UNIVERSITA' DEGLI STUDI DI BERGAMO
DIPARTIMENTO DI INGEGNERIA GESTIONALE
QUADERNI DEL DIPARTIMENTO[†]

Department of Economics and Technology Management

Working Paper

n. 12 – 2008

Sieve bootstrap for nonstationary panel factor models

by

Lorenzo Trapani

[†] Il Dipartimento ottempera agli obblighi previsti dall'art. 1 del D.L.L. 31.8.1945, n. 660 e successive modificazioni.

[§] L'accesso alla Collana dei Quaderni del Dipartimento di Ingegneria Gestionale è approvato dal Comitato di Redazione. I *Working Papers* della Collana costituiscono un servizio atto a fornire la tempestiva divulgazione dei risultati dell'attività di ricerca, siano essi in forma provvisoria o definitiva.

Sieve Bootstrap for Nonstationary Panel Factor Models

Lorenzo Trapani*
Cass Business School and Universita' di Bergamo

January 15, 2009

Abstract

This paper considers bootstrapping nonstationary panel factor models when possible time dependence is present in the factors dynamics. The analysis does not assume any specific DGP, and a sieve bootstrap algorithm is proposed to approximate the autocorrelation structure of the processes involved in the model. The conditions under which sieve bootstrap yields consistent estimators and test statistics are explored, and a selection rule for the order of the approximation of the AR dynamics is derived. Two main results are shown. First, an invariance principle for the partial sums of the bootstrap samples of the first differences of the estimated factors is shown to hold for large T and finite or large n. Secondly, it is proved that bootstrap estimates and test statistics are consistent only for $(n, T) \to \infty$, whilst the finite n case results in inconsistent bootstrap. Sieve bootstrap is shown to be consistent for the fixed n case only in presence of no serial correlation.

*Cass Business School, Faculty of Finance, 106 Bunhill Row, London EC1Y 8TZ, Tel.: +44 (0) 207 040 5260; email: L.Trapani@city.ac.uk

1 Introduction

The bootstrap is a popularly employed tool to approximate the sampling distribution of statistics. Bootstrapping can improve useful when a statistic is free from nuisance parameters, as this could lead to asymptotic refinements and better small sample properties; however, bootstrapping has also been widely employed with non pivotal statistics, as a way of overcoming the complications arising from estimating nuisance parameters. We refer to Horowitz (2001) for a comprehensive non technical survey. However, in order for bootstrapping to be applicable, it is necessary to show that (the partial sums of) the pseudo data generated by the bootstrapping algorithm follow the same distribution as the original data, and therefore that the distribution of bootstrap statistics is the same as the asymptotic distribution. This problem has been investigated in several recent contributions, e.g. Park (2002, 2003), Chang, Park and Song (2006). The main focus of these articles is proving a sieve bootstrap invariance principle for nonstationary time series building on strong approximations (see inter alia Sakhananeko, 1980).

This paper moves from a similar research question, thereby aiming to prove an invariance principle for sieve bootstrap samples. However, sieve bootstrap is not studied within a time series framework, but with respect to large nonstationary panel factor models. These models are popular in statistics, where the Lee and Carter (1992) model for mortality has been studied extensively, also leading to some applications of bootstrap (Haberman and Renshaw, 2000), and in econometrics - see e.g. Stock and Watson (1999), Bai (2004, 2005), Bai and Ng (2004) and Bai, Kao and Ng (2008). Panel factor models differ from the standard time series framework (e.g. cointegration), in that (1) they contain unobservable regressors, which have to be estimated as well as the other parameters, thereby affecting inference (and bootstrap as well), and (2) the asymptotics depends upon two indexes, the number of units n and the number of time observations T. In this context, the bootstrap should prove particularly useful, since the limiting distribution of statistics usually depends in a complicated way on nuisance parameters due to (1) non stationarity and (2) presence of latent variables which are usually proxied by generated regressors. Another important issue is the presence of serial correlation in the data. Although most applications of bootstrap to factor models consider the i.i.d. case (e.g. Haberman and Renshaw, 2000), this needs not be the case, and neglecting time dependence can lead to invalid bootstrap and therefore invalid inference. Thus, it is important to (1) derive a bootstrap scheme that allows for possible serial correlation and (2) prove under which conditions the bootstrap scheme is valid.

