RECENT DEVELOPMENTS IN LARGE DIMENSIONAL FACTOR ANALYSIS

Jushan Bai* Serena Ng †

May 8, 2008

Abstract

Econometric analysis of large dimensional factor models has been a heavily researched topic in recent years. This paper surveys the main theoretical results that relate to static factor models or dynamic factor models that can be cast in a static framework. Among the topics covered are how to determine the number of factors, how to conduct inference when estimated factors are used in regressions, how to assess the adequacy of observed variables as proxies for latent factors, how to exploit the estimated factors to test unit root tests and common trends, and how to estimate panel cointegration models. The fundamental result that justifies these analyses is that the method of asymptotic principal components consistently estimates the true factor space. We use simulations to better understand the conditions that can affect the precision of the factor estimates.

*Department of Economics, NYU, 19 West 4th St, New York, NY 10012. Email: Jushan.Bai@nyu.edu.
†Economics Department, Columbia University, 1019 International Affairs Building, 420 West 118th Street New York NY 10027. Email: Serena.Ng@columbia.edu
We are grateful to an anonymous referee for careful reading and constructive comments. This paper was presented by the second author as an invited lecture at the 2007 annual meeting of the Society of Computational Economics in Montreal. Financial support from the National Science Foundation (SES055127 and SES-0549978) is gratefully acknowledged.

1 Introduction

An inevitable fact as we move forward in time and as information technology improves is that data will be available for many more series and over an increasingly long span. While the availability of more data provides the opportunity to understand economic phenomena and anomalies better, researchers can also suffer from an information overload without some way to organize the data into an easy to interpret manner. In recent years, the analysis of large dimensional data has received the attention of theoretical and empirical researchers alike. The early focus has primarily been on the use of factor models as a means of dimension reduction. But the volume of research, both at the empirical and theoretical levels, has grown substantially. Empirical researchers have found it useful to extract a few factors from a large number of series in many forecasting and policy exercises. Theoretical researchers have taken up the challenge to extend standard factor analysis to allow the size of both dimensions of a panel data set to increase. The theoretical implications of using estimated factors in both estimation and inference are now better understood. Factor analysis plays a role not just in forecasting. In recent years, the factor structure has been incorporated into regression analysis to deal with cross-sectionally correlated errors and endogeneity bias.

This paper provides a survey of the main theoretical results for large dimensional factor models, emphasizing results that have implications for empirical work. We focus on the development of the static factor models, which are to be distinguished from dynamic factor models in ways to be made precise. Key results concerning large dynamic factor models are given in Forni et al. (2000), Forni et al. (2004), and Forni et al. (2005). Results concerning the use of factors in forecasting are discussed in Stock and Watson (2006), Banerjee et al. (2006) and Giannone et al. (2007). Here, our focus will be on the use of estimated factors in subsequent estimation and inference. While we will survey many of the analytical results that are of use to empirical researchers, a survey of empirical applications of large factor models will not be included. Surveys with heavier empirical focus can be found in Breitung and Eickmeier (2005) and Reichlin (2003). Suffice it to say that factor models have been used in forecasting of the conditional mean by Stock and Watson (2002b), Cristadoro et al. (2001), Artis et al. (2005), Marcellino et al. (2003), Schumacher (2005), Forni et al. (2001), den Reijer (2005), and many others. Boivin and Ng (2005) compared the use of dynamic and static factors in forecasting. Anderson and Vahid (2007) used the factor model to forecast volatility with jump components. A non-exhaustive list of policy analyses that adopt a factor approach includes Bernanke and Boivin (2003), Giannone et al. (2005a), Favero et al. (2005), Stock and Watson (2005), Giannone et al. (2005b), and Forni et al. (2003). Use of factors as conditioning information is discussed in the conditional risk-return analysis of Ludvigson and Ng (2007), and term structure analysis of Ludvigson and Ng (2005).

This survey, drawing heavily from our previous work, is organized to serve three purposes. First, the results are presented under a coherent and general set of assumptions. Situations that require stronger assumptions will be made clear as we go along. Second, results for stationary and non-stationary data are discussed separately, as they involve different assumptions and are used in different contexts. Third, consistent estimation of the factor space is fundamental to many of the results. We use simulations to study what are the main aspects of the data that affect the precision of the factor estimates.

2 Factor Models

We begin by setting up notation and making a distinction between a *static* and a *dynamic* factor model. Let N be the number of cross-section units and T be the number of time series observations. For $i = 1, \dots N, \;\; t = 1, \dots T,$[1] a *static* model is defined as

$$\begin{aligned} x_{it} &= \lambda_i' F_t + e_{it} \\ &= C_{it} + e_{it}. \end{aligned} \tag{1}$$

In factor analysis, e_{it} is referred to as the idiosyncratic error and λ_i is referred to as the factor loadings. This is a vector of weights that unit i put on the corresponding r (static) common factors F_t. The term $C_{it} = \lambda_i' F_t$ is often referred to as the common component of the model. Factor models arise naturally in economics. For example, x_{it} is the GDP growth rate for country i in period t, F_t is a vector of common shocks, λ_i is the heterogenous impact of the shocks, and e_{it} is the country-specific growth rate. In finance, x_{it} is the return for asset i in period t, and F_t is vector of systematic risks (or factor returns) and λ_i is the exposure to the factor risks, and e_{it} is the idiosyncratic returns.

Let $X_t = (x_{1t}, x_{2t}, \dots x_{Nt})'$, $F = (F_1, \dots F_T)'$, and $\Lambda = (\lambda_1, ..., \lambda_N)'$. In vector form, we have

$$X_t = \Lambda F_t + e_t \tag{2}$$

[1]By letting x_{it}^0 be the observed data, the model can be generalized to $x_{it} = (1 - \rho_i(L)) x_{it}^0 = \lambda_i' F_t + e_{it}$.

Let $X = (X_1', \ldots X_N')$ be a $T \times N$ matrix observations. The matrix representation of the factor model is

$$X = F\Lambda' + e, \tag{3}$$

where $e = (e_1', e_2', ..., e_N')$ is $T \times N$ (The upper case E is reserved for the expectation operator). Although the model specifies a static relationship between x_{it} and F_t, F_t itself can be a dynamic vector process that evolves according to

$$A(L)F_t = u_t \tag{4}$$

where $A(L)$ is a polynomial (possibly infinite order) of the lag operator. The idiosyncratic error e_{it} can also be a dynamic process. The assumptions to be stated below also permit e_{it} to be cross-sectionally correlated.

The static model is to be contrasted with a *dynamic* factor model, defined as

$$x_{it} = \lambda_i'(L)f_t + e_{it}$$

where $\lambda_i(L) = (1 - \lambda_{i1}L - \ldots - \lambda_{is}L^s)$ is a vector of dynamic factor loadings of order s. The term 'dynamic factor model' is sometimes reserved for the case when s is finite, whereas a 'generalized dynamic factor model' allows s to be infinite. In either case, the factors are assumed to evolve according to

$$f_t = C(L)\varepsilon_t,$$

where ε_t are iid errors. The dimension of f_t, denoted q, is the same as the dimension of ε_t. We can rewrite the model as

$$x_{it} = \lambda_i(L)'C(L)\varepsilon_t + e_{it}.$$

In the literature, $q = dim(\varepsilon_t)$ is referred to as the number of dynamic factors.

Both models have their origin in the statistics literature. Assuming F_t and e_t are uncorrelated and have zero mean, the covariance structure of the static model is given by

$$\Sigma = \Lambda\Lambda' + \Omega$$

where Σ and Ω are the $N \times N$ population covariance matrix of X_t and e_t respectively; the normalization $E(F_tF_t') = I_r$ is assumed. If Ω is diagonal, (1) is referred to as a strict factor model, see Chamberlain and Rothschild (1983). In classical factor analysis, F_t and e_t in (1) are generally assumed to be serially and cross-sectionally uncorrelated. Properties of such a model, under the assumptions that (i) e_t is iid over t; (ii) N is fixed as T tends to infinity (or vice versa); and (iii) both F_t and e_t are normally distributed, are well documented; see Lawley and Maxwell (1971), Anderson and Rubin (1956), and Anderson (1984), Classical factor analysis estimates Λ and the diagonal elements of Ω, with which factor scores F_t can also be estimated. The estimated score cannot be consistent since N is fixed. The limiting distribution is based on asymptotic normality for an estimator of Σ (e.g., the sample covariance matrix). For large N, this method of analysis is not applicable since Σ $(N \times N)$ is not consistently estimable.

Classical factor models have been widely used in psychology and other disciplines of the social sciences but less so in economics, perhaps because the assumption that the factors and errors are serially and cross-sectionally correlated do not match up well with economic data. The dynamic classical factor model maintains the assumption that the errors are independent across i but explicitly recognizes the fact that the data being analyzed are serially correlated. Sargent and Sims (1977) and Geweke (1977) were amongst the first to apply the dynamic factor approach to macroeconomic analysis.

A dynamic factor model with q factors can be written as a static factor model with r factors, where r is finite. However, the dimension of F_t will in general be different from the dimension of f_t since F_t includes the leads and lags of f_t. More generally, if we have q dynamic factors, we will end up with $r = q(s+1) \geq q$ static factors. Although knowledge of the dynamic factors is useful in some analysis such as precisely establishing the number of primitive shocks in the economy, it turns out that many econometric methods can be developed within the static framework. Consequently, the properties of the estimated static factors are much better understood from a theoretical standpoint. Empirically, the static and the dynamic factor estimates produce rather similar forecasts. From a practical perspective, the primary advantage of the static framework is that it is easily estimated using time domain methods and involves few choices of auxiliary parameters. Dynamic factor models are estimated using tools of frequency domain analysis, and the proper choice of the auxiliary parameters remains an issue requiring further research.

An important characteristic of a static model with r factors is that the largest r population eigenvalues of Σ increase with N, while the remaining eigenvalues of Σ, as well as *all* eigenvalues of Ω, are bounded. Intuitively, the information of the common component accumulates as we sum up the observations across i and therefore the eigenvalues of the population covariance matrix of the common component will increase with N. In contrast, the e_{it} are unit-specific errors and summing the errors across i does not lead to the same accumulation of information. In other words, the eigenvalues of Ω cannot increase without bound, as N increases. It is this difference in the property of the eigenvalues that distinguishes the common from the idiosyncratic component. If the eigenvalues of the common component increases with N, so will the population eigenvalues of Σ.

The large dimensional static factor model we consider differs from the classical factor model by relaxing the three mentioned assumptions. Work in this direction was initiated by J. Stock and M. Watson in the late nineties. Around that same time, assumptions of the classical dynamic factor model were also relaxed, notably by M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The efforts of these researchers were instrumental in advancing the theory and use of large dimensional dynamic factor models. Collectively, the new generation of factor models has come to be known as 'large dimensional approximate factor models'. By 'large', we mean that the sample size in both dimensions tends to infinity in the asymptotic theory. By an 'approximate' factor structure, we mean that the idiosyncratic errors are allowed to be 'weakly' correlated across i and t in a sense to be explained.

The only quantities that are observed in factor analysis are the data, x_{it}. Neither the factors, their loadings, nor the idiosyncratic errors are observed, and the factors and the loadings are not even separately identifiable.[2] Even estimation of classical factor models with the sample size fixed in one dimension can pose difficulties if one allows for heterogeneous variables, for example. Large dimensional factor models pose additional statistical problems that need to be solved. Whereas classical (static or dynamic) factor models can be estimated consistently by methods that rely on sample moments converging to population moments of fixed dimensions, this approach is no longer appropriate when the dimensions of these moment matrices are themselves increasing. The theory we explore below surrounds the new estimation and inferential results that are developed specifically for the large N and T environment. Results are presented for the principal components estimator, which is easy to compute.

[2]Connor et al. (2007) restrict the factor loadings to be unknown functions of some observable variables such that $\lambda_{ij} = g_j(z_i)$ $(j = 1, 2, .., .r)$. There are r unknown functions to be estimated. The estimation of this model will not be considered in this survey.

3 Principal Components and Related Identities

Under large N and large T, it is possible to estimate Λ and F simultaneously. That is, both Λ and F are treated as parameters. After obtaining Λ and F, the residual matrix e is also obtained from $e = X - F\Lambda'$. In contrast, classical factor analysis estimates Λ (under fixed N) and the covariance matrix of e_t, which is assumed to be diagonal. Given Λ, factor scores F_t are then estimated at the second stage. The estimate for F_t is not consistent under fixed N. Consider estimating (3),

$$X = F\Lambda' + e.$$

Clearly F and Λ are not separately identifiable. For an arbitrary $r \times r$ invertible matrix A, $F\Lambda' = FAA^{-1}\Lambda' = F^*\Lambda^{*\prime}$, where $F^* = FA$ and $\Lambda^* = \Lambda A^{-1\prime}$, the factor model is observationally equivalent to $X = F^*\Lambda^{*\prime} + e$. Hence restrictions are needed to uniquely fix F and Λ. Since an arbitrary $r \times r$ matrix has r^2 free parameters, we need r^2 number of restrictions. The normalization

$$F'F/T = I_r$$

provides $r(r+1)/2$ restrictions. The requirement of $\Lambda'\Lambda$ being diagonal gives $r(r-1)/2$ additional restrictions.[3] Combining normalization and diagonality we have r^2 restrictions, which will uniquely fix F and Λ (still up to a column sign change) given the product $F\Lambda'$. Alternatively, the normalization $\Lambda'\Lambda/N = I_r$ and $F'F$ being diagonal can be used. These restrictions are used in the principal components method. When k factors are estimated, k^2 restrictions are required.

The method of asymptotic principal components was first considered by Connor and Korajzcyk (1986) and Connor and Korajzcyk (1998) as an estimator of the factors in a large N, fixed T setup. For any given k not necessarily equal to the true number of factors r, the method of principal components (PC) constructs a $T \times k$ matrix of estimated factors and a corresponding $N \times k$ matrix of estimated loadings by solving the optimization problem

$$\min_{\Lambda^k, F^k} S(k), \quad \text{with} \quad S(k) = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \lambda_i^{k\prime} F_t^k)^2$$

subject to the normalization that $\Lambda^{k\prime}\Lambda^k/N = I_k$ and $F^{k\prime}F^k$ being diagonal, or $F^{k\prime}F^k/T = I_k$ and $\Lambda^{k\prime}\Lambda^k$ being diagonal.

[3] Diagonality in fact imposes $r(r-1)$ restrictions. Within the class of symmetric matrices, diagonality only imposes half as many restrictions.

Mechanically speaking, the estimates can be obtained in one of two ways. The first solution obtains if we concentrate out Λ^k. The problem is then identical to maximizing $\text{tr}(F^{k\prime}(X'X)F^k)$. The estimated factor matrix, $\widetilde{F}^k$, is $\sqrt{T}$ times the eigenvectors corresponding to the k largest eigenvalues of the $T \times T$ matrix XX'. Using the normalization $\widetilde{F}^{k\prime}\widetilde{F}/T = I_k$ yields $\widetilde{\Lambda}^{k\prime} = \widetilde{F}^{k\prime}X/T$. Note that in this case, $\widetilde{\Lambda}^{k\prime}\widetilde{\Lambda}^k$ being diagonal is automatically fulfilled. The solution for F^k that maximizes $\text{tr}(F^{k\prime}(X'X)F^k)$ is not unique under the restriction $F^{k\prime}F/T = I_k$. Any orthogonal rotation of a solution is also a solution. Choosing F^k to be the eigenvector fixes this rotational indeterminacy and, at the same time, makes $\Lambda^{k\prime}\Lambda^k$ diagonal.

The second solution obtains if we concentrate out F^k. Then the matrix of factor loadings $\bar{\Lambda}^k$ is $\sqrt{N}$ times the eigenvectors corresponding to the k largest eigenvalues of the $N \times N$ matrix $X'X$. Using the normalization that $\bar{\Lambda}^{k\prime}\bar{\Lambda}^k/N = I_k$ yields $\bar{F}^k = X\bar{\Lambda}^k/N$. The second set of calculations is computationally less costly when $T > N$ while the first is less intensive when $T < N$. In all except one case that will be stated below, our results do not depend on which of the two methods is used.

Let $\widetilde{V}^k$ denote the $k \times k$ diagonal matrix consisting of the first k largest eigenvalues of the matrix $XX'/(TN)$, arranged in decreasing order. Note that the matrices XX' and $X'X$ have identical nonzero eigenvalues. We first explore some identities among different estimators. First we have

$$\frac{\bar{F}^{k\prime}\bar{F}^k}{T} = \widetilde{V}^k \qquad (5)$$
$$\frac{\widetilde{\Lambda}^{k\prime}\widetilde{\Lambda}^k}{N} = \widetilde{V}^k$$

To see this, by the definition of the eigenvalues and eigenvectors, $(NT)^{-1}X'X\bar{\Lambda}^k = \bar{\Lambda}^k\widetilde{V}^k$. Left multiplying the transpose of $\bar{\Lambda}^k$, and using $\bar{\Lambda}^{k\prime}\bar{\Lambda}^k = N$ and $\bar{F}^k = X\bar{\Lambda}^k/N$, the first equality is obtained. The second equality follows from a similar argument (or simply by symmetry). Additional useful identities are

$$\bar{F}^k = \widetilde{F}^k(\widetilde{V}^k)^{1/2} \qquad (6)$$
$$\widetilde{\Lambda}^k = \bar{\Lambda}^k(\widetilde{V}^k)^{1/2}$$

To see this, by the definition of eigen relationship,

$$(\frac{1}{NT}X'X)\bar{\Lambda}^k = \bar{\Lambda}^k\,\widetilde{V}^k$$

Left multiplying X/N on each side of above and noting $\bar{F}^k = X\Lambda^k/N$, we have

$$(\frac{1}{NT}XX')\bar{F}^k = \bar{F}^k\widetilde{V}^k$$

This means that $\bar{F}^k$ is a $T \times k$ matrix consisting of the eigenvectors of the matrix XX'. By the uniqueness of eigenvector with probability 1 (up to a scale), each column of $\bar{F}^k$ is a scalar multiple of the corresponding column of $\widetilde{F}^k$ because $\widetilde{F}^k$ is also an eigenvector matrix. The squared length of each column in $\bar{F}^k$ is equal to the corresponding eigenvalue (see equation (5)), and the columns of $\widetilde{F}^k$ have unit length, it follows that $\bar{F}^k = \widetilde{F}^k(\widetilde{V}^k)^{1/2}$. The second equality in (6) follows from symmetry.

Throughout this paper, when k is equal to r (the true number of fators), we simply drop the subscript k to write $\widetilde{F}$, $\widetilde{\Lambda}$, and $\widetilde{V}$. Similarly, the superscript k will be dropped for the "bar" version of the estimates when $k = r$. The matrix $\widetilde{V}$ is identical to its bar version (does not depend on the estimation method).

