European University Institute

DEPARTMENT OF ECONOMICS

EUI Working Papers

ECO 2010/28
DEPARTMENT OF ECONOMICS

COMMON FACTORS IN NONSTATIONARY PANEL DATA WITH A DETERMINISTIC TREND – ESTIMATION AND DISTRIBUTION THEORY

Katarzyna Maciejowska

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE
DEPARTMENT OF ECONOMICS

Common Factors in Nonstationary Panel Data with a Deterministic Trend – Estimation and Distribution Theory

KATARZYNA MACIEJOWSKA

EUI Working Paper **ECO** 2010/28

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Common factors in nonstationary panel data with a deterministic trend - estimation and distribution theory

Katarzyna Maciejowska
EUI

June 8, 2010

Abstract

The paper studies large-dimention factor models with nonstationary factors and allows for deterministic trends and factors integrated of order higher then one.We follow the model specification of Bai (2004) and derive the convergence rates and the limiting distributions of estimated factors, factors loadings and common components. We discuss in detail a model with a linear time trend. We ilustrate the theory with an empirical exmple that studies the fluctuations of the real activity of U.S.economy. We show that these fluctuationas can be explained by two nonstationary factors and a small number of stationary factors. We test the economic interpretation of nonstationary factors.

JEL classification: C13, C33, C43
Keywords: Common-stochastic trends; Dynamic factors; Generalized dynamic factor models; Principal components; Nonstationary panel data

discusse

1 Motivation

In the last decade, one could observe a growing interest in models that can extract and use information from large sets of variables. One approach is based on an assumption that there exist common factors, which can explain the variables' comovement. The factor models have been shown useful in econometric modeling. There is a series of articles that demonstrate advantages of using factors in forecasting (Stock and Watson (2002a), Stock and Watson (2002b)) and impulse response analysis (Bernanke, Boivin, and Eliasz (2005), Kapetanios and Marcellino (2006))).

Recently, Stock and Watson (2005) adopts factor models for structural analysis. This article, together with other papers (Kapetanios and Marcellino (2006) and Forni, Giannone, Lippi, and Reichlin (2007)) discusses the possibility of integrating the factor methods into the SVAR framework. There is empirical evidence that factors can contribute to classical VAR analysis (see Bernanke, Boivin, and Eliasz (2005), Kapetanios and Marcellino (2006), Eickmeier (2009) and Forni and Gambetti (2008)).

So far, most of the research concentrates on modeling stationary panel data. Breitung and Eickmeier (2005) provides a comprehensive literature review of stationary dynamic factors models and their applications. There are, however, few articles that discuss the issue of common nonstationary trends. Bai (2004), Bai and Ng (2004) and Gonzalo and Granger (1995) describe estimation methods of nonstationary common components. Bai (2004) proposes information criteria, IPC, that allow for consistent estimation of the number of common trends and derives limiting distributions of estimated factors and common components. Banerjee and Marcellino (2008) discusses cointegration issues related to the existence of common trends and shows how the factor analysis can contribute to the existing literature. Eickmeier (2009) uses nonstationary factors in structural analysis of economic development of euro area countries.

The literature discusses two approaches in modeling nonstationary panels. The first one is based on the differenced data and was proposed by Bai and Ng (2004). This method allows for consistent estimation of nonstationary static factors and is independent from an integration order of the idiosyncratic component[1]. The second approach uses the data in levels and was introduced by Bai (2004). It is suitable for structural analysis because it directly estimates the dynamic nonstationary factors. The concept can also be easily integrated into the generalized dynamic factor models framework. Unfortunately, the results rely on the stationarity assumption of idiosyncratic errors, which is sometimes difficult to verify.

[1] The modeling strategy cannot be directly applied for structural analysis because it deals only with the static representation of the factor model. In order to recover dynamic factors, some additional steps have to be introduced, as in Eickmeier (2009).

In this paper, we follow the idea of Bai and Ng (2004) and extract factors from data in levels. We contribute to the existing literature by allowing for higher order dynamics in the data generating processes. We show that ignoring the time trend or $I(2)$ processes[2] leads to inconsistent estimation of factors and factors loadings. It has important implications for structural analysis and impulse responses. If it is not taken into consideration then some of the factor loadings grows to infinity and the relative importance of some shock increases unproportionaly. Moreover, we derive the convergence rates, the asymptotic distribution of factors, factor loadings and common components for a general model. The dynamics of the factors are summarized by a scaling matrix. It is chosen to ensure the convergence of the factors second moments. The results allow for the assessment of the accuracy of estimation procedure and for constructing confidence intervals around a rotation of true factors used in empirical analysis.

The theory is illustrated with an empirical example. We analyze a panel of 69 real variables describing the U.S. economy. We show that the data fluctuation can be summarized by a small number of common factors. Since most of the variables have a deterministic trend, then it is relevant to assume an existence of a factor with the time trend. The limiting distributions allows for testing if an interest rate, investments, a personal consumption and government spendings are the driving forces of the economy.

This paper is organized as follows: Section 2 describes the model and discusses the estimation issues. In Section 3, we derive the convergence rates and asymptotic distributions of estimates for a general model. Section 4 analyzes in more detail the model with $I(1)$ factors with a deterministic trend. In Section 5, we apply the approach to the panel measuring the real activity of U.S. economy. Finally, in Section 6, we summarize and conclude. The description of the data and proofs are provided in Appendix.

2 Model description and estimation

2.1 Model setup

Let us denote by X a $N \times T$ panel of observable variables. We use F_t^0, λ_i^0 and r to describe the true common factors, factor loadings and number of factors, respectively. Then for any $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$ it is assumed that

$$X_{it} = \lambda_i^{0\prime} F_t^0 + e_{it} \tag{1}$$

The residuals e_{it} are $I(0)$ error processes that can be serially correlated. F_t^0 is a $r \times 1$ vector of common factors and λ_i^0 is a $r \times 1$ vector of factor loadings.

Let X_i be $T \times 1$ vector of observations of the ith cross-section unit. Then

[2] A process X is $I(d)$ (integrated of order d) if d is a smallest number such that $(1-L)^d X$ is stationary. Here, L denotes a lag operator.

$$X_i = F^0 \lambda_i^0 + e_i$$

where $X_i = (X_{i1}, X_{i2}, \ldots, X_{iT})'$, $F^0 = \left(F_1^0, F_2^0, \ldots, F_T^0\right)'$ and $e_i = (e_{i1}, e_{i2}, \ldots, e_{iT})'$.

When it is needed, we will use the following notation

$$X = F^0 \Lambda_0' + e$$

where $\Lambda_0 = \left(\lambda_1^0, \lambda_2^0, \ldots, \lambda_N^0\right)'$ and e is a $N \times T$ matrix, $e = (e_1, ..., e_N)$.

In the model, we distinguish between common factors, F^0, and a common component, denoted by C. The common component is a $T \times N$ matrix that summarizes the total impact of the factors on the panel, defined as a product of factors and factor loadings

$$C = F^0 \Lambda_0'$$

The model setup is similar to the one described in Bai (2003) and Bai (2004). We do not assume any particular type of common factors. Thus, we allow for stationary, $I(1)$ or $I(2)$ factors with or without a deterministic time trend. It is assumed that a kth factor is generated by the following process

$$(1 - L)^d F_{kt}^0 = a_{kt} + u_{kt}$$

where L denotes the lag operator and d takes values $d \in \{0, 1, 2\}$. When $d = 0$ and $a_{kt} = a$ then the process is stationary, whereas for $d = 1$ or 2 the factors are nonstationary $I(1)$ or $I(2)$ processes, respectively. The a_{kt} denotes a deterministic component and u_{kt} is a stationary process. We define by u_t a $r \times 1$ vector of common shocks $u_t = (u_{1t}, ..., u_{rt})'$.

In this article, we are particularly interested in models with nonstationary factors of order not higher than one and a linear time trend. It this case either

$$(1 - L) F_t^0 = a + u_t$$

or

$$F_t^0 = at + u_t$$

Following Bai (2003), we assume that both dimensions of the panel increase to infinity $N, T \to \infty$. Throughout the paper the norm of a matrix is defined as $\|A\| = tr(A'A)^{1/2}$. We use I_r for a $r \times r$ identity matrix, $\lambda_i(A)$ for the ith largest eigenvalue of the square matrix A and $v_i(A)$ for the orthonormal eigenvector of the matrix A associated with the ith largest eigenvalue. Moreover, $[c]$ is a ceiling of the scalar c (it is the smallest integer number, such that $c \leq [c]$). We denote by $\to^p$ and $\to^d$ convergence in probability and distribution, respectively.

2.2 Assumptions

The following assumptions are used to derive the asymptotic properties of the estimators. Assumptions B-D are the same as in Bai (2003) and Bai (2004) and are discussed there in detail. We change Assumption A and Assumptions G-F in order to allow for factors with different dynamics.

Assumption A (Common factors)

1. $E\left\|u_t\right\|^{4+\delta} \leqslant M$ for some $\delta > 0$ and all $t \leqslant T$
2. $E\left\|F_1^0\right\|^4 \leqslant M$
3. The nonstationary $I(1)$ and $I(2)$ factors are not cointegrated.
4. There exists a diagonal scaling matrix D, which elements are functions of the time dimension T, such that for $T \to \infty$

$$D^{-1}F^{0\prime}F^0D^{-1} \to^d \Sigma$$

where Σ is a random matrix, which is positive definite with probability 1. Moreover, there exists $M \in \Re$ such that for all T

$$T\left\|D^{-2}\right\| \leq M$$

5. The maximum expected value of the normalized factors is bounded

$$\max_t E\left\|\sqrt{T}D^{-1}F_t^0\right\| \leq M$$

6. There exists a limit $\sqrt{T}D^{-1}F_t^0 \to^d F_\tau$ for $t/T = \tau$.

Assumption B (Heterogeneous factor loadings) The loading λ_i^0 is either deterministic, such that $\left\|\lambda_i^0\right\| \leqslant M$, or it is stochastic, such that $E\left\|\lambda_i^0\right\|^4 \leqslant M$. In both cases

$$\frac{1}{N}\sum_{i=1}^{N}\lambda_i^0\lambda_i^{0\prime} = \Lambda_0'\Lambda_0/N \to^p \Sigma_\Lambda$$

as $N \to \infty$ for some nonrandom, positive definite matrix Σ_Λ. Moreover, the matrix $\Sigma_\Lambda\Sigma$ has distinct eigenvalues with probability one.

Assumption C (Idiosyncratic component)

1. $E(e_{it}) = 0$ and $E\left|e_{it}\right|^8 \leqslant M$
2. $E(e_s'e_t/N) = \gamma_{NT}(s,t)$ with $\left|\gamma_{NT}(s,s)\right| \leqslant M$ for all s, and

$$\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\left|\gamma_{NT}(s,t)\right| \leqslant M$$

3. $E(e_{is}e_{jt}) = \pi_{ij,st}$ with $\left|\pi_{ij,tt}\right| \leqslant \left|\pi_{ij}\right|$ for some π_{ij} and for all t.

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\pi_{ij}\right| \leqslant M$$

4. $(NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}|\pi_{ij,st}|\leqslant M$

Assumption D $\{\lambda_i\}$, $\{e_t\}$, $\{u_t\}$ are mutually independent stochastic variables.

Assumptions A-D are necessary to prove the consistency of the estimators. Assumption A allows for factors with different dynamics. If all factors are stationary then the scaling matrix $D=\sqrt{T}I_r$, whereas if there are both stationary and nonstationary $I(1)$ and $I(2)$ factors without a time trend then D can be defined as follows

$$D=\left[\begin{array}{ccc} T^2 I_{r_2} & 0 & 0 \\ 0 & T I_{r_1} & 0 \\ 0 & 0 & \sqrt{T} I_{r_0} \end{array}\right] \tag{2}$$

where r_k denotes the number of $I(k)$ factors. In Bai (2004), there are only $I(0)$ and $I(1)$ factors and the scaling matrix takes the form

$$D=\left[\begin{array}{cc} T I_r & 0 \\ 0 & \sqrt{T} I_q \end{array}\right] \tag{3}$$

where r and q denotes the number of nonstationary and stationary common factors, respectively.

Remark 1 *If we allow for deterministic time trends then the scaling matrix D needs to be adjusted. Suppose the factors have a linear trend. Then, an element scaled by $T^{3/2}$ needs to be added to the diagonal of D. An exception is a model in which only the $I(2)$ factors have a linear (not quadratic) trend. In this case the scaling matrix remains unchanged as in (2). A model with $I(1)$ factors and the linear trend is discussed in detail in Section 4.*

In order to identify the number of nonstationary factors, we need to assume that they are not cointegrated. Otherwise, the space spanned by the factors could be described by the lower number of common trends G^0 and a stationary component. Hence, we would be able to reduce the number of nonstationary factors by substituting the corresponding vectors of F^0 by G^0 and the stationary term.

Assumption B is standard and is introduced to ensure that the factors load to infinitely many variables. It allows us to distinguish between a common component that is pervasive and an idiosyncratic component that has a limited effect. Hence, it ensures that the factor structure is identifiable. Assumption C describes a possible time and cross-sectional dependence of the idiosyncratic components. It is extensively discussed in Bai (2004). Assumption D excludes the correlation between the idiosyncratic and common shocks. It is not restrictive because in further analysis we allow for a dynamic structure of the factors.

In order to show a stronger result, we need to impose an additional Assumption E. It restricts the correlation of the idiosyncratic errors.

Assumption E

Let us denote $\bar{\gamma}_N(t,s)=E(|e_s'e_t/N|)$. Then there exists $M\leqslant\infty$ such that

1. For each t, $\sum_{s=1}^{T} |\bar{\gamma}_N (t,s)| \leqslant M$

2. For each i, $\sum_{j=1}^{N} |\pi_{ij}| \leqslant M$

Some moment conditions are introduced in Assumption F. The first two conditions F.1 and F.2 are needed to prove consistency and to compute the convergence rates. Finally, deriving the asymptotic distributions of estimators requires additional information about the limiting distribution of $N^{-1/2} \sum_{i=1}^{N} \lambda_i^0 e_{it}$ and $D^{-1} \sum F_t^0 e_{it}$. It is provided by Assumptions F.3 and F.4. If the loadings are deterministic then the Assumption F.3 follows from the Central Limit Theorem and the fact that the loadings are bounded. Otherwise, we assume, as in Bai (2004), that the limiting distribution of the first sum is normal.

Assumption F (Moments and Central Limit Theorem)

1. There exists $M < \infty$ such for every pair (s,t),

$$E \left| N^{-1/2} \sum_{i=1}^{N} [e_{it} e_{is} - E(e_{it} e_{is})] \right|^4 \leq M$$

2. There exists $M < \infty$ such that for any T

$$E \left| \frac{1}{T^{1/2}} \sum_{t=1}^{T} D^{-1} F_t^0 \Lambda_0' e_t \right|^2 \leq M$$

3. For each t as $N \to \infty$

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i^0 e_{it} \to^d N(0, \Gamma_t)$$

where $\Gamma_t = \lim_{N\to\infty} (1/N) \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i^0 \lambda_j^{0\prime} E(e_{it} e_{jt})$

4. For each i as $T \to \infty$ there exists a random variable W_i, such that

$$D^{-1} \sum_{t=1}^{T} F_t^0 e_{it} \to^d W_i$$

The distribution of the random variable W_i depends on the dynamics of the factors. If the kth factor is stationary or $I(1)$ with a time trend, then W_{ki} has a normal distribution, whereas if F_{kt} is $I(1)$ without deterministic trend then the distribution of W_{ki} is a functional of a Brownian motion, as in Bai and Ng (2004).

2.3 Estimation

Estimates of Λ and F are obtained by solving the optimization problem

$$\begin{aligned}\left(\tilde{\Lambda}, \tilde{F}\right) &= \arg\min_{\Lambda, F} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(X_{it} - \lambda_i F_t\right)^2 \\ &= \arg\min_{\Lambda, F} tr\left(\left(X - F\Lambda'\right)' \left(X - F\Lambda'\right)\right)\end{aligned}$$

where $X = \left(\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_N\right)$ and $F = \left(F_1, F_2, \ldots, F_T\right)'$. For any non-zero F the optimal loading matrix is

$$\tilde{\Lambda}' = \left(FF'\right)^{-1} F'X \tag{4}$$

and

$$X - F\tilde{\Lambda}' = \left(I_T - F\left(F'F\right)^{-1} F'\right) X$$

Define $P_F = F\left(F'F\right)^{-1} F'$. Then the optimal vector of factors F is

$$\begin{aligned}\tilde{F} &= \arg\min_{F} tr\left(\left(X - F\tilde{\Lambda}'\right)' \left(X - F\tilde{\Lambda}'\right)\right) \\ &= \arg\min_{F} tr\left(X'\left(I_T - P_F\right)'\left(I_T - P_F\right) X\right) \\ &= \arg\min_{F} tr\left(X'\left(I_T - P_F\right) X\right) \\ &= \arg\max_{F} tr\left(X' P_F X\right)\end{aligned}$$

In order to solve the above problem, we need to impose some normalization of the factors. It is standard to assume that the product of scaled factors gives the identity matrix,

$$D^{-1}F'FD^{-1} = I_r$$

Then

$$\begin{aligned}P_F &= F\left(F'F\right)^{-1\prime} F' \\ &= FD^{-2}F'\end{aligned}$$

and the problem is equivalent to maximizing

$$tr\left(X'FD^{-2}F'X\right) = tr\left(D^{-1}F'XX'FD^{-1}\right)$$

Thus, the estimated common factors $\tilde{F}$ are proportional to the eigenvectors v corresponding with the r largest eigenvalues of the $T \times T$ matrix XX'.

$$\tilde{F} = B \cdot v$$

The scaling matrix B is diagonal and is chosen to satisfy the normalization condition

$$I_r = D^{-1}F'FD^{-1} = D^{-1}Bv'vBD^{-1} = D^{-1}BBD^{-1}$$

Thus, $B = D$ and $\tilde{F}$ is D times the eigenvectors v

$$\tilde{F} = vD \tag{5}$$

The estimate of the loading matrix is obtained on the basis of (4) and is equal to

$$\tilde{\Lambda}' = D^{-2}\tilde{F}'X \tag{6}$$

The results correspond with the outcomes of Bai and Ng (2002) and Bai (2004) with $D = \sqrt{T}I_r$ or $D = TI_r$, respectively. In the first case, the estimated factors are the eigenvectors v multiplied by $\sqrt{T}$. In a model with $I(1)$ factors without drift, the estimators are $\tilde{F} = vT$. In the Generalized Factor Model (GFM) presented by Bai (2004), the scaling matrix is (3). Thus, the estimates of the nonstationary factors are the eigenvectors corresponding with the r largest eigenvalues multiplied by T, whereas the estimates of the stationary factors are the eigenvectors corresponding with the $r+1 : r+q$ largest eigenvalues multiplied by $\sqrt{T}$.