In this contribution, the validity of sieve bootstrap for (large) nonstationary panel factor models is shown. The main result is an invariance principle for the bootstrap samples, which is shown here to hold in the weak form (in probability). More specifically, we build on Sakhanenko's (1980) strong approximation and on the asymptotic theory derived in Bai (2004), to develop an invariance principle for the convergence of bootstrap partial sums of the residuals and of the estimated common factors to Brownian motions. The algortihm we propose is based on (1) decomposing the data into signal and noise by using the Principal Components estimator (PC); and (2) estimating a Vector AutoRegression (VAR) for the estimated common factors and the error term. The order of the VAR is shown to be dependent on n and T; based on this algorithm, an invariance principle for the pseudo innovations is proved. Since the estimation of the VAR roots is conducted using generated regressors, as an ancillary result we show that the asymptotic law of the estimated VAR coefficients is different than in the OLS case. Consistency is proved for the case whereby $(n, T) \to \infty$ jointly, with no need for restrictions on the rate of expansion between n and T.

Based on these two results, we show that the bootstrap is consistent for the case $(n, T) \to \infty$: thus, sieve bootstrap can be used to e.g. reduce the bias of the estimated loadings. An important result in the paper is that the bootstrap can achieve consistency only when both n and T are large, and that the order of truncation of the VAR depends on n and T as well.

The type of assumptions considered here are the same as in the previous literature: no further restrictions are required in order to implement the bootstrap. The results derived in this paper can be applied to prove the validity of various bootstrap statistics by applying the continuous mapping theorem.

The remainder of the paper is organised as follows. Section 2 lays out the model and discusses the main assumptions. Section 3 contains the bootstrap algorithm and the relevant asymptotic theory; conclusions are reported in Section 5. All proofs and derivations are in Appendix Finally, a word on notation. Throughout the paper, $\|A\|$ denotes the Euclidean norm of matrix A, $\sqrt{tr\left(A'A\right)}$, "$\rightarrow$" the ordinary limit, "$\Rightarrow$" weak convergence, "$\xrightarrow{p}$" convergence in probability. Stochastic processes such as $B\left(r\right)$ on $[0,1]$ are usually written as B, integrals such as $\int_0^1 B\left(r\right)dr$ as $\int B$ and stochastic integrals such as $\int_0^1 B\left(r\right)dB\left(r\right)$ as $\int BdB$.

2 Model and Assumptions

Consider the model

$$y_{it} = \lambda_i' F_t + u_{it}, \tag{1}$$

where $i = 1, ..., n$ and $t = 1, ..., T$. We assume that the (unobservable) factors F_t are a k-dimensional vector nonstationary process defined as

$$F_t = F_{t-1} + \varepsilon_t. \tag{2}$$

Model (1) has been considered in the early econometric literature by Chamberlain and Rotschild (1983). Recent developments on the (1) in terms of the estimation and inference on the loadings λ_i and the factors F_t have been derived by Bai (2004), Bai and Ng (2004), Kao, Trapani and Urga (2007a, 2007b). Estimation and inference in the stationary case have been studied in Bai and Ng (2002) and Bai (2003). Note that the determination of the number of common components k can be done up to some data driven procedure as designed in Bai (2004). Thus, we do not need to assume knowledge of k.

Henceforth, all the asymptotic theory will be studied for the case of both the cross-sectional and the time-series dimensions, n and T respectively, growing large. This is necessary for the identification (and therefore the consistent estimation) of both the loadings λ_i and the factors F_t. All limits will be derived for $(n, T) \rightarrow \infty$ jointly - we refer to Phillips and Moon (1999) for the definition of this mode of convergence. We also define, henceforth, $\delta_{nT} = \min\left\{\sqrt{n}, \sqrt{T}\right\}$, $C_{nT} = \min\{\sqrt{n}, T\}$ and $\varphi_{nT} = \min\left\{\sqrt{n}, \sqrt{T/\log T}\right\}$.