4 Theory: Stationary Data

To study the properties of this estimator, we let F_t^0 and λ_i^0 denote the true factors and the loadings, respectively. Let M be a generic constant. For stationary data, the following assumptions underlie much of the analysis:

Assumption F(0): $E\|F_t^0\|^4 \leq M$ and $\frac{1}{T}\sum_{t=1}^T F_t^0 F_t^{0\prime} \xrightarrow{p} \Sigma_F > 0$ for an $r \times r$ non-random matrix Σ_F.

Assumption L: λ_i^0 is either deterministic such that $\|\lambda_i^0\| \leq M$, or it is stochastic such that $E\|\lambda_i^0\|^4 \leq M$. In either case, $N^{-1}\Lambda^{0\prime}\Lambda^0 \xrightarrow{p} \Sigma_\Lambda > 0$ for an $r \times r$ non-random matrix Σ_Λ, as $N \to \infty$.

Assumption E:

a $E(e_{it}) = 0$, $E|e_{it}|^8 \leq M$.

b $E(e_{it}e_{js}) = \sigma_{ij,ts}$, $|\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$ for all (t,s) and $|\sigma_{ij,ts}| \leq \tau_{ts}$ for all (i,j) such that $\frac{1}{N}\sum_{i,j=1}^{N} \bar{\sigma}_{ij} \leq M$, $\frac{1}{T}\sum_{t,s=1}^{T} \tau_{ts} \leq M$, and $\frac{1}{NT}\sum_{i,j,t,s=1} |\sigma_{ij,ts}| \leq M$.

c For every (t,s), $E|N^{-1/2}\sum_{i=1}^{N}\Big[e_{is}e_{it} - E(e_{is}e_{it})\Big]|^4 \leq M$.

d For each t, $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \lambda_i e_{it} \xrightarrow{d} N(0,\Gamma_t)$, as $N \to \infty$ where

$$\Gamma_t = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} E(\lambda_i \lambda_j{}' e_{it} e_{jt}).$$

e For each i, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_t e_{it} \xrightarrow{d} N(0,\Phi_i)$ as $T \to \infty$ where

$$\Phi_i = \lim_{T\to\infty} T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} E(F_t^0 F_s^{0\prime} e_{is} e_{it}).$$

Assumption LFE: $\{\lambda_i\}$, $\{F_t\}$, and $\{e_{it}\}$ are three mutually independent groups. Dependence within each group is allowed.

Assumption IE: For all $t \leq T$, $i \leq N$, $\sum_{s=1}^{T} |\tau_{s,t}| \leq M$, and $\sum_{i=1}^{N} |\bar{\sigma}_{ij}| \leq M$.

Assumptions F(0) (stationary factors) and L are moment conditions on the factors and the loadings respectively and are standard in factor models. They ensure that the factors are non-degenerate and that each factor has a nontrivial contribution to the variance of X_t. Assumption E concerns the idiosyncratic errors. Part (b) allows for 'weak' time series and cross-section dependence in the idiosyncratic errors so that the model has an approximate instead of a strict factor structure. The notion of an approximate factor model is due to Chamberlain and Rothschild (1983) who showed in the context of asset returns that so long as the largest eigenvalue of Ω is bounded, the idiosyncratic errors are allowed to be mildly cross-sectionally correlated. A similar assumption was used by Connor and Korajzcyk (1986). Our restrictions on the errors allow weak correlation not just cross-sectionally, but also serially; heteroskeadsticity is also allowed. Now under covariance stationarity with $E(e_{it}e_{jt}) = \sigma_{ij}$ for all t, the largest eigenvalue of Ω is bounded by $\max_i \sum_{j=1}^{N} |\sigma_{ij}|$. By assuming that $\sum_{j=1}^{N} |\sigma_{ij}| \leq M$ for all i and N as in (b), the assumptions of an approximate factor model in the sense of Chamberlain and Rothschild (1983) are satisfied. As Heaton and Solo (2006) noted, the maximum eigenvalue bound of Chamberlain and Rothschild can be obtained even if the number of correlated series increases with N. Our assumption, also used by Stock and Watson (2002a), indeed allows the number of strongly cross-correlated errors to grow at a rate slower than $\sqrt{N}$, and in this sense, we permit a stronger degree of cross-section correlation in the errors than the approximate factor model of Chamberlain and Rothschild. Part (d) permits weak dependence between the factors and the idiosyncratic errors and falls short of requiring F_t to be independent of e_{it}.

Within group dependence in Assumption LFE means that F_t can be serially correlated, λ_i can be correlated over i, and e_{it} can have serial and cross-sectional correlations. None of these correlations can be too strong if Assumption (E) is to hold. However, we assume no dependence between the factor loadings and the factors, or between the factors and the idiosyncratic errors, which is the meaning of mutual independence among the three groups.

Assumption IE strengthens E. When e_{it} is independent over time, the first part of (IE) is equivalent to requiring $E(e_{it}^2) \leq M$ for all t. Similarly, under cross-section independence, the second part of (IE) is equivalent to $E(e_{it}^2) \leq M$ for all i. Thus, under time series and cross-section independence, both parts of (IE) are implied by (E).

Let $\widehat{F}^k = \bar{F}^k(\bar{F}^{k\prime}\bar{F}^k/T)^{1/2}$. By the relationship given in Section 3,

$$\widehat{F}^k = \bar{F}^k(\widetilde{V})^{1/2} = \widetilde{F}^k\widetilde{V}^k$$

And for $k = r$, dropping the superscript, we have $\widehat{F} = \widetilde{F}\widetilde{V}$. Equivalently, for every t,

$$\widehat{F}_t = \widetilde{V}\widetilde{F}_t$$

Let V be an $r \times r$ diagonal matrix with the eigenvalues of $\Sigma_\Lambda^{1/2}\Sigma_F\Sigma_\Lambda^{1/2}$ as its elements. The matrix $\widetilde{V}$ has been shown to have an invertible limit, given by V, see Bai (2003). Since $\widetilde{V}$ is invertible, $\widehat{F}$ and $\widetilde{F}$ are equivalent in the sense that knowing one will lead to the other.

Our main results concerning the factor estimates can be summarized as follows. The first result is for an arbitrary k, and the remaining results are for $k = r$. Throughout, we let $C_{NT} = \min[\sqrt{N}, \sqrt{T}]$.

4.1 Result A: Factor Space

A.1 For any fixed $k > 1$ and under Assumptions F(0), L, and E,

$$C_{NT}^2 \left(\frac{1}{T} \sum_{t=1}^{T} \|\widehat{F}_t^k - \bar{H}^{k\prime} F_t^0\|^2 \right) = O_p(1).$$

where $\bar{H}^k = (\widetilde{F}^{k\prime} F^0/T)(\Lambda^{0\prime}\Lambda^0/N)$ with $rank(\bar{H}^k) = \min(k, r)$. If, in addition, $\sum_{s=1}^{T} \tau_{s,t} \leq M$ for all t and T, then

$$C_{NT}^2 \|\widehat{F}_t^k - \bar{H}^{k\prime} F_t^0\|^2 = O_p(1)$$

for each t. Note that matrix H^k depends on both N and T, but not on t.

A.2 Let $H = \widetilde{V}^{-1}(\widetilde{F}' F^0/T)(\Lambda^{0\prime}\Lambda^0/N)$, $Q = V^{1/2}\Upsilon'\Sigma_\Lambda^{-1/2}$, where Υ are the eigenvectors corresponding to the eigenvalues V of the matrix $\Sigma_\Lambda^{-1/2}\Sigma_F\Sigma_\Lambda^{-1/2}$. Under Assumptions F(0), L, E and LFE,

a if $\sqrt{N}/T \to 0$, then for each t,

$$\sqrt{N}(\widetilde{F}_t - H' F_t^0) \xrightarrow{d} N(0, V^{-1} Q \Gamma_t Q' V^{-1}).$$

b if $\sqrt{T}/N \to 0$, then for each i,

$$\sqrt{T}(\widetilde{\lambda}_i - H^{-1}\lambda_i^0) \xrightarrow{d} N(0, (Q')^{-1}\Phi_i Q^{-1}).$$

A.3 Let $A_{it} = \lambda_i^{0\prime}\Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1}\lambda_i^0$ and $B_{it} = F_t^{0\prime}\Sigma_F^{-1}\Phi_i\Sigma_F^{-1}F_t^0$, where Φ_i is the variance of $T^{-1/2}\sum_{t=1}^{T} F_t^0 e_{it}$;

a Under Assumptions F(0), L, E, LFE, and IE

$$(N^{-1}A_{it} + T^{-1}B_{it})^{-1/2}(\widetilde{C}_{it} - C_{it}^0) \xrightarrow{d} N(0, 1)$$

b if $N/T \to 0$, then $\sqrt{N}(\widetilde{C}_{it} - C_{it}) \xrightarrow{d} N(0, A_{it})$;

c if $T/N \to 0$, then $\sqrt{T}(\widetilde{C}_{it} - C_{it}) \xrightarrow{d} N(0, B_{it})$.

A.4 Suppose Assumptions F(0), L, E, and LFE hold. Then

$$\max_{1 \leq t \leq T} \left\| \widetilde{F}_t - H' F_t^0 \right\| = O_p(T^{-1/2}) + O_p((T/N)^{1/2}).$$

As we can only estimate the space spanned by the factors, all results concerning the factor estimates are stated in terms of the difference between estimated and the true factor space,

$H^{k\prime}F_t^0$. Part 1, derived by Bai and Ng (2002), says that if we estimate k (not necessarily the same as r) factors, the average squared deviation between the k estimated factors and the space spanned by k of the true factors will vanish at rate $\min[N,T]$, or in other words, the smaller of the sample size in the two dimensions. Result A.1 is useful for determining the number of factors, which will be considered later on. The second part of A.1 imposes stronger condition on the error correlations.

Result A.1 has been extended by Anderson and Vahid (2007) to allow for jumps, an issue that is relevant in volatility analysis. The authors argue that jumps can distort the principal components estimates. Treating jumps as measurement error, an IV approach is used to correct for the bias. Their IV estimate of the factors are the eigenvectors of the covariance matrix of $\widehat{X}$, where $\widehat{X}$ is the orthogonal projection of X on lags of X.

Whereas A.1 provides the starting point for estimation, the asymptotic distribution is required to assess the sampling variability of the factor estimates. These results are derived by Bai (2003) and are summarized in A.2 and A.3. Essentially, for each t, $\widetilde{F}_t$ is $\sqrt{N}$ consistent for the true factor space while for each i, $\widetilde{\lambda}_i$ is $\sqrt{T}$ consistent for the space spanned by the true factor loadings. Although these results put restrictions between N and T, note that N and T pass to infinity jointly, not sequentially. It is clear that $\sqrt{N}/T \to 0$ and $\sqrt{T}/N \to 0$ are not strong conditions. The approximation should work well even for panels with just 50 units in each dimension. A.4 provides an upper bound on the maximum deviation of the estimated factors from the space spanned by the true ones. The upper bound goes to zero if $T/N \to 0$. A sharper bound is obtained in Bai and Ng (2008). If there exists $\ell \geq 4$ such that $E\|F_t\|^\ell \leq M$ and $E\|N^{-1/2}\sum_{i=1}^N \lambda_i^0 e_{it}\|^\ell \leq M$ for all t, then

$$\max_{1\leq t\leq T} \|\widetilde{F}_t - H'F_t^0\| = O_p(T^{-1+1/\ell}) + O_p(T^{1/\ell}/N^{1/2}). \tag{7}$$

For $\ell = 4$ as is assumed, the maximum deviation goes to zero provided that $T/N^2 \to 0$.

The sampling uncertainty of $\widetilde{F}_t$ is captured by $Avar(\widetilde{F}_t) = V^{-1}Q\Gamma_t Q'V^{-1}$, which can be rewritten as $V^{-1}QH(H^{-1}\Gamma_t H^{-1\prime})H'Q'V^{-1}$. Matrix $\widetilde{V}^{-1}$ is a consistent estimate for V^{-1}. Matrix Q is the limit of $\widetilde{F}'F^0/T$. An approximation for QH is obtained by substituting $\widetilde{F}$ for F^0H since $\widetilde{F}$ provides an approximation for F^0H. Note that $\widetilde{Q} = \widetilde{F}'\widetilde{F}/T$ is just an identity matrix by construction (the limit of QH is indeed an identity). It remains to obtain a consistent estimate for $H^{-1}\Gamma_t H^{-1\prime}$. Denote its consistent estimate by $\widetilde{\Gamma}_t$. This leads to

$$\widehat{Avar(\widetilde{F}_t)} = \widetilde{V}^{-1}\widetilde{\Gamma}_t\widetilde{V}^{-1}.$$

Several estimators of the $r \times r$ matrix $H^{-1}\Gamma_t H^{-1\prime}$ are possible. There is no need to estimate H. When estimating Γ_t, we have to use $\widetilde{\lambda}_i$ in place of λ_i because λ_i cannot be observed. Since $\widetilde{\lambda}_i = H^{-1}\lambda_i + o_p(1)$, H^{-1} is implicitly estimated.

Result B: Estimation of Γ_t Let $\widetilde{e}_{it} = x_{it} - \widetilde{\lambda}_i'\widetilde{F}_t$:

B.1 cross-sectionally independent but heterogeneous panels: let

$$\widetilde{\Gamma}_t = \frac{1}{N}\sum_{i=1}^{N} \widetilde{e}_{it}^2 \widetilde{\lambda}_i \widetilde{\lambda}_i'.$$

Under Assumption F(0), L, E, and LFE, we have $\|\widetilde{\Gamma}_t - H^{-1}\Gamma_t H^{-1\prime}\| \xrightarrow{p} 0$.

B.2 cross-sectionally independent and homogeneous panels: let

$$\widetilde{\Gamma}_t = \widetilde{\sigma}_e^2 \frac{1}{N}\sum_{i=1}^{N} \widetilde{\lambda}_i \widetilde{\lambda}_i'$$

where $\widetilde{\sigma}_e^2 = \frac{1}{NT - r(T+N-r)} \sum_{i=1}^{N}\sum_{t=1}^{T} \widetilde{e}_{it}^2$. The result of B.1 still holds.

B.3 cross-sectionally correlated but stationary panels: let

$$\widetilde{\Gamma} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \widetilde{\lambda}_i \widetilde{\lambda}_j' \frac{1}{T}\sum_{t=1}^{T} \widetilde{e}_{it}\widetilde{e}_{jt}.$$

Suppose Assumption F(0), L, E and LFE hold, and $E(e_{it}e_{jt}) = \sigma_{ij}$ for all t so that $\Gamma_t = \Gamma$ not depending on t. If n is such that with $n/\min[N,T] \to 0$, then

$$\|\widetilde{\Gamma} - H^{-1\prime}\Gamma H^{-1}\| \xrightarrow{p} 0.$$

The estimator defined in B.3 is referred to the CS-HAC estimator in Bai and Ng (2006a). It is robust to cross-section correlation and cross-section heteroskedasticity but requires the assumption of covariance stationarity. If there is cross-section correlation of an unknown form and the data have no natural ordering, estimation of Γ_t runs into problems analogous to the estimation of the spectrum in time series. Specifically, some truncation is necessary to obtain a consistent estimate of Γ_t. Here in a panel setting, we use the fact that if covariance stationarity holds, the time series observations will allow us to consistently estimate the cross-section correlations provided T is large. Furthermore, the covariance matrix of interest is of dimension $(r \times r)$ and can be estimated with $n < N$ observations. The assumption of covariance stationarity is not necessary for B.1 and B.2 since these assume cross-sectionally uncorrelated idiosyncratic errors.

4.2 Estimating the Number of Factors

Some economic models have a natural role for factors and thus determining the number of factors is of interest in its own right. For example, underlying the APT theory of Ross (1976) is the assumption that there are common risk factors across assets, while underlying consumer demand analysis is the notion that there are individual factors common across goods. From a statistical standpoint, being able to consistently estimate the number of factors enables the researcher to treat r as known, so that we can simply deal with the $r \times 1$ vector of factor estimates $\widetilde{F}_t$, instead of a sequence of factor estimates $\widetilde{F}_t^k$. The result concerning estimation of r can be summarized as follows:

Let $S(k) = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \widehat{\lambda}_i^{k\prime} \widehat{F}_t^k)^2$ be the sum of squared residuals (divided by NT) when k factors are estimated. Let $g(N,T)$ be a penalty function. Define the information criteria

$$PCP(k) = S(k) + k\,\bar{\sigma}^2\, g(N,T)$$

$$IC(k) = \ln(S(k)) + k\, g(N,T)$$

where $\bar{\sigma}^2$ is equal to $S(kmax)$ for a pre-specified value $kmax$. The second criterion does not depend on $\bar{\sigma}^2$. The estimator for the number of factors is defined as

$$\begin{aligned} \widehat{k}_{PCP} &= \underset{0 \le k \le kmax}{\text{argmin}}\ PCP(k) \\ \widehat{k}_{IC} &= \underset{0 \le k \le kmax}{\text{argmin}}\ IC(k) \end{aligned}$$

Result C: Number of Static Factors

C. 1 Suppose Assumptions F(0), L, E, and LFE hold. If (i) $g(N,T) \to 0$ and (ii) $C_{NT}^2 g(N,T) \to 0$ as $N, T \to \infty$, then

$$\text{prob}(\widehat{k}_{PCP} = r) \to 1, \text{ and } \text{prob}(\widehat{k}_{IC} = r) \to 1.$$

Result C is based on Bai and Ng (2002). The first condition in C.1 prevents under-fitting while the second condition prevents over-fitting. Similar looking conditions underlie well-known model selection procedures such as the AIC and the BIC that are often used in time series and cross section analysis. The notable and important difference is that our penalty factor $g(N,T)$ depends on the sample size in both dimensions, not just N or T. In problems for which $\sqrt{T}$ consistent estimates are available, condition (ii) would require divergence of $Tg(T)$. But in our particular panel setting, the convergence rate of the estimated factor space dictates that $\min[N,T]g(N,T)$ must diverge. Examples of $g(N,T)$ that satisfy the required conditions are

$$\begin{aligned} g_1(N,T) &= \frac{N+T}{NT}\ln(\frac{NT}{N+T}) \\ g_2(N,T) &= \frac{N+T}{NT}\ln C_{NT}^2 \\ g_3(N,T) &= \frac{\ln C_{NT}^2}{C_{NT}^2}. \end{aligned}$$

We have frequently used $g_2(N,T)$ in empirical work because it tends to be more stable. In our experience,

$$g_4(N,T) = (N+T-k)\ln(NT)/NT$$

has good properties especially when the errors are cross correlated. Strictly speaking, $g_4(N,T)$ fails condition (i) when $T = \exp(N)$ or $N = \exp(T)$. But these configurations of N and T do not seem empirically relevant. Thus, $g_4(N,T)$ should not be ruled out in practice.

Random matrix theory has recently been used to determine the number of factors. The idea is to exploit the largest and the smallest eigenvalue of large matrices whose properties are known for iid normal data. See Onatski (2005) and Kapetanios (2007). These methods use different and often stronger assumptions than the ones stated in the present paper. In simulations, the tests seem to have good properties when the factor structure is weak, or the idiosyncratic errors are highly serially or cross-sectionally correlated. Whether a factor model with these characteristics is a better characterization of the data remains debatable. It should be noted that a positive limit for $\Lambda'\Lambda/N$ still permits zero factor loadings for many series, implying weak factors for the corresponding series.