In further analysis, we consider also another normalization of factors and factor loadings. The following lemma defines so called normalized factors, $\hat{F}$, and normalized loadings, $\hat{\Lambda}$.

Lemma 2 *Define normalized factors $\hat{F} = N^{-1}X\tilde{\Lambda}$ and a normalized loading matrix $\hat{\Lambda}$ such that $\hat{F}\hat{\Lambda}' = \tilde{F}\tilde{\Lambda}$. Then*

$$\begin{aligned} \hat{\Lambda} &= \tilde{\Lambda}V_{NT}^{-1} \\ \hat{F} &= \tilde{F}V_{NT} \end{aligned}$$

where $V_{NT} = \tilde{V}_{NT}D^{-2}/N$ and $\tilde{V}_{NT}$ is the diagonal matrix consisting of the r largest eigenvalues of the matrix XX'.

This lemma shows how the two different estimators $\hat{F}$ and $\tilde{F}$ are related to each other. It is used to derive the asymptotic distribution of $\tilde{F}$ and to construct the confidence intervals around a rotation of the true factors.

3 Distribution theory

In this section, we present an asymptotic theory of estimated factors, factor loadings and a common component. Firstly, we discuss the consistency issue and derive the asymptotic distribution. Finally, we show how the confidence intervals of a rotation of the true factors can be constructed.

3.1 Consistency

Bai (2003) and Bai (2004) prove consistency of the estimators of stationary and random walk factors. They show that the mean squared errors of the estimated factors are $O_p\left(\max\left\{N^{-1}, T^{-1}\right\}\right)$ and $O_p\left(\max\left\{N^{-1}, T^{-2}\right\}\right)$, respectively. Using similar arguments, we show that the MSE of an estimated factors with a scaling matrix D is $O_p\left(\max\left\{N^{-1}, \left\|D^{-2}\right\|\right\}\right)$. Moreover, for a given time period t the error $\hat{F}_t - H'F_t^0$ is $O_p\left(N^{-1/2}\right) + O_p\left(\left\|D^{-1}\right\|\right)$.

Consider firstly the MSE of estimated factor.

Proposition 3 *Assume Assumptions A-D hold. There exists a nonsingular matrix $\tilde{H}$ and $\delta_{NT}^{-1} = \max\left\{N^{-1/2}, \left\|D^{-1}\right\|\right\}$ such that*

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t - \tilde{H}'F_t^0\right\|^2 = O_p\left(\delta_{NT}^{-2}\right)$$

The proposition states that the time average of a squared deviation between the estimated factors and the rotation of the true factors converges to zero with a growing sample size $N, T \to \infty$. The proposition is very important because it shows that the factors can be consistently estimated with a principle component method. The convergence rates are used to derive the asymptotic distribution of the estimators.

The result is in line with the existing literature. In a case of a model with stationary factors, the norm of the scaling matrix $\left\|D^{-1}\right\| = O_p\left(T^{-1/2}\right)$. Therefore, the convergence rate is $\delta_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$, as in Bai (2003). If we assume that all the factors are random walks without a drift then $\left\|D^{-1}\right\| = O_p\left(T^{-1}\right)$ and $\delta_{NT} = \min\left\{\sqrt{N}, T\right\}$. The convergence rate corresponds with the outcome presented in Bai (2004).

Finally, it is shown that for a given time period t the error converges to zero with a growing cross-sectional and time dimension. To prove the convergence rates we need to impose the more restrictive Assumption E.

Proposition 4 *Under Assumptions A-E the following holds for each t,*

$$\tilde{F}_t - \tilde{H}'F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(\left\|D^{-1}\right\|\right)$$

The convergence rate is the same as in Bai and Ng (2002) for stationary factors. Since we allow for different types of factors then the rate is lower then in Bai (2004), where only $I(1)$ factors without a trend are considered. It is, however, sufficient to derive the limiting distribution of factors.

Remark 5 *If we allow for only one type of nonstationary factors, for example $I(1)$ or $I(2)$ factors, then it is shown by Lemma 23 that*

$$\tilde{F}_t - \tilde{H}'F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right)$$

This is in line with the results of Bai (2004) for the $I(1)$ factors without a time trend, where

$$\tilde{F}_t - \tilde{H}' F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(T^{-3/2}\right)$$

3.2 Asymptotic distributions

We investigate the asymptotic distribution of the estimated factors, the factor loadings and the common component. Firstly, we describe a limiting behavior of V_{NT} and $D^{-2}\tilde{F}'F^0$.

Lemma 6 *Under assumptions A-E, as N, $T \to \infty$*

1. *There exists a random, diagonal, full rank with probability 1 matrix V such that $V_{NT} \to^d V$*
2. *There exists a random, positive definite, with probability 1 matrix Q such that*

$$D^{-2}\tilde{F}'F^0 = Q_{NT} \to^d Q$$

The lemma defines two matrices, V and Q, used to describe the asymptotic distribution of factors and factors loadings.

3.2.1 Limiting distribution of estimated common factors

The following proposition shows that the factor estimates are asymptotically normal. This property is used to construct the confidence intervals around the rotation of the true factors.

Proposition 7 *Under Assumptions A-F, as $N, T \to \infty$ and $N^{1/2}\left\|D^{-1}\right\| \to 0$ we have for each t*

$$\sqrt{N}\left(\tilde{F}_t - \tilde{H}'F_t^0\right) \to^d \Sigma_\Lambda^{-1} N\left(0, \Gamma_t\right)$$

where Σ_Λ and Γ_t are defined as in the Assumptions B and F.

The proposition requires restrictions on the relation between the cross-sectional and the time dimensions. If there are stationary factors the conditions say that $N/T \to 0$. In a case of a model with only nonstationary factors without the deterministic trend, the condition is $N/T^2 \to 0$. If there is only one type of factor, it can be shown that the condition is $N^{1/2}T^{-1/2}\left\|D^{-1}\right\|$ as in Bai (2004)[3].

The results will be used to construct the confidence intervals around a rotation of true factors.

[3] The condition follows directly from Lemma 23 and the proof of Proposition 7.

3.2.2 Limiting distribution of estimated factors loadings

In this section, we show that the estimated factor loadings converges to some random variable.

Proposition 8 *Under the Assumptions A-F, for each i, as $N, T \to \infty$ we have*

$$D\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) \to^d \left(\bar{H}\right)^{-1}\Sigma^{-1}W_i$$

with $\bar{H}$ is defined by Lemma 26. Σ and W_i are defined by Assumption A and F, respectively.

The actual limiting distribution of factor loadings depends on the dynamics of the factors. As shown in Bai (2003), if the factors are stationary then the matrix Σ converges to the factors variance-covariance matrix. On the other hand, if all factors are random walks without a drift then Σ is defined by a Brownian motion. Moreover, if we allow for other types of factors then the elements of the random matrix Σ may take different forms.

3.2.3 Limiting distribution of estimated common components

Let us denote the true and estimated common components[4] by $C_{it}^0 = F_t^0\lambda_i^0$ and $\hat{C}_{it} = \hat{F}_t\hat{\lambda}_i$, respectively. The limiting distribution of the estimates of the common component depends on the relation between the cross-sectional and time dimensions T/N.

Proposition 9 *Under Assumptions A-G as $N, T \to \infty$ it holds that*

1. *If $N/T \to 0$ then for each pair (i, t)*

$$\sqrt{N}\left(\hat{C}_{it} - C_{it}^0\right) \to^d \lambda_i^{0\prime}H^{-1\prime}QN\left(0, \Gamma_t\right)$$

where Γ_t is defined in Assumption F and Q is introduced in Lemma 6.

2. *If $T/N \to 0$ then for each pair (i, t) and $t = [\tau T]$*

$$\sqrt{T}\left(\hat{C}_{it} - C_{it}^0\right) \to^d F_\tau'\Sigma^{-1}W_i$$

where Σ and F_τ are defined in Assumption A and W_i is defined in Assumption F.

3. *If $N/T \to \pi$ then for each pair (i, t) and $t = [\tau T]$*

$$\sqrt{N}\left(\hat{C}_{it} - C_{it}^0\right) \to^d \lambda_i^{0\prime}H^{-1\prime}QN\left(0, \Gamma_t\right) + \sqrt{\pi}F_\tau'\Sigma^{-1}W_i$$

where Q, Γ_t, F_τ,Σ and W_i are defined as above.

[4] The estimated common component $\hat{C}_{it}$ does not depend on the normalization of common factors.

As noted by Bai (2004), the third case is the most useful in practice, because π can be estimated by the sample ratio N/T. Moreover, the distribution of the common components in cases (2) and (3) depends on the limiting distribution of F_τ. When the factor $\bar{F}_i^0$ is stationary then $F_{\tau i}$ is normally distributed. However, if the factor $\bar{F}_i^0$ is a $I(1)$ process without a deterministic drift then $F_{\tau i}$ is a Brownian motion process with a variance described by Bai and Ng (2004).

3.2.4 Confidence intervals

In the article, we interpret a scalar, observable variable R_t as a common factor if it is a linear combination of the true factors plus a constant.

$$R_t = \alpha + \beta' F_t^0$$

where α is a shift parameter and β is a $r \times 1$ vector that summarize the relation between R_t and F_t^0. We allow for both a rotation and a shift of the factors because neither R_t nor F_t^0 have to be zero mean processes and they may have different levels and scalings.

Consider the rotation of $\tilde{F}$ toward R_t described by the regression

$$\begin{aligned} R_t &= \alpha + \beta' \left(\tilde{H}^{-1\prime} \tilde{F}_t \right) + u_t \\ &= \alpha + \delta' \tilde{F}_t + u_t \end{aligned}$$

Let $\left(\hat{\alpha}, \hat{\beta}\right)$ be the least-square estimator of (α, β) and $\hat{R}_t = \hat{\alpha} + \hat{\beta}' \left(\tilde{H}^{-1\prime} \tilde{F}_t \right)$. From the identity $\delta' = \beta' \tilde{H}^{-1\prime}$ it follows that $\hat{\delta}' = \hat{\beta}' \tilde{H}^{-1\prime}$. If R_t is a common factor then the following proposition holds.

Proposition 10 *Under the Assumptions A-F and no cross-section correlation for the idiosyncratic errors, as $N, T \to \infty$ and $N^{1/2} \left\| D^{-1} \right\| \to 0$*

$$\sqrt{N} \left(\hat{R}_t - \alpha - \beta' F_t^0 \right) \to^d \hat{\delta} V^{-1} Q N \left(0, \Gamma_t\right)$$

where V, Q are defined in Lemma 6 and Γ_t is introduced in Assumption F.

Following Bai (2004), we will approximate the 95% confidence intervals as follows

$$\left(\hat{R}_t - 1.96 \sqrt{\tilde{S}_t^2 / N}, \hat{R}_t + 1.96 \sqrt{\tilde{S}_t^2 / N} \right) \tag{7}$$

where $\tilde{S}_t^2 = \left(\hat{\delta} V^{-1} Q \right) \Gamma_t \left(\hat{\delta} V^{-1} Q \right)'$.

Remark 11 *As stated in Bai (2003), the matrix*

$$\hat{\delta} V^{-1} D^{-2} \tilde{F}' F^0 \Gamma_t F^{0\prime} \tilde{F} D^{-2} V^{-1} \hat{\delta}'$$

involves the product of $F^0\Lambda_0$*, which can be consistently estimated with* $\tilde{F}\tilde{\Lambda}$*. Hence, it can be substituted by*

$$\hat{\delta}V^{-1}D^{-2}\tilde{F}'\tilde{F}\tilde{\Gamma}_t\tilde{F}'\tilde{F}D^{-2}V^{-1}\hat{\delta}' = \hat{\delta}V^{-1}\tilde{\Gamma}_tV^{-1}\hat{\delta}'$$

where

$$\tilde{\Gamma}_t = \lim_{N\to\infty}(1/N)\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{\lambda}_i\tilde{\lambda}_j'\hat{E}\left(e_{it}e_{jt}\right)$$

Remark 12 *Bai and Ng (2006) propose two types of estimators of the matrix* $\tilde{\Gamma}_t$ *that can be used for cross sectionally uncorrelated idiosyncratic errors* e_{it}

1. $\tilde{\Gamma}_t = \frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{it}^2\tilde{\lambda}_i\tilde{\lambda}_i'$
2. $\tilde{\Gamma} = \tilde{\sigma}_\varepsilon^2\frac{1}{N}\sum_{i=1}^{N}\tilde{\lambda}_i\tilde{\lambda}_i'$*, where* $\tilde{\sigma}_\varepsilon^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{e}_{it}^2$ *for errors with equal variances* $\sigma_{\varepsilon_i}^2 = \sigma_\varepsilon^2$*.*

$\tilde{\lambda}_i$ *and* $\tilde{e}_{it}$ *correspond to the estimates of* λ_i *and* e_{it}*.*

Remark 13 *If the observable variable* R_t *belongs to the panel (*$R_t = X_{it}$*) then the parameters* $\left(\hat{\alpha},\hat{\delta}\right)$ *can be replaced with* $\left(0,\hat{\lambda}_i\right)$*, where* $\hat{\lambda}_i$ *are estimated factor loadings.*

Remark 14 *In order to compute the confidence intervals, we need to ensure that the idiosyncratic errors have zero mean*[5]*. Otherwise* $E\left(\hat{R}_t - \alpha - \beta F_t^0\right) \neq 0$ *and* $\tilde{\Gamma}_t$ *will not be a consistent estimator of the variance-covariance matrix* Γ_t*.*

4 Model with $I(1)$ factors with a deterministic trend

So far, the literature considers only models with either stationary factors or common trends without deterministic drift. Since most of time series have both stochastic and deterministic trends, the theory does not match the needs of macroeconomic modeling. Thus, we believe that the model that allows for a deterministic trend is interesting, especially from an empirical point of view.

In this section, we discuss in more detail issues associated with an estimation of a factor model with a linear time trend. We address the problem of determining a number of common trends with a drift. We show the convergence rates, limiting distributions of factors, factor loadings and common components. Finally, we present the results in the context of a generalized factor model, as in Bai (2004).

[5] One possible way to construct idiosyncratic errors with zero mean is to remove the mean form the orginal data set.

4.1 Modeling the time trend vs. detrending the data

Once we decide, on the basis of analysis of the variables in panel, that the deterministic trend plays an important role in the model, we may consider two strategies. The first approach leaves the data unchanged and models the trend together with other factors. It is discussed in detail in the following sections. The second approach consists of two steps. Firstly, the data is detrended and secondly, the factors without trend are estimated as in Bai (2004). Its main disadvantage is that it requires either a precise parametrization of the time trend or a usage of some nonlinear filtering procedures. There is no agreement on which of the detrending methods should be used in the context. Therefore, we believe that our approach is a competitive alternative.

4.2 Number of common factors with a drift

The first issue is the number of identifiable common trends with a deterministic drift. We show that a model with $n > 1$ common factors with time trends can be represented as a model with only one factor with time trend and $n-1$ factors without a deterministic drift. Consider a system with n factors, $F_t = (F_{1t}, ..., F_{nt})'$, that depends both on the time trend and a stochastic, zero mean variable $\omega_t = (\omega_{1t}, ..., \omega_{nt})'$

$$\begin{aligned} F_t &= At + B\omega_t \\ &= [A_{n\times 1} : B_{n\times n}] \begin{bmatrix} t \\ \omega_t \end{bmatrix} \end{aligned}$$

The matrix $[A_{n\times 1} : B_{n\times n}]$ needs to have a rank n in order to ensure that all the factors are identifiable. Since the factors are assumed to follow a deterministic time trend, the vector A has to be non-zero. Then the system can be rewritten as follow

$$F_t = C\,[I_{n\times n} : D_{n\times 1}] \begin{bmatrix} t \\ \omega_t \end{bmatrix}$$

where C is a $n \times n$ full rank matrix and $I_{n\times n}$ is an identity matrix. Let us construct a new set of factors $\tilde{F}_t = C^{-1}F_t$. Then

$$\tilde{F}_t = [I_{n\times n} : D_{n\times 1}] \begin{bmatrix} t \\ \omega_t \end{bmatrix}$$

and

$$\begin{aligned} \tilde{F}_{1t} &= t + D_{11}\omega_{nt} \\ \tilde{F}_{2t} &= \omega_{1t} + D_{21}\omega_{nt} \\ &\vdots \\ \tilde{F}_{nt} &= \omega_{n-1t} + D_{n1}\omega_{nt} \end{aligned}$$

Thus, among the factors $\tilde{F}_t$ only the first one has a time trend. Moreover, if all the factors are nonstationary and noncointegrated then at least $n-1$ of the ω_t elements have to be $I(1)$ processes. We can order the elements of ω_t in such a way that only the last component ω_{nt} is allowed to be stationary. Depending on integration order of ω_{nt} the first factor $\tilde{F}_{1t}$ will be trend stationary (when ω_{nt} is $I(0)$) or a random walk with a drift (when ω_{nt} is $I(1)$).

We have shown that the factors F_t are a linear combination of $\tilde{F}_t$, where only one factor $\tilde{F}_{1t}$ has a time trend. Without loss of generality we can replace F_t with $\tilde{F}_t$. Therefore, in further analysis, we assume that there is only one common factor with a deterministic linear trend.

Remark 15 *The arguments are valid if the trend is not linear but is a function of time $f(t)$ and loads with weights A to the factors. Then, the factors F_t can be replaced with $\tilde{F}_t$, where only one of the elements of $\tilde{F}_t$ has a deterministic component and other elements have a zero mean.*

4.3 Static factor model

Let us first consider a static factor model with a single nonstationary factor with a deterministic time trend. Some of the restrictive assumptions on the total number of factors and the relation between factors and observable variables will be relaxed in the Section 4.4, where a generalized dynamic factor model will be discussed.

