Econometric Analysis of Large Factor Models

Jushan Bai[1,2] and Peng Wang[3]

[1]Department of Economics, Columbia University, New York, NY 10027; email: jb3064@columbia.edu

[2]School of Finance, Nankai University, Tianjin, China 300350

[3]Department of Economics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong; email: pwang@ust.hk

Annu. Rev. Econ. 2016. 8:53–80

The *Annual Review of Economics* is online at economics.annualreviews.org

This article's doi: 10.1146/annurev-economics-080315-015356

JEL codes: C01, C33, C38, C55

Keywords

high-dimensional data, factor-augmented regression, FAVAR, number of factors, interactive effects, principal components, regularization, Bayesian estimation

Abstract

Large factor models use a few latent factors to characterize the co-movement of economic variables in a high-dimensional data set. High dimensionality brings challenges as well as new insights into the advancement of econometric theory. Because of their ability to effectively summarize information in large data sets, factor models have been increasingly used in economics and finance. The factors, estimated from the high-dimensional data, can, for example, help improve forecasting, provide efficient instruments, control for nonlinear unobserved heterogeneity, and capture cross-sectional dependence. This article reviews the theory on estimation and statistical inference of large factor models. It also discusses important applications and highlights future directions.

1. INTRODUCTION

With the rapid development of econometric theory and methodologies on large factor models over the past decade, researchers are equipped with useful tools to analyze high-dimensional data sets, which have become increasingly available due to advancements in data collection techniques and information technology. High-dimensional data sets in economics and finance are typically characterized by both a large cross-sectional dimension N and a large time dimension T. Conventionally, researchers rely heavily on factor models with observed factors to analyze such a data set, such as the capital asset pricing model and Fama-French factor model for asset returns (Fama & French 1993) and affine models for bond yields. In reality, however, not all factors are observed. This poses both theoretical and empirical challenges to researchers. The development of modern theory on large factor models greatly broadens the scope of factor analysis, especially when one has high-dimensional data and the factors are latent. Such a framework has helped substantially in the literature of diffusion-index forecasting, business cycle comovement analysis, and economic linkage between countries, as well as improved causal inference through factor-augmented vector autoregression (FAVAR) and provided more efficient instrumental variables (IV).

Stock & Watson (2002a) find that if they first extract a few factors from the large data set and then use the factors to augment an autoregressive model, the model has much better forecasting performance than alternative univariate or multivariate models. Giannone et al. (2008) apply the factor model to conduct now-casting, which combines data of different frequencies and forms a forecast for real GDP. Bernanke et al. (2005) find that using a few factors to augment VAR helps to better summarize the information from different economic sectors and produces more credible impulse responses than the conventional VAR model. Boivin & Giannoni (2006) combine factor analysis with dynamic stochastic general equilibrium models, providing a framework for estimating dynamic economic models using large data sets. Such a methodology helps to mitigate the measurement error problem as well as the omitted variables problem during estimation. Using large dynamic factor models, Ng & Ludvigson (2009) identify important linkages between bond returns and macroeconomic fundamentals. Fan et al. (2011) base their high-dimensional covariance estimation on large approximate factor models, allowing sparse error covariance matrix after taking out common factors.

Apart from its wide applications, the large factor model also brings new insight to our understanding of nonstationary data. For example, the link between cointegration and common trend is broken in the setup of large (N, T). More importantly, the common trend can be consistently estimated regardless of whether individual errors are stationary or integrated processes. That is, the number of unit roots can exceed the number of series, yet common trends are well defined and can be extracted from the data.

This review aims to introduce the theory and various applications of large factor models. In particular, we examine basic issues related to high-dimensional factors models. We first introduce the factor models under the setup of large N and large T, with a special focus on how the factor space can be identified in the presence of weak correlations and heteroskedasticity. We then discuss the estimation of the large factor model via the method of principal components as well as the maximum likelihood approach, treating the number of factors as known. Next, we review the issue of determining the number of static and dynamic factors. We also study various applications of the large factor model, including factor-augmented linear regression and FAVAR, and how the framework of factor models can help deal with the many-IV problem. Some new developments in theory are also discussed. We start with recent advances in estimation and statistical tests of structural changes in large factor models. We then introduce static and dynamic panel data models with interactive fixed effects and possibly heterogeneous coefficients, which have been increasingly applied in empirical research. This review also examines the popular panel analysis of nonstationarity in the idiosyncratic and common components (PANIC), the ways in which structural restrictions help identify the factors and structural shocks, and the ways in which factor models may facilitate high-dimensional covariance estimation. Finally, we discuss the Bayesian approach to large factor models as well as its possible extensions. We conclude with a few potential topics that deserve future research.

2. LARGE FACTOR MODELS

Suppose we observe x_{it} for the i-th cross-sectional unit at period t. The large factor model for x_{it} is given by

$$x_{it} = \boldsymbol{\lambda}_i' \mathbf{F}_t + e_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1}$$

where the $r \times 1$ vector of factors $\mathbf{F}_t$ is latent, and the associated factor loadings $\boldsymbol{\lambda}_i$ are unknown. The model in Equation 1 can also be represented in vector form,

$$\mathbf{X}_t = \Lambda \mathbf{F}_t + \mathbf{e}_t, \tag{2}$$

where $\mathbf{X}_t = [x_{1t}, \ldots, x_{Nt}]'$ is $N \times 1$, $\Lambda = [\lambda_1, \ldots, \lambda_N]'$ is $N \times r$, and $\mathbf{e}_t = [e_{1t}, \ldots, e_{Nt}]'$ is $N \times 1$. Unlike short panel data studies (large N, fixed T) and multivariate time-series models (fixed N, large T), the large factor model is characterized by both large N and large T. The estimation and statistical inference are thus based on double asymptotic theory, in which both N and T converge to infinity. Such a large dimensional framework greatly expands the applicability of factor models to more realistic economic environments. For the identification of the space spanned by factors, weak correlations and heteroskedasticity are allowed along both the time dimension and the cross-sectional dimension for e_{it} without affecting the main properties of factor estimates. Let $E(e_{it}e_{js}) = \sigma_{ij,ts}$. Such weak correlations and heteroskedasticity can be defined by the following conditions as in Bai (2003, 2009):

$$|\sigma_{ij,ts}| \le \sigma_{ij} \text{ for all } (t,s), \text{ and } |\sigma_{ij,ts}| \le \tau_{ts} \text{ for all } (i,j),$$
$$\frac{1}{N}\sum_{i,j=1}^{N} \sigma_{ij} \le M, \quad \frac{1}{T}\sum_{t,s=1}^{T} \tau_{ts} \le M, \quad \text{and} \quad \frac{1}{NT}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T} |\sigma_{ij,ts}| \le M \quad \text{for some } M < \infty.$$

The weak correlations in errors give rise to what Chamberlain & Rothschild (1983) call the approximate factor structure. Weak correlations between the factors and idiosyncratic errors can be allowed as well, as in Bai (2003). In contrast, under a fixed T, identification of the factor space requires extra assumptions, such as asymptotic orthogonality and asymptotic homoskedasticity:

$$\text{plim} \frac{1}{N}\sum_{i=1}^{N} e_{it}e_{is} = \begin{cases} 0, & \text{for} \quad t \neq s, \\ \sigma^2, & \text{for} \quad t = s. \end{cases}$$

The high-dimensional framework also brings new insight into double asymptotics. For example, it is $C_N = \min\{N^{1/2}, T^{1/2}\}$ that determines the rate of convergence for the factor and factor loading estimates under large N and large T.

3. ESTIMATING THE LARGE FACTOR MODEL

The large factor model can be estimated using either the time-domain approach (mainly for static factor models) or frequency-domain approach (for dynamic factors). In this section, we focus on the time-domain approach. In particular, we mainly consider the principal components methods and maximum likelihood methods. Examples of the frequency-domain approach are provided by

Geweke (1977), Quah & Sargent (1993), and Forni et al. (2000, 2004, 2005). Below, we assume that the number of factors r is known. If r is unknown, it can be replaced by $\hat{k}$ using any of the information criteria discussed in the next section without affecting the asymptotic properties of the estimators.

Before estimating the model, we need to impose normalizations on the factors and factor loadings to pin down the rotational indeterminacy. This is because $\boldsymbol{\lambda}_i'\mathbf{F}_t = (\mathbf{A}^{-1}\boldsymbol{\lambda}_i)'(\mathbf{A}'\mathbf{F}_t)$ for any $r \times r$ full-rank matrix $\mathbf{A}$. Given that an arbitrary $r \times r$ matrix has r^2 degrees of freedom, we need to impose at least r^2 restrictions (order condition) to remove the indeterminacy. Let $\mathbf{F} = (\mathbf{F}_1, \ldots, \mathbf{F}_T)'$. Three commonly applied normalizations are PC1, PC2, and PC3 as follows. In PC1, $\frac{1}{T}\mathbf{F}'\mathbf{F} = \mathbf{I}_r$, $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}$ is diagonal with distinct entries. In PC2, $\frac{1}{T}\mathbf{F}'\mathbf{F} = \mathbf{I}_r$, the upper $r \times r$ block of Λ is lower triangular with nonzero diagonal entries. Finally, in PC3, the upper $r \times r$ block of $\boldsymbol{\Lambda}$ is given by $\mathbf{I}_r$.

PC1 is often imposed by maximum likelihood estimation (MLE) in classical factor analysis (see Anderson & Rubin 1956). PC2 is analogous to a recursive system of simultaneous equations. PC3 is linked to the measurement error problem such that the first observable variable x_{1t} is equal to the first factor f_{1t} plus a measurement error e_{1t}, and the second observable variable x_{2t} is equal to the second factor f_{2t} plus a measurement error e_{2t}, and so on. Bai & Ng (2013) discuss these restrictions in more detail.

Each preceding set of normalizations yields r^2 restrictions, meeting the order condition for identification (eliminating the rotational indeterminacy). These restrictions also satisfy the rank condition for identification (Bai & Wang 2014). The common components $\boldsymbol{\lambda}_i'\mathbf{F}_t$ have no rotational indeterminacy and are identifiable without restrictions.

3.1. The Principal Components Method

The principal components estimators for factors and factor loadings can be treated as outcomes of a least squares problem under normalization PC1. Estimators under normalizations PC2 and PC3 can be obtained by properly rotating the principal components estimators. Consider minimizing the sum of squares residuals under PC1,

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \boldsymbol{\lambda}_i'\mathbf{F}_t)^2 = \mathrm{tr}[(\mathbf{X} - \mathbf{F}\boldsymbol{\Lambda}')(\mathbf{X} - \mathbf{F}\boldsymbol{\Lambda}')'],$$

where $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_T)'$, a $T \times N$ data matrix. $\hat{\mathbf{F}}$ is given by the first r leading eigenvectors[1] of $\mathbf{XX}'$ multiplied by $T^{1/2}$, and $\hat{\boldsymbol{\Lambda}} = \mathbf{X}'\hat{\mathbf{F}}/T$ (see, e.g., Connor & Korajczyk 1986, Stock & Watson 1998). Bai & Ng (2008) discuss some equivalent ways of obtaining the principal components estimators.

Bai (2003) studies the asymptotic properties under large N and T. Under the general condition $\sqrt{N}/T \to 0$, one obtains $\sqrt{N}(\hat{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t^0) \xrightarrow{d} N(0, \mathbf{V}_F)$, where $\mathbf{H}$ is the rotation matrix, $\mathbf{F}_t^0$ is the true factor, and $\mathbf{V}_F$ is the estimable asymptotic variance. By symmetry, when $\sqrt{T}/N \to 0$, $\hat{\boldsymbol{\lambda}}_i$ is also asymptotically normal. Specifically, one obtains $\sqrt{T}(\hat{\boldsymbol{\lambda}}_i - \mathbf{H}'^{-1}\boldsymbol{\lambda}_i^0) \xrightarrow{\mathrm{d}} N(0, \mathbf{V}_\Lambda)$, for some $\mathbf{V}_\Lambda > 0$. The standard principal components estimates can be rotated to obtain estimates satisfying PC2 or PC3. The limiting distributions under PC2 and PC3 are obtained by Bai & Ng (2013). As for the common component $\boldsymbol{\lambda}_i'\mathbf{F}_t$, its limiting distribution requires no restriction on

[1] If there is an intercept in the model $\mathbf{X}_t = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_t + \mathbf{e}_t$, then the matrix $\mathbf{X}$ is replaced by its demeaned version $\mathbf{X} - (1, 1, \ldots, 1) \otimes \bar{\mathbf{X}} = (\mathbf{X}_1 - \bar{\mathbf{X}}, \mathbf{X}_2 - \bar{\mathbf{X}}, \ldots, \mathbf{X}_T - \bar{\mathbf{X}})$, where $\bar{\mathbf{X}} = T^{-1}\sum_{t=1}^{T}\mathbf{X}_t$.

the relationship between N and T, which is always normal. In particular, there exists a sequence of $b_{NT} = O_p\{\min\{N^{1/2}, T^{1/2}\}\}$ such that

$$b_{NT}(\hat{\boldsymbol{\lambda}}_i'\hat{\mathbf{F}}_t - \boldsymbol{\lambda}'_i\mathbf{F}_t) \xrightarrow{d} N(0, 1), \quad \text{as } N, T \to \infty.$$

Thus, the convergence rate for the estimated common components is $\min\{N^{1/2}, T^{1/2}\}$. This is the best rate possible.

3.2. The Generalized Principal Components

The principal components estimator is a least squares estimator [ordinary least squares (OLS)] and is efficient if $\boldsymbol{\Sigma}_e$ is a scalar multiple of an $N \times N$ identity matrix, that is, $\boldsymbol{\Sigma}_e = c\mathbf{I}_N$ for a constant $c > 0$. This is hardly true in practice; therefore generalized least squares (GLS) will give more efficient estimation. Consider the GLS objective function, assuming $\boldsymbol{\Sigma}_e$ is known:

$$\min_{\Lambda, F} \operatorname{tr}\left[(\mathbf{X} - \mathbf{F}\boldsymbol{\Lambda}')\boldsymbol{\Sigma}_e^{-1}(\mathbf{X} - \mathbf{F}\boldsymbol{\Lambda}')'\right].$$

Then the solution for $\mathbf{F}$ is given by the first r eigenvectors of the matrix $\mathbf{X}\boldsymbol{\Sigma}_e^{-1}\mathbf{X}'$, multiplied by $T^{1/2}$, and the solution for $\boldsymbol{\Lambda}$ is equal to $X'\hat{\mathbf{F}}/T$. The latter has the same expression as the standard principal components estimator.

