence of uninformative predictors. Under hard thresholding, a pretest procedure is used to decide whether a predictor should be kept or not. Under soft thresholding, the predictor ordering and selection is carried out using the least angle regression (LARS) algorithm developed by Efron et al. (2004).

Kelly and Pruitt (2015) proposed a three-pass regression filter with the ability to identify the subset of factors useful for forecasting a given target variable while discarding those that are target irrelevant but may be pervasive among predictors. The proposed procedure uses the covariances between the variables in the dataset and the proxies for the relevant latent factors, the proxies being observable variables either theoretically motivated or automatically generated.

Pass 1 The first pass captures the relationship between the predictors, $\boldsymbol{X}$, and $m \ll \min(n,T)$ factor proxies, $\boldsymbol{Z}$, by running a separate time series regression of each predictor, $\boldsymbol{x}_i$, on the proxies, $x_{it} = \alpha_i + \boldsymbol{z}_t'\boldsymbol{\beta}_i + \varepsilon_{it}$, for $i = 1 \dots n$, and retaining $\widehat{\boldsymbol{\beta}}_i$.

Pass 2 The second pass consists of T separate cross section regressions of the predictors, $\boldsymbol{x}_t$, on the coefficients estimated in the first pass, $x_{it} = \alpha_t + \widehat{\boldsymbol{\beta}}_i' \boldsymbol{f}_t + \varepsilon_{it}$, for $t = 1 \dots T$, and retaining $\widehat{\boldsymbol{f}}_t$. First-stage coefficient estimates map the cross-sectional distribution of predictors to the latent factors. Second-stage cross section regressions use this map to back out estimates of the factors, $\widehat{\boldsymbol{f}}_t$, at each point in time.

Pass 3 The third pass is a single time series forecasting regression of the target variable y_{t+1} on the predictive factors estimated in the second pass, $y_{t+1} = \alpha + \widehat{\boldsymbol{f}}_t' \boldsymbol{\beta} + \varepsilon_{t+1}$. The third-pass fitted value, $\widehat{y}_{t+1}$, is the forecast for period $t + 1$.

The automatic proxy selection algorithm is initialized with the target variable itself $\boldsymbol{z}_1 = \boldsymbol{y}$, and additional proxies based on the prediction error are added iteratively, $\boldsymbol{z}_{k+1} = \boldsymbol{y} - \widehat{\boldsymbol{y}}_k$ for $k = 1 \dots m - 1$, where k is the number of proxies in the model at the given iteration. The authors point out that partial least squares, futher analyzed by Groen and Kapetanios (2016), is a special case of the three-pass regression filter.

Bräuning and Koopman (2014) proposed a collapsed dynamic factor model where the factor estimates are established jointly by the predictors $\boldsymbol{x}_t$ and the target variable y_t. The procedure is a two-step process.

Step 1 The first step uses principal component analysis to reduce the dimension of a large panel of macroeconomic predictors as in Stock and Watson (2002a,b).

Step 2 In the second step, the authors use a state space model with a small number of parameters to model the principal components jointly with the target variable y_t. The principal components, $\boldsymbol{f}_{PC,t}$, are treated as dependent variables that are associated exclusively with the factors $\boldsymbol{f}_t$, but the factors $\boldsymbol{f}_t$ enter the equation for the target variable y_t. The unknown parameters are estimated by maximum likelihood, and the Kalman filter is used for signal extraction.

In contrast to the two-step method of Doz et al. (2011), this approach allows for a specific dynamic model for the target variable that may already produce good forecasts for y_t.

Several additional methods originating in the machine learning literature have been used to improve forecasting performance of dynamic factor models, including bagging (Inoue and Kilian, 2008) and boosting (Bai and Ng, 2009), which will be discussed in Chapters 14 and 16, respectively.

2.7 Hierarchical Dynamic Factor Models

If a panel of data can be organized into blocks using a priori information, then between- and within-block variation in the data can be captured by the hierarchical dynamic factor model framework formalized by Moench et al. (2013). The block structure helps to model covariations that are not sufficiently pervasive to be treated as common factors. For example, in a three-level model, the series i, in a given block b, at each time t can exhibit idiosyncratic, block specific, and common variation,

$$
\begin{aligned}
x_{ibt} &= \boldsymbol{\lambda}_{\boldsymbol{g}ib}(L)\boldsymbol{g}_{bt} + e_{xibt} \\
\boldsymbol{g}_{bt} &= \boldsymbol{\Lambda}_{\boldsymbol{f}b}(L)\boldsymbol{f}_t + \boldsymbol{e}_{\boldsymbol{g}bt} \\
\boldsymbol{\Phi}_{\boldsymbol{f}}(L)\boldsymbol{f}_t &= \boldsymbol{u}_t \ ,
\end{aligned}
$$

where variables x_{ibt} and x_{jbt} within a block b are correlated because of the common factors $\boldsymbol{f}_t$ or the block-specific shocks e_{gbt}, but correlations between blocks are possible only through $\boldsymbol{f}_t$. Some of the x_{it} may not belong to a block and could be affected by the common factors directly, as in the two-level model (2.5). The idiosyncratic components can be allowed to follow stationary autoregressive processes, $\phi_{xib}(L)e_{xibt} = \boldsymbol{\varepsilon}_{xibt}$ and $\boldsymbol{\Phi}_{\boldsymbol{g}b}(L)\boldsymbol{e}_{\boldsymbol{g}bt} = \boldsymbol{\varepsilon}_{\boldsymbol{g}bt}$.