Define by F_t a common nonstationary factor with a deterministic trend such that it is either $I(1)$ with a drift

$$F_t = a + F_{t-1} + u_t \tag{8}$$

or trend stationary.

$$F_t = at + u_t$$

with $a \neq 0$.

Since the factor has a time trend then it needs to be scaled by $T^{3/2}$. Hence, the scaling matrix $D = T^{3/2}$ and the limit of $D^{-1}F^{0\prime}F^0D^{-1} = T^{-3}\sum_{t=1}^{T}\left(F_t^0\right)^2$ equals a scalar $\Sigma = a^2/3$

$$\begin{aligned} T^{-3}\sum_{t=1}^{T}\left(F_t^0\right)^2 &= T^{-3}\sum_{t=1}^{T}a^2t^2 + o_p(1) \\ &= \sum_{t=1}^{T}a^2\left(\frac{t}{T}\right)^2\frac{1}{T} + o_p(1) \\ &\to \int_0^1 a^2x^2dx = a^2/3 \end{aligned}$$

For $t = [\tau T]$ the limit of F_t^0/T is $F_\tau = a\tau$. For a $I(1)$ factor

$$\begin{aligned} \frac{1}{T}F_t^0 &= a\frac{t}{T} + \frac{t}{T}\frac{1}{t}\sum_{s=1}^{t} u_s \\ &\to {}^p a\tau + \tau E u_t = a\tau \end{aligned}$$

and for a trend stationary factor

$$\begin{aligned} \frac{1}{T}F_t^0 &= a\frac{t}{T} + \frac{1}{T}u_t \\ &\to {}^p a\tau \end{aligned}$$

Moreover, it can be assumed that for each i, as $T \to \infty$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{T}F_t^0 e_{it} \to^d N(0, \Omega_i)$$

where $\Omega_i = \lim_{N\to\infty}(1/T)\sum_{t=1}^{T}\sum_{s=1}^{T} a^2 \frac{ts}{T^2} E(e_{it}e_{is})$. Thus, the variable W_i has a normal distribution.

Remark 16 *Suppose the deterministic trend is not linear and is described by a function $f(t)$. Then as long as*

$$0 < \lim_{T\to\infty} T^{-3}\sum_{t=1}^{T}(f(t))^2 < M$$

and

$$0 < \lim_{T\to\infty}\frac{1}{T}f(\tau T) < M$$

then the results hold and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{T}F_t^0 e_{it} \to^d N(0, \Omega_i)$$

The matrix Ω_i takes the following form form

$$\Omega_i = \lim_{N\to\infty}(1/T)\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{f(t)f(s)}{T^2}E(e_{it}e_{is})$$

4.3.1 Estimation

In Section 2, we derived estimators of the factor and factor loadings. Since $D = T^{3/2}$ then

$$\begin{aligned} \tilde{F} &= T^{3/2}v \\ \tilde{\Lambda}' &= T^{-3}\tilde{F}'X \end{aligned}$$

where $v = v_1(XX')$ is the eigenvector corresponding with the largest eigenvalue of the matrix XX'. Hence, the normalized factor and loadings can be computed as in Lemma 2, with V_{NT} being the largest eigenvalue of the matrix $XX'/(NT^3)$.

4.3.2 Convergence rates

The convergence rates can be computed on the basis of Proposition 3 and Lemma 23. Since $\left\|D^{-1}\right\| = T^{-3/2}$ then

$$\delta_{NT}^{-1} = \max\left\{N^{-1/2}, T^{-3/2}\right\}$$

and

$$\hat{F}_t - H'F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(T^{-2}\right)$$

The convergence rates are higher than in the model with only stationary factors or common trends without a drift.

4.4 Generalized dynamic factor model

Finally, consider the generalized dynamic factor model with both stationary and nonstationary factors

$$X_{it} = \lambda_i^r(L)\, F_t^r + \lambda_i^q(L)\, F_t^q + e_{it} \tag{9}$$

where $\lambda_i^r(L)$ and $\lambda_i^q(L)$ are lag polynomials corresponding to different types of factors: F_t^r is a $r \times 1$ vector of common nonstationary factors with the first factor having a time trend and F_t^q is a $q \times 1$ vector of stationary factors. Hence, in the generalized dynamic factor model we allow for more then one factor: there are r nonstationary and q stationary dynamic factors.

$$\begin{aligned} F_t^r &= A + F_{t-1}^r + u_t^r \\ F_t^q &= u_t^q \end{aligned}$$

with $A = (a, 0, \ldots, 0)'$. Following Bai (2004) and Forni, Hallin, Lippi, and Reichlin (2003), we assume

$$\begin{aligned} \lambda_i^r(L) &= \sum_{j=0}^{\infty} \lambda_{ij}^r L^j \\ \lambda_i^q(L) &= \sum_{j=0}^{\infty} \lambda_{ij}^q L^j \end{aligned}$$

with $\sum_{j=0}^{\infty} j\left|\lambda_{ij}^r\right| < \infty$ and $\sum_{j=0}^{\infty} j\left|\lambda_{ij}^q\right| < \infty$.

Since there are three types of factors the scaling matrix takes the form

$$D = \begin{bmatrix} T^{3/2} & 0 & 0 \\ 0 & T \cdot I_{r-1} & 0 \\ 0 & 0 & T^{1/2} \cdot I_q \end{bmatrix} \tag{10}$$

4.4.1 Static representation

The dynamic representation of the model (9) can not be directly estimated. In order to construct the estimators, we need to rewrite the model in the static form. Let us notice that (9) can be expressed as follows

$$\begin{aligned} X_{it} &= \lambda_i^r(L) F_t^r + \lambda_i^q(L) F_t^q + e_{it} \\ &= \varphi F_t^r + \phi^r(L) \Delta F_t^r + \lambda_i^q(L) F_t^q + e_{it} \end{aligned}$$

where the factors ΔF_t^r and F_t^q are stationary. In order to derive the asymptotic distributions, we need to approximate the model with finite order lag polynomials. Let us assume that $\phi^r(L)$, $\lambda_i^q(L)$ have an order p. Then the model can be written as

$$X_{it} = \varphi F_t^r + \Phi G_t \tag{11}$$

where $G_t = \left(\Delta F_t^r, \ldots, \Delta F_{t-p}^r, F_t^q, \ldots, F_{t-p}^q\right)'$ summarizes the stationary factors. Thus, the model has the static form that uniquely identifies the dynamic non-stationary factors[6] F_t^r. The representation (11) will be used in further analysis.

4.4.2 Estimation of the number of factors

In order to estimate the total number of factors, Bai (2004) proposes to use the data in first differences[7]. If the data are $I(1)$ then

$$\Delta X_{it} = \lambda_i^r(L) \Delta F_t^r + \lambda_i^q(L) \Delta F_t^q + \Delta e_{it} \tag{12}$$

and both ΔX_{it} and factors ΔF_t^r, ΔF_t^q are stationary. Therefore, the information criteria PC introduced by Bai and Ng (2002) can be applied. As stated in Bai (2004) the procedure allows for consistent estimation of the total number of factors (both stationary and nonstationary).

The second issue is determining the number of stationary and nonstationary factors separately. Bai (2004) shows that the number of nonstationary, dynamic factors can be estimated directly form the data in levels on the basis of representation (11). Bai and Ng (2004) constructs the information criteria IPC and proves their consistency for panels without a deterministic trend. In the paper, it is stated that the same information criteria can be used to estimate the total number of nonstationary factors regardless of the existence of the deterministic components and the order of integration. The number of stationary static factors, G_t, can be computed as the difference between the total number of factors and the number of nonstationary dynamic factors as in Bai (2004).

[6] The identification is achieved under the assumption of no cointegration between the nonstationary factors. See Bai (2004) for a discussion.

[7] The aim of the differencing is to ensure that the common factors are stationary. Therefore, the order of differencing should equal to the integration order of the data.

4.4.3 Estimation and convergence rates

Since the number of factors can be consistently estimated with the information criteria as in Bai and Ng (2002) and Bai (2004), then we assume that the true number of both stationary and nonstationary factors is known. The common factors can be estimated as follow

$$\tilde{F} = vD$$

where v are the eigenvectors corresponding with the $(r+q)$ largest eigenvalues of a matrix XX' and D is given by (10). Thus,

1. A nonstationary common trend with a drift is estimated as the eigenvector corresponding to the largest eigenvalue of the matrix XX' multiplied by $T^{3/2}$.
2. Nonstationary common trends without a drift are estimated as the eigenvectors corresponding to $2:r$ largest eigenvalues of the matrix XX' multiplied by T.
3. Stationary common trends are estimated as the eigenvectors corresponding to $(r+1):(r+q)$ largest eigenvalues of the matrix XX' multiplied by $T^{1/2}$.

Let V_{NT} be a diagonal matrix defined in Lemma 2. It has diagonal elements V_i such that

1. V_1 is the largest eigenvalue of the matrix XX'/NT^3.
2. $V_2, \ldots, V_r$ are the $2:r$ largest eigenvalues of the matrix XX'/NT^2.
3. $V_{(1+r)}, \ldots, V_{(r+q)}$ are the $(r+1):(r+q)$ largest eigenvalues of the matrix XX'/NT.

Finally, we present the convergence rates. Since $\left\|D^{-1}\right\| = O_p\left(T^{-1/2}\right)$ then the convergence rates are $\delta_{NT}^{-1} = \min\left\{N^{-1/2}, T^{-1/2}\right\}$ and

$$\hat{F}_t - H'F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\right)$$

5 Empirical example

In the paper, we study the behavior of 69 variables describing the real activity of US economy (an industrial production, components of the real GDP, two measures of the labor productivity and interest rates). The data are quarterly, spanning the period from January 1961 to September 2008. The description of the data is provided in the Appendix. Most of the variables in the panel are nonstationary and have both deterministic and stochastic trends.

5.1 Normalization

The literature on stationary panels underlines the need for data normalization. Usually, variables in panels are divided by their standard deviations. This approach ensures that all variables have equal input to the total variability of the panel. Therefore, the estimation method does not favour any of them. Moreover, the normalization does not change the theoretical results because it is associated with multiplying the data by a diagonal matrix that converges to an invertible matrix of asymptotic standard deviations.

This method cannot be directly applied for nonstationary panels because the standard deviations diverge to infinity. Thus, it will affect the asymptotic results of the estimation method. In order to normalize the data, we propose dividing them by

$$\sigma_i = \left(\sum_{t=1}^{T} \left(X_{it} - \mu_i \right)^2 / T^{n_i} \right)^{1/2}$$

where μ_i denotes the mean of the variable X_i and n_i is chosen to ensure that $\sigma_i = O_p(1)$ and that σ_i has a limit. For example, if a variable X_i is stationary then $n_i = 1$ and if X_i is an $I(1)$ process without a deterministic drift then $n_i = 1.5$. Finally, for a $I(1)$ variable X_i with a time trend there is $n_i = 2$.

The normalization ensures that the variables with the same type of dynamics have the same volatility. It has an intuitive interpretation for processes without time trends because it corresponds to a standard deviation. For data with a deterministic trend, the normalization guarantees that in the limit the slope of the trend equalize across the panel variables. Thus, it standardizes the main source of the volatility.

5.2 The number of factors

Firstly, we estimate the number of nonstationary factors using the IPC information criteria described by Bai (2004) and applied for data in levels. We assume that there are not more then ten common trends. Thus, we consider cases, in which $k_{\max} \leq 10$. The results are presented in Table 2 and indicate that there are either two or three nonstationary factors.

Finally, we estimate the number of factors from differenced data with the PC criteria described in Bai (2003). The criteria do not give conclusive results because they always choose the maximum permitted number of factors. It may indicate that either the model has a long lag structure or the cross sectional sample size is too small to provide correct estimates.

The literature discusses some alternative approaches that can be used to select the number of factors. Child (2006) provides a review of less formal, graphical methods that can be applied in this context. They are based on the eigenvalues of the panel correlation matrix. It can be seen that the sum of these eigenvalues equals the cross sectional dimension N. Therefore, the first approach is to look at the number of the eigenvalues larger then one and hence, above the average. This criterion indicates 18 common factors, which explains 83.25% of the total variability. As stated by Child (2006), the large cross sectional dimension leads to overestimation of the number of factors. Hence, we analyze the plot of the correlation matrix eigenvalues and use a Scree test[8]. The eigenvalues are presented in Figure 1 and indicate that there are around ten common factors. The plot starting from the eleventh eigenvalue is almost linear and decreases steadily to zero. The first ten common factors explain 67.85% of the total variability of the panel. The result is in line with the outcome of Stock and Watson (2005), which indicates the existence of nine static factors in the stationary panel describing US economy.

Since we cannot choose the total number of factors consistently, we check the robustness of the results with respect to the number of stationary factors. We will use, as a benchmark, a model with ten factors (three common trends and seven stationary factors).

5.3 Macroeconomic factors

Finally, we check whether some observable variables can be interpreted as common factors. Since the unobserved factors are consistently estimated then we can use a formal test described in Section 3. In order to construct the confidence intervals, we need to estimate the variance-covariance matrix Γ_t. We use the estimator applied in Bai (2004). It is constructed as follow

$$\Gamma_t = \frac{1}{N} \sum_{i=1}^{N} \tilde{e}_{it}^2 \tilde{\lambda}_i \tilde{\lambda}_i'$$

where $\tilde{\lambda}_i$ are the principle components estimates of the loadings matrices and $\tilde{e}_{it} = X_{it} - \tilde{\lambda}_i \tilde{F}_t$ are the idiosyncratic residuals.

5.3.1 Interest rate

In most of the macroeconomics literature, interest rates are one of the driving forces of the economy. In the analysis, we focus on the interest rate measured by Federal Funds rate (FF). We rotate the estimated factors toward FF by running the regression $FF_t = \alpha + \delta \tilde{F}_t + \varepsilon_t$. Next, we compute confidence intervals around fitted values (7) and the percentage of FF observations that remain outside the intervals. The results for different number factors are presented in Table 4. The outcomes indicate that for models with at least ten factors, all observations of FF remain inside the confidence intervals. Therefore, we cannot reject the hypothesis that the FF is one of the factors driving the economy. Figure 2 presents the observations of FF and the estimated confidence intervals for the benchmark model.

[8] The Scree test was introduced by Cattell (1966) and is based on the observation that the plot of correlation matrix eigenvalues for uncorrelated variables is almost flat and linearly converges to zero.

5.3.2 Private fixed investments vs. personal consumption expenditures

Next, we consider the hypothesis that investments play an important role in the economic development. Therefore, we examine if two measures of investments; real private fixed investments in nonresidential structures and residential permanent site structures, can be considered as common factors. We proceed as before and regress the variables on the estimated common factors. Next, we construct the confidence intervals as in (7) and compute the percentage of observations that remain outside the confidence intervals. The results are presented in the Table 4. They indicate that for sufficient number of factors both variables can be interpreted as common trends.

Unfortunately, for a benchmark model with ten common factors, around 22% of observations of the investments in nonresidential structures lay outside the confidence intervals. The variable and the confidence intervals are presented in Figure 3. Therefore, we consider another measure of nonresidential investments: the real private fixed investments in nonresidential commercial structures. For models with at least eight factors we can not reject the null that the variable is a common factor. Moreover, for models with at least eleven factors, we could not reject the hypothesis that both measures of investments in nonresidential structures are common trends. Thus, we conclude that they are the driving forces of the economy.

The outcomes for the investments in residential permanent site structures are more clear. For all considered models, at least 90% of observations stay inside the confidence intervals. Moreover, for a benchmark model only 6.28% of observations fall outside the intervals (Figure 4). Hence, we interpret the investments in residential site structure as a common factor.

Finally, we analyze whether different measures of real personal consumption expenditures can be interpreted as common trends. The outcomes indicate that the null hypothesis can be reject for all model setups. Thus, we do not find any results supporting the view that the personal consumption is a main driving force of the whole economy.

5.3.3 Government spendings

Since we do not find any arguments in favor of a hypothesis that the private real consumption expenditure can be interpreted as common factors, we test whether government spendings have an important effect on the economy. We consider two measurements of government spendings: real federal consumption expenditures and gross investments in national defence and nondefense sectors. We proceed as before and construct the confidence intervals. The percentage of variable observations that lay outside of the intervals are presented in Table 4. The results indicate that for a model with at least nine factors both variables can be interpreted as common factors. Figure 5 shows federal expenditures in national defence and the confidence intervals for the benchmark model. It can be noticed that almost all observations stay inside the intervals (only less then 2% are outside). Similar results are obtained for federal expenditures in nondefense sectors (Figure 6). The outcomes support the hypothesis that government spending have an impact on the whole economy.

6 Conclusions

This paper discusses the estimation methods of common factors with different types of dynamics. We generalize the existing methodology by allowing for other types of factors apart from stationary factors and common trends without a deterministic drift. In particular, we focus on nonstationary factors with a time trend. We believe that it is an important issue because most of the macroeconomic variables are subjected to a time trend. Thus, the data should be either detrended or the existence of a drift needs to be explicitly modeled. The model setup is similar to the generalized factor model presented in Bai (2004). Under some standard assumptions, we show that the common factors can be consistently estimated with a principal component method (under the assumption that both time and cross-sectional dimensions increase to infinity). Additionally, we derive convergence rates and asymptotic distributions of factors, factors loadings and common components. It allows us to construct the confidence intervals of a rotation of true factors and hence, to construct a formal test to verify if an observable variable can be interpreted as a common factor. We link the theory to the existing literature and present it as an extension to the work of Bai (2003) and Bai (2004).

The theory is illustrated with an empirical example. We analyze 69 macroeconomic variables describing the real part of the U.S. economy. We show that an interest rate, investments and government spendings can be interpreted as common factors, thus they are the driving forces of the economy. The results are in line with a macroeconomic literature. We do not find any arguments in favor of a hypothesis that personal consumption is also one of the common trends.