This is the generalized principal components estimator (GPCE) considered by Choi (2012). He shows that the GPCE of the common component has smaller variance than the principal components estimator. Using the GPCE-based factor estimates will also produce smaller variance of the forecasting error.

In practice, $\boldsymbol{\Sigma}_e$ is unknown and needs to be replaced by an estimate. The usual covariance matrix estimator $\frac{1}{T}\sum_{t=1}^{T}\hat{\mathbf{e}}_t\hat{\mathbf{e}}_t'$ based on the standard principal components residuals $\hat{\mathbf{e}}_t$ is not applicable as it is not a full-rank matrix, regardless of the magnitude of N and T. Even if every element of $\boldsymbol{\Sigma}_e$ can be consistently estimated by some way and the resulting estimator $\boldsymbol{\Sigma}_e$ is of full rank, the GPCE is not necessarily more accurate than the standard principal components estimator. Unlike standard inference with a finite dimensional weighting matrix [such as the generalized method of moments (GMM)], a mere consistency of $\hat{\boldsymbol{\Sigma}}_e$ is insufficient to obtain the limiting distribution of the GPCE. Even an optimally estimated Σ_e in the sense of Cai & Zhou (2012) is not enough to establish the asymptotic equivalence between the feasible and infeasible estimators. Thus, the high dimensionality of $\boldsymbol{\Sigma}_e$ makes a fundamental difference in terms of inference. Bai & Liao (2013) show that the true matrix $\boldsymbol{\Sigma}_e$ has to be sparse and its estimates should take into account the sparsity assumption. Sparsity does not require many zero elements, but many elements must be sufficiently small. Under the sparsity assumption, a shrinkage estimator of $\boldsymbol{\Sigma}_e$ (e.g., a hard thresholding method based on the residuals from the standard principal components method) will give a consistent estimation of $\boldsymbol{\Sigma}_e$. Bai & Liao (2013) derive the conditions under which the estimated $\boldsymbol{\Sigma}_e$ can be treated as known.

Other related methods include those by Breitung & Tenhofen (2011), who propose a two-step estimation procedure, which allows for heteroskedastic (diagonal $\boldsymbol{\Sigma}_e$) and serially correlated errors. They show that the feasible two-step estimator has the same limiting distribution as the GLS estimator. In finite samples, the GLS estimators tend to be more efficient than the usual principal components estimators. An iterated version of the two-step estimation method has also been proposed, which further improves efficiency in finite samples.

3.3. The Maximum Likelihood Method

For the factor model $\mathbf{X}_t = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_t + \mathbf{e}_t$, under the assumption that $\mathbf{e}_t$ is independent and identically distributed (i.i.d.) normal $N(0, \boldsymbol{\Sigma}_e)$ and $\mathbf{f}_t$ is i.i.d. normal $N(0, \mathbf{I}_r)$, then $\mathbf{X}_t$ is normal $N(\boldsymbol{\mu}, \boldsymbol{\Omega})$, where $\mathbf{\Omega} = \mathbf{\Lambda\Lambda}' + \mathbf{\Sigma}_e$, it follows that the likelihood function is

$$L(\mathbf{\Lambda}, \mathbf{\Sigma}_e) = -\frac{1}{N}\log|\mathbf{\Lambda\Lambda}' + \mathbf{\Sigma}_e| - \frac{1}{N}\mathrm{tr}(\mathbf{S}(\mathbf{\Lambda\Lambda}' + \mathbf{\Sigma}_e)^{-1}),$$

where $\mathbf{S} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{X}_t - \bar{\mathbf{X}})(\mathbf{X}_t - \bar{\mathbf{X}})'$ is the sample covariance matrix. The classical inferential theory of MLE is developed assuming N is fixed, and the sample size T goes to infinity (see Anderson 2003, Anderson & Rubin 1956, Lawley & Maxwell 1971). The basic assumption in classical factor analysis is that $\sqrt{T}(\mathbf{S} - \mathbf{\Omega})$ is asymptotically normal. This basic premise does not hold when N also goes to infinity. A new approach is required to obtain consistency and the limiting distribution under large N. Bai & Li (2012) derive the inferential theory assuming $\mathbf{\Sigma}_e$ is diagonal under various identification restrictions.

Normality assumption is inessential. The preceding likelihood is considered as the quasi-likelihood under non-normality. We point out that MLE is consistent even if N is fixed because the fixed-N case falls within the classical framework. In contrast, the principal components method is inconsistent under fixed N unless $\mathbf{\Sigma}_e$ is proportional to an identity matrix. GPCEs are hardly consistent when using residuals to estimate $\mathbf{\Sigma}_e$ because the residual $\hat{e}_{it} = x_{it} - \hat{\mu}_i - \hat{\mathbf{\lambda}}_i'\hat{\mathbf{f}}_t$ is not consistent for e_{it}. This follows because $\hat{\mathbf{f}}_t$ is not consistent for $\mathbf{f}_t$ under fixed N. MLE treats $\mathbf{\Sigma}_e$ as a parameter, which is jointly estimated with $\mathbf{\Lambda}$. MLE does not rely on residuals to estimate $\mathbf{\Sigma}_e$.

MLE for nondiagonal $\mathbf{\Sigma}_e$ is considered by Bai & Liao (2016). They assume $\mathbf{\Sigma}_e$ is sparse and use the regularization method to jointly estimate $\mathbf{\Lambda}$ and $\mathbf{\Sigma}_e$. Consistency is established for the estimated $\mathbf{\Lambda}$ and $\mathbf{\Sigma}_e$, but the limiting distributions remain unsolved, although the limiting distributions are conjectured to be the same as the two-step feasible GLS estimator in Bai & Liao (2013) under large N and T.

Given MLE for $\mathbf{\Lambda}$ and $\mathbf{\Sigma}_e$, the estimator for $\mathbf{f}_t$ is $\hat{\mathbf{f}}_t = (\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Sigma}}_e^{-1}\hat{\mathbf{\Lambda}})^{-1}\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Sigma}}_e^{-1}(\mathbf{X}_t - \bar{\mathbf{X}})$ for $t = 1, 2, \ldots, T$. This is a feasible GLS estimator of $\mathbf{f}_t$ in the model $\mathbf{X}_t = \mathbf{\mu} + \mathbf{\Lambda}\mathbf{f}_t + \mathbf{e}_t$. The estimated factor loadings have the same asymptotic distributions for the three different estimation methods (principal components, GPCE, and MLE) under large N and large T. But the estimated factors are more efficient under GPCE and MLE than standard principal components (see Choi 2012, Bai & Li 2012, Bai & Liao 2013). If time-series heteroskedasticity is of more concern, and especially when T is relatively small, then the role of $\mathbf{F}$ and $\mathbf{\Lambda}$ (also T and N) should be switched. Bai & Li (2012) consider the likelihood function for this setting.

The preceding discussion assumes the factors $\mathbf{f}_t$ are i.i.d. Doz et al. (2011, 2012) explicitly consider a finite-order VAR specification for $\mathbf{f}_t$, and then propose a two-step method or a quasi-maximum likelihood estimation procedure. The method is similar to MLE of a linear state space model. The main difference is that they initialize the estimation by using properly rotated principal components estimators. The factor estimates are obtained as either the Kalman filter or Kalman smoother. They show that estimation under independent Gaussian errors still leads to consistent estimators for the large approximate factor model, even when the true model has cross-sectional and time-series correlations in the idiosyncratic errors. Bai & Li (2016) study related issues for dynamic factors and cross-sectionally and serially correlated errors estimated by the maximum likelihood method.

4. DETERMINING THE NUMBER OF FACTORS

The number of factors is usually unknown, and thus alternative methods are available for estimation. In this section, we mainly focus on two types of methods. One is based on information criteria, and the other is based on the distribution of eigenvalues.

4.1. The Number of Static Factors

Bai & Ng (2002) treat the problem as that of model selection and propose a procedure that can consistently estimate the number of factors when N and T simultaneously converge to infinity. Let $\hat{\boldsymbol{\lambda}}_i^k$ and $\hat{\mathbf{F}}_t^k$ be the principal components estimators assuming that the number of factors is k. We may treat the sum of squared residuals (divided by NT) as a function of k:

$$V(k) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{it} - \hat{\boldsymbol{\lambda}}_i^{k\prime} \hat{\mathbf{F}}_t^k \right)^2 .$$

Define the following loss function:

$$\mathrm{IC}(k) = \ln(V(k)) + kg(N, T),$$

where the penalty function $g(N, T)$ satisfies two conditions: $g(N, T) \to 0$ and $\min\{N^{1/2}, T^{1/2}\} \cdot g(N, T) \to \infty$, as $N, T \to \infty$. Define the estimator for the number of factors as $\hat{k}_{IC} = \arg\min_{0 \le k \le k\max} \mathrm{IC}(k)$, where $k$max is the upper bound of the true number of factors r. Then consistency can be established under standard conditions on the factor model (Bai 2003, Bai & Ng 2002): $\hat{k}_{\mathrm{IC}} \xrightarrow{p} r$, as $N, T \to \infty$. Bai & Ng (2002) consider six formulations of the information criteria, which are shown to have good finite sample performance. We list three here:

$$\mathrm{IC}_1(k) = \ln(V(k)) + k \left(\frac{N+T}{NT} \right) \ln \left(\frac{NT}{N+T} \right),$$

$$\mathrm{IC}_2(k) = \ln(V(k)) + k \left(\frac{N+T}{NT} \right) \ln \left(C_{NT}^2 \right),$$

$$\mathrm{IC}_3(k) = \ln(V(k)) + k \left(\frac{\ln(C_{NT}^2)}{C_{NT}^2} \right).$$

The logarithm transformation in IC could be practically desirable, which avoids the need for a scaling factor in alternative criteria. Monte Carlo simulations show that all criteria perform well when both N and T are large. For the cases in which either N or T is small, and if errors are uncorrelated across units and time, the preferred criteria tend to be IC_1 and IC_2. Weak serial and cross-sectional correlations in errors do not alter the result; however, the relative performance of each criterion depends on the specific form of the correlation.

Some desirable features of the above method are worth mentioning. First, consistency is established without any restriction between N and T, and it does not rely on sequential limits. Second, the results hold under heteroskedasticity in both the time and cross-sectional dimensions, as well as under weak serial and cross-sectional correlations.

Based on large random matrix theory, Onatski (2009) establishes a test of k_0 factors against the alternative that the number of factors is between k_0 and k_1 ($k_0 < k \le k_1$). The test statistic is given by

$$R = \max_{k_0 < k \le k_1} \frac{\gamma_k - \gamma_{k+1}}{\gamma_{k+1} - \gamma_{k+2}},$$

where γ_k is the k-th largest eigenvalue of the sample spectral density of data at a given frequency. For macroeconomic data, the frequency could be chosen at the business cycle frequency. The basic idea of this approach is that under the null of k_0 factors, the first leading k_0 eigenvalues will be unbounded, while the remaining eigenvalues are all bounded. As a result, R will be bounded under the null, whereas it will explode under the alternative, making R asymptotically pivotal. The limiting distribution of R is derived under the assumption that T grows sufficiently faster than N, which turns out to be a function of the Tracy-Widom distribution.

Ahn & Horenstein (2013) propose two estimators, the eigenvalue ratio estimator and the growth ratio estimator, based on a simple calculation of eigenvalues. For example, the eigenvalue ratio estimator is defined as maximizing the ratio of two adjacent eigenvalues in decreasing order. The intuition of these estimators is similar to Onatski (2009, 2010), although their properties are derived under slightly different model assumptions. Although most of the literature assumes the number of factors to be fixed, H. Li et al. (2013) consider an increasing number of factors as the sample size increases. Such a situation might arise if there are structural breaks due to economic environmental changes, which could lead to new factors.

So far, above we consider only the case of strong factors. The implications of weak factors are investigated by Chudik et al. (2011) and Onatski (2011), who show that the principal components estimators might perform poorly in the presence of weak factors. Chudik et al. (2011) introduce the notions of weak and strong factors. Consider the factor loading $\boldsymbol{\lambda}_i = [\gamma_{i1}, \ldots, \gamma_{ir}]'$ and the factors $\mathbf{F}_t = [f_{1t}, \ldots, f_{rt}]'$ in Equation 1. The factor f_{lt} is said to be strong if $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}|\gamma_{il}| = K > 0$, and it is said to be weak if $\lim_{N\to\infty} \sum_{i=1}^{N}|\gamma_{il}| = K < \infty$. When factors are weak, Onatski (2011) proves that the principal components estimators became inconsistent. Under the special case of no temporal correlation in the idiosyncratic terms, one may identify the number of factors that are not orthogonal to their principal components estimators asymptotically. In the general case of weak cross-sectional and serial correlations in error terms, estimating the number of weak factors could be a potential future research topic. Our subsequent discussion is focused on strong factors.