If we were given data for production, employment, and consumption, then x_{i1t} could be one of the n_1 production series, x_{i2t} could be one of the n_2 employment series, and x_{i3t} could be one of the n_3 consumption series. The correlation between the production, employment, and consumption factors, $\boldsymbol{g}_{1t}, \boldsymbol{g}_{2t}, \boldsymbol{g}_{3t}$, due to economy-wide fluctuations, would be captured by $\boldsymbol{f}_t$. In a multicountry setting, a four-level hierarchical model could account for series-specific, country (subblock), region (block), and global (common) fluctuations, as in Kose et al. (2003). If the country and regional variations were not properly modeled, they would appear as either weak common factors or idiosyncratic errors that would be cross-correlated among series in the same region. Instead of assuming weak cross-sectional correlation as in approximate factor models, the hierarchical model explicitly specifies the block structure, which helps with the interpretation of the factors.

To estimate the model, Moench et al. (2013) extend the Markov Chain Monte Carlo method that Otrok and Whiteman (1998) originally applied to a single factor model. Let $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_{\boldsymbol{g}}, \boldsymbol{\Lambda}_{\boldsymbol{f}})$, $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_{\boldsymbol{f}}, \boldsymbol{\Phi}_{\boldsymbol{g}}, \boldsymbol{\Phi}_{\boldsymbol{x}})$, and $\boldsymbol{\Psi} = (\boldsymbol{\Psi}_{\boldsymbol{f}}, \boldsymbol{\Psi}_{\boldsymbol{g}}, \boldsymbol{\Psi}_{\boldsymbol{x}})$ denote the matrices containing the loadings, coefficients of the lag polynomials, and variances of the innovations, respectively. Organize the data into blocks, $\boldsymbol{x}_{bt}$, and get initial values for $\boldsymbol{g}_t$ and $\boldsymbol{f}_t$ using principal components; use these to produce initial values for $\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}$.

Step 1 Conditional on $\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}$, $\boldsymbol{f}_t$, and the data $\boldsymbol{x}_{bt}$, draw $\boldsymbol{g}_{bt}$ for all b.
Step 2 Conditional on $\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}$, and $\boldsymbol{g}_{bt}$, draw $\boldsymbol{f}_t$.
Step 3 Conditional on $\boldsymbol{f}_t$ and $\boldsymbol{g}_{bt}$, draw $\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}$, and return to **Step 1**.

The sampling of $\boldsymbol{g}_{bt}$ needs to take into account the correlation across blocks due to $\boldsymbol{f}_t$. As in previously discussed dynamic factor models, the factors and the loadings are not separately identified. To achieve identification, the authors suggest using lower triangular loading matrices, fixed variances of the innovations to the factors, $\boldsymbol{\Psi}_{\boldsymbol{f}}, \boldsymbol{\Psi}_{\boldsymbol{g}}$, and imposing additional restrictions on the structure of the lag polynomials. Jackson et al. (2016) survey additional Bayesian estimation methods. Under the assumption that common factors have a direct impact on x, but are uncorrelated with block-specific ones, so that $x_{ibt} = \boldsymbol{\lambda}_{\boldsymbol{f}ib}(L)\boldsymbol{f}_t + \boldsymbol{\lambda}_{\boldsymbol{g}ib}(L)\boldsymbol{g}_{bt} + e_{xibt}$, Breitung and Eickmeier (2016) propose a sequential least squares estimation approach and compare it to other frequentist estimation methods.

Kose et al. (2008) use a hierarchical model to study international business cycle comovements by decomposing fluctuations in macroeconomic aggregates of G-7

countries into a common factor across all countries, country factors that are common across variables within a country, and idiosyncratic fluctuations. Moench and Ng (2011) apply this model to national and regional housing data to estimate the effects of housing shocks on consumption, while Fu (2007) decomposes house prices in 62 U.S. metropolitan areas into national, regional and metro-specific idiosyncratic factors. Del Negro and Otrok (2008) and Stock and Watson (2008) use this approach to add stochastic volatility to their models.

2.8 Structural Breaks in Dynamic Factor Models

As evident from the discussion so far, the literature on dynamic factor models has grown tremendously, evolving in many directions. In the remainder of this chapter we will concentrate on two strands of research: dynamic factor models 1) with Markov-switching behavior and 2) with time-varying loadings. In both cases, the aim is to take into account the evolution of macroeconomic conditions over time, either through 1) non-linearities in the dynamics of the factors or 2) the variation of loadings, which measure the intensity of the links between each observable variable and the underlying common factors. This instability seemed indeed particularly important to address after the 2008 global financial crisis and the subsequent slow recovery. These two strands of literature have presented a number of interesting papers in recent years. In what follows, we briefly describe some of them, but we do not provide an exhaustive description of the corresponding literature.

2.8.1 Markov-switching dynamic factor models

One of the first uses of dynamic factor models was the construction of coincident indexes. The literature soon sought to allow the dynamics of the index to vary according to the phases of the business cycle. Incorporating Markov switching into dynamic factor models (MS-DFM) was first suggested by Kim (1994) and Diebold and Rudebusch (1996)[3]. Diebold and Rudebusch (1996) considered a single factor, that played the role of a coincident composite index capturing the latent state of the economy. They suggested that the parameters describing the dynamics of the factor may themselves depend on an unobservable two-state Markov-switching latent variable. More precisely, they modeled the factor the same way as Hamilton (1989) modeled US real GNP to obtain a statistical characterization of business cycle phases. In practice, Diebold and Rudebusch (1996) used a two-step estimation method since they applied Hamilton's model to a previously estimated composite index.