Onatski (2006b) developed a formal test for the number of factors in data with correlated Gaussian idiosyncratic errors. The idea is to test the slope of the scree plot to identify changes in curvature.[4] His analysis provides a formal justification for the use of the scree plot in determining the number of factors, and appears to be the first to obtain the asymptotic distribution of tests that determine the number of factors. Under the assumptions in

[4]A scree diagram is a plot of the ordered eigenvalues against the corresponding order number.

Onatski (2006b) and the large sample properties of random matrices, his proposed test is characterized by the Tracy-Widom distribution.

There are specific instances when knowledge of the number of dynamic factors, q, is useful. If the static factors F_t are driven by $q \le r$ primitive innovations (say, ε_t), then the innovations of F_t and ε_t are related by $u_t = R\varepsilon_t$. It then follows that $\Sigma_u = R\Sigma_\varepsilon R'$ has rank $q \le r$, and F_t can be represented as

$$A(L)F_t = R\varepsilon_t.$$

These features lead to two ways of determining q using time domain methods.

The first, from Stock and Watson (2005) and Amengual and Watson (2007), takes the dynamic factor model $x_{it} = \lambda_i' F_t + \rho_i(L)x_{it-1} + e_{it}$ as starting point. Rewriting $A(L) = I - A^+(L)L$ together with $A(L)F_t = R\varepsilon_t$ yields

$$x_{it} = \lambda_i' A^+(L)F_{t-1} + \rho_i(L)x_{it-1} + \lambda_i' R\varepsilon_t + e_{it}.$$

Then $w_t = \lambda_i' R\varepsilon_t + e_{it}$ has a factor structure. This factor structure has $q = dim(\varepsilon_t)$ dynamic factors. The PCP and IC criteria can be used to determine q, upon replacing w_{it} by a consistent estimate.

The second, developed in Bai and Ng (2007), starts with the premise that if the $r \times r$ matrix Σ_u has rank q, then the $r - q$ smallest eigenvalues are zero. Let $c_1 > c_2 \ldots > c_N$ be the ordered eigenvalues of Σ_u. Define

$$D_{1k} = \left(\frac{c_{k+1}^2}{\sum_{j=1}^r c_j^2} \right)^{1/2} \quad \text{and} \quad D_{2,k} = \left(\frac{\sum_{j=k+1}^r c_j^2}{\sum_{j=1}^r c_j^2} \right).$$

The test is based on the idea that when the true eigenvalues $c_{q+1}, ..., c_r$ are zero, $D_{1,k}$ and $D_{2,k}$ should be zero for $k \ge q$. To estimate Σ_u, a VAR model is fitted to the estimated factors $\bar{F}_t$ (but not $\widetilde{F}_t$). Let $\widehat{u}_t$ be the residuals from estimation of a VAR(p) in $\bar{F}_t$, and let $\widehat{\Sigma}_u = \frac{1}{T-p}\sum_{t=1}^T \widehat{u}_t \widehat{u}_t'$. Let $\widehat{c}_1 > \widehat{c}_2 \ldots \widehat{c}_r$ be the eigenvalues of $\widehat{\Sigma}_u$. Replace c_j in D_{1k} and $D_{2,k}$ by $\widehat{c}_j$ to obtain $\widehat{D}_{1k}$ and $\widehat{D}_{2k}$. For some $0 < m < \infty$ and $0 < \delta < 1/2$, let

$$M_{NT}(\delta) = \frac{m}{\min[N^{1/2-\delta}, T^{1/2-\delta}]}.$$

As there are sampling errors from estimation of Σ_u, $\widehat{D}_{1,k}$ and $\widehat{D}_{2,k}$ will not be zero exactly. The cut-off point of $M_{NT}(\delta)$ is used to account for estimation error. We now summarize the properties of these two tests.

Result D: Number of Dynamic Factors

D.1 Stock and Watson (2005). Let $\widehat{w}_{it}$ be the residuals from a restricted regression of x_{it} on the lags of $\widetilde{F}_t$ and x_{it}. Let $S(k)$ be the sum of squared residuals when k factors are estimated from the data matrix $\{\widehat{w}_{it}\}$. Let $\widehat{q} = \text{argmin}_k\, PCP(k)$ or $\widehat{q} = \text{argmin}_k\, IC(k)$. If (i) $g(N,T) \to 0$ and (ii) $C_{NT}^2 g(N,T) \to \infty$ as $N,T \to \infty$,

$$\text{prob}(\widehat{q} = q) \to 1$$

D.2 Bai and Ng (2007). Let $\widehat{q}_1$ be the smallest k such that $\widehat{D}_{1,k} < M_{NT}(\delta)$ and $\widehat{q}_2$ be the smallest k such that $\widehat{D}_{2,k} < M_{NT}(\delta)$. For $0 < m < \infty$ and $0 < \delta < 1/2$, as $N,T \to \infty$

$$P(\widehat{q} = q) \to 1$$

Both approaches enable q to be determined without having to estimate the spectrum of F. Note, however, that the Bai-Ng method assumes F_t is estimated by $\bar{F}_t$ (ie. with normalization $\Lambda'\Lambda/N = I_k$) and is not valid when $\widetilde{F}_t$ is used. The Stock-Watson method is flexible on which normalization is used to construct the principle component estimates.

Hallin and Liska (2007) proposed a method for determining the number of factors in a generalized dynamic factor model. The test is based on the eigenvalues $c_i(\theta)$ of the spectral density matrix of the data at frequency θ, denoted $\Sigma(\theta)$. Specifically, for $m = -M_T, \ldots M_T$, and $\theta_m = \frac{\pi m}{M_T + 1/2}$ with $M_T \to \infty$ and $M_T T^{-1} \to 0$, define

$$PCP(k) = \frac{1}{N} \sum_{i=k+1}^{N} \frac{1}{2M_T + 1} \sum_{m=-M_T}^{M_T} c_i(\theta_m) + k\bar{g}(N,T), \qquad 0 \le k \le q_{max}$$

and

$$IC(k) = \log \left[\frac{1}{N} \sum_{i=k+1}^{N} \frac{1}{2M_T + 1} \sum_{m=-M_T}^{M_T} c_i(\theta_m) \right] + k\bar{g}(N,T).$$

Under somewhat different assumptions from the ones considered here, the authors showed that if (i) $\bar{g}(N,T) \to 0$ and (ii) $\min[N, M_T^{-2}, M_T^{1/2} T^{1/2}]\bar{g}(N,T) \to \infty$, then $P(\widehat{q} = q) \to 1$. The authors suggest using $M_T = .5\sqrt{T}$ and penalties $\bar{g}_1(N,T) = (M_T^{-2} + M_T^{1/2} T^{-1/2} + N^{-1}) \log A_T$, $\bar{g}_2(N,T) = A_T$, and $\bar{g}_3(N,T) = A_T^{-1} \log(A_T)$ with $A_T = \min[N, M_T^2, M_T^{-1/2} T 1/2]$. Not surprisingly, their method outperforms $\widehat{D}_{1,k}$ and $\widehat{D}_{2,k}$ when the loadings have an autoregressive structure, which does not satisfy the assumption of Bai and Ng (2007). Implementation of their procedure requires careful selection of the auxiliary parameters. The authors suggest cross-validation as a possibility.

Work is on-going to find improved ways to estimate r. The issue is important as estimation and inference often hinge on precise estimation of r and q. One can expect new and improved tests to appear.

5 Applications

The motivation for considering factor analysis is to deal with large number of variables. Use of a small number of factors as conditioning variables is a parsimonious way to capture information in a data rich environment, and there should be efficient gains over (carefully) selecting a handful of predictors. There are three strands of research that evolve around this theme. The first concerns using the estimated factors as predictors; the second uses the estimated factors as improved instruments over observed variables. The third concerns testing the validity of observed proxies. These are discussed in the next three subsections.

5.1 Factor Augmented Regressions

The distinctive feature of a so-called 'factor-augmented' regression is to add factors estimated by the method of principal components to an otherwise standard regression:

$$\begin{aligned} y_{t+h} &= \alpha'\widetilde{F}_t + \beta' W_t + \varepsilon_{t+h} \\ &= \widetilde{z}'_{t+h}\delta + \varepsilon_{t+h} \end{aligned} \tag{8}$$

where W_t are predetermined variables (such as lags) that the researcher would include whether or not $\widetilde{F}_t$ is available. The following assumptions will be used.

Assumption FAR

a Let $z_t = (F_t' \quad W_t')'$, $E\|z_t\|^4 \le M$; $E(\varepsilon_{t+h}|y_t, z_t, y_{t-1}, z_{t-1}, \ldots) = 0$ for any $h > 0$; z_t and ε_t are independent of the idiosyncratic errors e_{is} for all i and s. Furthermore, $\frac{1}{T}\sum_{t=1}^T z_t z_t' \xrightarrow{p} \Sigma_{zz} > 0$

b $\frac{1}{\sqrt{T}}\sum_{t=1}^T z_t\varepsilon_{t+h} \xrightarrow{d} N(0, \Sigma_{zz,\varepsilon})$, where $\Sigma_{zz,\varepsilon} = \text{plim}\,\frac{1}{T}\sum_{t=1}^T(\varepsilon_{t+h}^2 z_t z_t')) > 0$.

The regression model given by (8) encompasses many applications of interest. If $h = 0$, (8) is simply a regression with generated regressors $\widetilde{F}_t$, and $\widetilde{z}_t'\widehat{\delta}$ is the estimated conditional mean of y_t. For example, if y_t is stock returns, then $\widetilde{z}_t'\widehat{\delta}$ is the estimated conditional mean of stock returns, and if y_t is the volatility of stock returns, then $\widetilde{z}_t'\widehat{\delta}$ is the estimated conditional volatility of stock returns with conditioning information $\widetilde{z}_t$. The ratio of the two estimates is the conditional Sharp ratio, which is useful in studying the risk-return trade-off of stock returns.

When $h > 0$, (8) is a forecasting equation and forms the basis of the so-called Diffusion Index (DI) forecasting methodology of Stock and Watson (2002a). The advantage of DI is that $\widetilde{F}_t$ drastically reduces the dimension of the information spanned by the predictors from N to a much smaller number as determined by the number of factors included. As such, it is capable of exploiting the information in a large panel of data while keeping the size of the forecasting model small. The DI framework is now used by various government agencies in a number of countries, as well as by independent and academic researchers alike.

In addition to forecasting, (8) also provides a framework for assessing predictability. For example, under the expectations hypothesis, excess bond returns should be unpredictable given the information available. Predictability tests are often sensitive to the choice of the predictors. A finding of non-predictability in a small information set may not be robust to changes in the set of predictors, or the extent to which the variable of interest is predictable. The factor augmented framework is well suited for these analysis because the inclusion of $\widetilde{F}_t$ brings the empirical problem closer to the conceptual problem of testing predictability with respect to 'all information available'. Not only can we assess if the 'usual suspects' W_t contain all relevant information (and that all other predictors are irrelevant), we can do so without much efficiency loss because $\widetilde{F}_t$ is of small dimension.

As written, equation (8) is a single equation. But multivariate models can also accommodate $\widetilde{F}_t$. More specifically, if y_t is a vector of m series, and F_t is a vector of r factors, a Factor Augmented Vector Autoregression of order p, or simply FAVAR(p), can be obtained.

$$\begin{aligned} y_{t+1} &= \textstyle\sum_{k=0}^{p} a_{11}(k) y_{t-k} + \sum_{k=0}^{p} a_{12}(k) \widetilde{F}_{t-k} + \varepsilon_{1t+1} \\ \widetilde{F}_{t+1} &= \textstyle\sum_{k=0}^{p} a_{21}(k) y_{t-k} + \sum_{k=0}^{p} a_{22}(k) \widetilde{F}_{t-k} + \varepsilon_{2t+1}, \end{aligned}$$

where $a_{11}(k)$ and $a_{21}(k)$ are coefficients on y_{t-k}, while $a_{12}(k)$ and $a_{22}(k)$ are coefficients on $\widetilde{F}_{t-k}$.

The following results validate estimation and inference based upon (8).

Result E: Linear Factor Augmented Regressions, Bai and Ng (2006a). Suppose Assumptions F(0), L, E, and FAR hold.

E1 If $\sqrt{T}/N \to 0$, then

$$\sqrt{T}(\widehat{\delta}-\delta)\xrightarrow{d}N\Big(0,\Sigma_\delta\Big)$$

where $\Sigma_\delta = \Phi_0'^{-1}\Sigma_{zz}^{-1}\Sigma_{zz,\varepsilon}\Sigma_{zz}^{-1}\Phi_0^{-1}$, $\Phi_0 = diag(V^{-1}Q\Sigma_\Lambda, I)$ is block diagonal, $V = \text{plim}\,\widetilde{V}$, $Q = \text{plim}\,\widetilde{F}'F/T$, and Σ_Λ defined in Assumption E. A consistent estimator for $Avar(\widehat{\delta}) = \Sigma_\delta$ is

$$\widehat{Avar(\widehat{\delta})} = \Big(\frac{1}{T}\sum_{t=1}^{T-h}\widehat{z}_t\widehat{z}_t'\Big)^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}_{t+h}^2\widehat{z}_t\widehat{z}_t'\Big)\Big(\frac{1}{T}\sum_{t=1}^{T-h}\widehat{z}_t\widehat{z}_t'\Big)^{-1}.$$

E2 Consider a p-th order vector autoregression in m observable variables y_t and r factors, $\widetilde{F}_t$, estimated by the method of principal components. Let $\widehat{z}_t = (y_t' \ \ldots y_{t-p}', \widetilde{F}_t', \ldots, \widetilde{F}_{t-p}')'$. Define $\widehat{Y}_t = (y_t', \widetilde{F}_t')'$ and let $\widehat{Y}_{jt}$ be the jth element of $\widehat{Y}_t$. For $j = 1, \ldots m+r$, let $\widehat{\delta}_j$ be obtained by least squares from regressing $\widehat{Y}_{jt+1}$ on $\widehat{z}_t$. Let $\widehat{\varepsilon}_{jt+1} = \widehat{Y}_{jt+1} - \widehat{\delta}_j'\widehat{z}_t$. If $\sqrt{T}/N \to 0$ as $N, T \to \infty$,

$$\sqrt{T}(\widehat{\delta}_j-\delta_j)\xrightarrow{d}N\bigg(0, \text{plim}(\frac{1}{T}\sum_{t=1}^{T}\widehat{z}_t\widehat{z}_t')^{-1}\bigg(\frac{1}{T}\sum_{t=1}^{T}(\widehat{\varepsilon}_{jt})^2\widehat{z}_t\widehat{z}_t'\bigg)(\frac{1}{T}\sum_{t=1}^{T}\widehat{z}_t\widehat{z}_t')^{-1}\bigg).$$

E3 Let $\widehat{y}_{T+h|T} = \widehat{\delta}'\widehat{z}_T$. If $\sqrt{N}/T \to 0$ and Assumptions F(0), L and E hold, then for any $h \geq 0$,

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

where $var(\widehat{y}_{T+h|T}) = \frac{1}{T}\widehat{z}_T'\,Avar(\widehat{\delta})\,\widehat{z}_T + \frac{1}{N}\widehat{\alpha}'\,Avar(\widetilde{F}_T)\,\widehat{\alpha}$.

Result E states that parameter estimates of equations involving $\widetilde{F}_{t+1}$, whether as regressand or regressors, are $\sqrt{T}$ consistent. Result E also shows how standard errors can be computed and provides a complete inferential theory for factor augmented regressions.

Conventionally generated regressors are obtained as the fitted values from a first step regression of some observed variable related to the latent variable of interest on a finite set of other observed regressors. As shown in Pagan (1984), sampling variability from the first step estimation is $O_p(1)$ in the second stage. Consequently, the standard errors of the second step parameter estimates must account for the estimation error from the first step. As indicated by Result E, no such adjustment is necessary when the generated regressors are $\widetilde{F}_t$ if $\sqrt{T}/N \to 0$. This is because the term that is $O_p(1)$ in a conventional setting is $O_p(\frac{\sqrt{T}}{\min[N,T]})$ in the factor augmented regression setting, and this term vanishes if $\sqrt{T}/N \to 0$. Note, however, that although the condition $\sqrt{T}/N \to 0$ is not stringent, it does put discipline on when estimated factors can be used in regression analysis.

Result E3 concerns the prediction interval for the conditional mean. There are two terms in $var(\widehat{y}_{T+h|T})$, and the overall convergence rate for $\widehat{y}_{T+h|T}$ is $\min[\sqrt{T}, \sqrt{N}]$. In a standard setting, $var(\widehat{y}_{T+h|T})$ falls at rate T, and for a given T, it increases with the number of observed predictors through a loss in degrees of freedom. With factor forecasts, the error variance decreases at rate $\min[N,T]$, and for a given T, efficiency improves with the number of predictors used to estimate F_t. A large N enables better estimation of the common factors and thus directly affects the efficiency of subsequent estimations involving $\widetilde{F}_t$. Note that if the estimation of F_t was based upon a fixed N, consistent estimation of the factor space would not be possible however large T becomes. That is to say, Result E will not apply if we simply use $\widetilde{F}_t$ to reduce the dimension of an already small set of predictors.

5.2 Extremum Estimation

The estimated factors can be used in estimation not just in linear models, but also when the models are non-linear in the parameters. Suppose we observe $z_t = (y_t, w_t, F_t)$ where y_t is typically the dependent variables and (w_t, F_t) is a set of explanatory variables ($t = 1, \dots T$). Consider the problem of estimating θ as

$$\widehat{\theta} = \operatorname*{argmax}_{\theta} Q_T(\theta), \qquad Q_T = \frac{1}{T}\sum_{t=1}^{T} m(z_t, \theta).$$

If z_t are iid with density function $f(z_t|\theta)$, the MLE can be written as $m(z_t, \theta) = \log f(z_t|\theta)$. For nonlinear regressions, $y_t = g(w_t, F_t, \theta) + \varepsilon_t$, the nonlinear squares method is equivalent to $m(z_t, \theta) = -[y_t - g(w_t, F_t, \theta)]^2$. Under some regularity conditions, $\widehat{\theta}$ can be shown to be $\sqrt{T}$ consistent and asymptotically normal. However, if F_t is not observed but we instead observe a large panel of data x_{it} with a factor structure, $x_{it} = \lambda_i' F_t + e_{it}$, we can replace F_t

by $\widetilde{F}_t$ in estimation. Replacing z_t by $\widetilde{z}_t = (y_t, w_t, \widetilde{F}_t)$ gives the feasible objective function

$$\widetilde{Q}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} m(\widetilde{z}_t, \theta)$$

and the estimator is defined as

$$\widetilde{\theta} = \text{argmax}\, \widetilde{Q}_T(\theta)$$

Let $h(z_t, \theta) = \frac{\partial m(z_t,\theta)}{\partial \theta}$ and $K(z_t, \theta) = \frac{\partial^2 m(z_t,\theta)}{\partial\theta\partial\theta'}$. The following assumptions are made:

Assumption EE

(i): $Eh(z_t; \theta_0) = 0$, $Eh(z_t; \theta) \neq 0$ for all $\theta \neq \theta_0$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h(z_t, \theta_0) \xrightarrow{d} N(0, \Sigma)$ where Σ is positive definite.