References

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

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

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

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

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

Banerjee, A. and M. Marcellino (2008). Factor-augmented error correction models. EUI Working Pepers ECO 2008/15.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, 387–422.

Breitung, J. and S. Eickmeier (2005). Dynamic factor models. Deutsche Bundesbank Discussion Paper, Series 1: Economic Studies, No 38/2005.

Cattell, R. B. (1966). The Scree test for the number of factors. *Multivariate Behavioral Research 1*, 245–276.

Child, D. (2006). *The essentials of factor analysis*. Continuum, New York.

Eickmeier, S. (2009). Comovements and heterogeneity in the euro area analyzed in a non-stationary dynamic factor model. *Journal of Applied Economietrics*.

Forni, M. and L. Gambetti (2008). The dynamic efects of monetary policy: A structural factor model approach. RECent Working Paper No. 26.

Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2007). Opening the black box - structural factor models woth large cross-sections. Working Paper Series No. 712, European Central Bank.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2003). The generalized dynamic factor model: Identification and estimation. LEM Papers Series 2003/13.

Gonzalo, J. and C. Granger (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics 13*(1), 27–35.

Kapetanios, G. and M. Marcellino (2006). Impulse response functions from structural dynamic factor models: A monte carlo evaluation. CEPR Discussion Paper No. 5621.

Lütkepohl, H. (1996). *Handbook of Matrices*. John Wiley & Sons, Chichester.

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

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

Stock, J. H. and M. W. Watson (2005). Implications of dynamic factors models for VAR analysis. NBER Working Papers No.11467.

7 Appendix: Data description and estimation results

The appendix lists the variables used in the empirical analysis and describes the applied transformation (column A in the following table). All variables are in levels and all but the Federal Funds rate are expressed in logarithms

Nr	Variable
1	Real Gross Domestic Product, Quantity Indexes; (2000=100,SA)
2	Real final sales to domestic purchasers; (2000=100,SA)
3	Real personal consumption expenditures; (2000=100, SA)
4	Real personal consumption expenditures: Durable goods; (2000=100, SA)
5	Real personal consumption expenditures: Motor vehicles and parts;(2000=100, SA)
6	Real personal consumption expenditures: Household equipment; (2000=100, SA)
7	Real personal consumption expenditures: Nondurable goods; (2000=100, SA)
8	Real personal consumption expenditures: Food; (2000=100, SA)
9	Real personal consumption expenditures: Clothing and shoes; (2000=100, SA)
10	Real personal consumption expenditures: Energy goods; (2000=100, SA)
11	Real personal consumption expenditures: Services; (2000=100, SA)
12	Real personal consumption expenditures: Housing; (2000=100, SA)
13	Real personal consumption expenditures: Household operation; (2000=100, SA)
14	Real personal consumption expenditures: Electricity and gas; (2000=100, SA)
15	Real personal consumption expenditures: Transportation; (2000=100, SA)
16	Real personal consumption expenditures: Medical care; (2000=100, SA)
17	Real personal consumption expenditures: Recreation;(2000=100, SA)
18	Real gross private domestic investment; (2000=100, SA)
19	Real private fixed investment; (2000=100, SA)
20	Real private fixed investment: Nonresidential: Structures; (2000=100, SA)
21	Real private fixed investment: Nonresidential: Commercial struct.;(2000=100, SA)
22	Real private fixed investment: Nonresidential: Manufacturing struct.; (2000=100,SA)
23	Real private fixed investment: Nonresidential: Power & communic. struct.; (2000=100, SA)
24	Real private fixed investment: Nonresidential: Mining struct.; (2000=100, SA)
25	Real private fixed investment: Nonresidential: Equipment and software; (2000=100, SA)
26	Real private fixed investment: Nonresidential: Information processing equipment and software; (2000=100, SA)
27	Real private fixed investment: Nonresidential: Software; (2000=100, SA)
28	Real private fixed investment: Nonresidential: Equipment and software: Industrial equipment;(2000=100, SA)
29	Real private fixed investment: Nonresidential: Equipment and software: Transportation equipment; (2000=100, SA)
30	Real private fixed investment: Residential: Structures; (2000=100, SA)
31	Real private fixed investment: Residential: Structures: Permanent site; (2000=100, SA)
32	Real private fixed investment: Residential: Structures: Permanent site: Single family; (2000=100, SA)
33	Real private fixed investment: Residential: Structures: Other structures; (2000=100, SA)

Nr	Variable
34	Real private fixed investment: Residential: Equipment; (2000=100, SA)
35	Real Exports; (2000=100, SA)
36	Real Exports: Goods; (2000=100, SA)
37	Real Exports: Services; (2000=100, SA)
38	Real Imports; (2000=100, SA)
39	Real Imports: Goods; (2000=100, SA)
40	Real Imports: Services; (2000=100, SA)
41	Real government consumption expenditures and gross investment; (2000=100, SA)
42	Real government consumption expenditures and gross investment: Federal; (2000=100, SA)
43	Real government consumption expenditures and gross investment: Federal: National defense; (2000=100, SA)
44	Real government consumption expenditures and gross investment: Federal: National defense: Consumption expenditures; (2000=100, SA)
45	Real government consumption expenditures and gross investment: Federal: National defense: Gross investment; (2000=100, SA)
46	Real government consumption expenditures and gross investment: Federal: Nondefense; (2000=100, SA)
47	Real government consumption expenditures and gross investment: Federal: Nondefense: Consumption expenditures; (2000=100, SA)
48	Real government consumption expenditures and gross investment: Federal: Nondefense: Gross investment; (2000=100, SA)
49	Real government consumption expenditures and gross investment: State and local;
50	Real government consumption expenditures and gross investment: State and local: Consumption expenditures; (2000=100, SA)
51	Real government consumption expenditures and gross investment: State and local: Gross investment, (2000=100, SA)
52	Industrial Production Index: Total index; (2000=100, SA)
53	Industrial Production Index: Final products and nonindustrial supplies;(2000=100, SA)
54	Industrial Production Index: Consumer goods; (2000=100, SA)
55	Industrial Production Index: Durable consumer goods; (2000=100, SA)
56	Industrial Production Index: Nondurable consumer goods; (2000=100, SA)
57	Industrial Production Index: Business equipment; (2000=100, SA)
58	Industrial Production Index: Defense and space equipment; (2000=100, SA)
59	Industrial Production Index: Materials; (2000=100, SA)
60	Industrial Production Index: Construction supplies; (2000=100, SA)
61	Industrial Production Index: Business supplies; (2000=100, SA)
62	Industrial Production Index: Mining NAICS=21; (2000=100, SA)
63	Industrial Production Index: Manufacturing (SIC); (2000=100, SA)
64	Output Per Hour of All Persons: Nonfarm Business Sector; Index (1992=100,SA)
65	Output Per Hour of All Persons: Business Sector; Index (1992=100,SA)
66	Federal Fund rate
67	1-Year Treasury Constant Maturity Rate
68	3-Year Treasury Constant Maturity Rate
69	5-Year Treasury Constant Maturity Rate

Table 2: Choice of the number of nonstationary dynamic factors, information criteria IPC

Inf. Criteria\$k_{\max}$	2	3	4	5	6	7	8	9	10
IPC_1	2	2	2	3	3	3	3	3	4
IPC_2	2	2	2	3	3	3	3	3	4
IPC_3	2	2	2	3	3	3	3	4	4

Table 3: Variable names and description

Name	Nr	Description
Con	3	Real personal consumption expenditures;
ConD	4	Real personal consumption expenditures: Durable goods;
ConND	7	Real personal consumption expenditures: Nondurable goods;
ConS	11	Real personal consumption expenditures: Services;
InvS	20	Real private fixed investment: Nonresidential: Structures;
InvCS	21	Real private fixed investment: Nonresidential: Commercial struct.;
InvRS	31	Real private fixed investment: Residential: Structures: Permanent site;
GovD	43	Real government consumption expenditures and gross investment: Federal: National defense;
GovND	46	Real government consumption expenditures and gross investment: Federal: Nondefense;
FF	66	Federal Fund rate

NOTE: Variable number corresponds with the ordering defined in the data description.

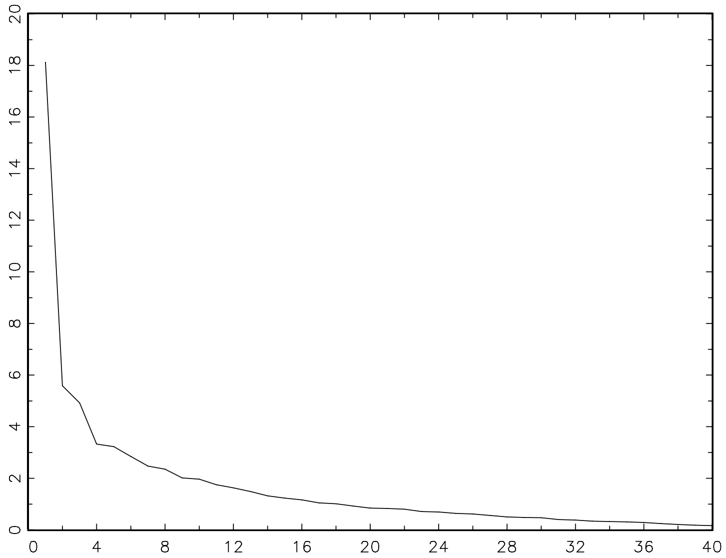

Figure 1: First largest eigenvalues of the panel correlation matrix.

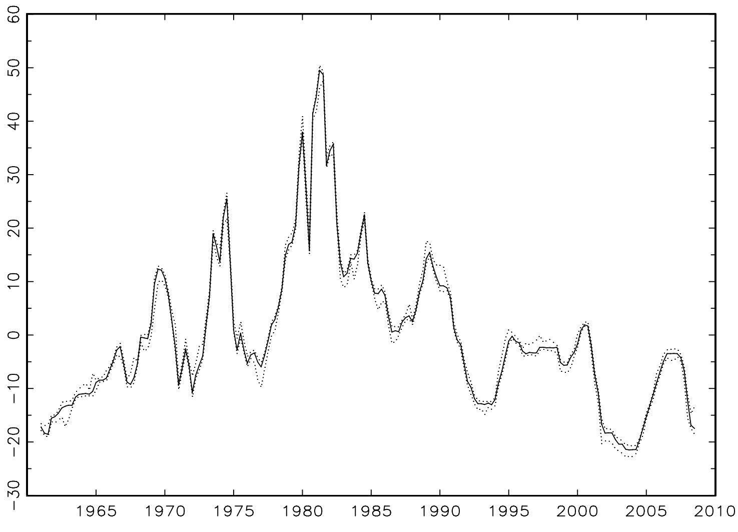

Figure 2: Federal Funds rate (solid line) and the confidence intervals (dotted lines) for a benchmark model with ten factors; significance level 5%; normalized data

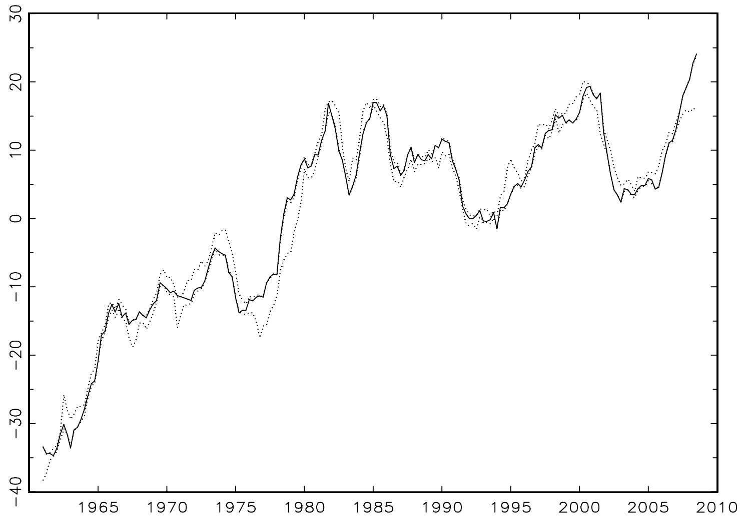

Figure 3: Real private fixed investments in nonresidential structures (solid line) and confidence intervals (dotted lines) for a benchmark model with ten factors; significance level 5%; normalized data.

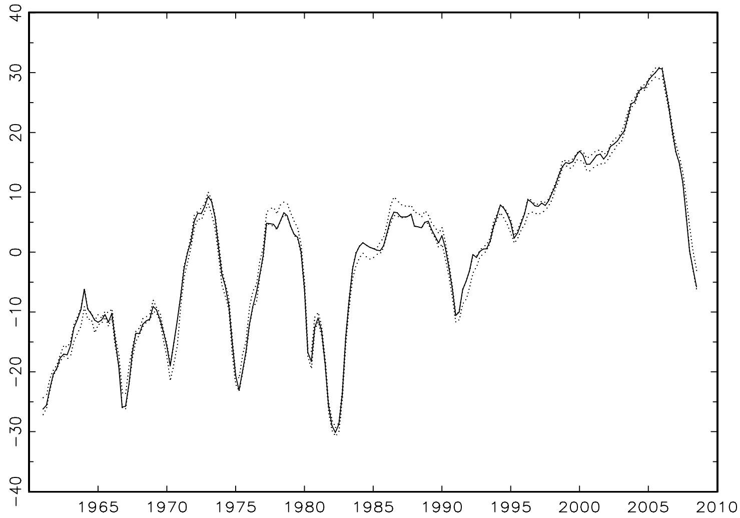

Figure 4: Real private fixed investments in residential permanent site structures (solid line) and confidence intervals (dotted lines) for a benchmark model with ten factors; significance level 5%; normalized data.

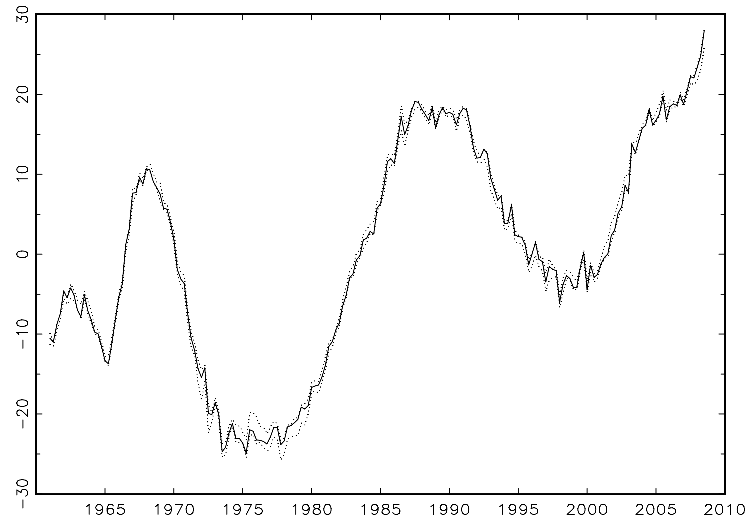

Figure 5: Real federal government consumption expenditures and gross investments in national defense (solid lines) and confidence intervals (dotted lines) for a benchmark model with ten factors; significance level 5%; normalized data.

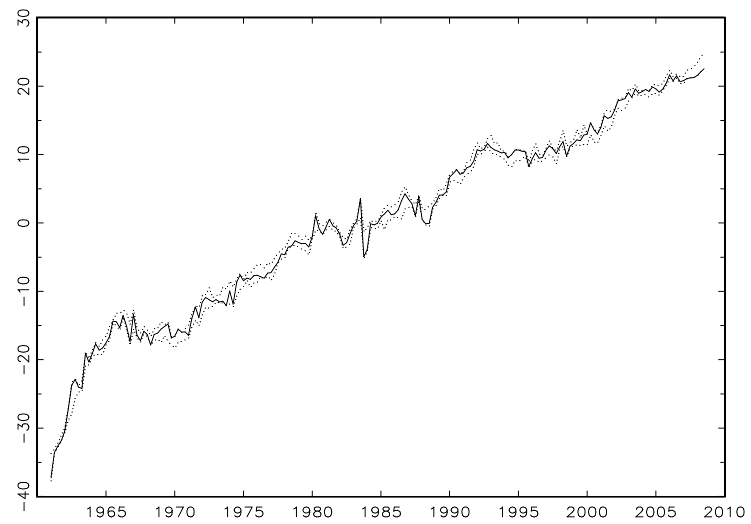

Figure 6: Real federal government consumption expenditures and gross investments in nondefence sectors (solid line) and confidence intervals (dotted lines) for a benchmark model with ten factors; significance level 5%; normalized data.

Table 4: Percentage of observations that remain outside confidence intervals for models with different number of factors

Variable Name	Number of factors 6	7	8	9	10	11	12	13	14
Con	51.83	53.40	18.85	24.61	31.41	25.13	26.70	31.94	28.27
ConD	72.25	72.77	72.77	71.73	74.35	63.87	49.74	38.22	27.23
ConND	80.10	65.97	34.56	38.22	42.93	39.27	45.55	48.17	45.55
ConS	14.14	17.80	15.18	17.23	18.32	23.04	36.70	30.37	32.46
InvS	78.53	44.50	52.88	23.56	21.99	0.00	0.00	0.00	0.00
InvCS	88.48	21.47	0.00	0.00	0.00	0.00	0.00	0.00	0.00
InvRS	6.28	7.33	8.90	4.71	6.28	0.00	0.00	0.00	0.00
GovD	3.67	9.95	12.04	1.57	1.57	4.71	3.14	5.76	7.33
GovND	82.72	71.22	81.67	4.19	3.66	1.57	1.57	0.00	0.00
FF	38.22	46.60	56.54	60.21	0.00	0.00	0.00	0.00	0.00

NOTE: Variable name corresponds with the description presented in Table 3.

8 Appendix: Proofs

8.1 General algebra results

In the following sections we use some general properties of the Euclidean norm

$$\|A\|^2 = tr\,(A'A)$$

The results can be found in Lütkepohl (1996).