Kim (1994) introduced a very general model in which the dynamics of the factors and the loadings may depend on a state-variable. Using the notation used so far in

[3] The working paper version appeared in 1994 in NBER Working Papers 4643.

the current chapter, his model was

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}_{S_t} \boldsymbol{f}_t + \boldsymbol{B}_{S_t} \boldsymbol{x}_t + \boldsymbol{e}_t$$
$$\boldsymbol{f}_t = \boldsymbol{\Phi}_{S_t} \boldsymbol{f}_{t-1} + \boldsymbol{\Gamma}_{S_t} \boldsymbol{x}_t + \boldsymbol{H}_{S_t} \boldsymbol{\varepsilon}_t$$

where $\boldsymbol{x}_t$ is a vector of exogenous or lagged dependent variables, and where the $\begin{pmatrix} \boldsymbol{e}_t \\ \boldsymbol{\varepsilon}_t \end{pmatrix}$'s are i.i.d. gaussian with covariance matrix $\begin{pmatrix} \boldsymbol{R} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{Q} \end{pmatrix}$. The underlying state variable S_t can take M possible values, and is Markovian of order one, so that $P(S_t = j|S_{t-1} = i, S_{t-2} = k, \ldots) = P(S_t = j|S_{t-1} = i) = p_{ij}$. He proposed a very powerful approximation in the computation of the Kalman filter and the likelihood, and estimated the model in one step by maximum likelihood[4]. Using this approximation, the likelihood can be computed at any point of the parameter space, and maximized using a numerical procedure. It must be however emphasized that such a procedure is applicable in practice only when the number of parameters is small, which means that the dimension of $\boldsymbol{x}_t$ must be small. Once the parameters have been estimated, it is possible to obtain the best approximations of $\boldsymbol{f}_t$ and S_t for any t using the Kalman filter and smoother, given the observations $\boldsymbol{x}_1 \ldots \boldsymbol{x}_t$ and $\boldsymbol{x}_1 \ldots \boldsymbol{x}_T$, respectively.

Kim (1994) and Chauvet (1998) estimated a one factor Markov switching model using this methodology. In both papers, the model is formulated like a classical dynamic factor model, with the factor following an autoregressive process whose constant term depends on a two-state Markov variable S_t:

$$\boldsymbol{x}_t = \boldsymbol{\lambda} f_t + \boldsymbol{e}_t \text{ and } \phi(L) f_t = \beta_{S_t} + \eta_t$$

where $\eta_t \sim$ i.i.d. $N(0,1)$ and S_t can take two values denoted as 0 and 1, which basically correspond to expansions and recessions: the conditional expectation of the factor is higher during expansions than during recessions. Kim and Yoo (1995) used the same four observable variables as the US Department of Commerce and Stock and Watson (1989, 1991, 1993), whereas Chauvet (1998) considered several sets of observable series over various time periods, taking into account different specifications for the dynamics of the common factor. Both papers obtained posterior recession probabilities and turning points that were very close to official NBER dates. Kim and Nelson (1998) proposed a Gibbs sampling methodology that helped to avoid Kim's (1994) approximation of the likelihood. This Bayesian approach also provided results that were very close to the NBER dates, and allowed tests of business cycle duration dependence. Chauvet and Piger (2008) compared the nonparametric dating algorithm given in Harding and Pagan (2003) with the dating obtained using Markov-switching models similar to those of Chauvet (1998). They showed that both approaches identify the NBER turning point dates in real time with reasonable accuracy, and identify the troughs with more timeliness than the NBER. But they

[4] Without this approximation, the Kalman filter would be untractable, since it would be necessary to take the M^T possible trajectories of $S_1, \ldots, S_T$. For further details, see Kim (1994) and the references therein.

found evidence in favor of MS-DFMs, which identified NBER turning point dates more accurately overall. Chauvet and Senyuz (2016) used a two-factor MS-DFM to represent four series; three related to the yield curve and the fourth, industrial production, representing economic activity. Their model allowed to analyze the lead-lag relationship between the cyclical phases of the two sectors.

Camacho et al. (2014) extended the Kim and Yoo (1995) approach to deal with mixed frequency and/or ragged-edge data. They ran simulations and applied their methodology to real time observations of the four variables used by the US Department of Commerce. They found evidence that taking into account all the available information in this framework yields substantial improvements in the estimated real time probabilities. Camacho et al. (2015) used the same approach and applied the method to a set of thirteen euro-area series. They obtained a non-linear indicator of the overall economic activity and showed that the associated business cycle dating is very close to the CEPR Committee's dating.

The one-step methods used in all these papers (most of them following Kim's (1994) approximation of the likelihood, others relying on Kim and Nelson's (1998) Gibbs sampling) have been successful in estimating MS-DFMs of very small dimensions. In order to estimate MS-DFM of larger dimensions, it is possible to take advantage of the two-step approach originally applied to a small number of variables by Diebold and Rudebusch (1996). Indeed, it is possible to use a two-step approach similar to the one of Doz et al. (2011) in a standard DFM framework: in the first step, a linear DFM is estimated by principal components, in the second step, a Markov-switching model, as in Hamilton (1989), is specified for the estimated factor(s) and is estimated by maximum likelihood. Camacho et al. (2015) compared this two-step approach to a one-step approach applied to a small dataset of coincident indicators. They concluded that the one-step approach was better at turning point detection when the small dataset contained good quality business cycle indicators, and they also observed a decreasing marginal gain in accuracy when the number of indicators increased. However other authors have obtained satisfying results with the two-step approach. Bessec and Bouabdallah (2015) applied MS-factor MIDAS models [5] to a large dataset of mixed frequency variables. They ran Monte Carlo simulations and applied their model to US data containing a large number of financial series and real GDP growth: in both cases the model properly detected recessions. Doz and Petronevich (2016) also used the two-step approach: using French data, they compared business cycle dates obtained from a one factor MS-DFM estimated on a small dataset with Kim's (1994) method, to the dates obtained using a one factor MS-DFM estimated in two steps from a large database. The two-step approach successfully predicted the turning point dates released by the OECD. As a complement, Doz and Petronevich (2017) conducted a Monte-Carlo experiment, which provided evidence that the two-step method is asymptotically valid for large N and T and provides good turning points prediction. Thus the relative performances of the one-step and two-step methods under the MS-DFM framework are worth exploring further, both from a turning point detection perspective and a forecasting perspective.