(ii): $E \sup_\theta \|h(z_t; \theta)\| < \infty$.

(iii)): $h(z; \theta)$ is twice continuously differentiable with respect to F_t and θ, and $K_0 = EK(z_t, \theta_0)$ is invertible.

(iv): For some $b_{NT} \to 0$

$$\sup_{\{F_t^*: \|F_t^* - F_t\| \leq b_{NT}, \forall t\}} \sup_{\theta \in \Theta} \frac{1}{T}\sum_{t=1}^{T} \|h(y_t, x_t, F_t^*, \theta)\|^2 = O_p(1) \tag{9}$$

(v): $\xi_t = (\partial/\partial F_t) h(z_t, \theta_0)$ is uncorrelated with e_{it} and $E\|\xi_t\|^2 \leq M$ for all t.

(vi): For $j = 1, 2, ..., p = dim(\theta)$, and for some $b_{NT} \to 0$,

$$\sup_{\{F_t^*: \|F_t^* - F_t\| \leq b_{NT}, \forall t\}} \sup_{\{\theta^*: \|\theta^* - \theta^0\| \leq b_{NT}\}} \frac{1}{T}\sum_{t=1}^{T} \|\frac{\partial^2 h_j(y_t, x_t, F_t^*, \theta^*)}{\partial\eta_1 \partial\eta_2'}\|^2 = O_p(1)$$

for $\eta_1, \eta_2 \in \{F_t, \theta\}$.

Assumptions EE(i)-(iii) are sufficient conditions for consistency and asymptotic normality when z_t is observable. Assumption EE(iv) guarantees consistency of $\widetilde{\theta}$ when F_t is replaced by $\widetilde{F}_t$. The remaining conditions are for asymptotic normality when using $\widetilde{F}_t$ in place of F_t.

Result F: Extreme Estimation Bai and Ng (2008).

Under Assumptions of EE, F(0), L, E, and $T^{5/8}/N \to 0$, we have

$$\sqrt{T}(\widetilde{\theta} - \theta_0) \xrightarrow{d} N(0, K_0^{-1}\Sigma K_0^{-1}).$$

The limiting distribution is the same as if F_t are observable. For linear models, only $T^{1/2}/N \to 0$ is required. A larger N is required for nonlinear models. Bai and Ng (2008) also consider a GMM estimator and obtain similar results.

5.3 Instrumental Variable Estimation,

A different use of $\widetilde{F}_t$ is as instrumental variables. Consider the regression model

$$\begin{aligned} y_t &= x_{1t}'\beta_1 + x_{2t}'\beta_2 + \varepsilon_t \\ &= x_t'\beta + \varepsilon_t. \end{aligned}$$

where $x_t = (x_{1t}', x_{2t}')'$ is $K \times 1$. The $K_1 \times 1$ regressors x_{1t} are exogenous or predetermined and may include lags of y_t. The $K_2 \times 1$ regressors x_{2t} are endogenous; $K = K_1 + k_2$. Suppose that

$$x_{2t} = \Psi' F_t + u_t$$

and $E(u_t\varepsilon_t) \neq 0$ and $E(F_t\varepsilon_t) = 0$. This implies that $E(x_{2t}\varepsilon_t) \neq 0$, and thus x_{2t} is endogenous. The least squares estimator of β will be inconsistent. The conventional treatment of endogeneity bias is to use lags of y, x_1 and x_2 as instruments for x_2. Our point of departure is to note that in this setting, F_t are the ideal instruments in the sense of satisfying both instrument exogeneity and instrument relevance. The reason why the moments $g_t = F_t\varepsilon_t$ are not used to estimate β is that F_t are not observed. We assume that there is a 'large' panel of valid instruments, $z_{1t}, \ldots z_{Nt}$ that are weakly exogenous for β and generated as follows:

$$z_{it} = \lambda_i' F_t + e_{it}. \tag{10}$$

The idea is to estimate F from the above factor model and then use the estimated F as instruments for x_{2t}. For a related study, see Kapetanios and Marcellino (2006a). The following analysis assumes no x_{1t} for simplicity. Otherwise, use $F_t^+ = (x_{1t}', F_t')'$ in place of F_t as instrument.

Assumption IV

a. $E(\varepsilon_t) = 0$, $E|\varepsilon_t|^{4+\delta} < \infty$ for some $\delta > 0$. The $r \times 1$ vector process $g_t(\beta^0) = F_t\varepsilon_t(\beta^0)$ satisfies $E[g_t(\beta^0)] \equiv E(g_t^0) = 0$ with $E(g_t(\beta)) \neq 0$ when $\beta \neq \beta^0$. Furthermore, $\sqrt{T}\bar{g}^0 \xrightarrow{d} N(0, S^0)$ where S^0 is the asymptotic variance of $\sqrt{T}\bar{g}^0$, and $\bar{g}^0 = \frac{1}{T}\sum_{t=1}^T F_t\varepsilon_t(\beta^0)$.

b. x_{1t} is predetermined such that $E(x_{1t}\varepsilon_t) = 0$.

c. $x_{2t} = \Psi' F_t + u_t$ with $\Psi'\Psi > 0$, $E(F_t u_t) = 0$, and $E(u_t\varepsilon_t) \neq 0$.

d. For all $i = 1, \ldots N$, $E(e_{it}u_t) = 0$, and $E(e_{it}\varepsilon_t) = 0$.

Assumption IV.c restricts consideration to the case of strong instruments, while IV.d assumes that z_{it} are valid instruments. Let $g_t = \widetilde{F}_t\varepsilon_t$, $\bar{g} = \frac{1}{T}\sum_{t=1}^T g_t$, and let $\breve{S} = \frac{1}{T}\sum_{t=1}^T(\breve{g}_t\breve{g}_t')$, with $\breve{g}_t = \widetilde{F}_t\breve{\varepsilon}_t$, and $\breve{\varepsilon}_t = y_t - x_t'\breve{\beta}_{FIV}$, where $\breve{\beta}_{FIV}$ is an initial consistent estimator for β. The two-step efficient estimator with $\widetilde{F}$ as instruments is defined as

$$\widehat{\beta}_{FIV} = \underset{\beta}{\operatorname{argmin}}\, \bar{g}(\beta)'\breve{S}^{-1}\bar{g}(\beta).$$

Result G: Factor IV

G.1 Under Assumptions F, L, E, LFE and FIV,

$$\sqrt{T}(\widehat{\beta}_{FIV} - \beta^0) \xrightarrow{d} N\Big(0, Avar(\widehat{\beta}_{FIV})\Big)$$

where $Avar(\widehat{\beta}_{FIV})$ is the probability limit of $(S_{\widetilde{F}x}(\widehat{S})^{-1}S_{\widetilde{F}x}')^{-1}$, which is therefore a consistent estimator for the asymptotic variance, and $S_{\widetilde{F}x} = \frac{1}{T}\sum_{t=1}^T \widetilde{F}_t x_t$.

G.2 Let z_2 be a subset of r of the N observed instruments $(z_{1t}, \ldots z_{Nt})$. Let $m_t = z_{2t}(y_t - x_t'\beta)$ with $\sqrt{T}\bar{m} \xrightarrow{d} N(0, Q)$. Let $\widehat{\beta}_{IV}$ be the minimizer of $\bar{m}'(\breve{Q})^{-1}\bar{m}$ with the property that $\sqrt{T}(\widehat{\beta}_{IV} - \beta^0) \xrightarrow{d} N(0, Avar(\widehat{\beta}_{IV}))$. Under the assumption that $E(e_{it}^2) > 0$ for all i in the subset of z_2,

$$Avar(\widehat{\beta}_{IV}) - Avar(\widehat{\beta}_{FIV}) \geq 0.$$

Result G is based on Bai and Ng (2006c). G.1 and G.2 say that not only do the instruments $\widetilde{F}_t$ yield $\sqrt{T}$ consistent and asymptotically normal estimates, but that the estimates so obtained are more efficient than using an equal number of observed instruments. The intuition is straightforward. The observed instruments are the ideal instruments contaminated with errors while $\widetilde{F}$ is consistent for the ideal instrument space. More efficient instruments thus lead to more efficient estimates. Pooling information across the observed variables washes out the noise to generate more efficient instruments for x_2. Result F1 is stated assuming that all instruments are valid, but this can be relaxed. Result F1 still holds if $\sum_{i=1}^{N} |E(\varepsilon_t e_{it})| \leq M < \infty$ for all N with M not depending on N, and $\sqrt{T}/N \to 0$. Thus in a data rich environment, use of invalid instruments does not preclude consistent estimation. However, use of invalid instruments will not yield consistent estimates if N is fixed; this highlights how a large N and T can open up new horizons for estimation and inference.

The factor estimates can also be used as instruments in a panel context.

$$y_{it} = x_{it}'\beta + \varepsilon_{it}$$

where x_{it} is $K \times 1$. We continue to maintain the assumption that

$$x_{it} = \Lambda_i' F_t + u_{it} = C_{it} + u_{it}$$

Assumption E-Panel Same as Assumption L and E with three changes. Part (i) holds with λ_i replaced by Λ_i; part (ii) holds with e_{it} replaced by each component of u_{it} (note that u_{it} is a vector).

Assumption IV-Panel:

a. $E(\varepsilon_{it}) = 0$, $E|\varepsilon_{it}|^{4+\delta} < M < \infty$ for all i, t, and for some $\delta > 0$; ε_{it} is independent over i and ε_{it} is independent of F_t and Λ_i.

b. $x_{it} = \Lambda_i' F_t + u_{it}$; $E(u_{it}\varepsilon_{it}) \neq 0$; u_{it} is independent over i.

c. $(NT)^{-1/2} \sum_{i=1}^{N} \sum_{t=1}^{T} C_{it}\varepsilon_{it} \xrightarrow{d} N(0, S)$, where

$$S = \lim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t,s=1}^{T} E(C_{it}C_{is}'\varepsilon_{it}\varepsilon_{is}),$$

which is the long run variance of $\xi_t = N^{-1/2} \sum_{i=1}^{N} C_{it}\varepsilon_{it}$.

The pooled two-stage least-squares estimator with $\widetilde{C}_{it}$ as instruments is

$$\widehat{\beta}_{PFIV} = \Big(\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{C}_{it}x_{it}'\Big)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{C}_{it}y_{it}.$$

Result G: Panel FIV . Under Assumptions E-Panel and IV-Panel,

G.3 (i) $\widehat{\beta}_{PFIV} - \beta^0 = O_p(T^{-1}) + O_p(N^{-1})$ and thus $\widehat{\beta}_{PFIV} \xrightarrow{p} \beta^0$.

(ii) If $T/N \to \tau > 0$, then

$$\sqrt{NT}(\widehat{\beta}_{PFIV} - \beta^0) \xrightarrow{d} N(\tau^{1/2}\Delta_1^0 + \tau^{-1/2}\Delta_2^0,\, \Omega)$$

where $\Omega = \text{plim}[S_{\widetilde{x}\widetilde{x}}]^{-1}S[S_{\widetilde{x}\widetilde{x}}']^{-1}$ with $S_{\widetilde{x}\widetilde{x}} = (NT)^{-1}\sum_{i=1}^{N}\widetilde{C}_{it}x_{it}'$, and Δ_1^0 and Δ_2^0 are bias terms.

iii Suppose ε_{it} are serially uncorrelated. Let $\widehat{\Delta} = \frac{1}{N}\widehat{\Delta}_1 + \frac{1}{T}\widehat{\Delta}_2$, where $\widehat{\Delta}_k$ is a consistent estimate of Δ_k $(k = 1, 2)$, see Bai and Ng (2006c). Suppose Assumptions F(0), L, E-Panel, LFE, and FAR-Panel hold. If $T/N^2 \to 0$, and $N/T^2 \to 0$, then

$$\sqrt{NT}(\widehat{\beta}_{PFIV} - \widehat{\Delta} - \beta^0) \xrightarrow{d} N(0, \Omega).$$

Result G.3 indicates that there can be no valid instrument in the conventional sense, yet, we can still consistently estimate the large simultaneous equations system. In a data-rich environment, the information in the data collectively permits consistent instrumental variable estimation under much weaker conditions on individual instruments. However, C_{it} is not observed, and biases arise from the estimation of C_{it}. More precisely, $\widetilde{C}_{it}$ contains elements of u_{it}, which are correlated with ε_{it}. This motivates the bias-corrected estimator stated in (iii). Although the biased estimator is reasonably precise, the bias correction terms, defined in Bai and Ng (2006c), are necessary for the t statistic to have the appropriate size.

5.4 Testing the Validity of Observed Proxies

Finance theory postulates that systematic risks should be priced. Many observed variables including inflation, term and premia, as well as variables constructed by Fama and French have been used as proxies for the latent risk factors. But are these factors guided by economic theory close to the statistical factors constructed in a data rich environment?

Let $G_t = (G_{1t}, \dots G_{mt})$ be a $m \times 1$ vector of observed variables that are thought to be useful proxies of a set of otherwise unobserved factors, F_t. The objective is to get a sense of how close F and G are. We consider two cases: (i) G_{jt} is an exact linear combination of F_t such that $G_{jt} = \delta_j' F_t$ for some vector δ_j; (ii) G_{jt} is a linear combination of F_t plus an error term such that $G_{jt} = \delta_j' F_t + \varepsilon_{jt}$. In the second case, we want to measure how big is the error ε_{jt} relative to $\delta_j' F_t$. We also want to make inference about the correlation between the vector G_t and F_t. But F_t is not observable, so we use $\widetilde{F}_t$ instead.

Let $\widehat{\gamma}_j$ be obtained from the least squares estimation of

$$G_{jt} = \gamma_j' \widetilde{F}_t + error.$$

Let $\widehat{\tau}_t(j)$ be the corresponding t statistic and let Φ_α be the α percentage point of the limiting distribution of $\widehat{\tau}_t(j)$.

Result H: Testing Validity of Observed Factors

H.1 Let $A(j) = \frac{1}{T}\sum_{t=1}^{T} I(|\widehat{\tau}_t(j)| > \Phi_\alpha)$ and $M(j) = \max_{1\le t\le T} |\widehat{\tau}_t(j)|$. Under the null hypothesis of exact linear combination $G_{jt} = \delta_j' F_t$ and $\sqrt{N}/T \to 0$ as $N, T \to \infty$, then $A(j) \to 2\alpha$. If, in addition, e_{it} is serially uncorrelated, $P(M(j) \le x) \approx [2\Phi(x) - 1]^T$, where $\Phi(x)$ is the cdf of a standard normal random variable.

H.2 Consider the null hypothesis $G_{jt} = \delta_j' F_t + \varepsilon_{jt}$. Let $\widehat{\varepsilon}_{jt} = G_{jt} - \widehat{G}_{jt}$. As $N, T \to \infty$ and with $s_{jt}^2 = T^{-1} \widetilde{F}_t' (T^{-1} \sum_{s=1}^{T} \widetilde{F}_s \widetilde{F}_s' \widehat{\varepsilon}_{js}^2)^{-1} \widetilde{F}_t + N^{-1} Avar(\widehat{G}_{jt})$, then for each t

$$\frac{\widehat{\varepsilon}_{jt} - \varepsilon_{jt}}{s_{jt}} \xrightarrow{d} N(0,1).$$

An estimate of $\frac{1}{N} Avar(\widehat{G}_{jt})$ is $N^{-1} \widehat{\gamma}_j' \widetilde{V}^{-1} \widetilde{\Gamma}_t \widetilde{V}^{-1} \widehat{\gamma}_j$, where $\widetilde{V}$ and $\widetilde{\Gamma}_t$ are defined in previous sections. Also define two overall statistics (not depending on t)

$$NS(j) = \frac{\widehat{\text{var}}(\widehat{\varepsilon}(j))}{\widehat{\text{var}}(\widehat{G}(j))} \quad \text{and } R^2(j) = \frac{\widehat{\text{var}}(\widehat{G}(j))}{\widehat{\text{var}}(G(j))}$$

where $\widehat{\text{var}}(\cdot)$ is simply the sample variance. Then $NS(j)$ should be close to zero and $R^2(j)$ should be close to one under the null hypothesis of an exact linear combination.

H.3 Let $\widetilde{\rho}_1^2 \dots, \widetilde{\rho}_p^2$ be the largest $p = \min[m, r]$ sample squared canonical correlations between $\widetilde{F}$ and G. Suppose that $(F_t', G_t')'$ are iid normally distributed and the true canonical correlation coefficient between F_t and G_t are given by $\rho_1^2, ..., \rho_p^2$. As $N, T \to \infty$

with $\sqrt{T}/N \to 0$,

$$\widetilde{z}_k = \frac{\sqrt{T}(\widetilde{\rho}_k^2 - \rho_k^2)}{2\widetilde{\rho}_k(1-\widetilde{\rho}_k^2)} \xrightarrow{d} N(0,1), \qquad k = 1, \ldots, \min[m, r].$$

These results are based on Bai and Ng (2006b). Both $A(j)$ and $M(j)$ are tests for exact factors. The $A(j)$ statistic allows G_{jt} to deviate from $\widehat{G}_{jt}$ for a specified number of time periods as specified by α. The $M(j)$ test is stronger and requires G_{jt} not to deviate from $\widehat{G}_{jt}$ by more than the sampling error at every t. As measurement error and time aggregation can be responsible for deviations between observed and latent factors, an approximate test is given in G.2. Instead of asking how large the measurement errors in the proxy variables are, the two overall statistics, $NS(j)$ and $R^2(j)$, provide a guide to the size of the measurement error. Under normality, how close is the set G_t to the set of latent factors can also be assessed by testing the canonical correlation coefficients. The normality assumption can be relaxed to iid elliptically distributed errors with a suitable rescaling of $\widetilde{z}_k$.

6 Panel Regression Models with a Factor Structure in the Errors

Consider the following model

$$Y_{it} = X_{it}'\beta + u_{it}$$

and

$$u_{it} = \lambda_i' F_t + \varepsilon_{it}$$

where X_{it} is a $p \times 1$ vector of observable regressors, β is a $p \times 1$ vector of unknown coefficients; u_{it} has a factor structure, but λ_i, F_t, and ε_{it} are all unobserved. We are interested in the estimation of β, the common slope coefficients.

We assume that ε_{it} are independent of the regressors. If both λ_i and F_t are also independent of the regressor X_{it}, the aggregate error u_{it} will also be independent of the regressor and the model can be estimated with pooled least squares. More efficient estimation could be obtained by GLS based on the factor error structure. However, we allow the regressors to be correlated with either the factors F_t or the factor loadings or both. In this case GLS will be inconsistent.