4.2. The Number of Dynamic Factors

The method of Bai & Ng (2002) considers only the static factor model, in which the relationship between x_{it} and F_t is static. In dynamic factor models, the lags of factors also directly affect x_{it}. The methods for static factor models can be readily extended to estimate the number of dynamic factors. Consider

$$x_{it} = \boldsymbol{\lambda}'_{i0}\mathbf{f}_t + \boldsymbol{\lambda}'_{i1}\mathbf{f}_{t-1} + \cdots + \boldsymbol{\lambda}'_{is}\mathbf{f}_{t-s} + e_{it} = \boldsymbol{\lambda}_i(L)'\mathbf{f}_t + e_{it}, \tag{3}$$

where $\mathbf{f}_t$ is $q \times 1$ and $\boldsymbol{\lambda}_i(L) = \boldsymbol{\lambda}_{i0} + \boldsymbol{\lambda}_{i1}L + \cdots \boldsymbol{\lambda}_{is}L^s$. Whereas Forni et al. (2000, 2004, 2005) consider the case with $s \to \infty$, as they also do in several of their subsequent studies, we focus on the case with a fixed s. The model in Equation 3 can be represented as a static factor model with $r = q(s+1)$ static factors,

$$x_{it} = \boldsymbol{\lambda}'_i\mathbf{F}_t + e_{it},$$

where

$$\boldsymbol{\lambda}_i = \begin{bmatrix} \boldsymbol{\lambda}_{i0} \\ \boldsymbol{\lambda}_{i1} \\ \vdots \\ \boldsymbol{\lambda}_{is} \end{bmatrix} \text{ and } \mathbf{F}_t = \begin{bmatrix} \mathbf{f}_t \\ \mathbf{f}_{t-1} \\ \vdots \\ \mathbf{f}_{t-s} \end{bmatrix}.$$

We refer to $\mathbf{f}_t$ as the dynamic factors and $\mathbf{F}_t$ as the static factors. Regarding the dynamic process for $\mathbf{f}_t$, we may use a finite-order VAR to approximate its dynamics. For example, $\mathbf{f}_t$ can follow VAR(h),

$$\boldsymbol{\Phi}(L)\mathbf{f}_t = \boldsymbol{\varepsilon}_t, \tag{4}$$

where $\boldsymbol{\Phi}(L) = \mathbf{I}_q - \boldsymbol{\Phi}_1 L - \cdots - \Phi_h L^h$. Then we may form the VAR(k) representation of the static factor $\mathbf{F}_t$, where $k = \max\{h, s+1\}$,

$$\boldsymbol{\Phi}_F(L)\mathbf{F}_t = \mathbf{u}_t,$$
$$\mathbf{u}_t = \mathbf{R}\boldsymbol{\varepsilon}_t,$$

with $\mathbf{\Phi}_F(L) = \mathbf{I}_{q(s+1)} - \mathbf{\Phi}_{F,1}L - \cdots - \Phi_{F,k}L^k$, and the $q(s+1) \times q$ matrix $\mathbf{R}$ is given by $\mathbf{R} = [\mathbf{I}_q, 0, \ldots, 0]'$. When $h = 1$, the construction of $\mathbf{\Phi}_{F,j}$ is trivial. When $h > 1$, one obtains

$$\mathbf{\Phi}_{F,1} = \begin{bmatrix} \mathbf{\Phi}_1 & 0 & \cdots & 0 \\ I_q & 0 & \cdots & 0 \\ \vdots & \ddots & \cdots & \vdots \\ 0 & \cdots & I_q & 0 \end{bmatrix}, \quad \mathbf{\Phi}_{F,2} = \begin{bmatrix} 0 & \mathbf{\Phi}_2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

For $1 < j \leq h$, $\mathbf{\Phi}_{F,j}$ can be constructed similarly to $\mathbf{\Phi}_{F,2}$. For $j > h$, one finds that $\mathbf{\Phi}_{F,j} = 0$.

We may see from the VAR representation that the spectrum of the static factors has rank q instead of $r = q(s+1)$. Given that $\mathbf{\Phi}_F(L)\mathbf{F}_t = \mathbf{R}\boldsymbol{\varepsilon}_t$, the spectrum of F at frequency ω is

$$\mathbf{S}_F(\omega) = \mathbf{\Phi}_F(e^{-i\omega})^{-1}\mathbf{R}\mathbf{S}_\varepsilon(\omega)\mathbf{R}'\mathbf{\Phi}_F(e^{i\omega})^{-1},$$

whose rank is q if $\mathbf{S}_\varepsilon(\omega)$ has rank q for $|\omega| \leq \pi$. This implies that $\mathbf{S}_F(\omega)$ has only q nonzero eigenvalues. Bai & Ng (2007) refer to q as the number of primitive shocks. Hallin & Liska (2007) estimate the rank of this matrix to determine the number of dynamic factors. Onatski (2009) also considers estimating q using the sample estimates of $\mathbf{S}_F(\omega)$.

Alternatively, we may first estimate a static factor model following Bai & Ng (2002) to obtain $\hat{\mathbf{F}}_t$. Next, we may estimate a VAR(p) for $\hat{\mathbf{F}}_t$ to obtain the residuals $\hat{u}_t$. Let $\hat{\mathbf{\Sigma}}_u = \frac{1}{T}\sum_{t=1}^T \hat{\mathbf{u}}_t\hat{\mathbf{u}}_t'$, which is positive semidefinite. Note that the theoretical moments $E(\mathbf{u}_t\mathbf{u}'_t)$ have rank q. We expect that we may estimate q using the information about the rank of $\hat{\mathbf{\Sigma}}_u$. The formal procedure is provided by Bai & Ng (2007).

Using a different approach, Stock & Watson (2005) consider the case of serially correlated measurement errors and transform the model such that the residual has a static factor representation with q factors. Bai & Ng's (2002) information criteria can then be directly applied to estimate q.

Amengual & Watson (2007) consider a similar transformation as Stock & Watson (2005) and derive the corresponding econometric theory for estimating q. They start from the static factor model in Equation 2, $\mathbf{X}_t = \mathbf{\Lambda}\mathbf{F}_t + \mathbf{e}_t$, and consider a VAR($p$) for $\mathbf{F}_t$,

$$\mathbf{F}_t = \sum_{i=1}^p \mathbf{\Phi}_i\mathbf{F}_{t-i} + \boldsymbol{\varepsilon}_t,$$
$$\boldsymbol{\varepsilon}_t = \mathbf{G}\boldsymbol{\eta}_t,$$

where G is $r \times q$ with full column rank, and η_t is a sequence of shocks with mean zero and variance I_q. The shock η_t is called the dynamic factor shock, whose dimension is called the number of dynamic factors. Let $\mathbf{Y}_t = \mathbf{X}_t - \sum_{i=1}^{p}\mathbf{\Lambda}\mathbf{\Phi}_i\mathbf{F}_{t-i}$ and $\mathbf{\Gamma} = \mathbf{\Lambda}\mathbf{G}$; then $\mathbf{Y}_t$ has a static factor representation with q factors,

$$\mathbf{Y}_t = \mathbf{\Gamma}\boldsymbol{\eta}_t + \mathbf{e}_t.$$

If $\mathbf{Y}_t$ is observed, q can be directly estimated using Bai & Ng's (2002) information criteria. In practice, $\mathbf{Y}_t$ needs to be estimated. Let $\hat{\mathbf{Y}}_t = \mathbf{X}_t - \sum_{i=1}^p \hat{\mathbf{\Lambda}}\hat{\mathbf{\Phi}}_i\hat{\mathbf{F}}_{t-i}$, where $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{F}}_t$ are principal components estimators from $\mathbf{X}_t$, and $\hat{\mathbf{\Phi}}_i$ is obtained by VAR(p) regression of $\hat{\mathbf{F}}_t$. Amengual & Watson (2007) show that Bai & Ng's (2002) information criteria, when applied to $\hat{\mathbf{Y}}_t$, can consistently estimate the number of dynamic factors q.

5. FACTOR-AUGMENTED REGRESSIONS

A popular application of the large factor model is factor-augmented regressions. For example, Stock & Watson (1999, 2002a) add a single factor to standard univariate autoregressive models and find that they provide the most accurate forecasts of macroeconomic time series such as inflation and industrial production among a large set of models. Bai & Ng (2006a) develop econometric theory for such factor-augmented regressions so that inference can be conducted.

Consider the following forecasting model for y_t:

$$y_{t+h} = \boldsymbol{\alpha}'\mathbf{F}_t + \boldsymbol{\beta}'\mathbf{W}_t + \varepsilon_{t+h}, \tag{5}$$

where $\mathbf{W}_t$ is the vector of a small number of observables including lags of y_t, and $\mathbf{F}_t$ is unobservable. Suppose there is a large number of series x_{it}, $i = 1, \ldots, N$, $t = 1, \ldots, T$, which has a large factor representation as Equation 1,

$$x_{it} = \boldsymbol{\lambda}_i'\mathbf{F}_t + e_{it}.$$

When y_t is a scalar, Equations 1 and 5 become the diffusion index forecasting model of Stock & Watson (2002b, 2006). Clearly, each x_{it} is a noisy predictor for y_{t+h}. Because $\mathbf{F}_t$ is latent, the conventional mean-squared optimal prediction of y_{t+h} is not feasible. Alternatively, consider estimating Equation 1 by the method of principal components to obtain $\hat{\mathbf{F}}_t$, which is a consistent estimator for $\mathbf{H}'\mathbf{F}_t$ for some rotation matrix $\mathbf{H}$, and then regress y_{t+h} on $\hat{\mathbf{F}}_t$ and $\mathbf{W}_t$ to obtain $\hat{\alpha}$ and $\hat{\beta}$. The feasible prediction for $y_{T+h|T} \equiv E(y_{T+h}|\Omega_T)$, where $\boldsymbol{\Omega}_T = [\mathbf{F}_T, \mathbf{W}_T, \mathbf{F}_{T-1}, \mathbf{W}_{T-1}, \ldots]$, is given by

$$\hat{y}_{T+h|T} = \hat{\boldsymbol{\alpha}}'\hat{\mathbf{F}}_T + \hat{\boldsymbol{\beta}}\mathbf{W}_T.$$

Let $\boldsymbol{\delta} \equiv (\boldsymbol{\alpha}'\mathbf{H}^{-1}, \boldsymbol{\beta}')'$ and $\varepsilon_{T+h|T} \equiv y_{T+h} - y_{T+h|T}$. Bai & Ng (2006a) show that when N is large relative to T (i.e., $\sqrt{T}/N \to 0$), $\hat{\boldsymbol{\delta}}$ will be $\sqrt{T}$-consistent and asymptotically normal. $\hat{y}_{T+h|T}$ and $\hat{\varepsilon}_{T+h|T}$ are $\min\{N^{1/2}, T^{1/2}\}$-consistent and asymptotically normal. For all cases, inference needs to take into account the estimated factors, except for the special case $T/N \to 0$. In particular, under standard assumptions for the large approximate factor model as in Bai & Ng (2002), when $\sqrt{T}/N \to 0$, we have

$$\hat{\boldsymbol{\delta}} - \boldsymbol{\delta} \xrightarrow{d} N(0, \boldsymbol{\Sigma}_\delta).$$

Let $\mathbf{z}_t = [\mathbf{F}'_t, \mathbf{W}'_t]'$, $\hat{\mathbf{z}}_t = [\hat{\mathbf{F}}_t', \mathbf{W}_t']'$, and $\hat{\varepsilon}_{t+h} = y_{t+h} - \hat{y}_{t+h|t}$. Then a heteroskedasticity-consistent estimator for $\boldsymbol{\Sigma}_\delta$ is given by

$$\hat{\boldsymbol{\Sigma}}_\delta = \left(\frac{1}{T}\sum_{t=1}^{T-h} \hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T-h} \hat{\varepsilon}_{t+h}^2\hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right)\left(\frac{1}{T}\sum_{t=1}^{T-h} \hat{\mathbf{z}}_t\hat{\mathbf{z}}_t'\right)^{-1}.$$

The key requirement of the theory is $\sqrt{T}/N \to 0$, which puts discipline on when estimated factors can be applied in the diffusion index forecasting models as well as the FAVAR, which we discuss in the next section.

If in addition, we assume $\sqrt{N}/T \to 0$, then

$$\frac{\hat{y}_{T+h|T} - y_{T+h|T}}{\sqrt{\operatorname{var}\left(\hat{y}_{T+h|T}\right)}} \xrightarrow{d} N(0, 1),$$

where

$$\operatorname{var}\left(\hat{y}_{T+h|T}\right) = \frac{1}{T}\mathbf{z}_T' \, A\mathrm{var}\left(\hat{\boldsymbol{\delta}}\right)\hat{\mathbf{z}}_T + \frac{1}{N}\hat{\boldsymbol{\alpha}}' A\mathrm{var}\left(\hat{\mathbf{F}}_T\right)\hat{\boldsymbol{\alpha}}.$$

An estimator for $A\mathrm{var}\,(\hat{\mathbf{F}}_T)$ is provided by Bai (2003). A notable feature of the limiting distribution of the forecast is that the overall convergence rate is given by $\min\{N^{1/2}, T^{1/2}\}$. Given that $\hat{\varepsilon}_{T+h} = \hat{y}_{T+h|T} - y_{T+h} = \hat{y}_{T+h|T} - y_{T+h|T} + \varepsilon_{T+h}$, if we further assume that ε_t is normal with variance σ_ε^2, then the forecasting error also becomes approximately normal:

$$\hat{\varepsilon}_{T+h} \sim N\left(0, \sigma_\varepsilon^2 + \operatorname{var}\left(\hat{y}_{T+h|T}\right)\right),$$

so confidence intervals can be constructed for the forecasts. Bai & Ng (2006b) developed methods for testing whether an observed macroeconomic or financial variable coincides with one of the underlying latent factors.

6. FACTOR-AUGMENTED VECTOR AUTOREGRESSION

VAR models have been widely applied in macroeconomic analysis. A central question regarding their use is how to identify structural shocks, which in turn depend on what variables are included in the VAR system. A small VAR usually cannot fully capture the structural shocks. In the meantime, including more variables in the VAR system could be problematic due to either the degree-of-freedom problem or the variable-selection problem. It has been a challenging job to determine which variables should be included in the system. There are several ways to overcome such difficulties, such as Bayesian VAR, initially considered by Doan et al. (1984), Litterman (1986), and Sims (1993), and global VAR, considered by Pesaran et al. (2004). This section focuses on another popular solution, FAVAR, originally proposed by Bernanke et al. (2005). FAVAR assumes that a large number of economic variables are driven by a small VAR, which can include both latent and observed variables. The dimension of structural shocks can be estimated instead of being assumed to be known and fixed.

Consider the case in which both the unobserved factors $\mathbf{F}_t$ and the observed factors $\mathbf{W}_t$ affect a large number of observed variables x_{it},

$$x_{it} = \boldsymbol{\lambda}_i'\mathbf{F}_t + \boldsymbol{\gamma}_t'\mathbf{W}_t + e_{it}, \tag{6}$$

and in which the vector $\mathbf{H}_t = \left[\mathbf{F}_t', \mathbf{W}_t'\right]'$ follows a VAR of finite order,

$$\boldsymbol{\Phi}(L)\mathbf{H}_t = \mathbf{u}_t,$$

where $\boldsymbol{\Phi}(L) = \boldsymbol{\Phi}_0 - \sum_{j=1}^{h} \boldsymbol{\Phi}_j L^h$, with $\boldsymbol{\Phi}_0$ being possibly not an identity matrix. Bernanke et al. (2005) propose two ways to analyze FAVAR. The first is based on a two-step principal components method, in which in the first step, the method of principal components is employed to form estimates of the space spanned by both $\mathbf{F}_t$ and $\mathbf{W}_t$. In the second step, various identification schemes, such as Cholesky ordering, can be applied to obtain estimates of latent factors $\hat{\mathbf{F}}_t$, which are treated as observed when conducting VAR analysis of $[\hat{\mathbf{F}}_t', \mathbf{W}_t']'$. Under suitable identification conditions, Bai et al. (2016) show that inferential theory can be developed for such a two-step estimator, which differs from a standard large factor model. Confidence bands for the impulse responses can be readily constructed using the theory therein. The second method involves a one-step likelihood approach, implemented by Gibbs sampling, which leads to joint estimation of both the latent factors and impulse responses. The two methods can complement each other, with the first one being computationally simple and the second providing possibly better inference in finite samples although with increased computational cost.