[5] Their model combines the MS-MIDAS model (Guérin and Marcellino, 2013) and the factor-MIDAS model (Marcellino and Schumacher, 2010).

2.8.2 Time varying loadings

Another strand of the literature deals with time-varying loadings. The assumption that the links between the economic variables under study and the underlying factors remain stable over long periods of time may be seen as too restrictive. If the common factors are driven by a small number of structural shocks, the observable variables may react to those structural shocks in a time varying fashion. Structural changes in the economy may also lead to changes in the comovements of variables, and in turn require adjustment in the underlying factor model. Such shifts have become particularly relevant after the 2008 financial crisis and the ensuing slow recovery. But the literature had dealt with similar questions even earlier, for instance during the Great Moderation. The issue is important since assuming constant loadings—when in fact the true relationships experience large structural breaks—can lead to several problems: overestimation of the number of factors, inconsistent estimates of the loadings, and deterioration of the forecasting performance of the factors.

The literature on this topic is rapidly growing, but the empirical results and conclusions vary across authors. We provide an overview of the literature, without covering it exhaustively. This overview can be roughly divided into two parts. In a first group of papers, the loadings are different at each point of time. In a second group of papers, the loadings display one break or a small number of breaks, and tests are proposed for the null hypothesis of no break.

The first paper addressing changes in the loadings is Stock and Watson (2002a). The authors allowed for small-amplitude time variations in the loadings. More precisely, they assumed that $x_{it} = \lambda_{it} f_t + e_{it}$ with $\lambda_{it} = \lambda_{it-1} + g_{it}\zeta_{it}$, where $g_{it} = O\left(T^{-1}\right)$ and ζ_{it} has weak cross-sectional dependence. They proved that PCA estimation of the loadings and factors is consistent, even though the estimation method assumes constant loadings. This result has been confirmed by Stock and Watson (2009), who analyzed a set of 110 US quarterly series spanning 1959 to 2006 and introduce a single break in 1984Q1 (start of the Great Moderation). They found evidence of instability in the loadings for about half of the series, but showed that the factors are well estimated using PCA on the full sample. Bates et al. (2013) further characterized the type and magnitude of parameter instability that is compatible with the consistency of PCA estimates. They showed that, under an appropriate set of assumptions, the PCA estimated factors are consistent if the loading matrix is decomposed as $\mathbf{\Lambda}_t = \mathbf{\Lambda}_0 + h_{nT}\boldsymbol{\xi}_t$ with $h_{nT} = O\left(T^{-1/2}\right)$, which strengthens the result of Stock and Watson (2002a), obtained for $h_{nT} = O\left(T^{-1}\right)$). They further showed that, for a given number of factors, if $h_{nT} = O\left(1/\min(n^{1/4}T^{1/2}, T^{3/4}\right)$ the estimated factors converge to the true ones (up to an invertible matrix) at the same rate as in Bai and Ng (2002) i.e. $1/\min(n^{1/2}, T^{1/2})$. However, they showed that, if the proportion of series undergoing a break is too high, usual criteria are likely to select two many factors.

In a Monte-Carlo experiment, Banerjee et al. (2007) demonstrated that consistent estimation of the factors does not preclude a deterioration of factor based forecasts. Stock and Watson (2009) pointed out that forecast equations may display even more instability than the factor loadings, and they assessed the importance of this instability

on the forecast performance. In particular, they showed that the best forecast results are obtained when the factors are estimated from the whole sample, but the forecast equation is only estimated on the sub-sample where its coefficients are stable.

Del Negro and Otrok (2008) proposed a DFM with time-varying factor loadings, but they also included stochastic volatility in both the factors and the idiosyncratic components. Their model aimed at studying the evolution of international business cycles, and their dataset consisted of real GDP growth for nineteen advanced countries. They considered a model with two factors: a world factor and a European factor. The model, with time varying loadings, can be written as

$$y_{it} = a_i + b_{it}^w f_t^w + b_{it}^e f_t^e + \varepsilon_{it} \ ,$$

where f_t^w and f_t^e are the world and European factors, and where $b_{it}^e = 0$ for non-European countries. The loadings are assumed to follow a random walk without drift: $b_{it} = b_{it-1} + \sigma_{\eta_i}\eta_{it}$, with the underlying idea that the sensitivity of a given country to the factors may evolve over time, and that this evolution is slow but picks up permanent changes in the economy. The factors and the idiosyncratic components have stochastic volatility, which allows variation in the importance of global/regional shocks and country-specific shocks. The model was estimated using Gibbs sampling. The results supported the notion of the Great Moderation in all the countries in the sample, notwithstanding important heterogeneity in the timing and magnitude, and in the sources (domestic or international) of this moderation. This is in line with features later highlighted by Stock and Watson (2012) (see below). Del Negro and Otrok (2008) also showed that the intensity of comovements is time-varying, but that there has been a convergence in the volatility of fluctuations in activity across countries.

Su and Wang (2017) considered a factor model where the number of factors is fixed, and the loadings change smoothly over time, using the following specification: $\lambda_{it} = \lambda_i(t/T)$ where λ_i is an unknown smooth function. They employed a local version of PCA to estimate the factors and the loadings, and obtained local versions of the Bai (2003) asymptotic distributions. They used an information criterion similar to the Bai and Ng (2002) information criteria and proved its consistency for large n and T. They also proposed a consistent test of the null hypothesis of constant loadings: the test statistic is a rescaled version of the mean square discrepancy between the common components estimated with time-varying loadings and the common components estimated by PCA with constant loadings, and it is asymptotically gaussian under the null. Finally, the authors also suggested a bootstrap version of the test in order to improve its size in finite samples. Their simulations showed that the information criteria work well, and that the bootstrap version of the test is more powerful than other existing tests when there is a single break at an unknown date. Finally, using the Stock and Watson (2009) dataset, they clearly rejected the null of constant loadings.