The above model with a factor error structure encompasses the fixed effect model as a special case. To see this, let $r = 2$, and for all i and all t, define

$$F_t = \begin{bmatrix} 1 \\ \xi_t \end{bmatrix} \quad \text{and} \quad \lambda_i = \begin{bmatrix} \alpha_i \\ 1 \end{bmatrix}.$$

Then

$$\lambda_i' F_t = \alpha_i + \xi_t.$$

Hence, the model reduces to

$$Y_{it} = X_{it}'\beta + \alpha_i + \xi_t + \varepsilon_{it},$$

which is the fixed effect model where the individual effects α_i and the time effects ξ_t enter the model additively instead of interactively.

The fixed effect model is usually estimated by the least squares dummy variable approach which treats α_i and ξ_t as parameters to be estimated. This suggests that for models with factor errors, we can also treat λ_i and F_t as parameters. The unknown parameters β, λ_i, and F_t can be estimated by simply minimizing the least squares objective function $\sum_{i=1}^N \sum_{t=1}^T (Y_{it} - X_{it}'\beta - \lambda_i' F_t)^2$. The model is, however, over-parameterized since $\lambda_i' F_t = \lambda_i A A^{-1} F_t$ for an arbitrary invertible matrix A. We need r^2 restrictions since an arbitrary invertible $r \times r$ matrix has r^2 free parameters. These restrictions are most clearly stated using vector representation of the model

$$Y_i = X_i\beta + F\lambda_i + \varepsilon_i$$

where for $i = 1, 2, ..., N$, $Y_i = (Y_{i1}, ..., Y_{iT})'$, $X_i = (X_{i1}, ..., X_{iT})'$ $(T \times k)$, and $F = (F_{i1}, ..., F_{iT})'$ $(T \times r)$. Also let $\Lambda = (\lambda_{i1}, ..., \lambda_{iN})'$ $(N \times r)$. The constraint

$$F'F/T = I$$

implies $r(r+1)/2$ restrictions since a symmetric matrix has $r(r+1)/2$ free parameters. The additional constraint that $\Lambda'\Lambda$ is a diagonal matrix (i.e., the off diagonal elements are zero) gives $r(r-1)/2$ restrictions since the off-diagonal elements of a symmetric matrix has $r(r-1)/2$ free parameters. These restrictions are similar to those in the pure factor model stated in previous sections. They are restated here for completeness.

Now consider minimizing the least squares objective function

$$SSR(\beta, F, \Lambda) = \sum_{i=1}^N (Y_i - X_i\beta - F\lambda_i)'(Y_i - X_i\beta - F\lambda_i)$$

subject to the constraint $F'F/T = I_r$ and $\Lambda'\Lambda$ being diagonal. Define the projection matrix

$$M_F = I_T - F(F'F)^{-1}F = I_T - FF'/T$$

The least squares estimator for β for each given F is simply

$$\widehat{\beta}(F) = \Big(\sum_{i=1}^{N} X_i' M_F X_i\Big)^{-1} \sum_{i=1}^{N} X_i' M_F Y_i$$

Given β, $W_i = Y_i - X_i\beta$ has a pure factor structure given by

$$W_i = F\lambda_i + \varepsilon_i$$

Define $W = (W_1, W_2, ..., W_N)$, a $T \times N$ matrix. Thus F is estimated as the first r eigenvectors associated with first r largest eigenvalues of the matrix

$$WW' = \sum_{i=1}^{N} W_i W_i' = \sum_{i=1}^{N} (Y_i - X_i\beta)(Y_i - X_i\beta)'.$$

Denote this estimate by $\widehat{F}(\beta)$. Then $\widehat{\Lambda}(\beta) = W'\widehat{F}(\beta)/T$. Therefore, given F, we can estimate β, and given β, we can estimate F. The final least squares estimator $(\widehat{\beta}, \widehat{F})$ is the solution to the following set of nonlinear equations

$$\widehat{\beta} = \Big(\sum_{i=1}^{N} X_i' M_{\widehat{F}} X_i\Big)^{-1} \sum_{i=1}^{N} X_i' M_{\widehat{F}} Y_i, \quad \text{and} \tag{11}$$

$$\Big[\frac{1}{NT} \sum_{i=1}^{N} (Y_i - X_i\widehat{\beta})(Y_i - X_i\widehat{\beta})'\Big] \widehat{F} = \widehat{F} V_{NT} \tag{12}$$

where V_{NT} is a diagonal matrix consisting of the r largest eigenvalues of the above matrix[5] in the brackets, arranged in decreasing order. The solution $(\widehat{\beta}, \widehat{F})$ can be simply obtained by iteration. Given $\widehat{F}$, we have $\widehat{\Lambda} = W'\widehat{F}/T$. An alternative estimation procedure is suggested by Pesaran (2006), in which time averages of the dependent variable and independent variables are included as additional regressors. These averages play the role of estimated common factors.

[5] We divide this matrix by NT to make V_{NT} have a proper limit. The scaling does not affect $\widehat{F}$.

Assumptions: EF-Panel Assume F(0), L, E and IE of Section 3 hold, together with one of the following assumptions

(i) ε_{it} are i.i.d for all i and t;

(ii) ε_{it} are correlated and heteroskedastic only in the cross-section dimension, and $T/N \to 0$;

(iii) ε_{it} are correlated and heteroskedastic only in the time dimension, and $N/T \to 0$.

Under the above assumption, the estimator $\widehat{\beta}$ has the following asymptotic distribution

$$\sqrt{NT}(\widehat{\beta} - \beta^0) = \left(\frac{1}{NT} \sum_{i=1}^{N} Z_i' Z_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} Z_i' \varepsilon_i + o_p(1)$$

where

$$Z_i = M_F X_i - \frac{1}{N} \sum_{i=1}^{N} a_{ik} M_F X_k$$

and $a_{ik} = \lambda_i'(\Lambda'\Lambda/N)^{-1}\lambda_k$. The right hand side of the representation does not depend on any estimated quantity.

Result I: Panel Data Models with Interactive Effects Under the assumption

$$\frac{1}{NT} \sum_{i=1}^{N} Z_i' Z_i \xrightarrow{p} D_0, \quad \text{and } \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} Z_i' \varepsilon_i \xrightarrow{d} N(0, D_Z),$$

I.1 If (i) of EF-Panel holds, then

$$\sqrt{NT}(\widehat{\beta} - \beta) \xrightarrow{d} N(0, D_0^{-1} D_Z D_0^{-1})$$

I.2 If (ii) or (iii) of EF-PANEL holds, and $T/N \to \rho > 0$,

$$\sqrt{NT}(\widehat{\beta} - \beta) \xrightarrow{d} N(\rho^{1/2} B_0 + \rho^{-1/2} C_0, D_0^{-1} D_Z D_0^{-1}).$$

If the ε_{it} are allowed to be correlated and heteroskedastic in both dimensions and if $T/N \to \rho > 0$, asymptotic bias exists. The expression of B_0 and C_0 are given in Bai (2005), who also derived the biased-corrected estimator. It is possible to test factor error structure against additive fixed effect. The model can be extended to

$$Y_{it} = X_{it}'\beta + \alpha_i + \xi_t + \tau_i'\delta_t + \varepsilon_{it}.$$

But in this model, $\tau_i'\delta_t$ must have a genuine factor structure in the sense that τ_i cannot be one for all i or F_t cannot be one for all t. The details are given by Bai (2005).

7 Theory: Non-Stationary Data

Consider the following data generating process:

$$\begin{aligned} X_{it} &= D_{it} + \lambda_i' F_t + e_{it} \\ (1-L)F_t &= C(L)\eta_t \\ e_{it} &= \rho_i\, e_{it-1} + \varepsilon_{it} \end{aligned} \tag{13}$$

where $D_{it} = \sum_{i=0}^{p} \delta_i t^i$ is the deterministic component. When $p = 0$, $D_{it} = \delta_i$ is the individual specific fixed effect, and when $p = 1$, an individual specific time effect is also present. When there is no deterministic term, D_{it} is null and we will refer to this as case $p = -1$.

Assumption F(1): (i) $\eta_t \sim iid(0, \Sigma_\eta)$, $E\|\eta_t\|^4 \leq M$, (ii) $\text{var}(\Delta F_t) = \sum_{j=0}^{\infty} C_j \Sigma_\eta C_j' > 0$, (iii) $\sum_{j=0}^{\infty} j\|C_j\| < M$; (iv) $C(1)$ has rank r_1, $0 \leq r_1 \leq r$; (v) $E\|F_0\| \leq M$.

Assumption E(1) $E|\varepsilon_{i0}| < M$ for all $i = 1, \ldots N$.

Under F(1) (non-stationary factors), the short-run variance of ΔF_t is positive definite but the long run variance can be reduced rank. Initial conditions are also imposed on the factors and the errors to permit asymptotic analysis.

The number of common stochastic trends is determined by r_1, the rank of $C(1)$. When $r_1 = 0$, ΔF_t is overdifferenced and the common factors are stationary.

In this model, X_{it} can be non-stationary when F_t has unit roots, or $\rho_i = 1$, or both. Clearly, if the common factors share a stochastic trend, X_{it} will all be non-stationary. But even if a series indeed has a common and trending component, the series may appear stationary if the idiosyncratic component is stationary.

7.1 Estimation of F_t when e_{it} may be I(1)

Suppose now that we observe only X_{it} and we do not know if it is the factors or the idiosyncratic errors are stationary. The possibility that e_{it} is non-stationary poses a serious problem for estimation, as any regression with a non-stationary error (observed or otherwise) is spurious. We now consider how to estimate the factors when we do not know a priori if e_{it} is stationary.

a Case $p = 0$: Let $\Delta X_{it} = \lambda_i' \Delta F_t + \Delta e_{it}$. Denote $x_{it} = \Delta X_{it}$, $f_t = \Delta F_t$, and $z_{it} = \Delta e_{it}$. Let $(\widehat{\lambda}_i, ..., \widehat{\lambda}_N)$ and $(\widehat{f}_1, ..., \widehat{f}_T)$ and $\widehat{z}_{it}$ for all i and t be the principal components estimates of λ_i, f_t, and z_{it}.

b Case $p = 1$: Let $x_{it} = \Delta X_{it} - \overline{\Delta X_i}$, $f_t = \Delta F_t - \overline{\Delta F}$, and $z_{it} = \Delta e_{it} - \overline{\Delta e_i}$. Let $\widehat{f}_t$ and $\widehat{\lambda}_i$ be obtained by applying the method of principal components to the differenced and demeaned data . Let $\widehat{F}_t = \sum_{s=2}^t \widehat{f}_s$ and $\widehat{e}_{it} = \sum_{s=2}^t \widehat{z}_{is}$.

Result A1 Let $\widehat{F}_t = \sum_{s=2}^t \widehat{f}_s$ and $\widehat{e}_{it} = \sum_{s=2}^t \widehat{z}_{is}$. Suppose f_t and λ_i satisfy Assumption F0, and z_{it} satisfies Assumption L and E. Then

$$\max_{1 \le t \le T} \left\| \widehat{F}_t - HF_t + HF_1 \right\| = O_p(T^{1/2} N^{-1/2}) + O_p(T^{-1/4}).$$

The basis of Result A1, developed in Bai and Ng (2004), is that $x_{it} = \lambda_i' f_t + z_{it}$ is a pure factor model satisfying Assumptions F0. By Result A0, $\widehat{f}_t, \widehat{\lambda}_i$ and $\widehat{z}_{it}$ are consistent for f_t, λ_i and z_{it}. Result A1 says that if $T/N \to 0$ as $N, T \to \infty$, then $\widehat{F}_t$ is uniformly consistent for HF_t. Pointwise convergence (for each given t) does not require $T/N \to 0$; the deviation of $\widehat{F}_t$ from HF_t is of order $\min[N^{-1/2}, T^{-1/2}]$. This result is quite remarkable, as it is obtained without knowledge of whether F_t or e_{it} is I(1) or I(0). This means that even if each cross-section equation is a spurious regression, the common stochastic trends are well defined and can be consistently estimated, if they exist. This is certainly not possible within the framework of traditional time series analysis in which N is fixed.

To see the method actually work, we simulate an I(1) common factor process $F_t = F_{t-1} + \eta_t$, and independent I(1) idiosyncratic errors $e_{it} = e_{i,t-1} + \varepsilon_{it}$ for $t = 1, 2..., T; i = 1, 2, ..., N$ $(N = 40, T = 100)$ to form $X_{it} = \lambda_i F_t + e_{it}$ with λ_i iid $N(0, 1)$. The observable processes X_{it} are not cointegrated because e_{it} are independent I(1) processes. Once the data are generated, we treat F_t and e_{it} as unobservable. We then estimate F_t by the difference-recumulating method discussed earlier. To demonstrate that the estimated common trend F_t is close to the actual F_t, we rotate $\widehat{F}_t$ towards F_t by running the regression $F_t = b\widehat{F}_t + error$. Figure 1 displays both F_t and $\widehat{b}\widehat{F}_t$. The estimated $\widehat{F}_t$ tracks F_t quite well. If the data are not differenced, one cannot expect consistent estimation of F_t, unless the data are I(0). To see this, we present, in Figure 2, the estimate without differencing the data. As can be seen, the estimate is unsatisfactory.

Result A1 pertains to the case when we do not know if F_t and e_{it} are stationary or not. The case in which F_t is a vector of integrated processes and e_{it} is stationary is considered by Bai (2004). He shows that the estimated common factors have a faster rate of convergence than the case in which F_t being I(0). He also derives the limiting distribution for the estimated common factors.

7.2 Unit Root Tests

The ability to estimate F_t when the data are non-stationary opens up new possibilities of re-examining hypotheses that have been difficult to test. For example, when x_{it} has a factor structure, non-stationarity can arise because F_t is I(1), or e_{it} is I(1), or both. When one of the components is stationary but the other is not, testing x_{it} for non-stationarity can be misleading. A more informative approach is to test F_t and e_{it} separately. Result A1 shows that this is possible. Consider the regression

$$\Delta\widehat{e}_{it} = d_0\widehat{e}_{it-1} + d_{i1}\Delta\widehat{e}_{it-1} + \ldots d_{ik}\Delta\widehat{e}_{it-k} + error. \tag{14}$$

Result J: PANIC (Panel Analysis of Nonstationarity in Idiosyncratic and Common components)

Suppose F(1), and E(1) hold. Let ADF_i^c be the t test for d_0 in (14) with k chosen such that $\frac{k^3}{\min[N,T]} \to 0$. Let $W_{\varepsilon i}$ be a standard Brownian and $V_{ei}(s)$ be a Brownian bridge.

J.1 case $p = 0$:

$$ADF_i^0 \Rightarrow \frac{\int_0^1 W_{\varepsilon i}(s)dW_{\varepsilon i}(s)}{(\int_0^1 W_{\varepsilon i}(s)^2 ds)^{1/2}}$$

J.2 case $p = 1$:

$$ADF_i^1 \Rightarrow \frac{-1}{2}\left(\int_0^1 V_{ei}(s)^2 ds\right)^{-1/2}$$

Result J says that the Said-Dickey-Fuller test can be used to test if $\widehat{e}_{it}$ is non-stationary, one series at a time. When $p = 0$, the test has the same limiting distribution as the usual Dickey-Fuller test without an intercept. When $p = 1$, the limiting distribution is not a DF type but is proportional to the reciprocal of a Brownian bridge. Result I suggests that other unit root and stationary tests, not just the ADF, can be used to test $\widehat{e}_{it}$. Our conjecture is that the limiting distribution when $p = 0$ will be the same as the one when an observed series is tested but without the intercept. When $p = 1$, the limiting distribution will likely be different.

Perhaps the most exploited aspect of the factor structure is that it enables construction of panel unit root tests in the presence of cross-section dependence. By assuming that the cross-section correlations are strong and are induced by the common factors, they can be directly removed from x_{it} to yield $\widehat{e}_{it}$. Thus, while pooling x_{it} will yield tests with size distortions, pooling $\widehat{e}_{it}$ is valid. Result K below considers three types of pooled tests.

Result K: Pooled Unit Root Tests

K.1 Suppose the e_{it} are iid across i. Let p_i be the p-values associated with $ADF_{\widehat{e}}^c$ or $ADF_{\widehat{e}}^\tau$. Then

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

K..2 Given $\widehat{e}_{it}$ from PANIC, define the PMSB (panel MSB) as

$$PMSB = \frac{\sqrt{N}\left(\text{tr}(\frac{1}{NT^2}\widehat{e}'\widehat{e}) - \widehat{\psi}\right)}{\sqrt{\widehat{\phi}_\varepsilon^4/K}} \tag{15}$$

where $\widehat{\psi} = \widehat{\omega}^2/2$ and $K = 3$ when $p = 0$, and $\widehat{\psi} = \widehat{\omega}^2/6$ and $K = 45$ when $p = 1$. Under the null hypothesis that all e_{it} are non-stationary and iid across i, then as $N, T \to \infty$ with $N/T \to 0$,

$$PMSB \xrightarrow{d} N(0,1).$$

Result K.1 pools the p values of the test and is especially appropriate when there is substantial heterogeneity in the data. Result K.2 exploits an important feature that distinguishes stationary from non-stationary processes, namely, that their sample moments require different rates of normalization in order to be bounded asymptotically. Assuming cross-section independence, Phillips and Ploberger (2002) proposed a point optimal test for panel unit root in the presence of incidental trends that has some resemblance to the Sargan-Bhargava test, first proposed in Sargan and Bhargava (1983). Stock (1990) modified the test to allow for serial correlation in the errors. The properties of the test are analyzed in Perron and Ng (1998). Our PMSB is explicitly motivated by the Sargan-Bhargava test and permits x_{it} to be cross-sectionally correlated. Standard methods that do not take into account common stochastic trends perform poorly, as documented by Banerjee et al. (2005). Other procedures that allow factor structures can be found in Moon and Perron (2004), Phillips and Sul (2003), Breitung and Das (2007), Pesaran (2007), among others. A recent survey on panel unit root and cointegration is provided by Breitung and Pesaran (2007).

Moon and Perron (2004) considered a panel unit root test that estimates the pooled autoregressive coefficient of the defactored data. The test, like P_ε, is also a test of the idiosyncratic errors. As such, a PANIC test of the pooled autoregressive coefficient in e_{it} can also be developed. Consider pooled OLS estimation of the model $\widehat{e}_{it} = \rho\widehat{e}_{it-1} + \varepsilon_{it}$.. where $\widehat{e}_{-1}$ and $\widehat{e}$ are $(T-2)\times N$ matrices. When $p=1$, we add an intercept and a linear trend in the above regression. Define the bias corrected estimator

$$\widehat{\rho}^{+} = \frac{\mathrm{tr}(\widehat{e}'_{-1}M_z\widehat{e} - NT\widehat{\psi})}{\mathrm{tr}(\widehat{e}'_{-1}M_z\widehat{e}_{-1})}$$

where $\widehat{\psi}$ is the bias correction estimated from $\widehat{\varepsilon} = M_z\widehat{e} - \widehat{\rho}M_z\widehat{e}_{-1}$, where $M_1 = I_{T-2} - z(z'z)z'$ with $z = (z_1, z_2, ..., z_{T-2})'$. Let $\widehat{\omega}^2 = \frac{1}{N}\sum_{i=1}^N \widehat{\omega}_i^2$, $\widehat{\omega}_i^2$ being the estimate for the long run variance of ε_{it}, computed from the residuals $\widehat{\varepsilon}_{it}$. Also let $\widehat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^N \widehat{\sigma}_i^2$ with $\widehat{\sigma}_i^2$ being the variance estimate for ε_{it}, and $\widehat{\phi}_\varepsilon^4 = \frac{1}{N}\sum_{i=1}^N \widehat{\omega}_i^4$.