A useful feature of FAVAR is that one can readily calculate the impulse response function of all variables to the fundamental shocks. For example, the impulse response of the observable $x_{i,t+h}$ with respect to the structural shock u_t is

$$\frac{\partial x_{i,t+h}}{\partial \mathbf{u}_t} = \left(\boldsymbol{\lambda}_t', \boldsymbol{\gamma}_i'\right)\mathbf{C}_h, \tag{7}$$

where $\mathbf{C}_h$ is the coefficient matrix for $\mathbf{u}_{t-h}$ in the vector moving average representation of $\mathbf{H}_t$,

$$\mathbf{H}_t = \boldsymbol{\Phi}(L)^{-1}\mathbf{u}_t = \mathbf{C}_0\mathbf{u}_t + \mathbf{C}_1\mathbf{u}_{t-1} + \mathbf{C}_2\mathbf{u}_{t-2} + \cdots.$$

The theory of estimation and inference for Equation 7 is provided by Bai et al. (2016). Forni et al. (2009) explore the structural implications of the factors and develop the corresponding econometric theory. Stock & Watson (2010) survey the application of dynamic factor models.

7. IV ESTIMATION WITH MANY INSTRUMENTAL VARIABLES

The IV method is fundamental to econometrics practice. It is useful when one or more explanatory variables are correlated with the error terms in a regression model, known as endogenous regressors. In this case, standard methods such as OLS are inconsistent. With the availability of IV, which are correlated with regressors but uncorrelated with errors, consistent estimation is achievable. Consider a standard setup

$$y_t = \mathbf{x}'_t \boldsymbol{\beta} + u_t, \quad t = 1, 2, \ldots, T, \tag{8}$$

where $\mathbf{x}_t$ is correlated with the error term u_t. Suppose there are N IV labeled as z_{it} for $i = 1, 2, \ldots, N$. Consider the two-stage least squares (2SLS), a special IV method. In the first stage, the endogenous regressor x_t is regressed on the IV

$$\mathbf{x}_t = \mathbf{c}_0 + \mathbf{c}_1 z_{1t} + \cdots + \mathbf{c}_N z_{Nt} + \eta_t, \tag{9}$$

the fitted value $\hat{\mathbf{x}}_t$ is used as the regressor in the second stage, and the resulting estimator is consistent for $\boldsymbol{\beta}$ for a small N. It is known that for large N, 2SLS can be severely biased. In fact, if $N \geq T$, then $\hat{\mathbf{x}}_t \equiv \mathbf{x}_t$, and 2SLS coincides with OLS, which is inconsistent. The problem lies in the overfitting in the first-stage regression. The theory of many-instruments bias has been extensively studied in the econometric literature (e.g., Hansen et al. 2008, Hausman et al. 2010). Within the GMM context, inaccurate estimation of a high-dimensional optimal weighting matrix is not the cause for many-moments bias. Bai & Ng (2010) show that even if the true optimal weighting matrix is used, inconsistency is still obtained under a large number of moments. In fact, with many moments, a sparse weighting matrix, such as an identity matrix, will give consistent estimation, as shown by Meng et al. (2011).

One solution to the many-IV problem is to assume that many of the coefficients in Equation 9 are zero (sparse) so that a regularization method such as LASSO (least absolute shrinkage and selection operator) can be used in the first-stage regression. Penalization prevents overfitting and picks up the relevant instruments (nonzero coefficients). These methods are considered by Ng & Bai (2009) and Belloni et al. (2012). In fact, any machine learning method that prevents in-sample overfitting in the first-stage regression will work.

The principal components method is an alternative solution and can be more advantageous than the regularization method (Bai & Ng 2010, Kapetanios & Marcellino 2010). It is well known that the principal components method is that of dimension reduction. The principal components are linear combinations of $z_{1t}, \ldots, z_{Nt}$. The high dimension of the IV can be reduced into a smaller dimension via the principal components method. If $z_{1t}, \ldots, z_{NT}$ are valid instruments, then any linear combination is also a valid IV, and so are the principal components. Interestingly, the principal components method does not require all the z's to be valid IV to begin with. Suppose that the regressors $\mathbf{x}_t$ and $z_{1t}, \ldots, z_{Nt}$ are driven by some common factors $\mathbf{F}_t$ such that

$$\mathbf{x}_t = \boldsymbol{\phi}'\mathbf{F}_t + \mathbf{e}_{xt},$$
$$z_{it} = \boldsymbol{\lambda}_i'\mathbf{F}_t + e_{it}.$$

Provided that the common shocks $\mathbf{F}_t$ are uncorrelated with errors u_t in Equation 8, then $\mathbf{F}_t$ is a valid IV because $\mathbf{F}_t$ also drives x_t. Although $\mathbf{F}_t$ is unobservable, it can be estimated from $\{z_{it}\}$ via the principal components. Even though e_{it} can be correlated with u_t for some i, so that not all z_{it}'s are valid IV, the principal components are valid IV. This is the advantage of the principal components method. The example provided by Meng et al. (2011) is instructive. Consider estimating the beta of an asset with respect to the market portfolio, which is unobservable. The market index, as a proxy, is measured with errors. But other assets' returns can be used as IV because all assets' returns are linked with the market portfolio. Thus, there exists a large number of instruments. Each individual asset can be a weak IV because the idiosyncratic returns can be large. But the principal components method will wash out the idiosyncratic errors, giving rise to a more effective IV.

The preceding setup can be extended into panel data with endogenous regressors:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}, \quad t = 1, 2, \ldots, T; \quad i = 1, 2, \ldots, N,$$

where x_{it} are correlated with the errors u_{it}. Suppose that x_{it} are driven by the common shocks $\mathbf{F}_t$ such that

$$x_{it} = \boldsymbol{\gamma}_i'\mathbf{F}_t + e_{it}.$$

Provided that the common components $c_{it} = \boldsymbol{\gamma}_i'\mathbf{F}_t$ are uncorrelated with u_{it}, outside IV are not needed. We can extract the common components c_{it} via the principal components estimation of γ_i and F_t to form $\hat{c}_{it} = \hat{\boldsymbol{\gamma}}_i'\hat{\mathbf{F}}_t$, and use $\hat{c}_{it}$ as IV. This method is considered by Bai & Ng (2010).

8. STRUCTURAL CHANGES IN LARGE FACTOR MODELS

In many economic applications, researchers have to be cautious about the potential structural changes in high-dimensional data sets. Parameter instability has been a pervasive phenomenon in time-series data (Stock & Watson 1996). Such instability could result from technological changes, preference shifts of consumers, and policy regime switching. Banerjee & Marcellino (2008) and Yamamoto (2016) provide simulations and empirical evidence that the forecasts based on estimated factors will be less accurate if the structural break in the factor loading matrix is ignored. In addition, the evolution of an economy might introduce new factors, while conventional factors might phase out. Econometric analysis of the structural change in large factor models is challenging because the factors are unobserved and factor loadings have to be estimated. Structural change can happen to the factor loadings or the dynamic process of factors, or both. Most of the theoretical challenge comes from the break in factor loadings, given that factors can be consistently estimated by principal components even if their dynamic process is subject to a break, whereas factor loadings are time invariant. In this section, we focus on the breaks in factor loadings. Consider a time-varying version of Equation 2,

$$\mathbf{X}_t = \boldsymbol{\Lambda}_t\mathbf{F}_t + \mathbf{e}_t,$$

where the time-varying loading matrix $\boldsymbol{\Lambda}_t$ might assume different forms. Bates et al. (2013) consider the following dynamic equation:

$$\boldsymbol{\Lambda}_t = \boldsymbol{\Lambda}_0 + h_{NT}\boldsymbol{\xi}_t,$$

where h_{NT} is a deterministic scalar that depends on (N, T), and $\boldsymbol{\xi}_t$ is an $N \times r$ stochastic process. Three examples are considered for ξ_t: white noise, random walk, and single break. Bates et al. then establish conditions under which the changes in the loading matrix can be ignored in the estimation of factors. Intuitively, the estimation and inference of the factor estimates are not affected if the size of the break is small enough. If the size of the break is large, however, the principal components factor estimators will be inconsistent.

In recent years, we have witnessed fast development in this area. Tests for structural changes in factor loadings of a specific variable are derived by Stock & Watson (2008), Breitung & Eickmeier (2011), and Yamamoto & Tanaka (2015). Chen et al. (2014) and Han & Inoue (2015) study tests for structural changes in the overall factor loading matrix. Corradi & Swanson (2014) construct joint tests for breaks in factor loadings and coefficients in factor-augmented regressions. Cheng et al. (2016) consider the determination of break date and introduce the shrinkage estimation method for factor models in the presence of structural changes. Current studies have not considered the case in which break dates are possibly heterogeneous across variables and the number of break dates might increase with the sample size. It would be interesting to study the properties of the principal components estimators and the power of the structural change tests under such scenarios.

Another branch of methods considers Markov regime switching in factor loadings (Kim & Nelson 1999). The likelihood function can be constructed using various filters. Then one may use either the maximum likelihood or Bayesian method to estimate factor loadings in different regimes, the regime probabilities, and the latent factors. Del Negro & Otrok (2008) develop a dynamic factor model with time-varying factor loadings and stochastic volatility in both the latent factors and idiosyncratic components. A Bayesian algorithm is developed to estimate the model, which is employed to study the evolution of international business cycles. The theoretical properties of such models remain to be studied under the large N–large T setup.

9. PANEL DATA MODELS WITH INTERACTIVE FIXED EFFECTS

There has been growing interest in panel data models with interactive fixed effects. Conventional methods assume additive individual fixed effects and time fixed effects. The interactive fixed effects allow possible multiplicative effects. Such a methodology has important theoretical and empirical relevance. Consider the following large N–large T panel data model

$$\begin{aligned} y_{it} &= \mathbf{X}_{it}'\boldsymbol{\beta} + u_{it}, \\ u_{it} &= \boldsymbol{\lambda}_i'\mathbf{F}_t + \varepsilon_{it}. \end{aligned} \tag{10}$$

We observe y_{it} and X_{it} but do not observe $\boldsymbol{\lambda}_i$, $\mathbf{F}_t$, and ε_{it}. The coefficient of interest is $\boldsymbol{\beta}$. Such models nest conventional fixed effects panel data models as special cases due to the following simple transformation:

$$\begin{aligned} y_{it} &= \mathbf{X}_{it}'\boldsymbol{\beta} + \alpha_i + \xi_t + \varepsilon_{it} \\ &= \mathbf{X}_{it}'\boldsymbol{\beta} + \boldsymbol{\lambda}_i'\mathbf{F}_t + \varepsilon_{it}, \end{aligned}$$

where $\boldsymbol{\lambda}_i = [1, \alpha_i]'$ and $\mathbf{F}_t = [\xi_t, 1]'$. In general, the interactive fixed effects allow a much richer form of unobserved heterogeneity. For example, $\mathbf{F}_t$ can represent a vector of macroeconomic common shocks, and $\boldsymbol{\lambda}_i$ captures individual i's heterogeneous response to such shocks.

The theoretic framework of Bai (2009) allows X_{it} to be correlated with $\boldsymbol{\lambda}_i$, $\mathbf{F}_t$, or both. Under the framework of large N and large T, we may estimate the model by minimizing a least squares objective function

$$\begin{aligned} \mathrm{SSR}(\boldsymbol{\beta}, \mathbf{F}, \boldsymbol{\Lambda}) &= \sum_{i=1}^{N} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{F}\boldsymbol{\lambda}_i)'(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{F}\boldsymbol{\lambda}_i) \\ \text{s.t.} \quad & \mathbf{F}'\mathbf{F}/T = \mathbf{I}_r, \ \boldsymbol{\Lambda}'\boldsymbol{\Lambda} \text{ is diagonal.} \end{aligned}$$

Although no closed-form solution is available, the estimators can be obtained by iterations. To begin, consider some initial values $\boldsymbol{\beta}^{(0)}$, such as least squares estimators from regressing $\mathbf{Y}_i$ on $\mathbf{X}_i$. Then perform principal components analysis for the pseudo-data $\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(0)}$ to obtain $\mathbf{F}^{(1)}$ and $\boldsymbol{\Lambda}^{(1)}$. Next, regress $\mathbf{Y}_i - \mathbf{F}^{(1)}\boldsymbol{\lambda}_i^{(1)}$ on $\mathbf{X}_i$ to obtain $\boldsymbol{\beta}^{(1)}$. Iterate such steps until convergence is achieved. Bai (2009) show that the resulting estimator $\hat{\boldsymbol{\beta}}$ is $\sqrt{NT}$-consistent. Given such results, the limiting distributions for $\hat{\mathbf{F}}$ and $\hat{\boldsymbol{\Lambda}}$ are the same as in Bai (2003) due to their slower convergence rates. The limiting distribution for $\hat{\boldsymbol{\beta}}$ depends on specific assumptions on the error term ε_{it} as well as on the ratio T/N. If $T/N \to 0$, then the limiting distribution of $\hat{\boldsymbol{\beta}}$ will be centered around zero, given that $E(\varepsilon_{it}\varepsilon_{js}) = 0$ for $t \neq s$, and $E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij}$ for all i, j, t.

Conversely if N and T are comparable such that $T/N \to \rho > 0$, then the limiting distribution will not be centered around zero, which poses a challenge for inference. Bai (2009) provides a bias-corrected estimator for β, whose limiting distribution is centered around zero. In particular, the bias-corrected estimator allows for heteroskedasticity across both N and T. Let $\hat{\boldsymbol{\beta}}$ be the bias-corrected estimator, and assume that $T/N^2 \to 0$ and $N/T^2 \to 0$, $E(\varepsilon_{it}^2) = \sigma_{it}^2$, and $E(\varepsilon_{it}\varepsilon_{js}) = 0$ for $i \neq j$ and $t \neq s$. Then, one obtains

$$\sqrt{NT}\left(\hat{\hat{\boldsymbol{\beta}}} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left(0, \boldsymbol{\Sigma}_{\beta}\right),$$

where a consistent estimator for Σ_{β} is also available in Bai (2009).