The following assumptions hold:

Assumption 1: *(time series and cross-sectional properties of u_{it})* the error term u_{it} admits an invertible $MA(\infty)$ approximation

$$u_{it} = D_i(L)\, e_t^{u(i)} = \sum_{j=0}^{\infty} D_{ij} e_{t-j}^{u(i)},$$

where:

(i) the $e_t^{u(i)}$s are iid (over i and t) random variables with $E\left[e_t^{u(i)}\right] = 0$ and $E\left|e_t^{u(i)}\right|^8 < \infty$;

(ii) $\sum_{j=0}^{\infty} D_{ij} L^j \neq 0$ for all $|L| \leq 1$ and $\sum_{j=0}^{\infty} j^s |D_{ij}| < \infty$ for some $s \geq 1$;

(iii) *(cross sectional dependence)* $E(u_{it}u_{jt}) = \tau_{ij}$ with $\sum_{i=1}^{n} |\tau_{ij}| \leq M$ for all j;

(iv) *(time series dependence)*

(a) $E\left|n^{-1/2}\sum_{i=1}^{n}[u_{is}u_{it} - E(u_{is}u_{it})]\right|^4 \leq M$ for every (t, s)

(b) $E[n^{-1}\sum_{i=1}^{n} u_{it}u_{is}] = \gamma_{s-t}$, $\left|\gamma_{s-t}\right| \leq M$ for all s and $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\left|\gamma_{s-t}\right| \leq M$;

(v) *(initial conditions)* $E|u_{i0}|^4 \leq M$.

Assumption 2: *(time series properties of ε_t)* ε_t admits an invertible $MA(\infty)$ approximation where $\varepsilon_t = C(L)\, e_t^F = \sum_{j=0}^{\infty} C_j L^j e_{t-j}^F$ with

(i) e_t^F is an iid k-dimensional vector random process with $E\left(e_t^F\right) = 0$, $E\left[e_t^F e_t^{F\prime}\right] = \Sigma_u$ and $E\left\|e_t^F\right\|^r < \infty$ for some $r > 4$;

(ii) *(FCLT and Law of the Iterated Logarithm)* as $T \to \infty$ it holds that

(a) $T^{-2}\sum_{t=1}^{T} F_t F_t' \to \int B_\varepsilon B_\varepsilon'$ where the vector Brownian motion B_ε has covariance matrix $\Sigma_{\Delta F} = \sum_{j=0}^{\infty} C_j \Sigma_u C_j'$, with $\Sigma_{\Delta F}$ a positive definite matrix and

(b) $\liminf_{T\to\infty} (\log\log T)\, T^{-2}\sum_{t=1}^{T} F_t F_t' = D$ where D is a nonrandom positive definite matrix;

(iii) $\sum_{j=0}^{\infty} j^s \|C_j\| < \infty$ for some $s \geq 1$;

(iv) *(initial conditions)* $E\|F_0\|^4 \leq M$.

Assumption 3: *(identifiability)* the loadings λ_i are

(i) either non random quantities such that $\|\lambda_i\| \leq M$, or random quantities such that $E\|\lambda_i\|^4 < \infty$;

(ii) either $n^{-1}\sum_{i=1}^{n} \lambda_i \lambda_i' = \Sigma_\Lambda$ if n is finite, or $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} \lambda_i \lambda_i' = \Sigma_\Lambda$, if $n \to \infty$ with Σ_Λ positive definite;

(iii) the eigenvalues of the matrix $\Sigma_\Lambda^{1/2}\Sigma_{\Delta F}\Sigma_\Lambda^{1/2}$ are distinct, and the eigenvalues of the stochastic matrix $\Sigma_\Lambda^{1/2}\int B_\varepsilon B_\varepsilon' \Sigma_\Lambda^{1/2}$ are distinct almost surely.