Breitung and Eickmeier (2011) were the first to consider the case where strong breaks may occur in the loadings. They noted that, in this case, the number of common factors has to be increased: for instance, in the case of a single break, two sets of factors are needed to describe the common component before and after

the break, which is tantamount to increasing the number of factors in the whole sample. For a known break date, they proposed to test the null hypothesis of constant loadings in individual series, using a Chow test, a Wald test or an LM test, with the PCA-estimated factors replacing the unknown factors. They also addressed the issue of a structural break at an unknown date: building on Andrews (1993), they proposed a Sup-LM statistic to test the null of constant loadings in an individual series. In both cases, autocorrelation in the factors and idiosyncratic terms are taken into account. Applying an LM-test, and using Stock and Watson (2005) US data, they found evidence of a structural break at the beginning of 1984 (start of the great Moderation). They also found evidence of structural breaks for the Euro-area, at the beginning of 1992 (Maastricht treaty) and the beginning of 1999 (stage 3 of EMU). Yamamoto and Tanaka (2015) noted that this testing procedure suffers from non-monotonic power, which is widespread in structural change tests. To remedy this issue, they proposed a modified version of the Breitung and Eickmeier (2011) test, taking the maximum of the Sup-Wald test statistics obtained from regressing the variable of interest on each estimated factor. They showed that this new test does not suffer from the non-monotonic power problem.

Stock and Watson (2012) addressed the issue of a potential new factor associated with the 2007Q4 recession and its aftermath. The authors used a large dataset of 132 disaggregated quarterly series, which were transformed to induce stationarity and subsequently "detrended" to eliminate low frequency variations from the data. They estimated six factors and the corresponding loadings by PCA over the period 1959Q1-2007Q3. The factors were extended over 2007Q4-2011Q2 by using the estimated loadings from the pre-recession period to form linear combinations of the observed variables after the onset of the recession. The extended factors available from 1959Q1 to 2011Q2 with constant loadings were denoted the "old factors". The authors showed that these old factors explain most of the variation in the individual time series, which suggests that there was no new factor after the financial crisis. They also tested the null hypothesis of a break in the loadings after 2007Q4, and showed that this hypothesis is rejected only for a small number of series. Further, they investigated the presence of a new factor by testing whether the idiosyncratic residuals display a factor structure, and concluded that there is no evidence of a new factor. Finally, they examined the volatility of the estimated innovations of the factors during different subperiods, and found evidence that the recession was associated with exceptionally large unexpected movements in the "old factors". Overall, they concluded that the financial crisis resulted in larger volatility of the factors, but neither did a new factor appear, nor did the response of the series to the factors change, at least for most series. These results are consistent with those obtained by Del Negro and Otrok (2008).

However, many users of DFMs have focused on the breaks-in-the-loadings scheme mentioned above, and several other tests have been proposed to test the null hypothesis of no break. Han and Inoue (2015) focused on the joint null hypothesis that all factor loadings are constant over time against the alternative that a fraction α of the loadings are not. Their test assumes that there is a single break at an unknown date that is identical for all series. They used the fact that, if the factors are estimated

from the whole sample, their empirical covariance matrix before the break will differ from their empirical covariance matrix after the break. They proposed a Sup-Wald and Sup-LM test, where the supremum is taken over the possible break dates. These tests were shown to be consistent even if the number of factors is overestimated.

Chen et al. (2014) proposed a test designed to detect big breaks at potentially unknown dates. As previously noticed by Breitung and Eickmeier (2011), in such a situation one can write a model with fixed loadings and a larger number of factors. The test is based on the behavior of the estimated factors before and after the break date if there is one. It relies on a linear regression of one of the estimated factors on the others, and tests for a structural break in the coefficients of this regression. If the potential break date is known, the test is a standard Wald test. If it is unknown, the test can be run using the Sup-LM or Sup-Wald tests which have been proposed by Andrews (1993). The authors' Monte-Carlo experiment showed that both tests perform well when $T \geq 100$, and have better power than the tests proposed by Breitung and Eickmeier (2011) or Han and Inoue (2015). The authors also showed that the Sup-Wald test generally behaves better than Sup-LM test in finite samples, and confirms that Bai-Ng's criteria overestimate the number of factors when there is a break. Finally, the authors applied the Sup-Wald test to the same dataset as Stock and Watson (2009): the null hypothesis of no break was rejected, and the estimated break date was around 1979-1980, rather than 1984, the break date chosen by Stock and Watson (2009) and usually associated with the start of the Great Moderation.

Cheng et al. (2016) considered the case where the number of factors may change at one, possibly unknown, break date, but adopted a different approach, based on shrinkage estimation. Since it is only the product of factors and loadings, the common component is uniquely identified in a factor model. The authors used a normalization that attributes changes in this product to changes in the loadings. The estimator is based on a penalized least-squares (PLS) criterion function, in which adaptive group-LASSO penalties are attached to pre-break factor loadings and to changes in the factor loadings. This PLS estimator shrinks the small coefficients to zero, but a new factor appears if a column of zero loadings turns into non-zero values after the break. The authors proved the consistency of the estimated number of pre- and post-break factors and the detection of changes in the loadings, under a general set of assumptions. Once the number of pre- and post-break factors has been consistently estimated, the break date can also be consistently estimated. The authors' Monte-Carlo experiment showed that their shrinkage estimator cannot detect small breaks, but is more likely to detect large breaks than Breitung and Eickmeier (2011), Chen et al. (2014), or Han and Inoue (2015). Finally, they applied their procedure to the same dataset as Stock and Watson (2012). The results provided strong evidence for a change in the loadings after 2007, and the emergence of a new factor that seems to capture comovements among financial series, but also spills over into real variables.