Ahn et al. (2001, 2013) study the model in Equation 10 under large N but fixed T. They employ the GMM method, which applies moments of zero correlation and homoskedasticity. Moon & Weidner (2016) consider the same model as in Equation 10 with lagged dependent variables as regressors. They devise a quadratic approximation of the profile objective function to show the asymptotic theory for the least square estimators and various test statistics. Moon & Weidner (2015) further extend their study by allowing an unknown number of factors. They show that the limiting distribution of the least square estimator is not affected by the number of factors used in the estimation, as long as this number is no smaller than the true number of factors. Lu & Su (2016) propose the method of adaptive group LASSO, which can simultaneously select the proper regressors and determine the number of factors.

Pesaran (2006) considers a slightly different setup with individual-specific slopes,

$$\begin{aligned} y_{it} &= \boldsymbol{\alpha}_t' \mathbf{d}_t + \mathbf{X}_{it}' \boldsymbol{\beta}_i + u_{it}, \\ u_{it} &= \boldsymbol{\lambda}_t' \mathbf{F}_t + \varepsilon_{it}, \end{aligned} \tag{11}$$

where $\mathbf{d}_t$ is observed common effects such as seasonal dummies. The unobserved factors and the individual-specific errors are allowed to follow arbitrary stationary processes. Instead of estimating the factors and factor loadings, Pesaran (2006) considers an auxiliary OLS regression. The proposed common correlated effects (CCE) estimator can be obtained by augmenting the model with additional regressors, which are the cross-sectional averages of the dependent and independent variables, in an attempt to control for the common factors. Define $\mathbf{z}_{it} = [y_{it}, \mathbf{X}_{it}']'$ as the collection of individual-specific observations. Consider weighted average of $\mathbf{z}_{it}$ as

$$\bar{\mathbf{z}}_{\omega t} = \sum_{i=1}^{N} \omega_i \mathbf{z}_{it},$$

with the weights that satisfy some very general conditions. For example, we may choose $\omega_i = 1/N$. Pesaran (2006) shows that the individual slope $\boldsymbol{\beta}_i$ can be consistently estimated through the following OLS regression of y_{it} on $\mathbf{d}_t$, $\mathbf{X}_{it}$, and $\bar{\mathbf{z}}_{\omega t}$ under the large N–large T framework,

$$y_{it} = \boldsymbol{\alpha}_i' \mathbf{d}_t + \mathbf{X}_{it}' \boldsymbol{\beta}_i + \bar{\mathbf{z}}_{\omega t}' \boldsymbol{\gamma}_i + \varepsilon_{it}. \tag{12}$$

If the coefficients $\boldsymbol{\beta}_i$ are assumed to follow a random coefficient model, then under more general conditions the mean $\boldsymbol{\beta} = E(\boldsymbol{\beta}_i)$ can be consistently estimated by a pooled regression of y_{it} on $\mathbf{d}_t$, $\mathbf{X}_{it}$, and $\bar{\mathbf{z}}_{\omega t}$ as long as N tends to infinity, regardless of large or small T,

$$y_{it} = \boldsymbol{\alpha}_i' \mathbf{d}_t + \mathbf{X}_{it}' \beta + \bar{\mathbf{z}}_{\omega t}' \boldsymbol{\gamma} + v_{it}. \tag{13}$$

Alternatively, the mean $\boldsymbol{\beta}$ can also be consistently estimated by taking the simple average of individual estimators from Equation 12, $\hat{\boldsymbol{\beta}}_{MG} = \frac{1}{N}\sum_{i=1}^{N} \hat{\boldsymbol{\beta}}_i^{CCE}$. Asymptotic normality for such CCE estimators can also be established and help conduct inference. CCE is easy to compute as an outcome of OLS, and no iteration is needed. With certain rank condition, desirable finite sample properties of CCE are also demonstrated.

Ando & Bai (2015b) also study heterogeneous panel models with interactive effects, in which the number of regressors can be large and the regularization method is used to select relevant regressors. Su & Chen (2013) and Ando & Bai (2015a) provide a formal test for homogenous coefficients in the model in Equation 11.

MLE of the model in Equation 10 is studied by Bai & Li (2014). They consider the case in which X_{it} also follows a factor structure and is jointly modeled. Heteroskedasticity is explicitly estimated jointly with other model parameters.

Chudik & Pesaran (2015) and Chudik et al. (2016) extend the CCE approach of Pesaran (2006) to a dynamic heterogeneous panel setup. A similar model setup is also studied by Song (2013), who examines the properties of Bai's (2009) estimator in the presence of lagged dependent variables as well as heterogeneous coefficients.

Another extension is investigated by Chudik et al. (2011), who introduces the notion of weak, semiweak, semistrong, and strong common factors. Such factors may be used to represent very general forms of cross-sectional dependence that are not captured by Pesaran (2006) or Bai (2009). Weak and strong factors are discussed in Section 4. The semistrong and semiweak factors can be defined in a similar way. Let α be a constant between zero and 1. Consider

$$\lim_{N\to\infty} \frac{1}{N^{\alpha}} \sum_{i=1}^{N} |\gamma_{il}| = K < \infty.$$

If the condition holds for $0 < \alpha < 1/2$ for some $1 \le l \le r$, then the factor f_{lt} is said to be semiweak. The semistrong, strong, and weak factors correspond to the cases of $1/2 \le \alpha < 1$, $\alpha = 1$, and $\alpha = 0$, respectively. The error term $u_{it} = \boldsymbol{\lambda}'_i \mathbf{F}_t + \varepsilon_{it}$ in the panel data model can include all four types of factors. Chudik et al. (2011) allow the number of factors r to increase with the sample size N. They prove that u_{it} is cross-sectionally weakly dependent if all factors are weak, semiweak, or semistrong and that u_{it} is cross-sectionally strongly dependent if there is at least one strong factor. In all cases, they show that the CCE mean group estimators and the CCE pooled estimator of Pesaran (2006) remain consistent. In Monte Carlo simulations in which the errors are subject to a finite number of unobserved strong factors and an infinite number of weak or semistrong factors, the CCE-type estimators showed little size distortions as compared to alternative estimators.

10. NONSTATIONARY PANEL

This section discusses other important topics, including the application of large factor models in nonstationary panel data, estimation, and inference of dynamic factor models.

10.1. Unit Roots and Cointegration

For nonstationary analysis, the large factor model brings new perspective for tests of unit roots. Consider the following data-generating process for x_{it}:

$$\begin{aligned} x_{it} &= c_i + \beta_i t + \boldsymbol{\lambda}_i' \mathbf{F}_t + e_{it}, \\ (1-L)\mathbf{F}_t &= \mathbf{C}(L)\mathbf{u}_t, \\ (1-\rho_i L)e_{it} &= D_i(L)\epsilon_{it}, \end{aligned} \tag{14}$$

where $\mathbf{C}(L)$ and $D_i(L)$ are polynomials of lag operators. The $r \times 1$ factor $\mathbf{F}_t$ has r_0 stationary factors and r_1 $I(1)$ components or common trends ($r = r_0 + r_1$). The idiosyncratic errors e_{it} could be either $I(1)$ or $I(0)$, depending on whether $\rho_i = 1$ or $|\rho_i| < 1$. PANIC by Bai & Ng (2004) develops an econometric theory for determining r_1 and testing $\rho_i = 1$ when neither $\mathbf{F}_t$ nor e_{it} is observed. The model in Equation 14 has many important features that are both of theoretical interests and of empirical relevance. For example, let x_{it} denote the real output for country i. Then x_{it} may be determined by the global common trend F_{1t}, the global cyclical component F_{2t}, and an idiosyncratic component e_{it} that could be either $I(0)$ or $I(1)$. PANIC provides a framework for the estimation and statistical inference for such components, which are all unobserved. Although conventional nonstationarity analysis looks at the unit root in x_{it} only, PANIC further explores whether possible unit roots are coming from common factors, idiosyncratic components, or both. Another very important feature of PANIC is that it allows weak cross-sectional correlation in idiosyncratic errors.

The initial steps of PANIC include transformations of the data so that the deterministic trend part is removed, and then it proceeds with principal components analysis. In the case of no linear trend ($\beta_i = 0$ for all i), a simple first differencing will suffice. We proceed with the example with a linear trend ($\beta_i \neq 0$ for some i). Let $\mathbf{\Delta} x_{it} = x_{it} - x_{i,t-1}$, $\overline{\mathbf{\Delta} x_i} = \frac{1}{T-1}\sum_{t=2}^{T}\mathbf{\Delta} x_{it}$, $\overline{\mathbf{\Delta} e_i} = \frac{1}{T-1}\sum_{t=2}^{T}\mathbf{\Delta} e_{it}$, and $\overline{\mathbf{\Delta F}} = \frac{1}{T-1}\sum_{t=2}^{T}\mathbf{\Delta F}_t$. After first differencing and removing the time average, one finds that the model in Equation 14 becomes

$$\mathbf{\Delta} x_{it} - \overline{\mathbf{\Delta} x_i} = \boldsymbol{\lambda}_i'\mathbf{f}_t + \mathbf{\Delta} e_{it} - \overline{\mathbf{\Delta} e_i}, \tag{15}$$

where $\mathbf{f}_t = \mathbf{\Delta F}_t - \overline{\mathbf{\Delta F}}$. Let $\hat{\boldsymbol{\lambda}}_i$ and $\hat{\mathbf{f}}_t$ be the principal components estimators of Equation 15. Let $\hat{z}_{it} = \mathbf{\Delta} x_{it} - \overline{\mathbf{\Delta} x_i} - \hat{\boldsymbol{\lambda}}_i'\hat{\mathbf{f}}_t$, $\hat{e}_{it} = \sum_{s=2}^{t}\hat{z}_{it}$, and $\hat{\mathbf{F}}_t = \sum_{s=2}^{t}\hat{\mathbf{f}}_t$. Then test statistics for the unit root in e_{it} and $\mathbf{F}_t$ can be constructed based on the estimates $\hat{e}_{it}$ and $\hat{\mathbf{F}}_t$. Let $\mathrm{ADF}_{\hat{e}}^{\tau}(i)$ denote the augmented Dickey-Fuller test using $\hat{e}_{it}$. The limiting distributions and critical values are provided by Bai & Ng (2004).

A few important properties of PANIC are worth mentioning. To begin, the test on idiosyncratic errors e_{it} can be conducted without knowing whether the factors are $I(1)$ or $I(0)$. Similarly, the test on factors is valid regardless of whether e_{it} is $I(1)$ or $I(0)$. Finally, the test on e_{it} is valid regardless of whether e_{jt} $(j \neq i)$ is $I(1)$ or $I(0)$. In fact, the limiting distribution of $ADF_{\hat{e}}^{\tau}(i)$ does not depend on the common factors. This helps to construct pooled panel unit root tests, which have improved power as compared to univariate unit root tests.

The literature on panel unit root tests has been growing fast. The early test in Quah (1994) requires strong homogeneous cross-sectional properties. Later tests by Levin et al. (2002) and Im et al. (2003) allow for heterogeneous intercepts and slopes but assume cross-sectional independence. Such an assumption is restrictive and tends to overreject the null hypothesis when violated. O'Connell (1998) provides a GLS solution to this problem under fixed N. However, such a solution does not apply when N tends to infinity, especially for the case in which $N > T$. The PANIC method allows strong cross-sectional correlation due to common factors, as widely observed in economic data, as well as weak cross-sectional correlation in e_{it}. In the same time, it allows heterogeneous intercepts and slopes. If we further assume that the idiosyncratic errors e_{it} are independent across i, and consider testing $H_0 : \rho_i = 1$ for all i, against $H_1 : \rho_i < 1$ for some i, a pool test statistic can be readily constructed. Let $p_{\hat{e}}^{\tau}(i)$ be the p value associated with $ADF_{\hat{e}}^{\tau}(i)$. Then, one obtains

$$P_{\hat{e}}^{\tau} = \frac{-2\sum_{i=1}^{N}\log p_{\hat{e}}^{\tau}(i) - 2N}{\sqrt{4N}} \xrightarrow{d} N(0,1).$$

The pooled test of the idiosyncratic errors can also be seen as a panel test of no cointegration, as the null hypothesis that $\rho_i = 1$ for all i holds only if no stationary combination of x_{it} can be formed.

Alternative methods to PANIC include that by Kapetanios et al. (2011), who derive the theoretical properties of the CCE estimator of Pesaran (2006) for panel regression with nonstationary common factors. The model slightly differs from PANIC in the sense that individual slopes can be studied. For example, they consider

$$y_{it} = \boldsymbol{\alpha}_i'\mathbf{d}_t + \mathbf{X}_{it}'\boldsymbol{\beta}_i + \boldsymbol{\lambda}'\mathbf{F}_t + \varepsilon_{it},$$

and the parameter of interest is $\boldsymbol{\beta} = E(\boldsymbol{\beta}_i)$. Another difference is that the individual error ε_{it} is assumed to be stationary. Kapetanios et al. show that the cross-sectional average-based CCE estimator is robust to a wide variety of data-generation processes and is not just restricted to stationary panel regression. Similar to that of Pesaran (2006), the method does not require knowledge of the number of unobserved factors. The only requirement is that the number of unobserved factors remains fixed as the sample size grows. The main results of Pesaran (2006) continue to hold in the case of nonstationary panel data. It has also been shown that the CCE has lower biases than the alternative estimation methods.

Although Kapetanios et al. (2011) do not focus on testing panel unit roots, the main idea of CCE can be applied for such a purpose. For a given individual i, Pesaran (2007) augments the Dickey-Fuller regression of y_{it} with cross-sectional averages $\bar{y}_{t-1} = \frac{1}{N}\sum_{i=1}^{N} y_{i,t-1}$ and $\boldsymbol{\Delta}\bar{y}_{t-1} = \bar{y}_{t-1} - \bar{y}_{t-2}$. Such auxiliary regressors help to account for cross-sectional dependence in error terms. The regression can be further augmented with $\boldsymbol{\Delta} y_{i,t-s}$ and $\boldsymbol{\Delta}\bar{y}_{t-s}$ for $s = 1, 2, \ldots,$ to handle possible serial correlation in the errors. The resulting augmented Dickey-Fuller statistic is referred to as the CADF statistic for individual i. The panel unit root test statistic is then computed as the average, $\mathrm{CADF} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{CADF}_i$. Given correlation among CADF_i, the limiting distribution of CADF is nonnormal. The CADF is shown to have good finite sample performance when only a single common factor is present. However, it shows size distortions in the case of multiple factors.