1. $\|A\| = \|A'\|$
2. $\|cA\| = |c|\,\|A\|$
3. Cauchy-Schwarz inequality

$$\|AB\| \leq \|A\|\,\|B'\| = \|A\|\,\|B\|$$

4. Parallelogram identity

$$\|A+B\|^2 + \|A-B\|^2 \leq 2\left(\|A\|^2 + \|B\|^2\right)$$

Thus,

$$\begin{aligned}\|A+B\|^2 &\leq 2\left(\|A\|^2 + \|B\|^2\right)\\ &\leq 2\left(\|A\| + \|B\|\right)^2\end{aligned}$$

and therefore,

$$\|A+B\| \leq \sqrt{2}\left(\|A\| + \|B\|\right)$$

Lemma 17 *(Eigenvalues and singular values results) Let us define by $\sigma_i(A)$ the ith largest singular value of a matrix A and by $\lambda_i(B)$ the ith largest eigenvalue of a square matrix B. Then, for any real $m \times n$ matrix A the following results holds*

1. *The matrices $A'A$ and AA' are square, symmetric and positive semidefinite*
2. *If $m \geq n$ then for $i \leq n$ there is $\lambda_i(AA') = \lambda_i(A'A)$*
3. *A and B are $m \times n$ matrices, with $r = \min\{m, n\}$ then for $1 \leq i, j, i+j-1 \leq r$*
$$\sigma_{i+j-1}(AB') \leq \sigma_i(A)\,\sigma_j(B)$$
4. *A is a $m \times n$ matrix, with $m \geq n$, B is a $n \times n$ square matrix then for $1 \leq i, j, i+j-1 \leq n$*
$$\sigma_{i+j-1}(AB') \leq \sigma_i(A)\,\sigma_j(B)$$

Proof. The results (1) and (3) are presented in Lütkepohl (1996). Consider (2). Since the matrices AA' and $A'A$ are symmetric and positive definite then $\lambda_i(AA') \geq 0$ and $\lambda_i(A'A) \geq 0$. Moreover $rk(AA') = rk(A'A) = r$ and r equals the number of the non-zero eigenvalues of the matrices AA' and $A'A$. Therefore, for all $i = 1, ..., r$ there is $\lambda_i(AA') > 0$ and $\lambda_i(AA') = \lambda_i(A'A)$ (see Lütkepohl (1996)). For $i > r$ we have $\lambda_i(AA') = \lambda_i(A'A) = 0$. Thus, $\lambda_i(AA') = \lambda_i(A'A)$.

Consider (4). It follows directly from the part (3). We can construct a $m \times n$ matrix $\bar{B}$ such that
$$\bar{B} = \begin{bmatrix} B \\ 0_{(m-n)\times n} \end{bmatrix}$$
and
$$\left(A\bar{B}'\right)'\left(A\bar{B}'\right) = \begin{bmatrix} (AB')'(AB') & 0 \\ 0 & 0 \end{bmatrix}$$
Then $\sigma_j(B) = \sigma_j\left(\bar{B}\right)$ and $\sigma_i(AB') = \sigma_i\left(A\bar{B}'\right)$ for any $i, j \leq n$. Therefore,
$$\begin{aligned} \sigma_{i+j-1}(AB') &= \sigma_{i+j-1}\left(A\bar{B}'\right) \\ &\leq \sigma_i(A)\,\sigma_j\left(\bar{B}\right) \\ &= \sigma_i(A)\,\sigma_j(B) \end{aligned}$$

■

8.2 Estimation

Proof of Lemma 2. The loadings matrix $\hat{\Lambda}$ satisfies the condition
$$\hat{F}\hat{\Lambda}' = \tilde{F}\tilde{\Lambda}'$$

Moreover, we know that

$$\tilde{\Lambda}' = D^{-2}\tilde{F}'X$$

and therefore

$$\hat{F}\hat{\Lambda}' = \tilde{F}\tilde{\Lambda}' = \frac{1}{T^3}\tilde{F}D^{-2}\tilde{F}'X$$

Thus,

$$\hat{F}'\hat{F}\hat{\Lambda}' = \hat{F}'\tilde{F}D^{-2}\tilde{F}'X$$

and

$$\hat{\Lambda}' = \left(\hat{F}'\hat{F}\right)^{-1}\hat{F}'\tilde{F}D^{-2}\tilde{F}'X \tag{13}$$

From definition of the normalized factor $\hat{F} = N^{-1}X\tilde{\Lambda}$ and the loadings $\tilde{\Lambda}' = D^{-2}\tilde{F}'X$ it follows that

$$\hat{F} = \frac{1}{N}X\tilde{\Lambda} = \frac{1}{N}\left(XX'\right)\tilde{F}D^{-2}$$

Let us denote by $\tilde{V}_{NT}$ the diagonal matrix consisting of the first r largest eigenvalues of the matrix XX' and $V_{NT} = D^{-2}\tilde{V}_{NT}/N$. Then by the fact that both V_{NT} and D are diagonal there is $\hat{F}'\tilde{F} = V_{NT}D^2$ and $\hat{F}'\hat{F} = V_{NT}^2D^2$

$$\begin{aligned}
\hat{F}'\tilde{F} &= \frac{1}{N}D^{-2}\tilde{F}'\left(XX'\right)\tilde{F} = \frac{D^{-2}}{N}\tilde{V}_{NT}D^2 = V_{NT}D^2 \\
\hat{F}'\hat{F} &= \left(\frac{1}{N}\right)^2 D^{-2}\tilde{F}'\left(XX'\right)\left(XX'\right)\tilde{F}D^{-2} = \frac{D^{-2}}{N}\tilde{V}_{NT}\frac{D^{-2}}{N}D^2 = V_{NT}^2D^2
\end{aligned}$$

Finally, from equation (13) the normalized loadings are $\hat{\Lambda} = \tilde{V}_{NT}^{-1}\tilde{\Lambda}$

$$\begin{aligned}
\hat{\Lambda}' &= \left(V_{NT}^2D^2\right)^{-1}\left(V_{NT}D^2\right)D^{-2}\tilde{F}'X = V_{NT}^{-1}D^{-2}\tilde{F}'X \\
&= V_{NT}^{-1}\tilde{\Lambda}'
\end{aligned}$$

Since $\hat{F}\hat{\Lambda}' = \tilde{F}\tilde{\Lambda}'$ then

$$\hat{F} = V_{NT}\tilde{F}$$

■

The following Lemma 18-19 discuss issues associated with the eigenvalues of matrix V_{NT}. They show that the matrix $V_{NT} = O_p(1)$.

Lemma 18 *Let us denote V_{NT}^* the diagonal matrix consisting of the first r largest eigenvalues of the matrix $F^0\left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}$ in the descending order multiplied by D^{-2}. Then $V_{NT}^* = O_p(1)$ and $\lim_{T,N\to\infty} V_{NT,i}^* > 0$, where $V_{NT,i}^*$ denotes the ith diagonal element of V_{NT}^*.*

Proof. The ith diagonal element of the matrix V_{NT}^* is the ith largest eigenvalue of the matrix

$$V_{NT,i}^* = \lambda_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)\frac{F^{0\prime}}{d_i}\right)$$

where $d_i = D_{ii}$. We show that

$$\lambda_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)\frac{F^{0\prime}}{d_i}\right) = O_p\left(1\right)$$

Let us first notice that since $i \leq r$. Then by Lemma 17

$$\begin{aligned}
\lambda_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)\frac{F^{0\prime}}{d_i}\right) &= \lambda_i\left(\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\frac{F^{0\prime}}{d_i}\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\right) \\
&= \sigma_i^2\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\right)
\end{aligned}$$

where $\sigma_i\left(A\right)$ denotes a ith largest singular value of a matrix A. From Lemma 17 it follows that

$$\begin{aligned}
\sigma_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\right) &\leq \sigma_i\left(\frac{F^0}{d_i}\right)\sigma_1\left(\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\right) \\
&= \sigma_i\left(\frac{F^0}{d_i}\right)\lambda_1\left(\frac{\Lambda_0'\Lambda_0}{N}\right)
\end{aligned}$$

We show that $\sigma_i\left(\frac{F^0}{d_i}\right) = O_p\left(1\right)$. Let us first notice that $\sigma_i\left(d_i^{-1}D\right) = 1$. Then by Lemma 17

$$\begin{aligned}
\sigma_i\left(\frac{F^0}{d_i}\right) &\leq \sigma_1\left(F^0D^{-1}\right)\sigma_i\left(d_i^{-1}D\right) \\
&= \sigma_1\left(F^0D^{-1}\right) \\
&\to {}^d\lambda_1\left(\Sigma\right)
\end{aligned}$$

and $\lambda_1\left(\Sigma\right) < M$ with probability 1. Since

$$\lambda_1\left(\frac{\Lambda_0'\Lambda_0}{N}\right) \to^p \lambda_1\left(\Sigma_\Lambda\right) < M$$

then

$$\lambda_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)\frac{F^{0\prime}}{d_i}\right) = O_p\left(1\right)$$

Finally, we show that $\lim_{T\to\infty}\lambda_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)\frac{F^{0\prime}}{d_i}\right) > 0$. By Lemma 17

$$\sigma_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\right)\sigma_1\left(\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{-1/2}\right) \geq \sigma_i\left(\frac{F^0}{d_i}\right)$$

Moreover,

$$\sigma_i\left(\frac{F^0}{d_i}\right)\sigma_{r-i+1}\left(d_iD^{-1}\right) \geq \sigma_r\left(F^0D^{-1}\right)$$

where $\sigma_{r-i+1}\left(d_i D^{-1}\right) = 1$ and $\sigma_r\left(F^0 D^{-1}\right) \to^p \lambda_r(\Sigma) > 0$. Thus,

$$\lim_{T\to\infty} \sigma_i\left(\frac{F^0}{d_i}\right) \geq \lambda_r(\Sigma) > 0$$

From Assumption B it follows that $\Lambda_0'\Lambda_0/N \to^p \Sigma_\Lambda$ and Σ_Λ is symmetric, positive definite. Thus

$$\begin{aligned}
\sigma_1\left(\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{-1/2}\right) &= \lambda_1\left(\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{-1}\right) \\
&= \lambda_r\left(\left(\frac{\Lambda_0'\Lambda_0}{N}\right)\right) \\
&\to {}^p\lambda_r(\Sigma_\Lambda)
\end{aligned}$$

where $0 < \lambda_r(\Sigma_\Lambda) < M$. Therefore,

$$\begin{aligned}
\lim_{N,T\to\infty} \sigma_i\left(\frac{F^0}{d_i}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\right) &\geq \lim_{N,T\to\infty} \frac{\sigma_i\left(\frac{F^0}{d_i}\right)}{\sigma_1\left(\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{-1/2}\right)} \\
&\geq \frac{\lambda_r(\Sigma)}{\lambda_r(\Sigma_\Lambda)} > 0
\end{aligned}$$

and

$$\lim_{T,N\to\infty} V_{NT,i}^* > 0$$

■

Lemma 19 *Under Assumptions A and F and N, T → ∞ the matrix* $V_{NT} = O_p(1)$.

Proof. From the model setup it follows that

$$\frac{XX'}{N} = F^0\left(\frac{\Lambda_0'\Lambda_0}{N}\right)F^{0\prime} + C$$

where C is a symmetric matrix

$$C = \frac{F^0\Lambda_0'e'}{N} + \frac{e\Lambda_0 F^{0\prime}}{N}$$

Let us denote $\lambda_i(A)$ the ith largest eigenvalue of the matrix A. By Lütkepohl (1996)

$$\lambda_i(A+B) \leq \lambda_i(A) + \lambda_{\max}(B)$$

for symmetric matrices A and B. Therefore,

$$\begin{aligned}
\lambda_i\left(\frac{XX'}{N}\right) &\leq \lambda_i\left(F^0\left(\frac{\Lambda_0'\Lambda_0}{N}\right)F^{0\prime}\right) + \lambda_{\max}(C) \\
&\leq \lambda_i\left(F^0\left(\frac{\Lambda_0'\Lambda_0}{N}\right)F^{0\prime}\right) + tr(C)
\end{aligned}$$

Thus,

$$V_{NT} \leq V_{NT}^{*} + tr\left(C\right) D^{-2}$$

From the definition of the trace operator and its properties (see Lütkepohl (1996))

$$\begin{aligned} tr\left(C\right) &= 2tr\left(\frac{e\Lambda_0 F^{0\prime}}{N}\right) \\ &= 2\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N} e_{it}\lambda_i'\right) F_t^0 \end{aligned}$$

Thus, by Assumptions A and F

$$\begin{aligned} \left\|tr\left(C\right) D^{-2}\right\| &= \left\|2\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N} e_{it}\lambda_i'\right) F_t^0 D^{-2}\right\| \\ &= \left\|2\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N} e_{it}\lambda_i'\right)\left(\sqrt{T}F_t^0 D^{-1}\right)\sqrt{T}D^{-1}\right\| \\ &\leq 2\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N} e_{it}\lambda_i'\right)\left(\sqrt{T}F_t^0 D^{-1}\right)\right\|\left\|\sqrt{T}D^{-1}\right\| \\ &= O_p\left(1\right) \end{aligned}$$

Hence, by Lemma 18

$$V_{NT} \leq V_{NT}^{*} + O_p\left(1\right) = O_p\left(1\right)$$

■

8.3 Consistency

In this section, we prove two important propositions: Propositions 3 and 4. They show that the factors can be consistently estimated and derive the corresponding convergence rates.

The following Lemmas 20 and 21 are needed to prove Proposition 3.

Lemma 20 *Under the assumptions A-C for all T and N there exists some $M < \infty$ such that*

1. $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_N\left(s,t\right)^2 \leqslant M$
2. $E\left\{\left(N^{-1/2}e_t'\Lambda_0\right)^2\right\} \leqslant M$
3. $E\left\|\left(NT\right)^{-1/2}\sum_{t=1}^{T}e_t'\Lambda_0\right\| \leqslant M$

Proof. Points (1) - (3) are proved in Bai(2004). ■

Lemma 21 *Under Assumptions A-C and* $N, T \to \infty$

1. $\left\| F^{0\prime}\tilde{F}D^{-2} \right\| = O_p(1)$
2. $\left\| e\Lambda_0 F^{0\prime}\tilde{F}D^{-2} \right\| = O_p\left(\sqrt{NT}\right)$
3. *Define a symmetric* $T \times T$ *matrix* Φ *by* $\Phi_{ts} = \gamma_N(t,s)$*, then* $\left\| \Phi\tilde{F}D^{-2} \right\| = O_p\left(\sqrt{T}\left\|D^{-1}\right\|\right)$
4. *Define a symmetric* $T \times T$ *matrix* Υ *as* $\Upsilon = ee' - \Phi$*, then* $\left\| \Upsilon\tilde{F}D^{-2} \right\| = O_p\left(\frac{T}{\sqrt{N}}\left\|D^{-1}\right\|\right)$

Proof. Consider (1). Let us denote

$$H = \frac{\Lambda_0' X' \tilde{F} D^{-2}}{N}$$

Then by Lemma 19 and Assumption B $\|H\| = O_p(1)$ because

$$\begin{aligned} \|H\| &= \left\| \frac{\Lambda_0' X \tilde{F}' D^{-2}}{N} \right\| \\ &\leq \left\| \frac{\Lambda_0'}{\sqrt{N}} \right\| \left\| \frac{X'\tilde{F}D^{-2}}{\sqrt{N}} \right\| \\ &= O_p(1)\, tr\left(\frac{D^{-2}\tilde{F}' X X' \tilde{F} D^{-2}}{N} \right) \\ &= O_p(1)\, tr\left(V_{NT}\right) = O_p(1) \end{aligned}$$

Moreover,

$$H = \frac{\Lambda_0'\Lambda_0 F^{0\prime}\tilde{F}D^{-2}}{N} + \frac{\Lambda_0' e' \tilde{F} D^{-2}}{N}$$

Then

$$\left\| \frac{\Lambda_0'\Lambda_0 F^{0\prime}\tilde{F}D^{-2}}{N} \right\| \leq \sqrt{2}\left(\|H\| + \left\| \frac{\Lambda_0' e' \tilde{F} D^{-2}}{N} \right\| \right)$$

We show that $\left\| \Lambda_0' e' \tilde{F} D^{-2}/N \right\|^2 = O_p(1)$. By Lemma 20

$$\begin{aligned} \left\| \Lambda_0' e' \tilde{F} D^{-2}/N \right\| &\leq \left\| \frac{\Lambda_0' e'}{\sqrt{NT}} \right\| \left\| \tilde{F} D^{-1} \right\| \left\| T^{1/2} D^{-1} N^{-1/2} \right\| \\ &= o_p(1) \end{aligned}$$

Thus, $\Lambda_0'\Lambda_0 F^{0\prime}\tilde{F}D^{-2}/N = O_p(1)$. Since $\Lambda_0'\Lambda_0/N$ converges to a positive definite matrix then it must be that $F^0\tilde{F}D^{-2} = O_p(1)$.