K.3 Consider the test statistics

$$\begin{aligned} P_a &= \frac{\sqrt{N}T(\widehat{\rho}^{+}-1)}{\sqrt{K_a\widehat{\phi}_\varepsilon^4/\widehat{\omega}^4}} \\ P_b &= \sqrt{N}T(\widehat{\rho}^{+}-1)\sqrt{\frac{1}{NT^2}\mathrm{tr}(\widehat{e}_{-1}M\widehat{e}'_{-1})K_b\frac{\widehat{\omega}^2}{\widehat{\phi}_\varepsilon^4}}. \end{aligned}$$

Then $P_a, P_b \xrightarrow{d} N(0,1)$, where the parameter values are given in the following table:

p	M_z	$\widehat{\psi}$	K_a	K_b
0	I_{T-2}	$\widehat{\lambda}_\varepsilon$	2	1
1	M_1	$-\widehat{\sigma}_\varepsilon^2/2$	15/4	4

Result K.3 is a test of whether the pooled autoregressive coefficient is unity, and was recently developed in Bai and Ng (2006d). The statistics $P_{a,b}$ are the analog of $t_{a,b}$ of Moon and Perron (2004), except that (i) the tests are based on PANIC residuals and (ii) the method of 'defactoring' of the data is different from MP. Improved versions of the tests can also be obtained by iteration. As discussed in Bai and Ng (2006d), iteration ensures that the estimate of ρ used to construct the long-run variance estimates coincides with the estimate of ρ that is being test for deviation from unity.

When $p = 0$, whether one tests the pooled autoregressive coefficient using P_a, P_b, or the divergence of the sample moments using PMSB, the results in terms of both size and power are quite similar. Both tests have better power than testing the pooled p values. However, we find in Bai and Ng (2006d) that P_e and PMSB are superior to tests of a common unit root against incidental trends. Moon et al. (2007) and Phillips and Ploberger (2002) also documented that panel unit root tests that rely on a pooled estimate of autoregressive parameter do not have good power properties against deterministic trends.

In a recent paper, Westerlund and Larsson (2007) point out that the asymptotic distribution of a test that pools the individual unit root tests is not centered around zero. While this suggests that the P_e (which pools p values instead of the tests) may have size distortion, simulations suggest that this is a problem in finite samples only when N is very small (less than 10). On the other hand, the P_e is still the test with the best properties when there are incidental trends.

7.3 Common Trends

If there are common factors in the data x_{it}, and one or more of the factors are non-stationary, any two series x_{it} and x_{jt} are cointegrated in the sense of Engle and Granger (1987). Result A1 shows that F_t can be consistently estimated without knowing if the data are stationary or not. Testing if $\widehat{F}_t$ is non-stationary, however, is more involved because $\widehat{F}_t$ is a vector process and there can be multiple unit roots. Define $\widehat{F}_t^c = \widehat{F}_t - \overline{\widehat{F}}$, $\overline{\widehat{F}} = (T-1)^{-1}\sum_{t=2}^T \widehat{F}_t$. For some $m \leq r$, let $\widehat{Y}_t^c = \widehat{\beta}_\perp' \widehat{F}_t^c$, where $\widehat{\beta}_\perp$ are the m eigenvectors associated with the m largest eigenvalues of $T^{-2}\sum_{t=2}^T \widehat{F}_t^c \widehat{F}_t^{c\prime}$. Let $\widehat{y}_t^c = \widehat{\Pi}(L)\widehat{Y}_t^c$ and $\widehat{v}_f^c$ be the smallest eigenvalue of $\widehat{\Phi}_f^c(m) = \frac{1}{2}[\sum_{t=2}^T(\widehat{y}_t^c \widehat{y}_{t-1}^{c\prime} + \widehat{y}_{t-1}^c \widehat{y}_t^{c\prime})](\sum_{t=2}^T y_{t-1}^c \widehat{y}_{t-1}^{c\prime})^{-1}$. Similarly define $\widehat{y}_t^\tau$ and $\widehat{F}_t^\tau$ when there is a time trend.

Result L: Testing for Common Trends Consider the null hypothesis that F_t has m stochastic trends with finite VAR representation.

L.1 (intercept model) Let W_m^c be a vector of m dimensional demeaned Brownian motions. Then $MQ_f^c(m) = T(\widehat{v}_f^c(m) - 1) \xrightarrow{d} v_*^c(m)$, where $v_*^c(m)$ is the smallest eigenvalue of

$$\Phi_*^c(m) = \frac{1}{2}[W_m^c(1)W_m^c(1)' - I_m][\int_0^1 W_m^c(s)W_m^c(s)'ds]^{-1}.$$

L.2 (linear trend case) Let W_m^τ be a vector of m dimensional detrended Brownian motions. Then $MQ_f^{tau}(m) = T(\widehat{v}_f^\tau(m) - 1) \xrightarrow{d} v_*^\tau(m)$, where $v_*^\tau(m)$ is the smallest eigenvalue of

$$\Phi_*^\tau(m) = \frac{1}{2}[W_m^\tau(1)W_m^\tau(1)' - I_m][\int_0^1 W_m^\tau(s)W_m^\tau(s)'ds]^{-1}.$$

Result L states that the number of common trends in F_t can be tested using $\widehat{F}_t$. The proposed statistic is a modified variation of Stock and Watson (1988). Their test is based on the idea that the real part of the smallest eigenvalue of an autoregressive coefficient matrix should be unity. Our modification ensures that the eigenvalues are always real and enables simplifications to the proofs.

7.4 Panel Unit Roots with Structural Breaks

The model has the same form as (13) except that the deterministic components have structural breaks. Bai and Carrion-i-Silvestre (2004) consider two specifications, referred to as Model 1 and Model 2, respectively,

$$\text{Model 1} \quad : \quad D_{i,t} = \mu_i + \sum_{j=1}^{l_i} \theta_{i,j} DU_{i,j,t} \tag{16}$$

$$\text{Model 2} \quad : \quad D_{i,t} = \mu_i + \beta_i t + \sum_{j=1}^{l_i} \theta_{i,k} DU_{i,j,t} + \sum_{k=1}^{m_i} \gamma_{i,k} DT_{i,k,t}^*, \tag{17}$$

where $DU_{i,j,t} = 1$ for $t > T^i_{a,j}$ and 0 elsewhere, and $DT^*_{i,k,t} = \left(t - T^i_{b,k}\right)$ for $t > T^i_{b,k}$ and 0 elsewhere, where $T^i_{a,j}$ and $T^i_{b,k}$ denote the j-th and k-th dates of the break for the i-th individual, $j = 1, \ldots, l_i$, $k = 1, \ldots, m_i$. There are l_i structural breaks affecting the mean and m_i structural breaks affecting the trend of the time series, where l_i is not necessarily equal to m_i.

The structural breaks are heterogeneous across individuals because (i) the break dates $T^i_{a,j}$ and $T^i_{b,k}$ are individual specific, (ii) the magnitude of shifts are also individual specific, (iii) each individual may have a different number of structural breaks, and (iv) the break points for the level and the slope can be at different times. While Model 1 is a special case of Model 2, the limiting distribution of the test statistic for Model 1 is not a special case for Model 2. Thus we will consider the two models separately.

To estimate the model, we follow Bai and Ng (2004). For model 1, differencing leads to

$$\Delta X_{it} = \Delta F_t' \pi_i + \Delta e^*_{it},$$

where

$$\Delta e^*_{it} = \Delta e_{it} + \sum_{j=1}^{\ell_i} \theta_{i,j} D(T^i_{a,j})_t, \tag{18}$$

with $D(T^i_{a,j})_t = 1$ if $t = T^i_{a,j}$ and $D(T^i_{a,j})_t = 0$ otherwise. In matrix notation, we can rewrite the above as

$$\Delta X_i = \Delta F \pi_i + \Delta e^*_i,$$

where $\Delta X_i = (\Delta X_{i,2}, \Delta X_{i,3}, \ldots, \Delta X_{i,T})'$ and $\Delta e^*_i = \left(\Delta e^*_{i,2}, \Delta e^*_{i,3}, \ldots, \Delta e^*_{i,T}\right)'$ are two $(T-1) \times 1$ vectors for the i-th individual, $\Delta F = [\Delta F_1,\ \Delta F_2,\ \ldots,\ \Delta F_T]'$ is a $(T-1) \times r$ matrix and $\pi_i = (\pi_{i,1}, \ldots, \pi_{i,r})'$ is the $(r \times 1)$-vector of loading parameters for the i-th individual, $i = 1, \ldots, N$.

We can rewrite the model more compactly as

$$x_i = f \pi_i + z_i, \tag{19}$$

where $x_i = \Delta X_i$, $f = \Delta F$, and $z_i = \Delta e^*_i$. Now f and z can be estimated by the principal components method. The unit root test statistic for each individual time series can be based on

$$\widehat{e}_{i,t} = \sum_{s=2}^{t} \widehat{z}_{i,s}.$$

A consistent unit root test is the modified Sargan-Bhargava (MSB) statistic, defined as

$$MSB_i = \frac{T^{-2}\sum_{t=1}^{T}\widehat{e}_{i,t}^2}{\widehat{\sigma}_i^2} \tag{20}$$

where $\widehat{\sigma}_i^2$ is a consistent estimator of the long-run variance of $e_{it} - \rho_i e_{i,t-1}$.

For Model 2, differencing leads to

$$\Delta X_{i,t} = \Delta F_t' \pi_i + \beta_i + \sum_{k=1}^{m_i} \gamma_{i,k} DU_{i,k,t} + \Delta e_{i,t}^*,$$

where $DU_{i,k,t} = 1$ when $t > T_{b,k}^i$ and 0 otherwise. The dummy variables result from differencing the broken trends, see (17). There is no loss of generality by assuming ΔF_t to have zero mean, otherwise, define $f_t = \Delta F_t - \tau$ with $\tau = E(\Delta F_t)$ and redefine the intercept as $\beta_i + \tau' \pi_i$. More compactly, we also have

$$x_i = f\pi_i + a_i \delta_i + z_i, \tag{21}$$

where x_i, f, and z_i are defined earlier; $\delta_i = (\beta_i, \gamma_{i,1}, ..., \gamma_{i,m_i})'$ and $a_i = (a_{i,2}, ..., a_{i,T})'$ with $a_{i,t} = (1, DU_{i,1,t}, ..., DU_{i,m_i,t})'$. Thus a_i is the matrix of regressors for the i-th cross section, and δ_i is the corresponding vector of coefficients.

The matrix a_i is not completely specified, as it depends on unknown break points in the slope. But a slight modification of the iterative approach considered in Section 4 can be used. Suppose the break points are known. Then we can start the iteration by first estimating δ_i with the least squares, i.e. $\widetilde{\delta}_i = (a_i' a_i)^{-1} a_i' x_i$. Because f has zero mean and may therefore be treated as a part of regression errors, $\widetilde{\delta}_i$ is in fact root-T consistent in the first step. Given $\widetilde{\delta}_i$, we estimate the factors and factor loadings based on $\widetilde{w}_i = x_i - a_i \widetilde{\delta}_i$ $(i = 1, 2..., N)$. Due to the good starting value for $\widetilde{\delta}_i$, convergence is rapid, where convergence is achieved when the successive change in sum of squared residuals is small than a prespecified small number. The iteration procedure is equivalent to simultaneously estimating f, π_i, and δ_i.

In general, we need to estimate the break points. Because ΔF_t has zero mean, we can regard $f\pi_i + ze_i$ as the overall disturbance. Then equation (21) is a simple model with mean breaks. We can apply the Bai-Perron dynamic programming algorithm to estimate the number and the location of the breaks. This is done equation by equation, unless one assumes common break dates across equations. Owing to the consistency and fast convergence rate for the estimated break points, the regression matrix a_i indeed can be treated as known. A formal argument is presented in Bai and Carrion-i-Silvestre (2004).

With breaks in slope, the estimation procedure consists of two steps. The first step estimates m_i and the break points $T_{b,k}^i$, $k = 1, \ldots, m_i$, to obtain $(\widehat{m}_i, \widehat{T}_{b,k}^i; k = 1, \ldots, m_i)$, from which we define the regressor matrix $\widehat{a}_i$. The second step estimates f, π_i, and δ_i $(i = 1, 2, ..., N)$ from the equation $x_i = f\pi_i + \widehat{a}_i\delta_i + z_i$ with an iteration procedure described earlier. Let $\widehat{f}$, $\widehat{\pi}_i$ and $\widehat{\delta}_i$ be the final estimates. The residual vector is given by

$$\widehat{z}_i = x_i - \widehat{f}\widehat{\pi}_i - \widehat{a}_i\widehat{\delta}_i.$$

The cumulative sum of $\widehat{z}_{i,t}$ gives $\widehat{e}_{i,t} = \sum_{s=2}^{t} \widehat{z}_{i,s}$. The MSB test based on the sequence $\widehat{e}_{i,t}$ is computed as in (20) for each i to test the null hypothesis that $\rho_i = 1$ in (13), against the alternative hypothesis that $|\rho_i| < 1$. The limiting distribution of the MSB statistics when there are common factors and multiple structural breaks is given in the following result.

Result M: Unit Root Test with Breaks Assume that the trend break points satisfy $T_{b,k}^i = [T\lambda_{i,k}]$, where $[x]$ denotes the largest integer less than or equal to x, and $\lambda_{i,k} \in (0,1)$, $k = 1, 2, ..., m_i$, $i = 1, 2..., N$. Under the null hypothesis that $\rho_i = 1$, the MSB test in (20) converges in distribution to

(1) Model 1: $MSB_i \Rightarrow \int_0^1 W_i^2(r)\, dr$

(2) Model 2: $MSB_i(\lambda_i) \Rightarrow \sum_{k=1}^{m_i+1} (\lambda_{i,k} - \lambda_{i,k-1})^2 \int_0^1 V_{i,k}^2(b)\, db$,

where $W_i(r)$ is the standard Brownian motion independent across i, $V_{i,k}(b) = W_{i,k}(b) - bW_{i,k}(1)$ is a Brownian bridge independent across i and k, $k = 1, \ldots, m_i + 1$, and $\lambda_{i,0} = 0$ and $\lambda_{i,m_i+1} = 1$, $i = 1, \ldots, N$.

Since the common factors F do not appear in the limiting distribution, the test statistics $MSB_i(\lambda)$ are asymptotically independent provided that e_{it} are independent over i. This implies that one can construct a valid pooled test statistic. Let ξ_i denote the limiting distribution. It is very easy to derive the mean and variance of ξ_i. which are given in Bai and Carrion-i-Silvestre (2005)). One way of pooling is through standardization

$$P_1 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{MSB_i(\lambda_i) - E(\xi_i)}{var(\xi_i)}.$$

The other way of pooling is the combination of individual p-values. Let p_i denote the the p-value for $MSB_i(\lambda_i)$. Define

$$P_2 = \frac{-2\sum_{i=1}^{N}\log(p_i) - 2N}{2\sqrt{N}}$$

Given a value of test statistic $MSB_i(\lambda_i)$, its p-value can be easily obtained via the response surfaces computed by Bai and Carrion-i-Silvestre (2004). The above discussion focuses on breaks in the mean and slops. Breaks in factor loadings are studied by Kao et al. (2007).

7.5 Panel Cointegration with Global Trends

Consider now the model

$$Y_{it} = X_{it}'\beta + u_{it}$$

where for $i = 1, ..., N$, $t = 1, ..., T$, Y_{it} is a scalar, $X_{it} = X_{it-1} + \nu_{it}$. If u_{it} is stationary and iid across i, we can then say the panel shares a common cointegrating vector, $(1, -\beta')$. It is straightforward to show that $\widehat{\beta}$ is superconsistent, even though its limiting distribution is, in general, non-standard. The assumption that u_{it} is cross-sectionally independent is strong. To remove this assumption, we assume

$$u_{it} = \lambda_i' F_t + e_{it}$$

where F_t is a $r \times 1$ vector of latent common factors, λ_i is a $r \times 1$ vector of factor loadings and e_{it} is the idiosyncratic error. If F_t is stationary, Y_{it} and X_{it} are still cointegrated in a panel sense. But if F_t is non-stationary, panel cointegration effectively occurs between Y_{it}, X_{it} and F_t. For obvious reasons, the model can be termed panel cointegration with global stochastic trends. Define

$$w_i = Y_i - X_i\beta = F\lambda_i + e_i.$$

Our proposed continuous updated estimator (Cup) for (β, F) is defined as

$$\left(\widehat{\beta}_{Cup}, \widehat{F}_{Cup}\right) = \underset{\beta,F}{\operatorname{argmin}}\, S_{NT}\left(\beta, F\right).$$

More precisely, $(\widehat{\beta}_{Cup}, \widehat{F}_{Cup})$ is the solution to the following two nonlinear equations

$$\widehat{\beta} = \left(\sum_{i=1}^{N} X_i' M_{\widehat{F}} X_i\right)^{-1} \sum_{i=1}^{N} X_i' M_{\widehat{F}} Y_i \qquad (22)$$

$$\widehat{F} V_{NT} = \left[\frac{1}{NT^2}\sum_{i=1}^{N}\left(Y_i - X_i\widehat{\beta}\right)\left(Y_i - X_i\widehat{\beta}\right)'\right]\widehat{F} \qquad (23)$$

where $M_{\widehat{F}} = I_T - T^{-2}\widehat{F}\widehat{F}'$ since $\widehat{F}'\widehat{F}/T^2 = I_r$, and V_{NT} is a diagonal matrix consisting of the r largest eigenvalues of the matrix inside the brackets, arranged in decreasing order. The estimator of $\widehat{\beta}$ has the same form as in the stationary case considered earlier. But the limiting distribution is different.

Result N: Panel Cointegration and Global Trends Define the bias-corrected Cup estimator as $\widehat{\beta}_{CupBC} = \widehat{\beta}_{Cup} - \frac{1}{T}\widehat{\phi}_{NT}$ where ϕ_{NT} is a bias correction term defined in Bai et al. (2006). Then $\sqrt{N}T(\widehat{\beta}_{CupBC} - \beta^0)\xrightarrow{d}N(0,\Sigma)$ for some positive definite matrix Σ.