Pesaran et al. (2013) extend the model of Pesaran (2007) to the case with multiple common factors. They propose a new panel unit root test based on a simple average of cross-sectionally augmented Sargan-Bhargava statistics. The basic idea is similar to the CCE estimator of Pesaran (2006), which exploits information of the unobserved factors shared by all observed time series. Pesaran et al. showed that the limit distribution of the tests is free from nuisance parameters given that the number of factors is no larger than the number of observed cross-sectional averages. The new test has the advantage that it does not require all the factors to be strong, in the sense of Bailey et al. (2016). Monte Carlo simulations show that the proposed tests have the correct size for all combinations of N and T considered, with power rising with N and T.

Bai & Carrion-I-Silvestre (2009) study the problem of unit root testing in the presence of multiple structural changes and common dynamic factors. Structural breaks represent infrequent regime shifts, whereas dynamic factors capture common shocks that drive the comovement of economic time series. They examine the modified Sargan-Bhargava test for unit roots in panel data and propose ways to handle multiple structural changes and dynamic factors. Properties of the modified Sargan-Bhargava test procedure are derived. Bai & Carrion-I-Silvestre (2013) propose statistics to test the null hypothesis of no cointegration in panel data that permit cross-sectional dependence, in which cross-sectional dependence is also captured by common factors. The common factors can be stationary or integrated processes. They allow the nonstationary regressors to be correlated with the unobserved common factors. Both endogenous regressors (correlated with the idiosyncratic errors) and strictly exogenous regressors are considered. The test statistics are shown to have limiting distributions independent of the common factors, making it possible to pool the individual statistics to gain power.

10.2. Estimating Nonstationary Factors

Studies by Bai & Ng (2002) and Bai (2003) assume the errors are $I(0)$. The PANIC method allows the estimation of factors under either $I(1)$ or $I(0)$ errors. Consider the case without linear trend

($\beta_i = 0$ for all i). The factor model after differencing is

$$\Delta x_{it} = \lambda_i' \Delta \mathbf{F}_t + \Delta e_{it}.$$

If e_{it} is $I(1)$, Δe_{it} is $I(0)$. If e_{it} is $I(0)$, then Δe_{it} is still $I(0)$, though overdifferenced. Under the assumption of weak cross-sectional and serial correlations in Δe_{it}, consistent estimators for $\Delta \mathbf{F}_t$ can be readily constructed.

When e_{it} is $I(0)$, estimating the original level equation already provides a consistent estimator for $\mathbf{F}_t$ (Bai 2003, Bai & Ng 2002). Although such estimators could be more efficient than the ones based on the differenced equations, they are not consistent when e_{it} is $I(1)$. An advantage of estimation based on differenced equations is that the factors in levels can still be consistently estimated. Define $\hat{\mathbf{F}}_t = \sum_{s=2}^{t} \widehat{\Delta \mathbf{F}}_s$ and $\hat{e}_{it} = \sum_{s=2}^{t} \widehat{\Delta e}_{is}$. Bai & Ng (2004) show that $\hat{\mathbf{F}}_t$ and $\hat{e}_{it}$ are consistent for $\mathbf{F}_t$ and e_{it}, respectively. In particular, uniform convergence can be established (up to a location shift factor),[2]

$$\max_{1 \leq t \leq T} \left\| \hat{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t + \mathbf{H}\mathbf{F}_1 \right\| = O_p\left(T^{1/2} N^{-1/2}\right) + O_p\left(T^{-1/4}\right).$$

Such a result implies that even if each cross-sectional equation is a spurious regression, the common stochastic trends are well defined and can be consistently estimated, given their existence, a property that is not possible within the framework of fixed-N time-series analysis.

Differencing the data is not necessary when the factors $\mathbf{F}_t$ are $I(1)$ but the errors are $I(0)$. The principal components method directly provides a consistent estimation of the common factors and factor loadings. The convergence rate for the estimated common factors is root N, but the rate for the estimated factor loadings has a faster rate of convergence of T. The inferential theory is derived by Bai (2004). Thus, more precise estimation is obtained when factors are $I(1)$. This is intuitive because the information noise ratio is high with $I(1)$ factors.

11. FACTOR MODELS WITH STRUCTURAL RESTRICTIONS

It is well known that factor models are only identified up to a rotation. Section 4 discusses three sets of restrictions, called PC1, PC2, and PC3. Each set provides r^2 restrictions, such that the static factor model is exactly identified. For dynamic factor models (Equation 3), Bai & Wang (2014, 2015) show that only q^2 restrictions are needed to identify the model, with q the number of dynamic factors. For example, to identify the dynamic factor model in Equation 3, we only need to assume that $\mathrm{var}(\boldsymbol{\varepsilon}_t) = \mathbf{I}_q$, and the $q \times q$ matrix $\boldsymbol{\Lambda}_{01} = [\boldsymbol{\lambda}_{10}, \ldots, \boldsymbol{\lambda}_{q0}]$ is a lower-triangular matrix with strictly positive diagonal elements.

In a number of applications, there might be more restrictions so that the dynamic factor model is overidentified. Bai & Wang (2014) provide general rank conditions for identification linked with q instead of r. In this section, we discuss some useful restrictions for both static and dynamic factor models.

11.1. Factor Models with Block Structure

The dynamic factor model with a multilevel factor structure has been increasingly applied to study the comovement of economic variables at different levels (see, e.g., Crucini et al. 2011, Gregory & Head 1999, Hallin & Liska 2011, Kose et al. 2003, Moench et al. 2013). Such a model imposes

[2]The upper bound can be improved (smaller) if one assumes Δe_{it} have higher-order bounded moments than is assumed by Bai & Ng (2004).

a block structure on the factor model so as to attach economic meaning to factors. For example, Kose et al. (2003) and a number of subsequent papers consider a dynamic factor model with a multilevel factor structure to characterize the comovement of international business cycles on the global, regional, and country level, respectively. We use an example with only two levels of factors, a world factor and a country-specific factor, to convey the main idea.

Consider C countries, each having an $n_c \times 1$ vector of country variables $\mathbf{X}_t^c$, $t = 1, \ldots, T$, $c = 1, \ldots, C$. Assume $\mathbf{X}_t^c$ is affected by a world factor $\mathbf{F}_t^W$ and a country factor $\mathbf{F}_t^c$, $c = 1, \ldots, C$, all factors being latent,

$$\mathbf{X}_t^c = \mathbf{\Lambda}_W^c \mathbf{F}_t^W + \mathbf{\Lambda}_C^c \mathbf{F}_t^c + \mathbf{e}_t^c, \tag{16}$$

where $\mathbf{\Lambda}_W^c$, $\mathbf{\Lambda}_C^c$ are the matrices of factor loadings of country c, and $\mathbf{e}_t^c$ is the vector of idiosyncratic error terms for country c's variables. Let $\mathbf{F}_t^C$ be a vector collecting all the country factors. We may assume that the factors follow a VAR specification,

$$\mathbf{\Phi}(L) \begin{bmatrix} \mathbf{F}_t^W \\ \mathbf{F}_t^C \end{bmatrix} = \begin{bmatrix} \mathbf{u}_t^W \\ \mathbf{U}_t^C \end{bmatrix}, \tag{17}$$

where the innovation to factors $[\mathbf{u}_t^W, \mathbf{U}_t^C]$ is independent of $\mathbf{e}_t^c$ at all leads and lags and is i.i.d. normal. Given some sign restriction, this special VAR specification allows one to separately identify the factors at different levels. Wang (2012) and Bai & Wang (2015) provide detailed identification conditions for such models. Under a static factor model setup, Wang (2012) estimates the model in Equation 16 using an iterated principal components method. The identification conditions therein assume that the world factors are orthogonal to all country factors, whereas country factors can be correlated with each other. Bai & Wang (2015) directly restrict the innovations of factors in Equation 17 and allow lags of factors to enter Equation 16. Such restrictions naturally allow all factors to be correlated with each other. A Bayesian method is then developed to jointly estimate the model in Equations 16 and 17. Alternative joint estimation method is still not available in the literature and would be an interesting topic to explore in the future.

Panel regression models with heterogeneous slope coefficients and with a block factor error structure are studied by Ando & Bai (2015b), who treat each block as a group. The case of unknown group membership is examined by Ando & Bai (2016b), who assume common slope coefficients across individuals or group-dependent coefficients. Heterogeneous slope coefficients with unknown group memberships are studied by Ando & Bai (2016a). These models are particularly useful in finance, in which explanatory variables represent observable risk factors, as in Chen et al. (1986). Classification analysis is required in estimating the unknown group memberships.

11.2. Cross-Equation Restrictions on Factor Loadings

In general, there may be overidentifying restrictions in addition to the exact identification conditions PC1–PC3. For example, the multilevel factor model has many zero blocks. Cross-equation restrictions may also be present. Consider the static factor representation in Equation 2 for the dynamic factor model in Equation 3,

$$\mathbf{X}_t = \mathbf{\Lambda} \mathbf{F}_t + \mathbf{e}_t,$$

where $\mathbf{F}_t = [\mathbf{f}_t, \mathbf{f}_{t-1}, \ldots, \mathbf{f}_{t-s}]'$, and $\mathbf{\Lambda} = [\mathbf{\Lambda}_0, \ldots, \mathbf{\Lambda}_s]$. Let $\mathbf{X}$ be the $(T - s - 1) \times N$ data matrix, $\mathbf{E}$ the $(T - s - 1) \times N$ matrix of the idiosyncratic errors, and $\mathbf{F}$ the $(T - s - 1) \times q(s + 1)$ matrix of the static factors. Then we have a matrix representation of the factor model

$$\mathbf{X} = \mathbf{F}\mathbf{\Lambda}' + \mathbf{E}, \text{ or } \operatorname{vec}(\mathbf{X}) = (\mathbf{I}_N \otimes \mathbf{F})\,\boldsymbol{\lambda} + \operatorname{vec}(\mathbf{E}), \tag{18}$$

where $\boldsymbol{\lambda} = \text{vec}\left(\boldsymbol{\Lambda}'\right)$. Consider the following restriction on the factor loadings:

$$\boldsymbol{\lambda} = \mathbf{B}\boldsymbol{\delta} + \mathbf{C}, \tag{19}$$

where $\boldsymbol{\delta}$ is a vector of free parameters with $\dim(\boldsymbol{\delta}) \le \dim(\boldsymbol{\lambda})$. In general, $\mathbf{B}$ and $\mathbf{C}$ are known matrices and vectors defined by either identifying restrictions or other structural model restrictions. In view of Equation 19, we may rewrite the restricted factor model in Equation 18 as

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \text{vec}(\mathbf{E}),$$

where $\mathbf{y} = \text{vec}(\mathbf{X}) - (\mathbf{I}_N \otimes \mathbf{F})\mathbf{C}$ and $\mathbf{Z} = [(\mathbf{I}_N \otimes \mathbf{F})\mathbf{B}]$. If we impose some distributional assumptions on the error terms, for example, $\text{vec}(\mathbf{E}|\mathbf{Z}) \sim N(0, \mathbf{R} \otimes \mathbf{I}_{T-s})$ for some $N \times N$ positive definite matrix $\mathbf{R}$, such models can be estimated using the Bayesian algorithm from Bai & Wang (2015).

11.3. Structural Vector Autoregression and Restricted Dynamic Factor Models

The dynamic factor models also bring new insight into the estimation of structural vector autoregression (SVAR) models with measurement errors. Consider a traditional SVAR given by

$$\mathbf{A}(L)\mathbf{Z}_t = \mathbf{a}_t,$$

where $\mathbf{Z}_t$ is a $q \times 1$ vector of economic variables, and $\mathbf{a}_t$ is the vector of structural shocks. Let

$$\mathbf{A}(L) = \mathbf{A}_0 - \mathbf{A}_1 L - \cdots - \mathbf{A}_p L^p,$$

with $\mathbf{A}_0 \ne \mathbf{I}_q$. The reduced form is given by

$$\mathbf{Z}_t = \mathbf{B}_1\mathbf{Z}_{t-1} + \cdots + \mathbf{B}_p\mathbf{Z}_{t-p} + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t = \mathbf{A}_0^{-1}\mathbf{a}_t$ and $\mathbf{B}_j = \mathbf{A}_0^{-1}\mathbf{A}_j$. Assume that we do not directly observe $\mathbf{Z}_t$ but observe Y_t:

$$\mathbf{Y}_t = \mathbf{Z}_t + \eta_t,$$

where η_t is the $q \times 1$ measurement error. In this case, it is difficult to estimate the SVAR model based on $\mathbf{Y}_t$. Assume that a large vector of other observed variables $\mathbf{W}_t$ is determined by

$$\mathbf{W}_t = \boldsymbol{\Gamma}_0\mathbf{Z}_t + \cdots + \Gamma_s\mathbf{Z}_{t-s} + \mathbf{e}_{wt}.$$

Let

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{Y}_t \\ \mathbf{W}_t \end{bmatrix}, \quad \mathbf{e}_t = \begin{bmatrix} \eta_t \\ \mathbf{e}_{wt} \end{bmatrix}, \quad \mathbf{f}_t = \mathbf{Z}_t,$$

$$\boldsymbol{\Lambda}_0 = \begin{bmatrix} \mathbf{I}_q \\ \boldsymbol{\Gamma}_0 \end{bmatrix}, \quad \boldsymbol{\Lambda}_j = \begin{bmatrix} 0 \\ \boldsymbol{\Gamma}_j \end{bmatrix}, \quad j \ne 0.$$

Then we have a structural dynamic factor model

$$\begin{aligned} \mathbf{X}_t &= \boldsymbol{\Lambda}_0\mathbf{f}_t + \cdots + \boldsymbol{\Lambda}_s\mathbf{f}_{t-s} + \mathbf{e}_t, \\ \mathbf{f}_t &= \mathbf{B}_1\mathbf{f}_{t-1} + \cdots + \mathbf{B}_p\mathbf{f}_{t-p} + \boldsymbol{\varepsilon}_t. \end{aligned} \tag{20}$$

According to Bai & Wang (2015), because the upper $q \times q$ block of Λ_0 is an identity matrix,

$$\boldsymbol{\Lambda}_0 = \begin{bmatrix} \mathbf{I_q} \\ * \end{bmatrix},$$

the model in Equation 20 is identified and can be analyzed using a Bayesian approach. In particular, without further assumptions, we are able to estimate $\mathbf{f}_t = \mathbf{Z}_t, \mathbf{B}(L), \boldsymbol{\Lambda}_i$, and $E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t\right) = \mathbf{A}_0^{-1}\left(\mathbf{A}'_0\right)^{-1}$.