Consider (2). From the first part of the lemma it follows that

$$\begin{aligned}\left\|e\Lambda_0 F^{0\prime}\tilde{F}D^{-2}\right\| &\leq \|e\Lambda_0\|\left\|F^{0\prime}\tilde{F}D^{-2}\right\| \\ &= O_p\left(\sqrt{NT}\right)\end{aligned}$$

Consider (3)

$$\left\|\Phi\tilde{F}D^{-2}\right\|^2 \leq \|\Phi\|\left\|\tilde{F}D^{-1}\right\|\left\|D^{-1}\right\|$$

By Lemma 20, the last component is $\|\Phi\|^2 = O_p(T)$ because

$$\begin{aligned}\|\Phi\|^2 &= \sum_{t=1}^{T}\sum_{s=1}^{T}(\gamma_N(t,s))^2 \\ &= T\left\{\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}(\gamma_N(t,s))^2\right\} \\ &= TO_p(1) = O_p(T)\end{aligned}$$

Thus,

$$\begin{aligned}\left\|\Phi\tilde{F}D^{-2}\right\| &= O_p\left(\left\|D^{-1}\right\|^2\right)O_p(1)\,O_p(T) \\ &= O_p\left(\sqrt{T}\left\|D^{-1}\right\|\right)\end{aligned}$$

Finally consider (4).

$$\left\|\Upsilon\tilde{F}D^{-2}\right\| \leq \|\Upsilon\|\left\|\tilde{F}D^{-1}\right\|\left\|D^{-1}\right\|$$

Under the Assumption F.1 the last component is $\|\Upsilon\|^2 = O_p\left(T^2/N\right)$ because

$$\begin{aligned}\|\Upsilon\|^2 &= \sum_{t=1}^{T}\sum_{s=1}^{T}\left(\frac{e_t'e_s}{N} - \gamma_N(t,s)\right)^2 \\ &= \frac{T^2}{N}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{\frac{1}{N^{1/2}}\left(e_t'e_s - E\left(e_t'e_s\right)\right)\right\}^2 \\ &= \frac{T^2}{N}O_p(1) = O_p\left(\frac{T^2}{N}\right)\end{aligned}$$

Thus,

$$\begin{aligned}\left\|\Upsilon\tilde{F}D^{-2}\right\| &= O_p\left(\left\|D^{-1}\right\|\right)O_p(1)\,O_p\left(\frac{T}{\sqrt{N}}\right) \\ &= O_p\left(\frac{T}{\sqrt{N}}\left\|D^{-1}\right\|\right)\end{aligned}$$

■

Proof of Proposition 3. Let us define a matrix H as in Lemma 21. The matrix H takes the form

$$H = \frac{\Lambda_0' X' \tilde{F} D^{-2}}{N}$$

Then it was shown that $\|H\| = O_p(1)$ and thus the matrix is well defined. The difference between the estimated factors and a rotation of the true factors can be expressed as follow

$$\begin{aligned} \hat{F} - F^0 H &= \frac{1}{N} X X' \tilde{F} D^{-2} - \frac{1}{N} F^0 \Lambda_0' X' \tilde{F} D^{-2} \qquad (14) \\ &= \frac{1}{N} \left\{ \left(F^0 \Lambda_0' + e \right) X' - F^0 \Lambda_0' X \right\} \tilde{F} D^{-2} \\ &= \frac{1}{N} \left\{ e \Lambda_0 F^{0\prime} + e e' \right\} \tilde{F} D^{-2} \\ &= \frac{1}{N} \left\{ e \Lambda_0 F^{0\prime} + N \Upsilon + N \Phi \right\} \tilde{F} D^{-2} \qquad (15) \end{aligned}$$

where Υ and Φ are defined as in Lemma 21.

$$\begin{aligned} \frac{1}{4T} \left\| \hat{F} - F^0 H \right\|^2 &\leq \frac{1}{4TN^2} \left\| \left\{ e \Lambda_0 F^{0\prime} + N \Upsilon + N \Phi \right\} \tilde{F} D^{-2} \right\|^2 \\ &\leq \frac{1}{TN^2} \left\| e \Lambda_0 F^{0\prime} \tilde{F} D^{-2} \right\|^2 + \frac{N^2}{TN^2} \left\| \Phi \tilde{F} D^{-2} \right\|^2 \\ &\quad + \frac{N^2}{TN^2} \left\| \Upsilon \tilde{F} D^{-2} \right\|^2 \end{aligned}$$

From Lemma 21 it follows that

$$\begin{aligned} \frac{1}{4T} \left\| \hat{F} - F^0 H \right\|^2 &= \frac{1}{TN^2} O_p(NT) + \frac{1}{T} O_p \left(T \|D\|^{-2} \right) + \frac{1}{T} O_p \left(\frac{T^2}{N} \|D\|^{-2} \right) \\ &= O_p \left(N^{-1} \right) + O_p \left(\|D\|^{-2} \right) + O_p \left(\frac{T}{N} \|D\|^{-2} \right) \end{aligned}$$

Under the assumption $T \|D\|^{-2} = O_p(1)$ we get

$$\frac{1}{T} \left\| \hat{F} - F^0 H \right\|^2 = O_p \left(N^{-1} \right) + O_p \left(\|D\|^{-2} \right)$$

and

$$\begin{aligned} \frac{1}{T} \left\| \tilde{F} - F^0 \tilde{H} \right\|^2 &= \frac{1}{T} \left\| \left(\hat{F} - F^0 H \right) V_{NT}^{-1} \right\|^2 \\ &\leq \frac{1}{T} \left\| \left(\hat{F} - F^0 H \right) V_{NT}^{-1} \right\| \left\| V_{NT}^{-1} \right\|^2 \\ &= O_p \left(\delta_{NT}^{-2} \right) \end{aligned}$$

■

Next, we show Lemma 22 and a proof of Proposition 4.

Lemma 22 *Under Assumptions A-E, N, $T \to \infty$ and $T \left\|D^{-1}\right\|^2 = O_p(1)$ for all t it holds*

1. $\left\|D^{-2}\tilde{F}'F^0\Lambda_0' e_t/N\right\| = O_p\left(N^{-1/2}\right)$

2. $\left\|e_t'e'\tilde{F}D^{-2}/N\right\| = O_p\left(\left\|D^{-1}\right\|\right)$

Proof. Consider (1). By Lemma 20 and Lemma 21

$$\begin{aligned}\left\|\frac{D^{-2}\tilde{F}'F^0\Lambda_0' e_t}{N}\right\| &\leq \left\|D^{-2}\tilde{F}'F^0\right\|\left\|\frac{\Lambda_0' e_t}{\sqrt{N}}\right\| N^{-1/2} \\ &= O_p\left(N^{-1/2}\right)\end{aligned}$$

Let us consider (2).

$$\left\|\frac{e_t'e'\tilde{F}D^{-2}}{N}\right\| \leq \left\|\frac{e_t'e'}{N}\right\|\left\|\tilde{F}D^{-1}\right\|\left\|D^{-1}\right\|$$

The second component by definition is $O_p(1)$. It is now shown that the first part is $\|e_t'e'/N\| = O_p(1)$.

$$\begin{aligned}\|e_t'e'/N\|^2 &= \frac{1}{N^2}\sum_{s=1}^{T}\left(e_t'e_s\right)^2 \\ &\leq \left(\frac{1}{N}\sum_{s=1}^{T}\left|e_t'e_s\right|\right)^2\end{aligned}$$

Moreover, by Assumption E

$$\begin{aligned}E\left|\frac{1}{N}\sum_{s=1}^{T}\left|e_t'e_s\right|\right| &= \sum_{s=1}^{T}E\left|\frac{e_t'e_s}{N}\right| \\ &= \sum_{s=1}^{T}\bar{\gamma}_N(t,s) \\ &= O_p(1)\end{aligned}$$

Therefore,

$$\begin{aligned}\left\|\frac{e_t'e'\tilde{F}D^{-2}}{N}\right\| &= O_p(1)\left\|D^{-1}\right\| \\ &= O_p\left(\left\|D^{-1}\right\|\right)\end{aligned}$$

■

Proof of Proposition 4. Form equation (14) it follows that

$$\hat{F}_t - H'F_t^0 = \frac{1}{N}\left\{e_t\Lambda_0 F^{0\prime} + e_t'e'\right\}\tilde{F}D^{-2}$$

Thus, from Lemma 22 we get

$$\hat{F}_t - H'F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(\left\|D^{-1}\right\|\right)$$

Since $\tilde{F}_t - \tilde{H}'F_t^0 = V_{NT}^{-1}\left(\hat{F}_t - \tilde{H}'F_t^0\right)$ then also

$$\tilde{F}_t - \tilde{H}'F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(\left\|D^{-1}\right\|\right)$$

■

The following Lemma 23 is a counterpart of the Proposition 4 for models with only one type of nonstationary factors.

Lemma 23 *If there is only one type of factors (hence, $D = T^d I_r$) and $d \geq 1$ then for N, $T \to \infty$*

1. $\left\|e_t'e'\tilde{F}D^{-2}/N\right\| = O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right) + O_p\left(N^{-1/2}\right)$
2. $\tilde{F}_t - \tilde{H}'F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right)$

Proof. Consider (1). Since $D = T^d I_r$ then $e_t'e'\tilde{F}D^{-2}/N = e_t'e'\tilde{F}/\left(T^{2d}N\right)$ and

$$e_t'e'\tilde{F}/\left(T^{2d}N\right) = N\Upsilon_t\tilde{F}/\left(T^{2d}N\right) + N\Phi_t\tilde{F}/\left(T^{2d}N\right)$$

where $\Phi_t = \left(\gamma_N(t,1), ..., \gamma_N(t,T)\right)$ and $\Upsilon_t = e_te'/N - \Phi_t$.

We show that the first component $N\Upsilon_t\tilde{F}/\left(T^{2d}N\right) = O_p\left(N^{-1/2}T^{3/2-2d}\delta_{NT}^{-1}\right) + O_p\left(T^{1/2-d}N^{-1/2}\right)$.

$$\begin{aligned}
\frac{N\Upsilon_t\tilde{F}}{NT^{2d}} &= \frac{1}{T^{2d}}\sum_{s=1}^{T}\left(\frac{e_t'e_s}{N} - \gamma_N(s,t)\right)\tilde{F}_s \\
&= \frac{1}{T^{2d}}\sum_{s=1}^{T}\left(\frac{e_t'e_s}{N} - \gamma_N(s,t)\right)\left(\tilde{F}_s - \tilde{H}'F_s^0\right) + \frac{1}{T^{2d}}\sum_{s=1}^{T}\left(\frac{e_t'e_s}{N} - \gamma_N(s,t)\right)F_s^0
\end{aligned}$$

The first part is $O_p\left(N^{-1/2}T^{3/2-2d}\delta_{NT}^{-1}\right)$ by Assumption F and Proposition 3 because

$$\begin{aligned}
\frac{1}{T^{2d}}\sum_{s=1}^{T}\left(\frac{e_t'e_s}{N} - \gamma_N(s,t)\right)\left(\tilde{F}_s - \tilde{H}'F_s^0\right) &\leqslant \frac{1}{N^{1/2}T^{2d-3/2}}\left(\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s - \tilde{H}'F_s^0\right)^2\right)^{1/2} \\
&\quad \times\frac{1}{T}\sum_{s=1}^{T}\left|N^{-1/2}\sum_{i=1}^{N}\left[e_{it}e_{is} - E\left(e_{it}e_{is}\right)\right]\right| \\
&= \frac{1}{N^{1/2}T^{2d-1/2}}O_p\left(\delta_{NT}^{-1}\right)O_p(1) \\
&= O_p\left(N^{-1/2}T^{3/2-2d}\delta_{NT}^{-1}\right)
\end{aligned}$$

Since for all t, $E\left|F_t^0/T^{d-1/2}\right| = O_p(1)$, it follows that

$$
\begin{aligned}
E\left(\frac{1}{T^{2d}}\sum_{s=1}^{T}\left(\frac{e_t'e_s}{N}-\gamma_N(s,t)\right)F_s^0\right) &\leqslant \frac{1}{T^{d-1/2}N^{1/2}}\max_{1\leqslant s\leqslant T}E\left|\frac{F_s^0}{T^{d-1/2}}\right| \\
&\quad \times E\left(\frac{1}{T}\sum_{s=1}^{T}\left|\frac{1}{N^{1/2}}\sum_{i=1}^{N}\left[e_{it}e_{is}-E(e_{it}e_{is})\right]\right|\right) \\
&= \frac{1}{T^{d-1/2}N^{1/2}}O_p(1)\,O_p(1) \\
&= O_p\left(T^{1/2-d}N^{-1/2}\right)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{N\Upsilon_t\tilde{F}}{NT^3} &= O_p\left(N^{-1/2}T^{3/2-2d}\delta_{NT}^{-1}\right)+O_p\left(T^{1/2-d}N^{-1/2}\right) \\
&= O_p\left(N^{-1/2}\right)+O_p\left(T^{1/2-d}N^{-1/2}\right)
\end{aligned}
$$

Next, we prove that $N\Phi_t\tilde{F}/\left(T^{2d}N\right) = O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right)$.

$$
\begin{aligned}
\frac{N\Phi_t\tilde{F}}{NT^3} &= \frac{1}{T^{2d}}\sum_{s=1}^{T}\gamma_{NT}(t,s)\,\tilde{F}_s \\
&= \frac{1}{T^{2d}}\sum_{s=1}^{T}\gamma_{NT}(t,s)\left(\tilde{F}_s-\tilde{H}'F_s^0\right)+\frac{\tilde{H}'}{T^{2d}}\sum_{s=1}^{T}\gamma_{NT}(t,s)\,F_s^0
\end{aligned}
$$

The first expression is $O_p\left(T^{1/2-2d}\right)$ by Assumption E.1 and Proposition 3

$$
\begin{aligned}
\frac{1}{T^{2d}}\sum_{s=1}^{T}\gamma_{NT}(t,s)\left(\tilde{F}_s-\tilde{H}'F_s^0\right) &\leqslant \frac{1}{T^{2d-1/2}}\sum_{s=1}^{T}\left|\bar{\gamma}_{NT}(t,s)\right|\left(\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s-hF_s^0\right)^2\right)^{1/2} \\
&= \frac{1}{T^{2d-1/2}}O_p(1)\,O_p\left(\delta_{NT}^{-1}\right) \\
&= O_p\left(T^{1/2-2d}\right)
\end{aligned}
$$

The second expression is $O_p\left(T^{-1/2-d}\right)$ because

$$
\frac{1}{T^{2d}}\sum_{s=1}^{T}\gamma_{NT}(t,s)\,F_s^0 \leqslant \frac{1}{T^{d+1/2}}\sum_{s=1}^{T}\left|\frac{F_s^0}{T^{d-1/2}}\right|\left|\bar{\gamma}_{NT}(t,s)\right|
$$

Since for all t, $E\left|F_t^0/T\right| = O_p(1)$ then by Assumption E.1

$$
\begin{aligned}
E\left(\frac{1}{T^{d+1/2}}\sum_{s=1}^{T}\left|\frac{F_s^0}{T^{d-1/2}}\right|\left|\gamma_{NT}(t,s)\right|\right) &\leqslant \frac{1}{T^{d+1/2}}\max_{1\leqslant s\leqslant T}E\left|\frac{F_s^0}{T^{d-1/2}}\right|\sum_{s=1}^{T}\left|\gamma_{NT}(t,s)\right| \\
&= \frac{1}{T^{d+1/2}}O_p(1)\,O_p(1) \\
&= O_p\left(T^{-1/2-d}\right)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\frac{N\Phi_t \tilde{F}}{NT^3} &= O_p\left(T^{1/2-2d}\right) + O_p\left(T^{-1/2-d}\right) \\
&= O_p\left(T^{-1/2-d}\right) \\
&= O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
e_t' e' \tilde{F} D^{-2}/N &= O_p\left(N^{-1/2}\right) + O_p\left(T^{1/2-d}N^{-1/2}\right) + T^{-1/2}\left\|D^{-1}\right\| \\
&= O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right)
\end{aligned}
$$

Consider (2). From Lemma 22 and the above point it follows that

$$
\begin{aligned}
\hat{F}_t - H' F_t^0 &= \frac{1}{N}\left\{e_t \Lambda_0 F^{0\prime} + e_t' e'\right\} \tilde{F} D^{-2} \\
&= O_p\left(N^{-1/2}\right) + O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right) \\
&= O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right)
\end{aligned}
$$

Since $\tilde{F}_t - \tilde{H}' F_t^0 = V_{NT}^{-1}\left(\hat{F}_t - \tilde{H}' F_t^0\right)$ then also

$$
\tilde{F}_t - \tilde{H}' F_t^0 = O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\left\|D^{-1}\right\|\right)
$$

■

8.4 Asymptotic distribution

In this section, we derive the limiting distribution of the discussed estimators. Firstly, we show some general results and prove Lemma 6. Next, we discuss separately the issues associated with derivation of asymptotic distributions of the estimators of factors, factor loadings and common components.

Lemma 24 *Under Assumptions A-F, as N, $T \to \infty$,*

$$
\left\|N^{-1} D^{-2} \tilde{F}' \left(XX'\right) \tilde{F} D^{-2} - N^{-1} D^{-2} \tilde{F}' F^0 \left(\Lambda_0' \Lambda_0\right) F^{0\prime} \tilde{F} D^{-2}\right\|^2 = o_p(1)
$$

Proof. Let us denote

$$
b_{NT} = N^{-1} D^{-2} \tilde{F}' \left(XX'\right) \tilde{F} D^{-2} - N^{-1} D^{-2} \tilde{F}' F^0 \left(\Lambda_0' \Lambda_0\right) F^{0\prime} \tilde{F} D^{-2}
$$

Then

$$
\begin{aligned}
b_{NT} &= N^{-1}D^{-2}\tilde{F}'e\Lambda_0 F^{0\prime}\tilde{F}D^{-2} + N^{-1}D^{-2}\tilde{F}'F^0\Lambda_0'e'\tilde{F}D^{-2} + N^{-1}D^{-2}\tilde{F}'ee'\tilde{F}D^{-2} \\
&= D^{-2}\tilde{F}'\left(e\Lambda_0 F^{0\prime}\tilde{F}D^{-2}/N + F^0\Lambda_0'e'\tilde{F}D^{-2}/N + ee'\tilde{F}D^{-2}/N\right) \\
&= D^{-2}\tilde{F}'\left(\hat{F} - F^0 H\right) + D^{-2}\tilde{F}'e\Lambda_0 F^{0\prime}\tilde{F}D^{-2}/N
\end{aligned}
$$

Thus, by Proposition 3

$$
\begin{aligned}
\|b_{NT}\|/\sqrt{2} &\leq \left\|D^{-2}\tilde{F}'\left(\hat{F} - F^0 H\right)\right\| + \left\|D^{-2}\tilde{F}'e\Lambda_0 F^{0\prime}\tilde{F}D^{-2}/N\right\| \\
&\leq \sqrt{T}\left\|D^{-1}\right\|\left\|D^{-1}\tilde{F}'\right\|\left(\frac{1}{T}\left\|\hat{F} - F^0 H\right\|\right)^{1/2} + \left\|D^{-1}\right\|\left\|D^{-1}\tilde{F}'\right\|\left\|e\Lambda_0 F^{0\prime}\tilde{F}D^{-2}/N\right\| \\
&\leq O_p(1)\,O_p\left(\delta_{NT}^{-1}\right) + O_p\left(\left\|D^{-1}\right\|\right)O_p(1)\,O_p\left(N^{-1/2}\right) \\
&= O_p\left(\delta_{NT}^{-1}\right)
\end{aligned}
$$

Hence

$$
\left\|N^{-1}D^{-2}\tilde{F}'\left(XX'\right)\tilde{F}D^{-2} - N^{-1}D^{-2}\tilde{F}'F^0\left(\Lambda_0'\Lambda_0\right)F^{0\prime}\tilde{F}D^{-2}\right\| = o_p(1)
$$

■

Proof of Lemma 6. Consider (1). From Lemma 24 it follows that

$$
\left\|D^{-2}\tilde{F}'\left(XX'/N\right)\tilde{F}D^{-2} - D^{-2}\tilde{F}'F^0\left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}\tilde{F}D^{-2}\right\|^2 = o_p(1)
$$

Let us denote V_{NT}^* the diagonal matrix consisting of the r largest eigenvalues of the matrix $F^0\left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}$ multiplied by D^{-2} and F^*, the corresponding eigenvectors. Let us assume that $D^{-1}F^{*\prime}F^*D^{-1} = I$. Then

$$
\left\|D^{-2}\tilde{F}'F^0\left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}\tilde{F}D^{-2} - D^{-2}F^{*\prime}F^0\left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}F^*D^{-2}\right\|^2 = o_p(1)
$$

and $V_{NT} = V_{NT}^* + o_p(1)$. Moreover, the diagonal elements of V_{NT}^* are equal to the eigenvalues of the matrix $\left(F^{0'}F^0\right)\left(\Lambda_0'\Lambda_0/N\right)$ divided by $D^{-2}$and V_{NT}^* converges to V, where $V_{ii} = \lim_{N,T\to\infty} V_{NT,i}^* > 0$ by Lemma 18.