The result is based on Bai et al. (2006). Because of asymptotic normality, the usual chi-square test for hypothesis on the coefficient β can be performed. The Cup estimator is obtained by iteratively solving for $\widehat{\beta}$ and $\widehat{F}$ using (22) and (23). It is a non-linear estimator even though linear least squares estimation is involved at each iteration. The CupBC estimator is consistent with a limiting distribution that is centered at zero as long as $(N,T)\to\infty$ and $\frac{N}{T}\to 0$. A fully modified estimator can also be obtained along the lines of Phillips and Hansen (1990). The resulting bias-corrected CupFM estimator is also $\sqrt{N}T$ consistent and asymptotically normally. An extension to include incidental trends such as

$$Y_{it} = \mu_i + \gamma_i t + X_{it}'\beta + \lambda_i' F_t + e_{it}$$

is also possible. The presence of incidental trends will affect the limiting covariance matrix Σ, but asymptotic normality and and thus chi-squared distribution for hypothesis testing on β remain valid.

7.6 Testing Panel Cointegration with Global Trends

Panel cointegration imposes the restriction that e_{it} in (24) are all I(0). This restriction can be tested. Bai and Carrion-i-Silvestre (2005) consider testing the null hypothesis of no cointegration versus cointegration. More specifically, consider testing $H_0 : e_{it} \sim I(1)$ versus $H_1 : e_{it} \sim I(0)$ in the following heterogeneous slope panel cointegration model:

$$Y_{it} = \mu_i + \gamma_i t + X_{it}'\beta_i + \lambda_i' F_t + e_{it}. \tag{24}$$

We assume both X_{it} and F_t are I(1). That is, $X_{it} = X_{i,t-1} + v_{it}$ and $F_t = F_{t-1} + \eta_t$. As shown in Bai and Carrion-i-Silvestre (2005), there is no loss of generality for testing purpose by assuming $(\Delta X'_{it}, \Delta F'_t) = (v'_{it}, \eta'_t)$ is uncorrelated with Δe_{it}. This implies that the product $\Delta X_{it} \Delta e_{it}$ will be a sequence of zero mean random variables. As long as Δe_{it} and ΔX_{it} have weak serial correlations, the product $\Delta X_{it} \Delta e_{it}$ is also serially correlated but only weakly. So we can assume $T^{-1/2} \sum_{t=1}^{T} \Delta X_{it} \Delta e_{it}$ to be $O_p(1)$. Similarly, $T^{-1/2} \sum_{t=1}^{T} \Delta F_t \Delta e_{it}$ is also $O_p(1)$.

Again as in Bai and Ng (2004), differencing in the intercept only case leads to

$$\Delta Y_{it} = \Delta X'_{it} \beta_i + \lambda'_i \Delta F_t + \Delta e_{it}$$

In the absence of correlation between ΔX_{it} and ΔF_t, we can first estimate β_i by least squares, treating $\lambda_i F_t + \Delta e_{it}$ as the regression error. The lack of correlation between ΔX_{it} and the error $\lambda_i \Delta F_t + \Delta e_{it}$ makes it possible to obtain $\sqrt{T}$ consistent estimation of β_i. We can then estimate Λ and F by the method of principal components treating $\Delta X_{it} \beta_i + \Delta e_{it}$ as the idiosyncratic errors. But when ΔX_{it} and ΔF_t are correlated, the two step procedure will be inconsistent; we must estimate the unknown quantities simultaneously. This can be achieved by the iteration procedure outline in Result N, with F_t replaced by $f_t = \Delta F_t$.

Let $\widetilde{z}_i = y_i - x_i \widetilde{\beta}_i - \widetilde{f} \widetilde{\lambda}_i$, and define $\widetilde{e}_{it} = \sum_{s=2}^{t} \widetilde{z}_{is}$. The MSB statistic is again computed as

$$MSB_i = \frac{T^{-2} \sum_{t=2}^{T} \widetilde{e}_{it}^2}{\widetilde{\sigma}_i^2}$$

where $\widetilde{\sigma}_i^2$ is an estimator for the long-run variance of $e_{it} - \rho e_{it-1}$. The estimator is obtained from the residuals $\widetilde{\varepsilon}_{it} = \widetilde{e}_{it} - \widehat{\rho}_i \widetilde{e}_{it-1}$ by the Newey-West (1987) procedure.

For the linear trend case, we difference and demean the data as in PANIC, and $f_t = \Delta F_t - \overline{\Delta F}$.

Result O: Testing Panel Cointegration Under the null hypothesis that $\rho_i = 1$:

O.1 for the intercept only case,

$$MSB_i \xrightarrow{d} \int_0^1 W_i(r)^2 dr$$

where W_i is a Brownian motion.

O.2 for the linear trend case

$$MSB_i \xrightarrow{d} \int_0^1 B_i(r)^2 dr$$

where B_i is Brownian bridge.

It is important to highlight that the limiting distributions do not depend on X_{it} and F_t. Thus the limiting distributions are independent over i provided that e_{it} are cross-sectionally independent. As a result, a valid pooled test statistic for panel cointegration can be constructed. One way of pooling is based on standardization. Let $U_i = \int_0^1 W_i(r)^2$ and $V_i = \int_0^1 B_i(r)^2 dr$. Using $EU_i = 1/2$, $var(U_i)$, $E(V_i) = 1/6$, and $var(V_i) = 1/45$, we have, for the intercept only case, $\sqrt{3}N^{-1/2}\sum_{i=1}^N(MSB_i - 1/2)\xrightarrow{d}N(0,1)$, and for the linear trend case, $\sqrt{45}N^{-1/2}\sum_{i=1}^N(MSB_i - 1/6)\xrightarrow{d}N(0,1)$. Pooling based on p-values can also be easily constructed.

An LM type test for panel cointegration that also permits structural breaks is considered by Westerlund and Edgerton (2007).

8 How Precise are the Factor Estimates?

The key to being able to use the principal component estimates as though they were the observed factors is that the factor space can be estimated precisely as N and T tend to infinity. In essence, if the principal components can estimate the space spanned by the true factors as N and T increase, and the sample principal components precisely estimate the population principal components as N gets large, the sample principal components will consistently estimate the space spanned by the factors. In simulations, we found that when the errors are iid over i and t, $\min[N, T]$ can be as small as 30 and the number of factors can be estimated with almost perfect certainty, suggesting that the factor space can indeed be estimated with high precision even with rather small samples. Figure 3 illustrates this for stationary and non-stationary factors.

Can it be possible that increasing N does not improve the precision of the factor estimates? Boivin and Ng (2006) argued that this is possible if the additional data are uninformative about the factor structure. This can arise if the idiosyncratic error variances are large, or if the factor loadings of the additional data are small. More generally, our theory for approximate factor models permits heteroskedasticity, some cross-section and some serial correlation in the errors. Errors in which such effects are strong would be incompatible with Assumption E. When the importance of the idiosyncratic error is magnified, it will become more difficult to separate out the common from the idiosyncratic component in the data, and data with these characteristics cannot be ruled out in practice.

Efficiency issues surrounding the principal components estimator are best understood in terms of OLS versus GLS. It is well known that when the errors are spherical and other conditions of the Guass-Markov theorem are met, OLS is the most efficient amongst the class of linear unbiased estimators. However, when the errors are non-spherical, a GLS estimator that exploits information in the structure of the error variance is more efficient. In the present context, the principal components estimator is based on an unweighted objective function that minimizes the sum of squared residuals. No consideration is given as to whether a particular observation has errors that may be correlated over t or i. Inefficient estimates can also be thought of as a consequence of not exploiting the moments implied by the idiosyncratic errors. However, unlike in simple regressions when feasible FGLS is usually possible, a weighted principal components estimator that weights the errors inversely by Ω is not possible. This is because the estimate of Ω, or the sample covariance of $\widetilde{e}$, cannot be full rank if $\widetilde{e}$ are the errors associated with data generated by $r \geq 1$ factors. Indeed, $\widetilde{\Omega}$ has a rank of $\min[T, N] - r$ making $\widetilde{\Omega}$ non-invertible. To remedy this problem, several suggestions have been made to estimate weighted principal components by solving the problem

$$\min_{F_1,\ldots F_T,\Lambda} \sum_{t=1}^{T} (X_t - \Lambda F_t)'\widehat{\Omega}(X_t - \Lambda F_t),$$

giving $\widetilde{\Lambda}$ as the eigenvectors of $\widehat{\Omega}^{-1/2}\widehat{\Sigma}_X\widehat{\Omega}^{-1/2\prime}$. The methods differ in the choice of $\widehat{\Omega}$. Jones (2001) and Boivin and Ng (2006) use the diagonal of $\widetilde{\Omega}$ obtained from the unweighted estimation, while Forni et al. (2005) set $\widehat{\Omega} = \widehat{\Sigma} - \widehat{\Sigma}_{\widetilde{C}}$, where $\widehat{\Sigma}_{\widetilde{C}}$ is the covariance matrix of the unweighted estimated common component $\widetilde{C}_{it}$. Boivin and Ng (2006) also considered a priori rules that set the off-diagonal elements of $\widehat{\Omega}$ to be either zeros or ones depending on whether the error of a series is deemed correlated with other errors. However, none of these methods are optimal in any formal sense.

Let eig_r^z be the r-th largest eigenvalue of the population covariance matrix of the N variables, z. Recall that in theory, eig_r^x should diverge as N increases while eig_{r+1}^x should be bounded. Theory also says that eig_r^x/eig_1^e should diverge since the numerator increases with N and the denominator is bounded. With weak loadings, correlated or large errors, eig_r^x will not be 'too different' from eig_{r+1}^x. Heaton and Solo (2006) interpret eig_r^x/eig_{r+1}^x as an indicator of signal to noise. The eigenvalue conditions, while conceptually simple, are not easily translated into practical use because with one set of observational data, we cannot assess how the eigenvalues increase with N.

To better understand the precision of the factor estimates under more general conditions, we consider a monte carlo experiment. The data are generated according to the dynamic factor model. For $i = 1, \ldots N$ and $t = 1, \ldots T$,

$$x_{it} = \lambda_i'(L)f_t + \sigma_i e_{it}$$

where σ_i^2 is set so that for a pre-specified $R_i^2 \sim U[R_L^2, R_U^2]$, $\frac{\text{var}(\lambda_i(L)'f_t)}{\text{var}(e_{it})} = \frac{R_i^2}{1-R_i^2}$ on average. In the simulations, we fix R_U^2 to .8. The factor loadings $\lambda_i(L) = \lambda_{i0} + \lambda_{i1}L + \ldots \lambda_i L^s$ are generated with $\lambda_{ij} \sim N(0,1)$ for $j = 0, \ldots s$. When $s = 0$, we have $r = q = 1$ static factor. When $s > 0$, we have $r = q(s+1)$ static factors but $q = 1$ dynamic factor. The single common factor f_t and the idiosyncratic errors evolve according to

$$\begin{aligned}(1-\rho_f L)f_t &= u_t, \quad u_t \sim N(0,1) \\ (1-\rho_e L)e_{it} &= \varepsilon_{it}, \quad E(\varepsilon_t \varepsilon_t') = \Omega.\end{aligned}$$

The error variance matrix Ω is an identity matrix of order N when the errors are cross-sectionally uncorrelated. Otherwise, it is a positive definite correlation matrix such that a fraction N_c of the N^2 elements of Ω are non-zero. The parameters of the simulations are

- (N, T)=(20,50), (50,100), (100,50), (100,100), (50,200), (100,200);
- $s = 0, 1$;
- ρ_f= 0, .4, .8 ;
- ρ_e=0, U(0, .5), or U(.4, .8)
- R_L^2=.1, .35, .6;
- N_c= 0, .15, .3;

For a given s, there are 81 configurations for each sample size, giving a total of 486 configurations. We consider 1000 replications for each configuration. In each run, we keep track of the eigenvalues of $\Sigma_{xx} = x'x/NT$ and of $\Omega = e'e/(NT)$. Let eig_r^x be average of the r-th largest eigenvalue of the matrix Σ_{xx} over 1000 replications, and let eig_1^e be corresponding the largest eigenvalue of Ω. We will use $EIG_{A,B}(a, b)$ to denote the ratio of the a-th largest eigenvalue of the covariance matrix of A to the b-th largest eigenvalue of the covariance matrix of B. We also keep track of FIT, which is the R^2 from a regression of $\widetilde{F}_t$ on F_t and a constant. When $\widehat{F}_t$ is two-dimensional (as is the case when $s = 1$), we regress each of the $\widetilde{F}_t$ on F_t and then average the two R^2.

To evaluate the precision of the factor estimates and to compactly summarize the results, we perform a response surface analysis by regressing FIT on the sample size, and other potential determinants of FIT. Let $C_{NT} = \min[\sqrt{N}, \sqrt{T}]$ and $\bar{R}^2$ be importance of the common component. The estimates and the robust t statistics are reported in Table 1.

Several features should be noted. First, the precision falls with the number of factors. Second, while FIT improves with $\min[\sqrt{N}, \sqrt{T}]$, the effect is not statistically significant. The variation of FIT is primarily explained by $EIG_{x,x}(r+1, r)$, the ratio of the $r+1$-th to the r-th eigenvalue of Σ_x. Under the assumption that the factors are strong, this ratio should tend to zero in the population as N increases. However, if the factors are weak, such as when the idiosyncratic errors are large or correlated, or when the loadings are weak, this ratio may be non-negligible. As suggested by Onatski (2006a) and Heaton and Solo (2006), the larger this ratio, the less precise will be the factor estimates. This ratio, along with the corresponding quadratic term, explain close to 80% of the variation in FIT. Additional explanatory power is provided by $EIG_{e,x}(1, 1)$, which also measures the size of the idiosyncratic component. Thus the relative importance of the idiosyncratic component is indeed an important determinant of the precision of FIT.

The above results suggest that the principal component estimator can indeed be sensitive to the possibility of weak loadings, large error variances, and cross-section correlation all culminating in a large eig_{r+1}^x that is no longer small relative to eig_r^x as theory assumes. Two questions arise. Is this a situation of empirical relevance, and are there alternatives?

As for the first question, the availability of a lot of data should never be taken to mean the availability of a lot of data informative about the factor structure. Economic reasoning can usually help screen out data that are overly similar to other series already included. For example, it should not come as a surprise that the idiosyncratic error of two series, one being a finer disaggregation of another, are correlated. Boivin and Ng (2006) found that the errors IPC and IPCD (industrial production of consumer goods and of durable goods) are strongly correlated, and a case can be made to include only one series. Pre-screening of the importance of the common component in each series can also go a long way in removing the uninformative variables from the panel of data used in subsequent estimation. Admittedly, pretesting has its consequences. But improving the precision of the factor estimates may well justify the cost.

What are the alternatives to the principal components estimator? For static factor models, the maximum likelihood estimator via EM (expectation and maximization) algorithm has been considered by many authors, for example, Rubin and Thayer (1982), Lehmann and Modest (1988), and Ghahramani and Hinton (1996). Dynamic factor models fit into the state space framework. Under large N, a dynamic factor model is that of a state space model with a small number of state variables but a large number of measurement equations. Naturally, the state space method (especially the EM algorithm combined with the Kalman smoother) can be used to estimate the model, see Watson and Engle (1983), Quah and Sargent (1992), and Shumway and Stoffer (2000). A subspace algorithm is considered by Kapetanios and Marcellino (2006b). Other estimators that have been considered involve re-weighting the data matrix prior to extracting the principal components estimator as discussed earlier. These methods are ad-hoc. Recently, Doz et al. (2007) considered the quasi-maximum likelihood estimator as an alternative. The estimator allows for heterogeneous errors and serially correlated factors. Although similar in spirit to the estimator considered in Anderson (1984), the theory that Doz et al. (2007) developed assumes that N and T are both large. Their QMLE estimator is based on the Kalman filter and reduces to an iteratively reweighted principal components estimator when the data are assumed to be iid. The weighting serves to downweigh the noisy series and goes some way in resolving the weak loading/large error variance problem. The advantage of the estimator is that it allows restrictions to be imposed on the factor structure, whereas the principal components estimator is best thought of as an unrestricted reduced form type estimator. The limitation, however, is that QMLE is based on an approximation to the 'approximate factor model', where the approximate model is in fact a strict factor model that assumes the errors are cross-sectionally uncorrelated. Therefore, information in the correlated errors is never taken into account, just like the unweighted estimator. Although their estimator appears to work well for a low degree of cross-section correlation, it is unclear if the properties hold when the data with a high noise to signal ratio (in the sense of eig^x_r/eig^x_{r+1} not tending to zero) are being analyzed. It therefore appears that no frequentist method has yet provided a satisfactory solution to the problem of correlated errors.

Can Bayesian methods solve the problem? At the moment, Bayesian analysis of dynamic factor models still do not allow for cross-sectionally correlated errors. In preparing the monte carlo study, we also obtained Bayesian factor estimates using the method discussed in Otrok and Whiteman (1998), as well as the one described in Kim and Nelson (2000). The Bayesian methods give posterior means very similar results to the principal component estimates, but are tremendously more time consuming to compute with little to no gain in precision. Thus, while the Bayesian methods can be alternatives to the principal components approach, they cannot (so far) be justified on precision grounds. Coming up with alternatives to the principal components estimator remains very much an important area for research.

9 Conclusion

This paper has surveyed some of the theoretical results relating to the use of principal components as estimated factors in empirical work. The estimator is simple to compute and can estimate well the factor space when conditions required for consistent estimation are satisfied. In practice, the data may be at odds with some of these conditions. New simulation results provided in this paper show that the factor estimates can be severely compromised when the data have a weak factor structure in the sense that $\text{eig}_r^x/\text{eig}_{r+1}^x$ does not tend to zero. This problem may not arise frequently in practice, but it also cannot be ruled out. A user can alleviate the problem to some extent by using only data that are truly informative about the factor structure. However, how to deal with cross-sectionally correlated errors, which is a genuine feature of an approximate factor model, remains an unresolved issue. Thus, although much work has been accomplished in this research, much more remains to be done.

References

Amengual, D. and Watson, M. 2007, Consistent Estimation of the Number of Dynamic Factors in Large N and T Panel, *Journal of Business and Economic Statistics* **25:1**, 91–96.

Anderson, H. and Vahid, F. 2007, Forecasting the Volatility of Australian Stock Returns: Do Common Factors Help?, *Journal of the American Statistical Association* **25:1**, 75–90.

Anderson, T. W. 1984, *An Introduction to Multivariate Statistical Analysis*, New York: Wiley.

Anderson, T. W. and Rubin, H. 1956, Statistical Inference in Factor Analysis, *in* J. Neyman (ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. V, Berkeley: University of California Press, pp. 114–150.

Artis, M., Banerjee, A. and Marcellino, M. 2005, Factor Forecasts for the U.K., *Journal of Forecasting* **24**, 279–298.

Bai, J. 2003, Inferential Theory for Factor Models of Large Dimensions, *Econometrica* **71:1**, 135–172.

Bai, J. 2004, Estimating Cross-Section Common Stochastic Trends in Non-Stationary Panel Data, *Journal of Econometrics* **122**, 137–183.