We may also incorporate additional structural restrictions (such as long-run restrictions) as in standard SVAR analysis to identify $\mathbf{A}_0$. The impulse responses to structural shocks can be obtained as $\partial \mathbf{Y}_{t+k}/\partial \mathbf{a}_t = \partial \mathbf{Z}_{t+k}/\partial \mathbf{a}_t = \partial \mathbf{f}_{t+k}/\partial \mathbf{a}_t$ and

$$\frac{\partial \mathbf{W}_{t+k}}{\partial \mathbf{a}_t} = \left\{ \begin{array}{ll} \mathbf{\Gamma}_0 \dfrac{\partial \mathbf{f}_{t+k}}{\partial \mathbf{a}_t} + \cdots + \Gamma_s \dfrac{\partial \mathbf{f}_{t+k-s}}{\partial a_t}, & k \geq s, \\ \mathbf{\Gamma}_0 \dfrac{\partial \mathbf{f}_{t+k}}{\partial \mathbf{a}_t} + \cdots + \Gamma_k \dfrac{\partial \mathbf{f}_t}{\partial \mathbf{a}_t}, & k < s, \end{array} \right\}$$

where the partial derivative is taken for each component of $\mathbf{a}_t$ when $\mathbf{a}_t$ is a vector.

12. HIGH-DIMENSIONAL COVARIANCE ESTIMATION

The variance-covariance matrix plays a key role in inferential theories of high-dimensional factor models as well as in various applications in finance and economics. Using an observed factor model of large dimensions, Fan et al. (2008) examine the impacts of covariance matrix estimation on optimal portfolio allocation and portfolio risk assessment. Fan et al. (2011) further study the case with unobserved factors. By assuming a sparse error covariance matrix, they allow cross-sectional correlation in errors in the sense of an approximate factor model. An adaptive thresholding technique is employed to account for the fact that the idiosyncratic components are unobserved. Consider the vector representation of the factor model in Equation 2,

$$\mathbf{X}_t = \mathbf{\Lambda} \mathbf{F}_t + \mathbf{e}_t,$$

which implies the following covariance structure:

$$\mathbf{\Sigma}_X = \mathbf{\Lambda} \mathrm{cov}\left(\mathbf{F}_t\right) \mathbf{\Lambda}' + \mathbf{\Sigma}_e,$$

where $\mathbf{\Sigma}_X$ and $\mathbf{\Sigma}_e = \left(\sigma_{ij}\right)_{N\times N}$ are covariance matrices of $\mathbf{X}_t$ and $\mathbf{e}_t$, respectively. Assume $\mathbf{\Sigma}_e$ is sparse instead of diagonal, and define

$$m_T = \max_{i\leq N} \sum_{j\leq N} 1\left(\sigma_{ij} \neq 0\right).$$

The sparsity assumption puts an upper bound assumption on m_T in the sense that

$$m_T^2 = o\left(\frac{T}{r^2 \log(N)}\right).$$

In this formulation, the number of factors r is allowed to be large and grows with T. Using principal components estimators under the normalization $\frac{1}{T}\sum_{t=1}^{T} \mathbf{F}_t \mathbf{F}_t' = \mathbf{I}_r$, one can decompose the sample covariance of $\mathbf{X}_t$ as

$$\mathbf{S}_X = \hat{\mathbf{\Lambda}} \hat{\mathbf{\Lambda}} + \sum_{i=r+1}^{N} \hat{\mu}_i \hat{\boldsymbol{\xi}}_i \hat{\boldsymbol{\xi}}_i',$$

where $\hat{\mu}_i$ and $\hat{\boldsymbol{\xi}}_i$ are the i-th leading eigenvalues and eigenvectors of S_X, respectively. In the high-dimensional setup, the sample covariance might be singular and provides a poor estimator for the population covariance. For example, when $N > T$, the rank of $\mathbf{S}_X$ can never exceed T, whereas the theoretical covariance Σ_X always has rank N. To overcome this problem, we may apply the thresholding technique to the component $\sum_{i=r+1}^{N} \hat{\mu}_i \hat{\boldsymbol{\xi}}_i \hat{\boldsymbol{\xi}}_i'$, which yields a consistent estimator of $\mathbf{\Sigma}_e$, namely $\hat{\mathbf{\Sigma}}_e$. Finally, the estimator for $\mathbf{\Sigma}_X$ is defined as

$$\hat{\mathbf{\Sigma}}_X = \hat{\mathbf{\Lambda}} \hat{\Lambda}' + \hat{\mathbf{\Sigma}}_e,$$

which is always of full rank and can be shown to be a consistent estimator for $\mathbf{\Sigma}_X$.

The adaptive thresholding technique is easy to implement. Denote $\sum_{i=r+1}^{N}\hat{\mu}_i\hat{\boldsymbol{\xi}}_i\hat{\boldsymbol{\xi}}_i' = (\hat{\sigma}_{ij})_{N\times N}$ and $\hat{e}_{it} = x_{it} - \hat{\boldsymbol{\lambda}}_i\hat{\mathbf{F}}_t$. Define

$$\hat{\theta}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{e}_{it}\hat{e}_{jt} - \hat{\sigma}_{ij}\right)^2,$$

$$s_{ij} = \hat{\sigma}_{ij} 1\left(|\hat{\sigma}_{ij}| \geq \sqrt{\hat{\theta}_{ij}}\omega_T\right),$$

where $\omega_T = Cr\sqrt{\frac{\log N}{T}}$ for some positive constant C. In practice, alternative values of C can be assumed to check the robustness of the outcome.

The estimated $\hat{\Sigma}_e$ can be used to obtain more efficient estimation of Λ and F based on the generalized principal components method (Bai & Liao 2013). Alternatively, the unknown parameters of Λ and Σ_e can be jointly estimated by the maximum likelihood method (Bai & Liao 2016). The resulting estimator $\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Lambda}}' + \hat{\boldsymbol{\Sigma}}_e$ from MLE is a direct estimator for the high-dimensional covariance matrix. A survey on high-dimensional covariance estimation is given by Bai & Shi (2011).

13. BAYESIAN METHOD TO LARGE FACTOR MODELS

With fast growing computing power, Markov chain Monte Carlo (MCMC) methods are increasingly used in estimating large-scale models. The Bayesian estimation method, facilitated by MCMC techniques, naturally incorporates identification restrictions such as those from structural VAR into the estimation of the large factor model. The statistical inference on impulse responses can be readily constructed from the Bayesian estimation outcome.

The Bayesian approach has been considered, for example, by Kose et al. (2003), Bernanke et al. (2005), Del Negro & Otrok (2008), Crucini et al. (2011), and Moench et al. (2013). The basic idea is to formulate the dynamic factor model as a state space system with structural restrictions. Initially, we set up the priors for factor loadings and the VAR parameters for factors. We also specify the prior distribution of factors for initial periods. Then Carter & Kohn's (1994) multimove Gibbs sampling algorithm can be adopted for estimation. One of the key steps is to use a Kalman filter or other filters to form a conditional forecast for the latent factors. The main advantage of the Bayesian method is that researchers can incorporate prior information into the estimation of the large factor model, and the outcome is the joint distribution of both model parameters and latent factors. Inference for impulse responses is an easy by-product of the procedure. The computational intensity is largely related to the number of dynamic factors, which is small, but only slightly affected by the dimension of the data.

The Bayesian approach can naturally incorporate structural restrictions. For example, consider the dynamic factor model with linear restrictions on the factor loadings, such as Equations 18 and 19. We may rewrite the restricted factor model in Equation 18 as

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \mathrm{vec}\,(\mathbf{E}), \quad \mathrm{vec}\,(\mathbf{E}|\mathbf{Z}) \sim N\left(0, \mathbf{R}\otimes\mathbf{I}_{T-s}\right),$$

where $\mathbf{y} = \mathrm{vec}\,(\mathbf{X}) - (\mathbf{I}_N \otimes \mathbf{F})\,\mathbf{C}$ and $\mathbf{Z} = [(\mathbf{I}_N \otimes \mathbf{F})\,\mathbf{B}]$. Impose the Jeffreys prior for $\boldsymbol{\delta}$ and $\mathbf{R}$:

$$p(\boldsymbol{\delta}, \mathbf{R}) \propto |\mathbf{R}|^{-(N+1)/2},$$

which implies the following conditional posterior:

$$\boldsymbol{\delta}|\mathbf{R}, \mathbf{X}, \mathbf{F} \sim N\left(\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\left(\mathrm{vec}\left(\hat{\boldsymbol{\Lambda}}'\right) - \mathbf{C}\right), \left(\mathbf{B}'\left(\mathbf{R}^{-1}\otimes\mathbf{F}'\mathbf{F}\right)\mathbf{B}\right)^{-1}\right), \tag{21}$$

with $\hat{\Lambda} = \mathbf{X}'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}$. Thus, we may draw $\boldsymbol{\delta}$ according to Equation 21 and construct the associated loading matrix $\boldsymbol{\Lambda}(\boldsymbol{\delta})$. In the meantime, we may draw $\mathbf{R}$ according to an inverse-Wishart distribution

$$\mathbf{R}|\mathbf{X}, \mathbf{F} \sim \text{invWishart}(\mathbf{S}, T - s + N + 1 - q(s+1)),$$

where $\mathbf{S} = (\mathbf{X} - \mathbf{F}\hat{\boldsymbol{\Lambda}}')'(\mathbf{X} - \mathbf{F}\hat{\boldsymbol{\Lambda}}')$.

There are still some challenges to the Bayesian approach. First, sometimes there is little guidance as to how to choose the prior distribution. Bai & Wang (2015) employ the Jeffreys priors to account for the lack of a priori information about model parameters. It remains an open question as to how theoretical properties of the posterior distribution are affected by an alternative choice of priors. Second, usually the number of restrictions for identification is small and fixed, whereas the number of parameters grows with sample size in both dimensions. This might lead to weak identification and poor inference. Some shrinkage method, such as the application of Minnesota-type priors, might help to mitigate such problems. One might also incorporate overidentifying restrictions to improve estimation efficiency. Third, model selection for the large factor model using Bayes factors is generally computationally intensive. Some simple-to-compute alternative model selection methods, such as provided by Y. Li et al. (2013, 2014), might be considered.

14. CONCLUDING REMARKS

This review provides an introduction to some recent developments in the theory and applications of large factor models. There are still lots of open and interesting issues that await future research. For example, almost all current studies focus on linear factor models and rely on information from covariance matrix for estimation. Introducing nonlinearities into the large factor model could be relevant to a number of potential applications. Freyberger (2012) introduces interactive fixed effects into nonlinear panel regression and identifies important differences between linear and nonlinear regression results. Su et al. (2015) develop a consistent nonparametric test for linear versus nonlinear models in the presence of interactive effects. A similar area has been largely unexplored and could provide potential future research topics. Other examples include the theory on discrete choice models with factor error structure, quantile regression with interactive fixed effects (Ando & Tsay 2011, Harding & Lamarche 2014), factors from higher moments of the data, nonlinear functions of factors, and nonbalanced panel data sets or those with missing observations. In terms of estimation, one may also study alternative methods, such as the adaptive thresholding techniques of Fan et al. (2011), which is useful in situations with many zero factor loadings. For the FAVAR model, one may consider one-step maximum likelihood estimators and compare their theoretical properties with the two-step estimators of Bai et al. (2016). General inferential theory for large factor models with structural restrictions, especially overidentification, is also an important area to explore, and may help estimate macroeconomic models.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Jushan Bai's research is partially supported by the National Science Foundation under grant SES1357598. Peng Wang's research is partially supported by the Hong Kong Research Grant Council under project number 693413.

LITERATURE CITED

Ahn S, Horenstein A. 2013. Eigenvalue ratio test for the number of factors. *Econometrica* 81:1203–27

Ahn S, Lee Y, Schmidt P. 2001. GMM estimation of linear panel data models with time-varying individual effects. *J. Econom.* 102:219–55

Ahn S, Lee Y, Schmidt P. 2013. Panel data models with multiple time-varying individual effects. *J. Econom.* 174:1–14

Amengual D, Watson M. 2007. Consistent estimation of the number of dynamic factors in large N and T panel. *J. Bus. Econ. Stat.* 25:91–96

Anderson TW. 2003. *An Introduction to Multivariate Statistical Analysis*. New York: Wiley. 3rd ed.

Anderson TW, Rubin H. 1956. Statistical inference in factor analysis. *Proc. 3rd Berkeley Symp. Math. Stat. Probab.*, Vol. 5: *Contrib. Econom. Ind. Res. Psychom.*, ed. J Neyman, pp. 111–50. Berkeley: Univ. Calif. Press

Ando T, Bai J. 2015a. A simple new test for slope homogeneity in panel data models with interactive effects. *Econ. Lett.* 136:112–17

Ando T, Bai J. 2015b. Asset pricing with a general multifactor structure. *J. Financ. Econom.* 13:556–604

Ando T, Bai J. 2016a. Clustering huge number of financial time series: a panel data approach with high-dimensional predictors and factor structures. *J. Am. Stat. Assoc.* In press

Ando T, Bai J. 2016b. Panel data models with grouped factor structure under unknown group membership. *J. Appl. Econom.* 136:163–91

Ando T, Tsay R. 2011. Quantile regression models with factor-augmented predictors and information criterion. *Econom. J.* 14:1–24

Bai J. 2003. Inferential theory for factor models of large dimensions. *Econometrica* 71:135–72

Bai J. 2004. Estimating cross-section common stochastic trends in nonstationary panel data. *J. Econom.* 122:137–83

Bai J. 2009. Panel data models with interactive fixed effects. *Econometrica* 77:1229–79

Bai J, Carrion-I-Silvestre JL. 2009. Structural changes, common stochastic trends, and unit roots in panel data. *Rev. Econ. Stud.* 76:471–501

Bai J, Carrion-I-Silvestre JL. 2013. Testing panel cointegration with unobservable dynamic common factors that are correlated with the regressors. *Econom. J.* 16:222–49

Bai J, Li K. 2012. Statistical analysis of factor models of high dimension. *Ann. Stat.* 40:436–65

Bai J, Li K. 2014. Theory and methods of panel data models with interactive effects. *Ann. Stat.* 42:142–70

Bai J, Li K. 2016. Maximum likelihood estimation and inference for approximate factor models of high dimension. *Rev. Econ. Stat.* 98:298–309

Bai J, Li K, Lu L. 2016. Estimation and inference of FAVAR models. *J. Bus. Econ. Stat.* In press

Bai J, Liao Y. 2013. *Statistical inferences using large estimated covariances for panel data and factor models*. Work. Pap., Columbia Univ., New York

Bai J, Liao Y. 2016. Efficient estimation of approximate factor models via penalized maximum likelihood. *J. Econom.* 191:1–18

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

Bai J, Ng S. 2004. A PANIC attack on unit roots and cointegration. *Econometrica* 72:1127–77

Bai J, Ng S. 2006a. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica* 74:1133–50

Bai J, Ng S. 2006b. Evaluating latent and observed factors in macroeconomics and finance. *J. Econom.* 131:507–37

Bai J, Ng S. 2007. Determining the number of primitive shocks in factor models. *J. Bus. Econ. Stat.* 25:52–60

Bai J, Ng S. 2008. *Large Dimensional Factor Analysis.* Boston: Now Publ.