Assumption 4:

(i) $\{\varepsilon_t\}$, $\{u_{it}\}$ and $\{\lambda_i\}$ are three mutually independent groups;

(ii) F_0 is independent of $\{u_{it}\}$ and $\{e_{it}\}$.

COMMENTS

Model (1) is a standard panel cointegration model, as employed, among others, by Lee and Carter (1992), Kao, Trapani and Urga (2007b) and Bai, Kao and Ng (2007). The model considers a common trends representation and it does not require prior knowledge of the number of latent stochastic trends k, which in this context plays the role of the rank of cointegration for the vector $[y_{1t}, ..., y_{nt}]$. This can be determined a priori in principle using various techniques depending on whether both indexes n and T tend to infinity or only one - we refer to Bai and Ng (2002) and Onatski (2006) for the cases whereby $(n, T) \to \infty$ and to Lewbel (1991), Donald (1997) and Cragg and Donald (1997) for cases where $\max\{n, T\} \to \infty$ and $\min\{n, T\}$ is fixed. A problem that arises in this framework is that neither the loadings λ_i nor the factors F_t can be observed, and therefore an estimation technique should be employed that relies solely upon the dependent variables y_{it}, thereby treating both λ_i and F_t as parameters.

Assumption 1 is similar to Assumption C in Bai (2004, p. 141), the only difference being the summability requirement for the AR coefficients. Particularly, conditions *(i)* and *(ii)* allow to establish an invariance principle for the partial sums of the bootstrap value from the general linear process u_{it}. Note that Assumption 1*(i)* is slightly more stringent than Assumption 3.1 in Park (2002, p. 474), where only $E|\eta_{it}|^r < \infty$ for $r \geq 4$ is assumed. In this paper, invariance principles are obtained only in their weak (in probability) form, and therefore $r = 4$ would suffice in principle; however, assuming $r > 4$ is needed (here and in the next Assumption) in order for inferential theory to hold. Part *(ii)* of the assumption is needed to be able to approximate the $AR(\infty)$ polynomial with a finite autoregressive representation - see e.g. Hannan and Kavalieris (1986). This is needed in order to prove consistency of the estimated factors and loadings - see Bai (2004). Assumptions *(iii)* and *(iv)* are not needed for the proof of the bootstrapping algorithm, but they are sufficient conditions in order for and they allow for some (limited) cross-sectional and time-series dependence in the error term u_{it}. Such generalizations (cross dependence and serial correlation) are possible only in a panel data environment where both n and T tend to infinity. See Bai (2003) for a discussion, albeit related to the stationary case. Part *(v)* is a standard initial condition requirement for the ordinary CLT to hold.

Assumption 2 mimics Assumption A in Bai (2003) and is required in order for (a) the dimension of the factor space to be estimated consistently and (b) the asymptotic theory for the estimated factors to hold. Assumption *(i)* is enough for both purposes and it is also the same requirement as in Park (2002, see Assumption 3.1(a)); part *(iii)* of the assumption plays the same role as Assumption 1*(ii)*. In our case, the asymptotic theory results for the estimated factors F_t will only allow for "in probability" instead of "almost sure" versions of the invariance principle, and thus in principle assuming $r = 4$ would be sufficient. Assumption *(ii)* is merely a set of sufficient conditions needed for the identification of k via information criteria (the Law of the Iterated Logarithm) and the asymptotics of the estimated F_t; note that it would be possible to have more primitive assumptions to allow for the Law of the Iterated Logarithm and the FCLT to hold.

Assumption 3 and Assumption 4 are standard requirements needed to develop the asymptotics for the estimates of λ_i and F_t. We refer to Bai (2004) for further discussions.