Corradi and Swanson (2014) looked at the consequences of instability in factor-augmented forecast equations. Forecast failures can result from instability in the loadings, instability in the regression coefficients of forecast equations, or both. They built a test for the joint null hypothesis of structural stability of factor loadings and factor augmented forecast equation coefficients. The test statistic is based on

the difference between the sample covariance of the forecasted variable and the factors estimated on the whole sample, and the sample covariance of the forecasted variable and the factors estimated using a rolling window estimation scheme. The number of factors is fixed according to the Bai and Ng (2002) criterion, and is thus overestimated if there is a break in the loadings. Under a general set of assumptions, and if $\sqrt{T}/n \to 0$, the test statistics based on the difference between the two sample covariances has an asymptotic χ^2 distribution under the null. Using this test on an empirical dataset analogous to Stock and Watson (2002a), the authors rejected the null of stability for six forecasted variables (in particular GDP) but did not reject the null for four others.

Baltagi et al. (2017) also addressed the issue of a single break in the number of factors and/or the factor loadings at an unknown date. The number of factors is fixed on the whole sample, without taking the break into account, and the estimation of the break point relies on the discrepancy between the pre- and post-break second moment matrices of the estimated factors. Once the break point is estimated, the authors showed that the number of factors and the factor space are consistently estimated on each sub-sample at the same rate of convergence as in Bai and Ng (2002).

Ma and Su (2018) considered the case where the loadings exhibit an unknown number of breaks. They proposed a three-step procedure to detect the breaks if any, and identify the dates when they occur. In the first step, the sample is divided into $J + 1$ intervals, with $T \gg J \gg m$, where m is an upper bound for the number of breaks, and a factor model is estimated by PCA on each interval. In the second step a fused group Lasso is applied to identify intervals containing a break. In the third step, a grid search allows to determine each break inside the corresponding interval. The authors proved that this procedure consistently estimates the number of breaks and their location. Using this method on Stock and Watson (2009) dataset, they identified five breaks in the factor loadings for the 1959-2006 period.

2.9 Conclusion

This chapter reviews the literature on dynamic factor models and several extensions of the basic framework. The modeling and estimation techniques surveyed include static and dynamic representation of small and large scale factor models, non-parametric and maximum likelihood estimation, estimation in the time and frequency domain, accommodating datasets with missing observations, and regime switching and time varying parameter models.

References

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

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(1):91–96.

Anderson, T. (1984). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, 2nd edition.

Anderson, T. and Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Econometrics, Industrial Research, and Psychometry*.

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, pages 821–856.

Aruoba, S. B., Diebold, F. X., and Scotti, C. (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics*, 27(4):417–427.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics*, 122(1):137–183.

Bai, J. and Li, K. (2012). Statistical analysis of factor models of high dimension. *The Annals of Statistics*, 40(1):436–465.

Bai, J. and Li, K. (2016). Maximum likelihood estimation and inference for approximate factor models of high dimension. *Review of Economics and Statistics*, 98(2):298–309.

Bai, J. and Liao, Y. (2013). Statistical inferences using large estimated covariances for panel data and factor models. *Unpublished*.

Bai, J. and Liao, Y. (2016). Efficient estimation of approximate factor models via penalized maximum likelihood. *Journal of Econometrics*, 191(1):1–18.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2004). A panic attack on unit roots and cointegration. *Econometrica*, 72(4):1127–1177.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, 25(1):52–60.

Bai, J. and Ng, S. (2008a). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317.

Bai, J. and Ng, S. (2008b). Large dimensional factor analysis. *Foundations and Trends in Econometrics*, 3(2):89–163.

Bai, J. and Ng, S. (2009). Boosting diffusion indices. *Journal of Applied Econometrics*, 24(4):607–629.

Bai, J. and Wang, P. (2016). Econometric analysis of large factor models. *Annual Review of Economics*, 8:53–80.

Baltagi, B. H., Kao, C., and Wang, F. (2017). Identification and estimation of a large factor model with structural instability. *Journal of Econometrics*, 197(1):87–100.

Banbura, M. and Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29(1):133–160.

Bańbura, M. and Rünstler, G. (2011). A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting*, 27(2):333–346.

Banerjee, A., Marcellino, M., and Masten, I. (2007). Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. In Rapach, D. and Wohar, M., editors, *Forecasting in the Presence of Structural Breaks and Model Uncertainty*. Emerald Group Publishing.

Barigozzi, M. (2018). *Dynamic Factor Models.* Lecture notes.

Bartholomew, D. J. (1987). *Latent Variable Models and Factors Analysis*. Oxford University Press.

Bates, B. J., Plagborg-Möller, M., Stock, J. H., and Watson, M. W. (2013). Consistent factor estimation in dynamic factor models with structural instability. *Journal of Econometrics*, 177(2):289–304.

Bessec, M. and Bouabdallah, O. (2015). Forecasting GDP over the business cycle in a multi-frequency and data-rich environment. *Oxford Bulletin of Economics and Statistics*, 77(3):360–384.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.

Bräuning, F. and Koopman, S. J. (2014). Forecasting macroeconomic variables using collapsed dynamic factor analysis. *International Journal of Forecasting*, 30(3):572–584.

Breitung, J. and Eickmeier, S. (2006). Dynamic factor models. *Allgemeines Statistisches Archiv*, 90(1):27–42.

Breitung, J. and Eickmeier, S. (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, 163(1):71–84.

Breitung, J. and Eickmeier, S. (2016). Analyzing international business and financial cycles using multi-level factor models: A comparison of alternative approaches. In *Dynamic Factor Models*, pages 177–214. Emerald Group Publishing Limited.

Breitung, J. and Tenhofen, J. (2011). GLS estimation of dynamic factor models. *Journal of the American Statistical Association*, 106(495):1150–1166.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*, volume 36. Siam.