Consider (2). It can be shown that

$$
\begin{aligned}
D^{-1}\tilde{H}'F^{0\prime}F^0\tilde{H}D^{-1} &= D^{-1}\tilde{F}'\tilde{F}D^{-1} + o_p(1) \\
&= I + o_p(1)
\end{aligned}
$$

Since $\tilde{H} = \left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}\tilde{F}D^{-2}V_{NT}^{-1} + o_p(1)$, it holds that

$$
D^{-3}V_{NT}^{-1}\tilde{F}'F^{0\prime}\left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}F^0\left(\Lambda_0'\Lambda_0/N\right)^{0\prime}F^{0\prime}\tilde{F}V_{NT}^{-1}D^{-3} = I + o_p(1)
$$

and

$$
D^{-3}V_{NT}^{-1/2}\tilde{F}'F^{0\prime}\left(\Lambda_0'\Lambda_0/N\right)F^{0\prime}F^0\left(\Lambda_0'\Lambda_0/N\right)^{0\prime}F^{0\prime}\tilde{F}V_{NT}^{-1/2}D^{-3} = V_{NT} + o_p(1)
$$

Let us denote

$$R_{NT} = \left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2} Q_{NT} V_{NT}^{-1/2}$$

From the definition of Q_{NT} and Lemma 24 it follows that $R_{NT}'R_{NT} = I + o_p(1)$. Then the equation can be transformed into

$$D^{-1}R_{NT}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2} F^{0\prime}F^0 \left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2} R_{NT}D^{-1} = V_{NT} + o_p(1)$$

If the matrix D has all diagonal elements equal then it is straightforward that

$$R_{NT}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2} D^{-1}F^{0\prime}F^0 D^{-1}\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2} R_{NT} = V_{NT} + o_p(1)$$

and R_{NT} converges in distribution to the eigenvectors of the matrix $\Sigma_\Lambda^{1/2}\Sigma\Sigma_\Lambda^{1/2}$. Since the eigenvalues of the matrix $\Sigma_\Lambda^{1/2}\Sigma\Sigma_\Lambda^{1/2}$ are distinct then R is unique. Thus $Q = \Sigma_\Lambda^{-1/2}RV^{1/2}$ and Q is positive definite with probability 1.

If D has different elements on the diagonal then

$$R_i = \lim_{N,T\to\infty} v_i\left(\left(F^{0\prime}F^0/D_{ii}^2\right)\left(\Lambda_0'\Lambda_0/N\right)\right)$$

where $v_i(A)$ denotes the eigenvector of matrix A corresponding with the ith largest eigenvalue. ■

8.4.1 Limiting distribution of estimated common factors

The following Lemma 25 is used in the proof of Proposition 7.

Lemma 25 *Under Assumptions A-F, for N, $T \to \infty$*

$$\sqrt{N}\left(\hat{F}_t - H'F_t^0\right) \to^d Q'N\left(0, \Gamma_t\right)$$

Proof. Under the assumption $N\left\|D^{-2}\right\| \to 0$ by Proposition 4, we have

$$\sqrt{N}\left(\hat{F}_t - H'F_t^0\right) = O_p(1) + O_p\left(\left\|D^{-1}\right\| N^{1/2}\right)$$

Thus, the limiting distribution is defined by the first term $e_t\Lambda_0F^{0\prime}\tilde{F}D^{-2}/N$ and

$$\begin{aligned}\sqrt{N}\left(\hat{F}_t - H'F_t^0\right) &= \frac{D^{-2}\tilde{F}'F^0\Lambda_0'e_t}{\sqrt{N}} + o_p(1) \\ &= D^{-2}\tilde{F}'F^0\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it} + o_p(1)\end{aligned}$$

By Assumption F and Lemma 6

$$\sqrt{N}\left(\hat{F}_t - H'F_t^0\right) \to^d Q'N\left(0, \Gamma_t\right)$$

where Q is independent of $N\left(0,\Gamma_t\right)$ since it depends only on the common components that are independent from idiosyncratic disturbances. ■

Proof of Proposition 7. Under the Lemma 2 and Lemma 25

$$\begin{aligned}\sqrt{N}\left(\tilde{F}_t - \tilde{H}'F_t^0\right) &= V_{NT}^{-1}\sqrt{N}\left(\hat{F}_t - H'F_t^0\right) \\ &\to {}^d V^{-1} Q N\left(0, \Gamma_t\right)\end{aligned}$$

■

8.4.2 Limiting distribution of estimated factors loadings

Firstly, in Lemmas 26-28, we present some general results that are needed to prove Proposition 8. Then we present a proof of Proposition 8.

Lemma 26 *Under the assumption $A-F$ for N, $T \to \infty$,*

$$\bar{H} = D\tilde{H}D^{-1} = O_p\left(1\right)$$

and

$$\bar{H}\bar{H}' \to^d \Sigma^{-1}$$

Proof. Let us first notice that

$$\begin{aligned}D^{-1}\tilde{F}'\tilde{F}D^{-1} &= D^{-1}\tilde{H}'F^{0\prime}F^0\tilde{H}D^{-1} + D\tilde{H}'F^{0\prime}\left(\tilde{F} - F^0\tilde{H}\right)D^{-1} \\ &\quad + D^{-1}\left(\tilde{F} - F^0\tilde{H}\right)'F^0\tilde{H}D^{-1} \\ &\quad + TD^{-1}\frac{1}{T}\left(\tilde{F} - F^0\tilde{H}\right)'\left(\tilde{F} - F^0\tilde{H}\right)D^{-1}\end{aligned}$$

By Proposition 3 and Assumption A

$$\begin{aligned}\left\|TD^{-1}\frac{1}{T}\left(\tilde{F} - F^0\tilde{H}\right)'\left(\tilde{F} - F^0\tilde{H}\right)D^{-1}\right\| &\leq \frac{1}{T}\left\|\tilde{F} - F^0\tilde{H}\right\|^2 T\left\|D^{-2}\right\| \\ &= O_p\left(\delta_{NT}^{-2}\right) = o_p\left(1\right)\end{aligned}$$

Since $D^{-1}\tilde{F}'\tilde{F}D^{-1} = I_r = O_p\left(1\right)$, then

$$\bar{H}'\Sigma_{NT}\bar{H} + \bar{H}'B + B'\bar{H} = O_p\left(1\right) \tag{16}$$

where $\bar{H} = DHD^{-1}$, $\Sigma_{NT} = D^{-1}F^{0\prime}F^0D^{-1}$ and $B = D^{-1}F^{0\prime}\left(\tilde{F} - F^0\tilde{H}\right)D^{-1}$. Firstly, we show that $\|B\| = O_p\left(1\right)$. By Proposition 3 and Assumption A we have

$$\begin{aligned}\|B\| &= \left\|D^{-1}F^{0\prime}\left(\tilde{F}-F^{0}\tilde{H}\right)D^{-1}\right\| \\ &\leq \left\|D^{-1}F^{0\prime}\right\|\left(\frac{1}{T}\left\|\tilde{F}-F^{0}\tilde{H}\right\|^{2}\right)^{1/2}\left\|D^{-1}\right\|\sqrt{T} \\ &= O_{p}\left(\delta_{NT}^{-2}\right)=o_{p}\left(1\right)\end{aligned}$$

Since $\Sigma_{NT}=O_p\left(1\right)$ and $B=o_p\left(1\right)$, then from the properties of the quadratic form (16) it follows that $\bar{H}'=O_p\left(1\right)$. Then

$$\bar{H}'B+B'\bar{H}=o_p\left(1\right)$$

and

$$\bar{H}'\Sigma_{NT}\bar{H}=I+o_p\left(1\right)$$

Thus, by Assumption A

$$\begin{aligned}\bar{H}\bar{H}' &= \Sigma_{NT}^{-1}+o_p\left(1\right) \\ &\rightarrow {}^{d}\Sigma^{-1}\end{aligned}$$

■

Lemma 27 *Under Assumptions A-F, for N, $T\rightarrow\infty$*

1. $D^{-1}F^{0\prime}\left(\tilde{F}-F^{0}\tilde{H}\right)=O_p\left(\delta_{NT}^{-1}\right)$
2. $D^{-1}\tilde{F}'\left(\tilde{F}-F^{0}\tilde{H}\right)=O_p\left(\delta_{NT}^{-1}\right)$

Proof. Consider (1). As noted by Bai (2004)

$$\begin{aligned}D^{-1}F^{0\prime}\left(\tilde{F}-F^{0}\tilde{H}\right) &= \sum_{t=1}^{T}D^{-1}F_t^{0}\left(\tilde{F}_t-\tilde{H}'F_t^{0}\right)' \\ &\leq \max_t\left(\sqrt{T}D^{-1}F_t^{0}\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left|\tilde{F}_t-\tilde{H}'F_t^{0}\right|'\end{aligned}$$

Moreover,

$$\left(\sum_{t=1}^{T}\left|\tilde{F}_t-\tilde{H}'F_t^{0}\right|'\right)\left(\sum_{t=1}^{T}\left|\tilde{F}_t-\tilde{H}'F_t^{0}\right|'\right)'\leq 2\sum_{t=1}^{T}\left(\tilde{F}_t-\tilde{H}'F_t^{0}\right)'\left(\tilde{F}_t-\tilde{H}'F_t^{0}\right)$$

Thus, by Proposition 3

$$\left\|\sum_{t=1}^{T}\left|\tilde{F}_t-\tilde{H}'F_t^{0}\right|'\right\|\leq\left\|\tilde{F}-F^{0}\tilde{H}\right\|=O_p\left(\delta_{NT}^{-1}\sqrt{T}\right)$$

and under Assumption A

$$D^{-1}F^{0\prime}\left(\tilde{F}-F^0\tilde{H}\right)=O_p\left(\delta_{NT}^{-1}\right)$$

Part (2) follows directly from (1)

$$\begin{aligned}D^{-1}\tilde{F}'\left(\tilde{F}-F^0\tilde{H}\right) &= D^{-1}\left(\tilde{F}-F^0\tilde{H}\right)'\left(\tilde{F}-F^0\tilde{H}\right)+D^{-1}\tilde{H}F^{0\prime}\left(\tilde{F}-F^0\tilde{H}\right)\\ &= TD^{-1}O_p\left(\delta_{NT}^{-2}\right)+O_p\left(\delta_{NT}^{-1}\right)\\ &= O_p\left(\delta_{NT}^{-1}\right)\end{aligned}$$

■

Lemma 28 *Under Assumptions A-E, for N, $T\to\infty$, we have for each i*

$$\left(\tilde{\lambda}_i-\tilde{H}^{-1}\lambda_i^0\right)=O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)+O_p\left(\left\|D^{-1}\right\|\right)$$

Proof. Let us consider an expression for $\tilde{\lambda}_i$. Form the definition of $\tilde{\Lambda}'=D^{-2}\tilde{F}'X$ it follows that

$$\begin{aligned}\tilde{\lambda}_i &= D^{-2}\tilde{F}'\bar{X}_i\\ &= D^{-2}\tilde{F}'\left(F_0\lambda_i^0+\bar{e}_i\right)\\ &= D^{-2}\left(\tilde{F}'F_0\right)\lambda_i^0+D^{-2}\left(\tilde{F}'\bar{e}_i\right)\end{aligned}$$

Since $D^{-2}\tilde{F}'\tilde{F}=I$ and $F^0=F^0+\tilde{F}\tilde{H}^{-1}-\tilde{F}\tilde{H}^{-1}$ it follows

$$\begin{aligned}\tilde{\lambda}_i &= D^{-2}\tilde{F}'\tilde{F}\tilde{H}^{-1}\lambda_i^0+D^{-2}\tilde{F}'\left(F^0-\tilde{F}\tilde{H}^{-1}\right)\lambda_i^0+D^{-2}\left(\tilde{F}'\bar{e}_i\right)\\ &= \tilde{H}^{-1}\lambda_i^0+D^{-2}\tilde{F}'\left(F^0-\tilde{F}\tilde{H}^{-1}\right)\lambda_i^0+D^{-2}\left(\tilde{F}'\bar{e}_i\right)\end{aligned}$$

Hence,

$$\tilde{\lambda}_i-\tilde{H}^{-1}\lambda_i^0=D^{-2}\tilde{F}'\left(F^0\tilde{H}-\tilde{F}\right)\tilde{H}^{-1}\lambda_i^0+D^{-2}\tilde{F}'\bar{e}_i$$

The first part is $O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)$. By Lemma 27

$$\begin{aligned}\left\|D^{-2}\tilde{F}'\left(F^0\tilde{H}-\tilde{F}\right)\tilde{H}^{-1}\right\| &= \left\|D^{-2}\tilde{F}'\left(F^0\tilde{H}-\tilde{F}\right)\right\|\left\|\tilde{H}^{-1}\right\|\\ &= O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)\end{aligned}$$

From Assumption B it follows that $\lambda_i^0=O_p(1)$. Therefore,

$$D^{-2}\tilde{F}'\left(F^0\tilde{H}-\tilde{F}\right)\tilde{H}^{-1}\lambda_i^0=O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)$$

The second part can be decomposed as follows

$$D^{-2}\tilde{F}'\bar{e}_i=D^{-2}\left(\tilde{F}-F^0\tilde{H}\right)'\bar{e}_i+D^{-2}\tilde{H}'F^{0\prime}\bar{e}_i$$

By Proposition 3 the first expression $\left\| D^{-2}\left(\tilde{F}-F^{0}\tilde{H}\right)'\right\| = O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)$ because

$$\begin{aligned}\left\| D^{-2}\left(\tilde{F}-F^{0}\tilde{H}\right)'\right\| &= \left\|D^{-2}\right\|\sqrt{T}\left(\frac{1}{T}\left\|\tilde{F}-F^{0}\tilde{H}\right\|^{2}\right)^{1/2} \\ &= O_p\left(\left\|D^{-2}\right\|\sqrt{T}\right)O_p\left(\delta_{NT}^{-1}\right)O_p(1) \\ &= O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)\end{aligned}$$

Since $\bar{e}_i = O_p(1)$ then $D^{-2}\left(\tilde{F}-F^{0}\tilde{H}\right)'\bar{e}_i = O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)$. The second expression is $D^{-2}\tilde{H}'F^{0\prime}\bar{e}_i = O_p\left(\left\|D^{-1}\right\|\right)$

$$\begin{aligned}D^{-2}\sum_{i=1}^{T}F_t^{0}e_{it} &\leqslant D^{-1}\max\left\|\sqrt{T}D^{-1}F_t^{0}\right\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left|e_{it}\right| \\ &= O_p\left(\left\|D^{-1}\right\|\right)\end{aligned}$$

Thus,

$$\begin{aligned}\frac{\tilde{F}'\bar{e}_i}{T^3} &= O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)+O_p\left(\left\|D^{-1}\right\|\right) \\ &= O_p\left(\left\|D^{-1}\right\|\right)\end{aligned}$$

Finally,

$$\hat{\lambda}_i - H^{-1}\lambda_i^{0} = O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\sqrt{N}\right)+O_p\left(\left\|D^{-1}\right\|\right)$$

■

Proof of Proposition 8. By Lemma 28, we have

$$D\left(\tilde{\lambda}_i-\tilde{H}^{-1}\lambda_i^{0}\right) = O_p\left(\delta_{NT}^{-1}\right)+O_p(1)$$

Thus, the limiting distribution of $D\left(\tilde{\lambda}_i-\tilde{H}^{-1}\lambda_i^{0}\right)$ is determined by the last term $F^{0\prime}\bar{e}_i$. Therefore,

$$\begin{aligned}D\left(\tilde{\lambda}_i-\tilde{H}^{-1}\lambda_i^{0}\right) &= DD^{-2}\tilde{H}'F^{0\prime}\bar{e}_i+o_p(1) \\ &= \bar{H}'\frac{1}{\sqrt{T}}\sum_{t=1}^{N}\sqrt{T}D^{-1}F_t^{0}e_{it}+o_p(1)\end{aligned}$$

As discussed in Bai (2003), by Lemma 26

$$\bar{H}\bar{H}' \to^{d} \Sigma^{-1}$$

Thus

$$\bar{H}' \to^{d} \bar{H}^{-1}\Sigma^{-1}$$

where $\bar{H}$ is defined in Lemma 26. Therefore, by Assumption G there is

$$D\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) \to^d \bar{H}^{-1}\Sigma^{-1}N\left(0, \Omega_i\right)$$

■

Corollary 29 *Under the Assumption A-F, for N, $T \to \infty$*

$$D\left(\hat{\lambda}_i - H^{-1}\lambda_i^0\right) \to^d V^{-1}\bar{H}^{-1}\Sigma^{-1}N\left(0, \Omega_i\right)$$