The estimation theory (based on Principal Components) for λ_i and F_t is studied in Bai (2004). Particularly, after deriving the number of common components k using the information criteria proposed in Bai (2004), the common factors F_t can be estimated as $\hat{F}_t$, where $\hat{F}_t$ is T times the eigenvectors corresponding to the k largest eigenvalues of matrix YY' where $Y = [y_1, ..., y_n]'$ with $y_i = [y_{i1}, ..., y_{iT}]'$. Then λ_i can be estimated running the the OLS estimator in a linear regression with y_{it} as dependent variable and the estimated factors $\hat{F}_t$ as regressors, viz.:

$$y_{it} = \lambda_i' \hat{F}_t + \hat{u}_{it}. \tag{3}$$

It is well known that λ_i and F_t are not directly identifiable since they are identifiable only up to a transformation. Therefore, instead of estimating the factors F_t (or the loadings λ_i), what one does by employing the principal component estimator is to estimate the space spanned by them up to a $k \times k$ transformation matrix, say H, thereby finding HF_t instead of F_t and $H^{-1}\lambda_i$ instead of λ_i. Whilst this issue is important, we henceforth assume (for the purpose of notational simplicity) that H is a $k \times k$ identity matrix. The implications of λ_i and F_t being identifiable only up to a nonsingular transformation will be discussed after Theorem 3.

It is important to note here that replacing the true, unobservable factors F_t with their estimates $\hat{F}_t$ alters the error term u_{it} in (1), so that now

$$\hat{u}_{it} = u_{it} + \lambda_i' \left(F_t - \hat{F}_t \right). \tag{4}$$

Thus, one would get $\hat{\lambda}_i = \left[\sum_{t=1}^{T} \hat{F}_t \hat{F}_t' \right]^{-1} \left[\sum_{t=1}^{T} \hat{F}_t y_{it} \right]$.

3 Bootstrapping

This section contains a description of the bootstrapping algorithm, the algorithm itself and an intuitive argument of the proof.

Since (1) is a cointegrating regression, one may apply the framework of Chang, Park and Song (2006) to its observable counterpart (3), and therefore carry out the bootstrapping algorithm to the vector $\left[\Delta \hat{F}_t', \hat{u}_{it} \right]'$. This would impose a unit root in the boostrap version of $\hat{F}_t$, which is needed in order for the bootstrap to be consistent as shown by Park (2003). Note that bootstrapping the whole vector would be necessary if some correlation were assumed between u_{it} and ε_t. In our case, Assumption 4 rules out any endogeneity and thus it is not strictly necessary to bootstrap the vector $\left[\Delta \hat{F}_t', \hat{u}_{it} \right]'$ and one may equivalently think of bootstrapping separately $\Delta \hat{F}_t$ and $\hat{u}_{it}$.

Henceforth, we shall use the vector $\xi_{it} = [\Delta F_t', u_{it}]'$, and we shall indicate its $AR(\infty)$ representation as $\xi_{it} = \sum_{j=1}^{\infty} \beta_j \xi_{it-j} + e_{it}$, also denoting $1 - \sum_{j=1}^{\infty} \beta_j$ as $\beta(1)$.

3.1 The generation of the bootstrap sample

The presence of autoregressive dynamics in ΔF_t and u_{it} entails the use of a bootstrapping algorithm that preserves the autocorrelation structure over time. The algorithm we propose is the sieve bootstrap (see Buhlmann, 1997, and Park, 2002), which is based on approximating the infinite AR polynomials $C(L)$ and $D_i(L)$ by truncating them at lag q, so that

$$\Delta F_t = \sum_{j=1}^{q} \alpha_{q,j} \Delta F_{t-j} + e_{qt}^{F},$$

and

$$u_{it} = \sum_{j=1}^{q} \gamma_{q,j}^{(i)} u_{it-j} + e_{qt}^{u(i)}.$$

The choice of q depends on the values of n and T and it is discussed in the following assumption.

Assumption 5: *As* $(n, T) \to \infty$, $q \to \infty$ *with* $q = o(\varphi_{nT})$.