Bai J, Ng S. 2010. Instrumental variable estimation in a data rich environment. *Econom. Theory* 26:1577–606

Bai J, Ng S. 2013. Principal components estimation and identification of static factors. *J. Econom.* 176:18–29

Bai J, Shi S. 2011. Estimating high dimensional covariance matrices and its applications. *Ann. Econ. Finance* 12:199–215

Bai J, Wang P. 2014. Identification theory for high dimensional static and dynamic factor models. *J. Econom.* 178:794–804

Bai J, Wang P. 2015. Identification and Bayesian estimation of dynamic factor models. *J. Bus. Econ. Stat.* 33:221–40

Bailey N, Kapetanios G, Pesaran M. 2016. Exponent of cross-sectional dependence: estimation and inference. *J. Appl. Econom.* In press

Banerjee A, Marcellino M. 2008. *Forecasting macroeconomic variables using diffusion indexes in short samples with structural change*. CEPR Work. Pap. 6706, Cent. Econ. Policy Res., Washington, DC

Bates B, Plagborg-Moller M, Stock J, Watson M. 2013. Consistent factor estimation in dynamic factor models with structural instability. *J. Econom.* 177:289–304

Belloni A, Chen D, Chernozhukov V, Hansen C. 2012. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80:2369–429

Bernanke B, Boivin J, Eliasz P. 2005. Measuring monetary policy: a factor augmented vector autoregressive (FAVAR) approach. *Q. J. Econ.* 120:387–422

Boivin J, Giannoni M. 2006. *DSGE models in a data-rich environment*. NBER Work. Pap. 12772

Breitung J, Eickmeier S. 2011. Testing for structural breaks in dynamic factor models. *J. Econom.* 163:71–84

Breitung J, Tenhofen J. 2011. GLS estimation of dynamic factor models. *J. Am. Stat. Assoc.* 106:1150–66

Cai T, Zhou H. 2012. Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Stat.* 40:2389–420

Carter C, Kohn R. 1994. On Gibbs sampling for state space models. *Biometrika* 81:541–53

Chamberlain G, Rothschild M. 1983. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* 51:1281–304

Chen L, Dolado J, Gonzalo J. 2014. Detecting big structural breaks in large factor models. *J. Econom.* 180:30–48

Chen NF, Roll R, Ross S. 1986. Economic forces and the stock market. *J. Bus.* 59:383–403

Cheng X, Liao Z, Schorfheide F. 2016. Shrinkage estimation of high-dimensional factor models with structural instabilities. *Rev. Econ. Stud.* In press

Choi I. 2012. Efficient estimation of factor models. *Econom. Theory* 28:274–308

Chudik A, Mohaddes K, Pesaran MH, Raissi M. 2016. Long-run effects in large heterogeneous panel data models with cross-sectionally correlated errors. In *Essays in Honor of Aman Ullah*, ed. RC Hill, G Gonzalez-Rivera, T-H Lee. Adv. Econom. 36. Bingley, UK: Emerald Insight. In press

Chudik A, Pesaran MH. 2015. Long-run effects in large heterogeneous panel data models with cross-sectionally correlated errors. *J. Econom.* 188:393–420

Chudik A, Pesaran MH, Tosetti E. 2011. Weak and strong cross-section dependence and estimation of large panels. *Econom. J.* 14:45–90

Connor G, Korajczyk RA. 1986. Performance measurement with the arbitrage pricing theory. *J. Financ. Econ.* 15:373–94

Corradi V, Swanson NR. 2014. Testing for structural stability of factor augmented forecasting models. *J. Econom.* 182:100–18

Crucini M, Kose M, Otrok C. 2011. What are the driving forces of international business cycles? *Rev. Econ. Dyn.* 14:156–75

Del Negro M, Otrok C. 2008. *Dynamic factor models with time-varying parameters: measuring changes in international business cycles*. Staff Rep. 326, Fed. Reserve Bank New York

Doan T, Litterman R, Sims C. 1984. Forecasting and policy analysis using realistic prior distributions. *Econom. Rev.* 3:1–100

Doz C, Giannone D, Reichlin L. 2011. A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *J. Econom.* 164:188–205

Doz C, Giannone D, Reichlin L. 2012. A quasi-maximum likelihood approach for large approximate dynamic factor models. *Rev. Econ. Stat.* 94:1014–24

Fama E, French K. 1993. Common risk factors in the returns on stocks and bonds. *J. Financ. Econ.* 33:3–56

Fan J, Fan Y, Lv J. 2008. High dimensional covariance matrix estimation using a factor model. *J. Econom.* 147:186–97

Fan J, Liao Y, Mincheva M. 2011. High dimensional covariance matrix estimation in approximate factor models. *Ann. Stat.* 39:3320–56

Forni M, Giannone D, Lippi M, Reichlin L. 2009. Opening the black box: structural factor models with large cross sections. *Econom. Theory* 25:1319–47

Forni M, Hallin M, Lippi M, Reichlin L. 2000. The generalized dynamic factor model: identification and estimation. *Rev. Econ. Stat.* 82:540–54

Forni M, Hallin M, Lippi M, Reichlin L. 2004. The generalized dynamic factor model: consistency and rates. *J. Econom.* 119:231–55

Forni M, Hallin M, Lippi M, Reichlin L. 2005. The generalized dynamic factor model: one-sided estimation and forecasting. *J. Am. Stat. Assoc.* 100:830–39

Freyberger J. 2012. *Nonparametric panel data models with interactive fixed effects*. Work. Pap., Univ. Wisconsin–Madison

Geweke J. 1977. The dynamic factor analysis of economic time series. In *Latent Variables in Socio-Economic Models*, ed. DJ Aigner, AS Goldberger, pp. 365–83. Amsterdam: North-Holland

Giannone D, Reichlin L, Small D. 2008. Nowcasting: the real-time informational content of macroeconomic data. *J. Monet. Econ.* 55:665–76

Gregory A, Head A. 1999. Common and country-specific fluctuations in productivity, investment, and the current account. *J. Monet. Econ.* 44:423–51

Hallin M, Liska R. 2007. The generalized dynamic factor model: determining the number of factors. *J. Am. Stat. Assoc.* 102:603–17

Hallin M, Liska R. 2011. Dynamic factors in the presence of blocks. *J. Econom.* 163:29–41

Han X, Inoue A. 2015. Tests for parameter instability in dynamic factor models. *Econom. Theory* 31:1117–52

Hansen C, Hausman J, Newey WK. 2008. Estimation with many instrumental variables. *J. Bus. Econ. Stat.* 26:398–422

Harding M, Lamarche C. 2014. Estimating and testing a quantile regression model with interactive effects. *J. Econom.* 178:101–13

Hausman JA, Newey WK, Woutersen T, Chao JC, Swanson NR. 2010. *Instrumental variable estimation with heteroskedasticity and many instruments*. Work. Pap., Univ. Maryland and Mass. Inst. Technol.

Im K, Pesaran M, Shin Y. 2003. Testing for unit roots in heterogeneous panels. *J. Econom.* 115:53–74

Kapetanios G, Marcellino M. 2010. Factor-GMM estimation with large sets of possibly weak instruments. *Comput. Stat. Data Anal.* 54:2655–75

Kapetanios G, Pesaran M, Yamagata T. 2011. Panels with non-stationary multifactor error structures. *J. Econom.* 160:326–48

Kim C, Nelson C. 1999. *State Space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Applications*. Cambridge, MA: MIT Press

Kose M, Otrok C, Whiteman C. 2003. International business cycles: world, region, and country-specific factors. *Am. Econ. Rev.* 93:1216–39

Lawley DN, Maxwell AE. 1971. *Factor Analysis as a Statistical Method*. Amsterdam: Elsevier

Levin A, Lin C, Chu C. 2002. Unit root tests in panel data: asymptotic and finite sample properties. *J. Econom.* 108:1–24

Li H, Li Q, Shi Y. 2013. *Determining the number of factors when the number of factors can increase with sample size*. Work. Pap., Dep. Econ., Texas A & M Univ., College Station

Li Y, Yu J, Zeng T. 2014. A new approach to Bayesian hypothesis testing. *J. Econom.* 178:602–12

Li Y, Zeng T, Yu J. 2013. *Robust deviance information criterion for latent variable models*. Work. Pap., Singapore Manag. Univ.

Litterman R. 1986. Forecasting with Bayesian vector autoregressions: five years of experience. *J. Bus. Econ. Stat.* 4:25–38

Lu X, Su L. 2016. Shrinkage estimation of dynamic panel data models with interactive fixed effects. *J. Econom.* 190:148–75

Meng JG, Hu G, Bai J. 2011. OLIVE: a simple method for estimating betas when factors are measured with error. *J. Financ. Res.* 34:27–60

Moench E, Ng S, Potter S. 2013. Dynamic hierarchical factor models. *Rev. Econ. Stat.* 95:1811–17

Moon H, Weidner M. 2015. Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* 83:1543–79

Moon H, Weidner M. 2016. Dynamic linear panel regression models with interactive fixed effects. *Econ. Theory*. In press

Ng S, Bai J. 2009. Selecting instrumental variables in a data rich environment. *J. Time Ser. Econom.* 1:1–34

Ng S, Ludvigson S. 2009. Macro factors in bond risk premia. *Rev. Financ. Stud.* 22:5027–67

O'Connell P. 1998. The overvaluation of purchasing power parity. *J. Int. Econ.* 44:1–19

Onatski A. 2009. A formal statistical test for the number of factors in the approximate factor models. *Econometrica* 77:1447–79

Onatski A. 2010. Determining the number of factors from the empirical distribution of eigenvalues. *Rev. Econ. Stat.* 92:1004–16

Onatski A. 2011. *Asymptotics of the principal components estimator of large factor models with weakly influential factors*. Work. Pap., Univ. Cambridge, Cambridge, UK

Pesaran M. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74:967–1012

Pesaran M. 2007. A simple panel unit root test in the presence of cross-section dependence. *J. Appl. Econom.* 22:265–312

Pesaran M, Schuermann T, Weiner S. 2004. Modeling regional interdependencies using a global error-correcting macroeconometric model. *J. Bus. Econ. Stat.* 22:129–62

Pesaran M, Smith L, Yamagata T. 2013. Panel unit root tests in the presence of a multifactor error structure. *J. Econom.* 175:94–115

Quah D. 1994. Exploiting cross-section variations for unit root inference in dynamic panels. *Econ. Lett.* 44:1–9

Quah D, Sargent T. 1993. A dynamic index model for large cross sections. In *Business Cycles, Indicators and Forecasting*, ed. JH Stock, MW Watson, pp. 285–310. Cambridge, MA: Natl. Bur. Econ. Res.

Sims C. 1993. A nine-variable probabilistic macroeconomic forecasting model. In *Business Cycles, Indicators and Forecasting*, ed. JH Stock, MW Watson, pp. 179–204. Cambridge, MA: Natl. Bur. Econ. Res.

Song M. 2013. *Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications*. Work. Pap., Korea Inst. Finance, Seoul

Stock J, Watson M. 1996. Evidence on structural instability in macroeconomic time series relations. *J. Bus. Econ. Stat.* 14:11–30

Stock J, Watson M. 1998. *Diffusion indexes*. NBER Work. Pap. 6702

Stock J, Watson M. 1999. Forecasting inflation. *J. Monet. Econ.* 44:293–335

Stock J, Watson M. 2002a. Forecasting using principal components from a large number of predictors. *J. Am. Stat. Assoc.* 97:1167–79

Stock J, Watson M. 2002b. Macroeconomic forecasting using diffusion indexes. *J. Bus. Econ. Stat.* 20:147–62

Stock J, Watson M. 2005. *Implications of dynamic factor models for VAR analysis*. NBER Work. Pap. 11467

Stock J, Watson M. 2006. Macroeconomic forecasting using many predictors. In *Handbook of Economic Forecasting*, Vol. 1, ed. G Elliott, C Granger, A Timmermann, pp. 515–54. Amsterdam: North-Holland

Stock J, Watson M. 2008. Forecasting in dynamic factor models subject to structural instability. In *The Methodology and Practice of Econometrics: A Festschrift in Honour of Professor David F. Hendry*, ed. J Castle, N Shephard, pp. 1–57. New York: Oxford Univ. Press

Stock J, Watson M. 2010. Dynamic factor models. In *Oxford Handbook of Economic Forecasting*, ed. MP Clements, DF Hendry. New York: Oxford Univ. Press. doi: 10.1093/oxfordhb/9780195398649.013.0003

Su L, Chen Q. 2013. Testing homogeneity in panel data models with interactive fixed effects. *Econom. Theory* 29:1079–135

Su L, Jin S, Zhang Y. 2015. Specification test for panel data models with interactive fixed effects. *J. Econom.* 186:222–44

Wang P. 2012. *Large dimensional factor models with a multi-level factor structure: identification, estimation, and inference*. Work. Pap., Hong Kong Univ. Sci. Technol.

Yamamoto Y. 2016. Forecasting with non-spurious factors in U.S. macroeconomic time series. *J. Bus. Econ. Stat.* 34:81–106

Yamamoto Y, Tanaka S. 2015. Testing for factor loading structural change under common breaks. *J. Econom.* 189:187–206