Proof. By Lemma 2 and Proposition 8

$$\begin{aligned} D\left(\hat{\lambda}_i - H^{-1}\lambda_i^0\right) &= V_{NT}^{-1}D\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) \\ &\to {}^dV^{-1}\bar{H}^{-1}\Sigma^{-1}N\left(0, \Omega_i\right) \end{aligned}$$

■

8.4.3 Limited distribution of estimated common components

Let us denote $C_{it}^0 = F_t^{0\prime}\lambda_i^0$ and $\hat{C}_{it} = \hat{F}_t'\hat{\lambda}_i$. The asymptotic distribution of common components follows from the above Proposition 7 and 8.
Proof of Proposition 9. From the definition of $\hat{C}_{it}$ and C_{it}^0, we get

$$\hat{C}_{it} - C_{it}^0 = \left(\tilde{F}_t - \tilde{H}'F_t^0\right)'\tilde{H}^{-1}\lambda_i^0 + \tilde{F}_t'\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right)$$

By Proposition 4 and Assumption B we have that

$$\left(\tilde{F}_t - \tilde{H}'F_t^0\right)'H^{-1}\lambda_i^0 = O_p\left(N^{-1/2}\right) + O_p\left(\left\|D^{-1}\right\|\right)$$

Finally, by Proposition 8 and Lemma 28

$$\begin{aligned} \tilde{F}_t'\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right) &= \tilde{F}_t'D^{-1}\sqrt{T}D\left(\tilde{\lambda}_i - \tilde{H}^{-1}\lambda_i^0\right)T^{-1/2} \\ &= O_p\left(T^{-1/2}\right) + O_p\left(\delta_{NT}^{-1}T^{-1/2}\right) = O_p\left(T^{-1/2}\right) \end{aligned}$$

1. If $N/T \to 0$ then $N^{1/2}\left\|D^{-1}\right\| \to 0$ and

$$\begin{aligned} \sqrt{N}\left(\hat{C}_{it} - C_{it}^0\right) &= O_p\left(1\right) + O_p\left(N^{1/2}T^{-1/2}\right) \\ &= O_p\left(1\right) + o_p\left(1\right) \end{aligned}$$

   Thus, by Proposition 4

$$\begin{aligned} \sqrt{N}\left(\hat{C}_{it} - C_{it}^0\right) &= \lambda_i^{0\prime}\tilde{H}^{-1\prime}\sqrt{N}\left(\tilde{F}_t - \tilde{H}'F_t^0\right) + o_p\left(1\right) \\ &\to {}^d\lambda_i^{0\prime}\left(V^{-1}Q\Sigma_\Lambda\right)^{-1}V^{-1}QN\left(0, \Gamma_t\right) \\ &= \lambda_i^{0\prime}\Sigma_\Lambda N\left(0, \Gamma_t\right) \end{aligned}$$

2. If $T/N \to 0$ then

$$\begin{aligned}\sqrt{T}\left(\hat{C}_{it}-C_{it}^{0}\right) &= O_p\left(T^{1/2}N^{-1/2}\right)+O_p(1)\\ &= o_p(1)+O_p(1)\end{aligned}$$

By Proposition 8 and under assumption $t/T=\tau$

$$\begin{aligned}\sqrt{T}\left(\hat{C}_{it}-C_{it}^{0}\right) &= \tilde{F}_t'\sqrt{T}D^{-1}D\left(\tilde{\lambda}_i-\tilde{H}^{-1}\lambda_i^0\right)+o_p(1)\\ &= F_t^{0\prime}\tilde{H}\sqrt{T}D^{-1}D\left(\tilde{\lambda}_i-\tilde{H}^{-1}\lambda_i^0\right)+o_p(1)\\ &= F_t^{0\prime}\sqrt{T}D^{-1}\bar{H}D\left(\tilde{\lambda}_i-\tilde{H}^{-1}\lambda_i^0\right)+o_p(1)\\ &\to {}^{d}F_\tau'\bar{H}\left(\bar{H}\right)^{-1}\Sigma^{-1}W_i\\ &= F_\tau'\Sigma^{-1}W_i\end{aligned}$$

3. If $N/T \to \pi$ and $t/T=\tau$

$$\begin{aligned}\sqrt{N}\left(\hat{C}_{it}-C_{it}^{0}\right) &= O_p(1)+\sqrt{\pi}O_p(1)\\ &= \lambda_i^{0\prime}H^{-1\prime}\sqrt{N}\left(\hat{F}_t-H'F_t^0\right)+\sqrt{\pi}F_t^{0\prime}\sqrt{T}D^{-1}D\left(\hat{\lambda}_i-H^{-1}\lambda_i^0\right)+o_p(1)\\ &\to {}^{d}\lambda_i^{0\prime}\Sigma_\Lambda N\left(0,\Gamma_t\right)+\sqrt{\pi}F_\tau'\Sigma^{-1}W_i\end{aligned}$$

■

8.4.4 Confidence intervals

Consider the rotation of $\tilde{F}$ towards an observable variable R_t described by the regression

$$R_t=\alpha+\beta\left(\tilde{H}^{-1}\tilde{F}_t\right)+error$$

Let $\left(\hat{\alpha},\hat{\beta}\right)$ be the least-squares estimator of (α,β) and $\hat{R}_t=\hat{\alpha}+\hat{\beta}\left(\tilde{H}^{-1}\tilde{F}_t\right)$.

In Lemma 30 we show some properties of the factor estimators that are used in the proof of Proposition 10.

Lemma 30 *Under Assumptions A-E and* $T\left\|D^{-2}\right\|\leq M$ *we have for* N, $T\to\infty$

1. *If* $N^{1/2}T^{-1/2}\left\|D^{-1}\right\|\to 0$ *then*

$$\left\|N^{1/2}T^{-1/2}D^{-1}\tilde{F}'\left(\tilde{F}-F^0\tilde{H}\right)\right\|=N^{1/2}T^{-1/2}O_p\left(\delta_{NT}^{-1}\right)=o_p(1)$$

2. *If* $N^{1/2}T^{-1/2}\left\|D^{-1}\right\| \to 0$ *then*

$$\left\|N^{1/2}T^{-1}\sum_{t=1}^{T}\left(\tilde{F}_t - \tilde{H}'F_t^0\right)\right\| = N^{1/2}T^{-1/2}O_p\left(\delta_{NT}^{-1}\right) = o_p\left(1\right)$$

3. $\left\|D^{-1}T^{1/2}\tilde{F}_t\right\| = O_p\left(1\right)$

Proof. Consider (1). Let us notice that

$$\left\|N^{1/2}T^{-1/2}D^{-1}\tilde{F}'\left(\tilde{F} - F^0\tilde{H}\right)\right\| = \left\|D\right\|T^{-1/2}\left\|N^{1/2}D^{-2}\tilde{F}'\left(\tilde{F} - F^0\tilde{H}\right)\right\|$$

By Lemma 27

$$\begin{aligned}\left\|N^{1/2}D^{-2}\tilde{F}'\left(\tilde{F} - F^0\tilde{H}\right)\right\| &= O_p\left(N^{1/2}\right)\left(O_p\left(\delta_{NT}^{-2}T\left\|D^{-2}\right\|\right) + O_p\left(\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)\right)\\ &= O_p\left(N^{1/2}\delta_{NT}^{-2}T\left\|D^{-2}\right\|\right) + O_p\left(N^{1/2}\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)\\ &= O_p\left(N^{1/2}\delta_{NT}^{-2}\right) + O_p\left(N^{1/2}\left\|D^{-1}\right\|\delta_{NT}^{-1}\right)\\ &= O_p\left(\delta_{NT}^{-1}\right) + O_p\left(\left\|D^{-1}\right\|\right) = o_p\left(1\right)\end{aligned}$$

Thus,

$$\begin{aligned}\left\|N^{1/2}T^{-1/2}D^{-1}\tilde{F}'\left(\tilde{F} - F^0\tilde{H}\right)\right\| &= \left\|D\right\|T^{-1/2}o_p\left(1\right)\\ &= o_p\left(1\right)\end{aligned}$$

Consider (2).

$$\begin{aligned}\left\|\sum_{t=1}^{T}\left(\tilde{F}_t - \tilde{H}'F_t^0\right)\right\|^2 &= tr\left(\left(\sum_{t=1}^{T}\left(\tilde{F}_t - \tilde{H}'F_t^0\right)\right)\left(\sum_{t=1}^{T}\left(\tilde{F}_t - \tilde{H}'F_t^0\right)\right)'\right)\\ &\leq tr\left(2\sum_{t=1}^{T}\left(\tilde{F}_t - \tilde{H}'F_t^0\right)\left(\tilde{F}_t - \tilde{H}'F_t^0\right)'\right)\\ &= 2tr\left(\left(\tilde{F} - F^0\tilde{H}\right)'\left(\tilde{F} - F^0\tilde{H}\right)\right)\\ &= 2\left\|\tilde{F} - F^0\tilde{H}\right\|^2 = O_p\left(T\delta_{NT}^{-2}\right)\end{aligned}$$

Thus,

$$\left\|N^{1/2}T^{-1}\sum_{t=1}^{T}\left(\tilde{F}_t - \tilde{H}'F_t^0\right)\right\| = N^{1/2}T^{-1/2}O_p\left(\delta_{NT}^{-1}\right)$$

Consider (3). $D^{-1}T^{1/2}\tilde{F}_t$ can be decomposed into two parts

$$\left\|D^{-1}T^{1/2}\tilde{F}_t\right\| = \left\|D^{-1}T^{1/2}\left(\tilde{F}_t - \tilde{H}'F_t^0\right)\right\| + \left\|D^{-1}T^{1/2}\tilde{H}'F_t^0\right\|$$

By Proposition 4, $\left\|\tilde{F}_t - \tilde{H}'F_t^0\right\| = o_p(1)$. Moreover, by Assumption A

$$\begin{aligned}\left\|D^{-1}T^{1/2}\tilde{F}_t\right\| &= o_p(1) + \left\|\bar{H}'\right\| \left\|D^{-1}T^{1/2}F_t^0\right\| \\ &= o_p(1) + O_p(1) = O_p(1)\end{aligned}$$

■

Proof of Proposition 10. One can express $\hat{R}_t - \alpha - \beta F_t^0$ as follows

$$\begin{aligned}\hat{R}_t - \alpha - \beta F_t^0 &= \hat{\alpha} + \hat{\beta}\left(\tilde{H}^{-1\prime}\tilde{F}_t\right) - \alpha - \beta F_t^0 \\ &= (\hat{\alpha} - \alpha) + \left(\hat{\beta} - \beta\right)\left(\tilde{H}^{-1\prime}\tilde{F}_t\right) + \beta\tilde{H}^{-1\prime}\left(\tilde{F}_t - \tilde{H}'F_t^0\right)\end{aligned}$$

Thus,

$$\sqrt{N}\left(\hat{R}_t - \alpha - \beta F_t^0\right) = \sqrt{N}(\hat{\alpha} - \alpha) + \sqrt{N}\left(\hat{\beta} - \beta\right)\left(\tilde{H}^{-1\prime}\tilde{F}_t\right) + \sqrt{N}\beta\tilde{H}^{-1\prime}\left(\tilde{F}_t - \tilde{H}'F_t^0\right)$$

It can be shown that the first two terms are $o_p(1)$. Let us denote $Z_t = \left[1, \left(\tilde{H}^{-1}\tilde{F}_t\right)'\right]$ and a $T \times (1+r)$ matrix $Z' = [Z_1', \ldots, Z_T']$. We write ι to describe a $T \times 1$ vector $\iota' = [1, ..., 1]$. The parameter vector $\psi = (\alpha, \beta)'$ is estimated with the least-squares method. Thus, $\hat{\psi} = (Z'Z)^{-1} Z'R$. Under the null $R_t = \alpha + \beta F_t^0 = \alpha + \beta\left(\tilde{H}^{-1\prime}\tilde{F}_t\right) + \beta\tilde{H}^{-1\prime}\left(\tilde{H}'F_t^0 - \tilde{F}_t\right)$ and in matrix notation $R = Z\psi + \left(F^0\tilde{H}' - \tilde{F}\right)\tilde{H}^{-1}\beta'$. Therefore,

$$\begin{aligned}\hat{\psi} &= (Z'Z)^{-1} Z'R \\ &= (Z'Z)^{-1} Z'Z\psi + (Z'Z)^{-1} Z'\left(F^0\tilde{H}' - \tilde{F}\right)\tilde{H}^{-1}\beta' \\ &= \psi + (Z'Z)^{-1} Z'\left(F^0\tilde{H}' - \tilde{F}\right)\tilde{H}^{-1}\beta'\end{aligned}$$

So

$$\hat{\psi} - \psi = (Z'Z)^{-1} Z'\left(F^0\tilde{H}' - \tilde{F}\right)\tilde{H}^{-1}\beta'$$

Let us define a $(1+r) \times (1+r)$ diagonal matrix

$$D_T = \begin{bmatrix} T^{1/2} & 0 \\ 0 & D \end{bmatrix}$$

where D_T is the scaling matrix. Then

$$\left(\hat{\psi} - \psi\right) = D_T^{-1}MD_T^{-1}Z'\left(F^0\tilde{H}' - \tilde{F}\right)\tilde{H}^{-1}\beta'$$

with $M = \left(D_T^{-1}Z'ZD_T^{-1}\right)^{-1} = O_p(1)$. Let us denote the blocks of the matrix M as follow

$$M = \begin{bmatrix} M_{11} & M_{1F} \\ M_{F1} & M_{FF} \end{bmatrix}$$

where M_{11} is a 1×1 matrix and M_{FF} is a $r \times r$ matrix.

This implies that by Lemma 27 and Lemma 30 $\left\| \sqrt{N} (\hat{\alpha} - \alpha) \right\| = o_p(1)$

$$\begin{aligned}
\left\| \sqrt{N} (\hat{\alpha} - \alpha) \right\| / \sqrt{2} &= \left\| N^{1/2} T^{-1/2} M D_T^{-1} Z' \left(F^0 \tilde{H}' - \tilde{F} \right) \tilde{H}^{-1} \beta' \right\| / \sqrt{2} \\
&\leq \left\| N^{1/2} T^{-1/2} M_{11} T^{-1/2} \iota' \left(F^0 \tilde{H}' - \tilde{F} \right) \tilde{H}^{-1} \beta' \right\| \\
&\quad + \left\| N^{1/2} T^{-1/2} M_{1F} D^{-1} \tilde{H}^{-1\prime} \tilde{F}' \left(F^0 \tilde{H}' - \tilde{F} \right) \tilde{H}^{-1} \beta' \right\| \\
&= \| M_{11} \| \left\| \frac{N^{1/2}}{T} \iota' \left(F^0 \tilde{H}' - \tilde{F} \right) \right\| \left\| \tilde{H}^{-2} \beta' \right\| \\
&\quad + \| M_{1F} \| \left\| \frac{N^{1/2}}{T^{1/2}} D^{-1} \tilde{F}' \left(F^0 \tilde{H}' - \tilde{F} \right) \right\| \left\| \tilde{H}^{-2} \beta' \right\| \\
&= O_p(1) \, o_p(1) \, O_p(1) + O_p(1) \, o_p(1) \, O_p(1) \\
&= o_p(1)
\end{aligned}$$

By Lemma 30 $\sqrt{N} \left(\hat{\beta} - \beta \right) \left(\tilde{H}' \tilde{F}_t \right) = o_p(1)$

$$\begin{aligned}
\left\| \sqrt{N} \left(\hat{\beta} - \beta \right) \left(\tilde{H}' \tilde{F}_t \right) \right\| / \sqrt{2} &= \left\| N^{1/2} D^{-1/2} M D_T^{-1} Z' \left(F^0 \tilde{H}' - \tilde{F} \right) \tilde{H}^{-1} \beta' \tilde{H}' \tilde{F}_t \right\| / \sqrt{2} \\
&\leq \left\| N^{1/2} D^{-1} M_{F1} T^{-1/2} \iota' \left(F^0 \tilde{H}' - \tilde{F} \right) \tilde{H}^{-1} \beta' \tilde{H}' \tilde{F}_t \right\| \\
&\quad + \left\| N^{1/2} D^{-1} M_{FF} D^{-1} \tilde{H}^{-1\prime} \tilde{F}' \left(F^0 \tilde{H}' - \tilde{F} \right) \tilde{H}^{-1} \beta' \tilde{H}' \tilde{F}_t \right\| \\
&= \| M_{F1} \| \left\| \frac{N^{1/2}}{T} \iota' \left(F^0 \tilde{H}' - \tilde{F} \right) \right\| \left\| D^{-1} T^{1/2} \tilde{F}_t \right\| \left\| \tilde{H}^{-3} \beta' \right\| \\
&\quad + \| M_{FF} \| \left\| \frac{N^{1/2}}{T^{-1/2}} D^{-1} \tilde{F}' \left(F^0 \tilde{H}' - \tilde{F} \right) \right\| \left\| D^{-1} T^{1/2} \tilde{F}_t \right\| \left\| \tilde{H}^{-2} \beta' \right\| \\
&= o_p(1)
\end{aligned}$$

Therefore,

$$\sqrt{N} \left(\hat{R}_t - \alpha - \beta F_t^0 \right) = o_p(1) + \sqrt{N} \beta \tilde{H}^{-1\prime} \left(\tilde{H}' F_t^0 - \tilde{F}_t \right)$$

Since $\left(\hat{\beta} - \beta \right) = o_p(1)$ and $\sqrt{N} \left(\tilde{F}_t - \tilde{H}' F_t^0 \right) = O_p(1)$, then β can be replaced with $\hat{\beta}$ and

$$\sqrt{N} \left(\hat{R}_t - \alpha - \beta F_t^0 \right) = o_p(1) + \sqrt{N} \hat{\beta} \tilde{H}^{-1\prime} \left(\tilde{H}' F_t^0 - \tilde{F}_t \right)$$

Finally, by Proposition 7

$$\begin{aligned}
\sqrt{N} \left(\hat{R}_t - \alpha - \beta F_t^0 \right) &\rightarrow {}^d \hat{\beta} \tilde{H}^{-1\prime} V^{-1} Q N(0, \Gamma_t) \\
&= \hat{\delta} V^{-1} Q N(0, \Gamma_t)
\end{aligned}$$

■

European University Institute
European University Press