Assumption 5 requires that the order of truncation of the AR polynomial be large and it contains an upper bound on the rate of expansion of q; note that, in order for q to diverge, it is necessary that botht the time series dimension T and the cross-sectional dimension n be large. No restrictions on the rate of expansion between n and T are required. Assumption 5 states also that, as long as $q \to \infty$, no lower bounds are required and thus q is allowed to grow as slowly as required. No indications as to the optimal choice of q are provided; however, one could think of selecting the order of truncation q using some information criterai such as e.g. AIC or BIC, under the restriction that the maximum lag allowed for be of order $o(\varphi_{nT})$. As it will be shown hereafter, the condition that $q \to \infty$ is needed in order for $\hat{\alpha}_{q,j}$ and $\hat{\gamma}_{q,j}^{(i)}$ (the estimates of $\alpha_{q,j}$ and $\gamma_{q,j}^{(i)}$ respectively) to be consistent estimators of α_j and $\gamma_j^{(i)}$.

Consider a statistic based on $\hat{\lambda}_i$, say $\varphi\left(\hat{\lambda}_i; \hat{F}_t\right)$, where φ is a continuous transformation and the presence of $\hat{F}_t$ is introduced only to emphasize that $\hat{\lambda}_i = \hat{\lambda}_i\left(\hat{F}_t\right)$. The algorithm which we propose in order to generate the pseudo sample $\xi_{it,b}^{*}$ for each iteration b is based on the following steps; note that this sieve bootstrap algorithm mimics the one proposed by Chang, Park and Song (2006), albeit for the case of a cointegration regression with the standard VAR representation. The main differences here are *(i)* the presence of unobservable variables in (1) and *(ii)* the double-index asymptotics, in that both the cross-sectional and the time series dimensions of the panel are allowed to tend to infinity. See also the discussion in Section XX.

Step 1. *(PC estimation)*

(1.1) Determine the number of common trends k using the criteria in Bai (2004), or an equivalent information criterion.

(1.2) Estimate λ_i and F_t in (1) using the PC estimator. Particularly, estimate F_t as $\hat{F}_t$, where $\hat{F}_t$ is T times the k largest eigenvalues of matrix YY'; and λ_i as $\hat{\lambda}_i = \left[\sum_{t=1}^{T} \hat{F}_t \hat{F}_t'\right]^{-1} \left[\sum_{t=1}^{T} \hat{F}_t y_{it}\right]$, the OLS estimate in (3).

(1.3) Generate $\hat{u}_{it} = y_{it} - \hat{\lambda}_i' \hat{F}_t$ and define $\xi_{it}^q = \left[\Delta \hat{F}_t', \hat{u}_{it}\right]'$.

Step 2. *(sieve estimation)*

(2.1) Estimate $\beta_{q,j}$ (obtaining $\hat{\beta}_{q,j}$) applying OLS[1] to

$$\xi_{it}^q = \sum_{j=1}^{q} \beta_{q,j} \xi_{it-j}^q + e_{it}^q. \tag{5}$$

(2.2) Compute the OLS residuals from (5) as

$$\hat{e}_{it}^q = \xi_{it} - \sum_{j=1}^{q} \hat{\beta}_{q,j} \xi_{it-j}. \tag{6}$$

Step 3. *(sieve bootstrap)* for B iterations (each iteration denoted using subscript b where necessary)

(3.1) *(resampling)*

[1]The theory discussed here relies on employing the OLS estimator for $\beta_{q,j}$. However, other techniques could be employed as well. A possible example is the Yule-Walker estimator, which could prove useful since it is known to constrain the estimated AR polynomial to be stationary. As also pointed out in Park (2002), this would not change the results derived herein, and thus, for the sake of a concise discussion, we focus our attention solely on the OLS estimator.

(3.1.a) Center the residuals (6) around their mean, as

$$\bar{e}_{it}^{q} = \hat{e}_{it}^{q} - \frac{1}{T}\sum_{t=1}^{T}\hat{e}_{it}^{q}.$$

(3.1.b) Draw (with replacement) T values from $\{\bar{e}_{it}^{q}\}_{t=1}^{T}$ to obtain the bootstrap sample $\left\{e_{it,b}^{*}\right\}_{t=1}^{T}$.