Burns, A. F. and Mitchell, W. C. (1946). *Measuring Business Cycles*. NBER.

Camacho, M., Perez-Quiros, G., and Poncela, P. (2014). Green shoots and double dips in the euro area: A real time measure. *International Journal of Forecasting*, 30(3):520–535.

Camacho, M., Perez-Quiros, G., and Poncela, P. (2015). Extracting nonlinear signals from several economic indicators. *Journal of Applied Econometrics*, 30(7):1073–1089.

Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica*, pages 1305–1323.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, pages 1281–1304.

Chauvet, M. (1998). An econometric characterization of business cycle dynamics with factor structure and regime switching. *International Economic Review*, pages 969–996.

Chauvet, M. and Piger, J. (2008). A comparison of the real-time performance of business cycle dating methods. *Journal of Business & Economic Statistics*, 26(1):42–49.

Chauvet, M. and Senyuz, Z. (2016). A dynamic factor model of the yield curve components as a predictor of the economy. *International Journal of Forecasting*, 32(2):324–343.

Chen, L., Dolado, J. J., and Gonzalo, J. (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics*, 180(1):30–48.

Cheng, X., Liao, Z., and Schorfheide, F. (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *Review of Economic Studies*, 83(4):1511–1543.

Choi, I. (2012). Efficient estimation of factor models. *Econometric Theory*, 28(2):274–308.

Connor, G. and Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics*, 15(3):373–394.

Connor, G. and Korajczyk, R. A. (1988). Risk and return in an equilibrium apt: Application of a new test methodology. *Journal of Financial Economics*, 21(2):255–289.

Connor, G. and Korajczyk, R. A. (1993). A test for the number of factors in an approximate factor model. *Journal of Finance*, 48(4):1263–1291.

Corradi, V. and Swanson, N. R. (2014). Testing for structural stability of factor augmented forecasting models. *Journal of Econometrics*, 182(1):100–118.

Darné, O., Barhoumi, K., and Ferrara, L. (2014). Dynamic factor models: A review of the literature. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 2.

Del Negro, M. and Otrok, C. (2008). Dynamic factor models with time-varying parameters: measuring changes in international business cycles. *Federal Reserve Bank of New York, Staff Report*.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*, pages 1–38.

Diebold, F. X. (2003). Big data dynamic factor models for macroeconomic measurement and forecasting. In M. Dewatripont, L. H. and Turnovsky, S., editors, *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress of the Econometric Society*, pages 115–122.

Diebold, F. X. and Rudebusch, G. D. (1996). Measuring business cycles: A modern perspective. *Review of Economics and Statistics*, pages 67–77.

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(1):188–205.

Doz, C., Giannone, D., and Reichlin, L. (2012). A quasi–maximum likelihood approach for large, approximate dynamic factor models. *Review of Economics and Statistics*, 94(4):1014–1024.

Doz, C. and Petronevich, A. (2016). Dating business cycle turning points for the french economy: An MS-DFM approach. In *Dynamic Factor Models*, pages 481–538. Emerald Group Publishing Limited.

Doz, C. and Petronevich, A. (2017). On the consistency of the two-step estimates of the MS-DFM: A Monte Carlo study. Working paper.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, 32(2):407–499.

Engle, R. and Watson, M. (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, 76(376):774–781.

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(5):603–620.

Fiorentini, G., Galesi, A., and Sentana, E. (2018). A spectral EM algorithm for dynamic factor models. *Journal of Econometrics*, 205(1):249–279.

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009). Opening the black box: Structural factor models with large cross sections. *Econometric Theory*, 25(5):1319–1347.

Forni, M., Giovannelli, A., Lippi, M., and Soccorsi, S. (2018). Dynamic factor model with infinite-dimensional factor space: Forecasting. *Journal of Applied Econometrics*, 33(5):625–642.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics*, 82(4):540–554.

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(471):830–840.

Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2015). Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of Econometrics*, 185(2):359–371.

Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2017). Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis. *Journal of Econometrics*, 199(1):74–92.

Forni, M. and Lippi, M. (2011). The general dynamic factor model: One-sided representation results. *Journal of Econometrics*, 163(1):23–28.

Fu, D. (2007). National, regional and metro-specific factors of the us housing market. Federal Reserve Bank of Dallas Working Paper*, (0707)*.

Fuleky, P. and Bonham, C. S. (2015). Forecasting with mixed frequency factor models in the presence of common trends. *Macroeconomic Dynamics*, 19(4):753–775.

Geweke, J. (1977). *The dynamic factor analysis of economic time series*. Latent variables in socio-economic models.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.

Groen, J. J. and Kapetanios, G. (2016). Revisiting useful approaches to data-rich macroeconomic forecasting. *Computational Statistics & Data Analysis*, 100:221–239.

Guérin, P. and Marcellino, M. (2013). Markov-switching MIDAS models. *Journal of Business & Economic Statistics*, 31(1):45–56.

Hallin, M. and Lippi, M. (2013). Factor models in high-dimensional time series a time-domain approach. *Stochastic Processes and their Applications*, 123(7):2678–2695.

Hallin, M. and Liska, R. (2007). The generalized dynamic factor model: Determining the number of factors. *Journal of the American Statistical Association*, 102:603–617.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, pages 357–384.

Han, X. and Inoue, A. (2015). Tests for parameter instability in dynamic factor models. *Econometric Theory*, 31(5):1117–1152.

Harding, D. and Pagan, A. (2003). A comparison of two business cycle dating methods. *Journal of Economic Dynamics and Control*, 27(9):1681–1690.

Harvey, A. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge Univ Pr.

Inoue, A. and Kilian, L. (2008). How useful is bagging in forecasting economic time series? A case study of US consumer price inflation. *Journal of the American Statistical Association*, 103(482):511–522.