Bai, J. and Carrion-i-Silvestre, J.L., 2004, Structural Changes, Common Stochastic Trends, and Unit Roots in Panel Data, unpublished manuscript, revised 2007.

Bai, J. and Carrion-i-Silvestre, J.L., 2005, Testing Panel Cointegration with Unobservable Dynamic Common Factors, Department of Economics, New York University, unpublished manuscript.

Bai, J. and Ng, S. 2002, Determining the Number of Factors in Approximate Factor Models, *Econometrica* **70:1**, 191–221.

Bai, J. and Ng, S. 2004, A PANIC Attack on Unit Roots and Cointegration, *Econometrica* **72:4**, 1127–1177.

Bai, J. and Ng, S. 2006a, Confidence Intervals for Diffusion Index Forecasts and Inference with Factor-Augmented Regressions, *Econometrica* **74:4**, 1133–1150.

Bai, J. and Ng, S. 2006b, Evaluating Latent and Observed Factors in Macroeconomics and Finance, *Journal of Econometrics* **113:1-2**, 507–537.

Bai, J. and Ng, S. 2006c, Instrumental Variables in a Data Rich Environment, unpublished manuscript, Department of Economics, Columbia University http://www.columbia.edu/ sn2294/papers/iv.pdf

Bai, J. and Ng, S. 2006d, Panel Unit Root Tests with Cross-Section Dependence, unpublished manuscript, Department of Economics, Columbia University. http://www.columbia.edu/ sn2294/papers/newpanic.pdf

Bai, J. and Ng, S. 2007, Determining the Number of Primitive Shocks, *Journal of Business and Economic Statistics* **25:1**, 52–60.

Bai, J. and Ng, S. 2008, Extremum Estimation when the Predictors are Estimated from Large Panels, unpublished manuscript, Department of Economics, Columbia University. http://www.columbia.edu/ sn2294/papers/probit.pdf

Bai, J., Kao, C. and Ng, S. 2006, Panel Cointegration with Global Stochastic Trends, unpublished manuscript, Department of Economics, University of Michigan. http://www.columbia.edu/ sn2294/papers/bkn.pdf

Banerjee, A., Marcellino, M. and Osbat, C. 2005, Testing for PPP: Should we use Panel Methods?, *Empirical Economics* **30**, 77–91.

Banerjee, A., Masten, I. and Massimiliano, M. 2006, Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples With Structural Change, *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, Elsevier.

Bernanke, B. and Boivin, J. 2003, Monetary Policy in a Data Rich Environment, *Journal of Monetary Economics* **50:3**, 525–546.

Boivin, J. and Ng, S. 2005, Undertanding and Comparing Factor Based Forecasts, *International Journal of Central Banking* **1:3**, 117–152.

Boivin, J. and Ng, S. 2006, Are More Data Always Better for Factor Analysis, *Journal of Econometrics* **132**, 169–194.

Breitung, J. and Das, S. 2007, Testing for Unit Root in Panels with a Factor Structure, *Econometric Theory* Vol 24, 88-108.

Breitung, J. and Eickmeier, S. 2005, Dynamic Factor Models, Deutsche Bundesbank Discussion Paper 38/2005.

Breitung, J. and Pesaran, H. M. 2007, Unit Roots and Cointegration in Panels, *in* L. Matyas and P. Sevestre (eds), *The Econometrics of Panel Data*, Klver Academic Press.

Chamberlain, G. and Rothschild, M. 1983, Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets, *Econometrica* **51:5**, 1281–1304.

Connor, G. and Korajzcyk, R. 1986, Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis, *Journal of Financial Economics* **15**, 373–394.

Connor, G. and Korajzcyk, R. 1998, Risk and Return in an Equilibrium APT Application of a New Test Methodology, *Journal of Financial Economics* **21**, 225–289.

Connor, Hagmann, G. M. and Linton, O. 2007, Efficient Estimation of a Semiparametric Characteristic-based Factor Model of Security Returns, Department of Economics, London School of Economics, unpublished manuscript.

Cristadoro, R., Forni, M., Reichlin, L. and Giovanni, V. 2001, A Core Inflation Index for the Euro Area, unpublished manuscript, www.dynfactor.org.

den Reijer, A. H. J. 2005, Forecasting Dutch GDP using Large Scale Factor Models, DNB Working Papers 028. http://ideas.repec.org/p/dnb/dnbwpp/028.html

Doz, C., Giannone, D. and Reichlin, L. 2007, A Quasi-Maximum Likelihood Approach for Large Approximate Dynamic Factor Models, European Central Bank Working Paper Series 674. http://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp674.pdf

Engle, R. F. and Granger, C. 1987, Cointegration and Error-Correction: Representation, Estimation, and Testing, *Econometrica* **55:2**, 251–276.

Favero, C., Marcellino, M. and Neglia, F. 2005, Principal Components at Work: the Empirical Analysis of Monetary Policy with Large Datasets, *Journal of Applied Econometrics* **20**, 603–620.

Forni, M., Giannone, D., Lippi, M. and Reichlin, L. 2003, Opening the Black Box: Identifying Shocks and Propagation Mechanisms in VAR and Factor Models, unpublished manuscript.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. 2000, The Generalized Dynamic Factor Model: Identification and Estimation, *Review of Economics and Statistics* **82:4**, 540–554.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. 2001, Do Financial Variables Help in Forecasting Inflation and Real Activity in the Euro Area, manuscript, http://homepages.ulb.ac.be/ lreichli/financial.pdf.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. 2004, The Generalized Factor Model: Consistency and Rates, *Journal of Econometrics* **119**, 231–255.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. 2005, The Generalized Dynamic Factor Model, One Sided Estimation and Forecasting, *Journal of the American Statistical Association* **100**, 830–840.

Geweke, J. 1977, The Dynamic Factor Analysis of Economic Time Series, *in* D. J. Aigner and A. S. Goldberger (eds), *Latent Variables in Socio Economic Models*, Amsterdam: North Holland.

Ghahramani, Z. and Hinton, G. E. 1996, The EM Algorithm for Mixtures of Factor Analyzers, *Department of Computer Science, University of Toronto.* Technical Report CRG-TR-96-1,.

Giannone, D., Reichlin, L. and Sala, L. 2005a, Monetary Policy in Real Time, *Macroeconomic Annual* **19**, 161–200.

Giannone, D., Reichlin, L. and Sala, L. 2005b, VARs, Common Factors and the Empirical Validation of Equilibrium Business Cycle Models, *Journal of Econometrics* **127:1**, 257–279.

Giannone, D., Reichlin, L. and Small, D. 2007, Nowcasting: The Real-Time Informational Content of Macroeconomic Data, forthcoming in *Journal of Monetary Economics*.

Hallin, M. and Liska, R. 2007, Determining the Number of Factors in the General Dynamic Factor Model, *Journal of the American Statistical Association* **102**, 603–617.

Heaton, C. and Solo, V. 2006, Estimation of Approximate Factor Models: Is it Important to have a Large Number of Variables, Department of Economics, Macquarie University, Research Papers series 0605. http://www.econ.mq.edu.au/research/2006/HeatonEstimtnOfApproxFactorModels.pdf

Jones, C. 2001, Extracting Factors from Heteroskedsatic Asset Returns, *Journal of Financial Economics* **62:2**, 293–325.

Kao, C., Trapani, L. and Urga, G. 2007, Modelling and Testing for Structural Breaks in Panels with Common Stochastic Trends, Cass Business School, unpublished manuscript.

Kapetanios, G. 2007, A Testing Procedure for Determining the Number of Factors in Approximate Factor Models with Large Datasets, forthcoming in *Journal of Business and Economic Statistics*.

Kapetanios, G. and Marcellino, M. 2006a, Factor-GMM Estimation with Large Sets of Possibly Weak Instruments, unpublished manuscript.

Kapetanios, G. and Marcellino, M. 2006b, A Parametric Estimation Methods for Dynamic Factor Models of Large Dimensions, CEPR WP 5620.

Kim, C. and Nelson, C. 2000, *State Space Models with Regime Switching*, Cambridge, MA: MIT Press.

Lawley, D. N. and Maxwell, A. E. 1971, *Factor Analysis in a Statistical Method*, London: Butterworth.

Lehmann, B. N. and Modest, D. 1988, The Empirical Foundations of the Arbitrage Pricing Theory, *Journal of Financial Economics* **21**, 213–254.

Ludvigson, S. and Ng, S. 2005, Macro Factors in Bond Risk Premia, NBER Working Paper 11703.

Ludvigson, S. and Ng, S. 2007, The Empirical Risk Return Relation: A Factor Analysis Approach, *Journal of Financial Economics* **83**, 171–222.

Marcellino M, Stock J.H. and Watson M., (2003), Macroeconomic forecasting in the Euro area: country specific versus Euro wide information, European Economic Review, 47, 1-18.

Moon, R. and Perron, B. 2004, Testing For a Unit Root in Panels with Dynamic Factors, *Journal of Econometrics* **122:1**, 81–126.

Moon, R., Perron, B. and Phillips, P. 2007, Incidental Trends and the Power of Panel Unit Root Tests, *Journal of Econometrics* **141:2**, 416–459.

Onatski, A. 2005, Determining the Number of Factors From Empirical Distribution of Eigenvalues, Department of Economics, Columbia University, Discussion Paper 0405-19. http://www.columbia.edu/cu/economics/discpapr/DP0405-19.pdf

Onatski, A. 2006a, Asymptotic Distribution of the Principal Components Estimator of Large Factor Models when Factors are Relatively Weak, unpublished manuscript, Department of Economics, Columbia University. http://www.columbia.edu/ ao2027/inference33.pdf

Onatski, A. 2006b, A Formal Statistical Test for the Number of Factors in Approximate Factor Models, unpublished manuscript, Department of Economics, Columbia University.

Otrok, C. and Whiteman, C. 1998, Bayesian Leading Indicators: Measuring and Predicting Economic Conditions in Iowa, *International Economic Review* **39:4**, 997–1014.

Pagan, A. 1984, Econometric Issues in the Analysis of Regressions with Generated Regressors, *International Economic Review* **25**, 221–247.

Perron, P. and Ng, S. 1998, An Autoregressive Spectral Density Estimator at Frequency Zero for Nonstationarity Tests, *Econometric Theory* **14**, 560–603.

Pesaran, M.H. 2007, A Simple Unit Root Test in the Presence of Cross-Section Dependence, *Journal of Applied Economics* **22:2**, 265–312.

Pesaran, M.H., 2006, Estimation and Inference in Large Heterogeneous panels with a Multifactor Error Structure, *Econometrica* **74**, 967–1012.

Phillips, P.C.B. and Sul, D. 2003, Dynamic Panel Estimation and Homogeneity Testing Under Cross-Section Dependence, *Econometrics Journal* **6:1**, 217–259.

Phillips, P. C. B. and Hansen, B. E. 1990, Statistical Inference in Instrumental Variables Regression with I(1) Processes, *Review of Economic Studies* **57**, 99–125.

Phillips, P. C. B. and Ploberger, W. 2002, Optimal Testing for Unit Roots in Panel Data. unpublished manuscript, Department of Economics, University of Rochester.

Quah, D. and Sargent, T. 1992, A dynamic index model for large cross sections, *Federal Reserve Bank of Minneapolis,* Discussion Paper 77.

Reichlin, L. 2003, Factor Models in Large Cross Sections of Time Series, *in* S. T. M. Dewatripoint, L. P. Hansen (ed.), *Advances in Economics and Econometrics: Theory and Applications, Vol. 111, 8th World Congress of the Econometric Society*, Cambridge: Cambridge University Press.

Ross, S. 1976, The Arbitrage Theory of Capital Asset Pricing, *Journal of Finance* **13**, 341–360.

Rubin, D. R. and Thayer, D. T. 1982, EM Algorithms for ML Factor Analysis, *Psyhometrika* **47**, 69–76.

Sargan, J. D. and Bhargava, A. 1983, Testing for Residuals from Least Squares Regression Being Generated by Gaussian Random Walk, *Econometrica* **51**, 153–174.

Sargent, T. and Sims, C. 1977, Business Cycle Modelling without Pretending to have too much a Priori Economic Theory, *in* C. Sims (ed.), *New Methods in Business Cycle Research*, Minneapolis: Federal Reserve Bank of Minneapolis.

Schumacher, C. 2005, Forecasting German GDP Using Alternative Factor Models Based on Large Datasets. Bundesbank Discussion paper 24-2005.

Shumway, R. and Stoffer, D. 2000, *Time Series Analsyis and its Applications*, New York: Springer.

Stock, J.H. 1990, A Class of Tests for Integration and Cointegration, Department of Economics, Harvard University, unpublished manuscript.

Stock, J.H. and Watson, M.W. 1988, Testing for Common Trends, *Journal of the American Statistical Association* **83**, 1097–1107.

Stock, J.H. and Watson, M.W. 2002a, Forecasting Using Principal Components from a Large Number of Predictors, *Journal of the American Statistical Association* **97**, 1167–1179.

Stock, J.H. and Watson, M.W. 2002b, Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business and Economic Statistics* **20:2**, 147–162.

Stock, J. H. and Watson, M.W. 2005, Implcations of Dynamic Factor Models for VAR Analysis, *NBER Working Paper 11467*.

Stock, J.H. and Watson, M.W. 2006, Forecasting with Many Predictors, *Handbook of Economic Forecasting*, North Holland: Elsevier.

Watson, M. and Engle, R. 1983, Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC, and Varying Coefficient Regression Models, *Journal of Econometrics* **23**, 385–400.

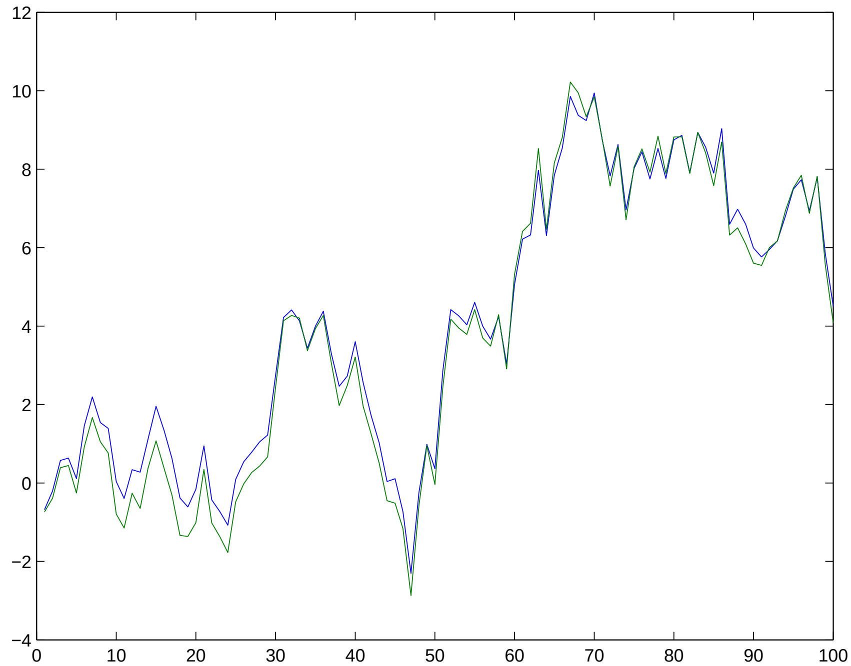
Westerlund, J. and Edgerton, D. 2007, Simple Tests for Cointegration in Dependent Panels with Structural Breaks" Lund University Department of Economics Working Paper No. 2006:13 Available at SSRN: http://ssrn.com/abstract=1080598.

Westerlund, J. and Larsson, R. 2007, A Note on the Pooling of Individual PANIC Unit Root Tests, unpublished manuscript, Department of Economics, Lund University. http://www.nek.lu.se/NEKfng/panic.pdf.

Table 1 Dependent variable: FIT

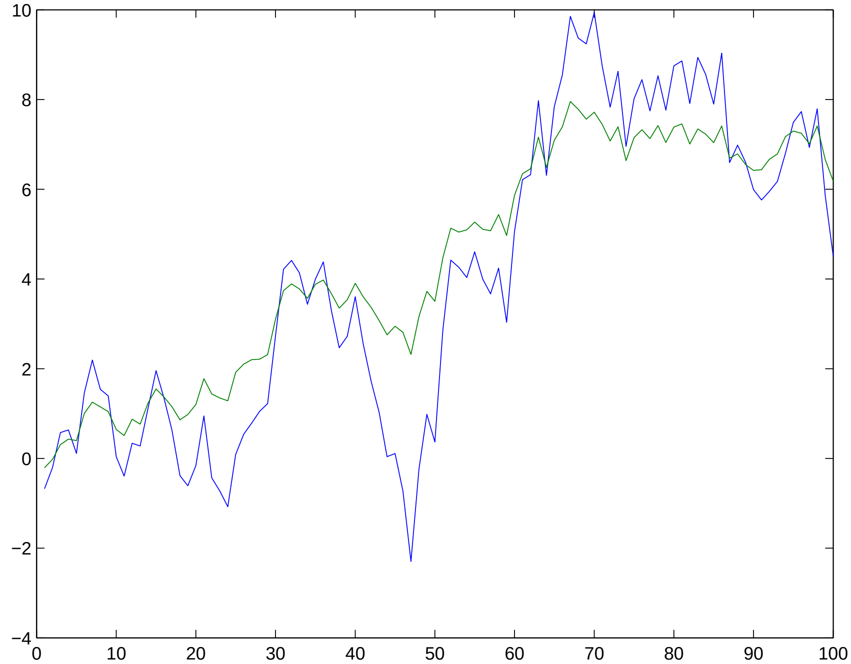
Regressor	$\widehat{\beta}$	$t_{\widehat{\beta}}$	$\widehat{\beta}$	$t_{\widehat{\beta}}$	$\widehat{\beta}$	$t_{\widehat{\beta}}$	$\widehat{\beta}$	$t_{\widehat{\beta}}$
	$r=1$				$r=2$			
constant	0.974	21.244	1.000	66.855	0.958	15.048	0.988	37.176
C_{NT}^{-1}	0.158	0.238	0.219	1.066	-0.257	-0.299	0.022	0.061
C_{NT}^{-1}	-4.086	-1.819	-3.030	-4.250	-3.196	-1.184	-1.499	-1.307
$EIG_{x,x}(k+1,k)$			-0.116	-1.700			0.286	7.681
$EIG_{e,x}(1,1)$			0.025	7.906			-0.019	-5.231
$EIG_{x,x}(k+1,k)^2$			-0.952	-6.564			-1.007	-19.892
$EIG_{e,x}(1,1))^2$			-0.003	-10.694			-0.000	-0.214
$\bar{R}^2$	.246		.927		.121		0.8454	

Figure 1: Estimated common trend from a large spurious system: differenced data approach



Blue: actual F_t; green: estimated F_t

Figure 2: Estimated common trend from a large spurious system: level data approach



Blue: actual F_t; green: estimated F_t

Figure 3: Actual and estimated stationary and nonstationary common factors with stationary idiosyncratic errors ($N = 30, T = 30$)