(3.2) *(generation of the sieve bootstrap sample)*

(3.2.a) Generate recursively the bootstrap sample $\left\{\xi_{it,b}^{*}\right\}_{t=1}^{T}$ as

$$\xi_{it,b}^{q*} = \sum_{j=1}^{q}\hat{\beta}_{q,j}\xi_{it-j,b}^{q*} + e_{it,b}^{*}, \tag{7}$$

using as initialization $\left\{\xi_{iq}^{q*}, ..., \xi_{i1}^{q*}\right\} = \left\{\xi_{iq}^{q}, ..., \xi_{i1}^{q}\right\}$.

(3.2.b) Integrate the first k elements of $\left\{\xi_{it,b}^{q*}\right\}_{t=1}^{T}$, say $\left\{\Delta F_{t,b}^{*}\right\}_{t=1}^{T}$, to generate F_t^* as

$$F_{t,b}^{*} = F_0^{*} + \sum_{j=1}^{t}\Delta F_{j,b}^{*},$$

where the initialization is $F_0^* = F_0$.[2]

(3.2.c) Generate the bootstrap sample $\left\{y_{it,b}^{*}\right\}_{t=1}^{T}$.[3]

3.2 Bootstrap asymptotics

In this section, we shall prove that the bootstrap approximation $\varphi_b\left(\lambda_{i,b}^{*}; F_{t,b}^{*}\right)$ is consistent, i.e. that it has the same asymptotic law as the sample counterpart $\varphi\left(\hat{\lambda}_i; \hat{F}_t\right)$. To being with, let the partial sums of the process $e_{it} = \left[e_t^{u(i)}, e_t^{F\prime}\right]'$ be defined as $W_T(r) = T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor} e_{it}$. Then Assumptions 1 and 2 ensure that the classical FCLT holds, and therefore $W_T(r) \xrightarrow{d} W(r)$ where $W(r)$ is a standard $(k+1)$-dimensional Brownian motion. This convergence

[2]Initialising F_0 could be done either by setting the (unobservable) value of F_0 equal to the first value of the estimated process F_t, say $\hat{F}_0$, or alternatively setting $F_0^* = T^{-1}\sum_{t=1}^{T}\hat{F}_t$.

[3]Generating $y_{it,b}^{*}$ will be discussed in Section .

is in the weak form, and it holds in the space of cadlag functions $D[0,1]$ endowed with the Euclidean norm $\|\cdot\|$. The weak convergence result can anyway be strengthened by defining (on a suitable space) a copy of $W_T(r)$, say $W_T'(r)$, which has the same distribution as $W_T(r)$ and can be chosen such that (see Sakhanenko, 1980)

$$P\{\|W_T'(r) - W(r)\| \geq \delta\} \leq M_r T^{1-r/2} E\|e_{it}\|^r, \tag{8}$$

where $\delta > 0$, $r > 2$ and M_r is an absolute constant depending only on r. Such results are known as "strong approximations" and they ensure that $W_T'(r)$, and therefore $W_T(r)$ which has the same distribution, converge almost surely to $W(r)$. That (8) holds in our case is immediate in light of Assumptions 1 and 2, since r is assumed to be (at least) bigger than 4 in there. Strong approximations entail that, as long as one can prove that $E\|e_{it}\|^r < \infty$ for some $r > 2$, then the FCLT holds. Depending on whether one can prove that $T^{1-r/2}E\|e_{it}\|^r \to 0$ in probability or almost surely, the invariance principle is said to hold in the weak or strong form respectively. Consider now the bootstrap sample $\{e_{it}^*\}_{t=1}^T$, where dependence on b has been supressed. Then $\{e_{it}^*\}_{t=1}^T$ is an i.i.d. sample from the empirical distribution of $\{\hat{e}_{it}\}_{t=1}^T$ defined on the probability space induced by the bootstrap. Let P^* be the measure in this probability space; then we shall denote convergence in probability and in distribution in the bootstrap space with respect to P^* as $\stackrel{p^*}{\rightarrow}$ and $\stackrel{d^*}{\rightarrow}$ respectively.