Jackson, L. E., Kose, M. A., Otrok, C., and Owyang, M. T. (2016). Specification and estimation of Bayesian dynamic factor models: A Monte Carlo analysis with an application to global house price comovement. In *Dynamic Factor Models*, pages 361–400. Emerald Group Publishing Limited.

Jones, C. S. (2001). Extracting factors from heteroskedastic asset returns. *Journal of Financial Economics*, 62(2):293–325.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, 32(4):443–482.

Jungbacker, B. and Koopman, S. J. (2014). Likelihood-based dynamic factor analysis for measurement and forecasting. *Econometrics Journal*, 18(2):C1–C21.

Jungbacker, B., Koopman, S. J., and van der Wel, M. (2011). Maximum likelihood estimation for dynamic factor models with missing data. *Journal of Economic Dynamics and Control*, 35(8):1358–1368.

Kelly, B. and Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, 186(2):294–316.

Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60(1-2):1–22.

Kim, C.-J. and Nelson, C. R. (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *Review of Economics and Statistics*, 80(2):188–201.

Kim, M.-J. and Yoo, J.-S. (1995). New index of coincident indicators: A multivariate Markov switching factor model approach. *Journal of Monetary Economics*, 36(3):607–630.

Koopman, S. and Harvey, A. (2003). Computing observation weights for signal extraction and filtering. *Journal of Economic Dynamics and Control*, 27(7):1317–1333.

Kose, M. A., Otrok, C., and Whiteman, C. H. (2003). International business cycles: World, region, and country-specific factors. *American Economic Review*, pages 1216–1239.

Kose, M. A., Otrok, C., and Whiteman, C. H. (2008). Understanding the evolution of world business cycles. *Journal of International Economics*, 75(1):110–130.

Lawley, D. N. and Maxwell, A. E. (1971). *Factor Analysis as Statistical Method*. Butterworths, 2nd edition.

Li, H., Li, Q., and Shi, Y. (2017). Determining the number of factors when the number of factors can increase with sample size. *Journal of Econometrics*, 197(1):76–86.

Lütkepohl, H. (2014). A survey, Structural vector autoregressive analysis in a data rich environment.

Ma, S. and Su, L. (2018). Estimation of large dimensional factor models with an unknown number of breaks. *Journal of Econometrics*, 207(1):1–29.

Magnus, J. R. and Neudecker, H. (2019). *Matrix differential calculus with applications in statistics and econometrics*. Wiley.

Marcellino, M. and Schumacher, C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for german GDP. *Oxford Bulletin of Economics and Statistics*, 72(4):518–550.

Mariano, R. and Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18(4):427–443.

Moench, E. and Ng, S. (2011). A hierarchical factor analysis of us housing market dynamics. *Econometrics Journal*, 14(1):C1–C24.

Moench, E., Ng, S., and Potter, S. (2013). Dynamic hierarchical factor models. *Review of Economics and Statistics*, 95(5):1811–1817.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5):1447–1479.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Review of Economics and Statistics*, 92(4):1004–1016.

Otrok, C. and Whiteman, C. H. (1998). Bayesian leading indicators: measuring and predicting economic conditions in Iowa. *International Economic Review*, pages 997–1014.

Peña, D. and Yohai, V. J. (2016). Generalized dynamic principal components. *Journal of the American Statistical Association*, 111(515):1121–1131.

Pinheiro, M., Rua, A., and Dias, F. (2013). Dynamic factor models with jagged edge panel data: Taking on board the dynamics of the idiosyncratic components. *Oxford Bulletin of Economics and Statistics*, 75(1):80–102.

Proietti, T. and Moauro, F. (2006). Dynamic factor analysis with non-linear temporal aggregation constraints. *Journal of The Royal Statistical Society Series C*, 55(2):281–300.

Reis, R. and Watson, M. W. (2010). Relative goods' prices, pure inflation, and the Phillips correlation. *American Economic Journal: Macroeconomics*, 2(3):128–57.

Sargent, T. J. and Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New Methods in Business Cycle Research*, 1:145–168.

Shumway, R. H. and Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3(4):253–264.

Spearman, C. (1904). General intelligence objectively determined and measured. *The American Journal of Psychology*, 15(2):201–292.

Stock, J. H. and Watson, M. (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual*, pages 351–394.

Stock, J. H. and Watson, M. (1991). A probability model of the coincident economic indicators. *Leading Economic indicators: New approaches and forecasting records*, 66.

Stock, J. H. and Watson, M. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20(2):147–162.

Stock, J. H. and Watson, M. (2006). Forecasting with many predictors. *Handbook of Economic Forecasting*, 1:515.

Stock, J. H. and Watson, M. (2008). The evolution of national and regional factors in US housing construction. In *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*.

Stock, J. H. and Watson, M. (2009). Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*, 173:205.

Stock, J. H. and Watson, M. W. (1993). A procedure for predicting recessions with leading indicators: econometric issues and recent experience. In *Business Cycles, Indicators and Forecasting*, pages 95–156. University of Chicago Press.

Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factor models for VAR analysis. Technical report, National Bureau of Economic Research.

Stock, J. H. and Watson, M. W. (2011). Dynamic factor models. In *Handbook on Economic Forecasting*. Oxford University Press.

Stock, J. H. and Watson, M. W. (2012). Disentangling the channels of the 2007-09 recession. *Brookings Papers on Economic Activity*, 1:81–135.

Stock, J. H. and Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. *Handbook of Macroeconomics*, 2:415–525.

Su, L. and Wang, X. (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics*, 198(1):84–101.

Watson, M. W. and Engle, R. F. (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23(3):385–400.

Yamamoto, Y. and Tanaka, S. (2015). Testing for factor loading structural change under common breaks. *Journal of Econometrics*, 189(1):187–206.