In order to prove the consistency of the bootstrapping algorithm, we shall first provide will need the following preliminary Lemmas.

Lemma 1 *Let Assumptions 1-5 hold; then, as* $(n,T) \to \infty$

$$E^*\left\|e_{it,b}^*\right\|^r < \infty, \tag{9}$$

$$E^*\left\|e_{it,b}^*\right\|^r = E\|e_{it}\|^r + O_p\left(q^{-rs}\right) + O_p\left(q^r\delta_{nT}^{-r}\right) + O_p\left(q^r\varphi_{nT}^{-r}\right) + o_p(1) \tag{10}$$

for some $r > 4$.

This result is useful to prove an invariance principle for the partial sums of $e^*_{it,b}$ using (8). Note that the type of invariance principle that we shall be able to prove is in the weak form, since (9) holds in probability and not almost surely. The condition that $q \to \infty$ is necessary in order for (9) to hold, although it is not required for (10). Thus, as also shown in the proof with greater detail, the condition that $n \to \infty$ is not needed in order for (9) to hold, whilst it is needed in order for (10) to be valid.

Lemma 1 and (8) entail

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} e^*_{it,b} \xrightarrow{d^*} W(r),$$

where $W(r)$ is a $(k+1)$-dimensional standard Brownian motion. In order for this result to be extended to the bootstrap sample $\{\xi^{q*}_{it,b}\}_{t=1}^{T}$, we need the following result as well.

Lemma 2 *Under Assumptions 1-5, we have, as* $(n, T) \to \infty$

$$\max_{1 \leq j \leq q} \left\| \hat{\beta}_{q,j} - \beta_j \right\| = O_p\left(\sqrt{\frac{\log T}{T}}\right) + O_p\left(\frac{1}{C_{nT}}\right) + o_p\left(\frac{1}{q^s}\right) = O_p\left(\frac{1}{\varphi_{nT}}\right) + o_p\left(\frac{1}{q^s}\right). \tag{11}$$

Lemma 2 states that $\hat{\beta}_{q,j}$ is a uniformly consistent estimator of β_j. The rate $O_p\left(\sqrt{\log T/T}\right)$ is a well-known result in time series analysis (see e.g. Theorem 2.1 in Hannan and Kavalieris, 1986); the term $O_p\left(C_{nT}^{-1}\right)$ arises from the fact that $\hat{\beta}_{q,j}$ is obtained from a regression where the latent variables F_t and ΔF_t are replaced by their estimated counterparts $\hat{F}_t$ and $\Delta \hat{F}_t$. Thus, the $O_p\left(C_{nT}^{-1}\right)$ term arises from the estimation error in estimating ΔF_t. Equation (11) is a joint limit result, in that $(n, T) \to \infty$ are allowed to both go to infinity with no particular ordering; also, no restrictions are required on the rate of expansion n/T. Note that in this case the condition that $n \to \infty$ is pivotal: allowing for fixed n would lead to $\max_{1 \leq j \leq q} \left\| \hat{\beta}_{q,j} - \beta_j \right\| = O_p(1)$, thereby making $\hat{\beta}_{q,j}$ not (uniformly) consistent. Thus, even though an invariance principle for the partial sums of $u^*_{it,b}$ and $\Delta F^*_{t,b}$ still holds, the long run covariance matrices can no longer be estimated consistently.

Assumption 5 is only needed to put a bound onto the choice of q here; as the proof shows with greater detail, note that, as in "classical" time series analysis without generated regressors, the condition $q \to \infty$ is necessary in order to make the approximation error $\left\| \beta_{q,j} - \beta_j \right\|$ negligible.

The condition that both n and T must be large is somehow counterintuitive in light of the literature on nonstationary panel factor models (see Bai, 2004). It is known that consistent estimation of the loadings λ_i only requires that $T \to \infty$ (see Bai, 2004); however, when bootstrapping the limiting distribution of $\hat{\lambda}_i$, validity of sieve bootstrap is achieved only if $(n, T) \to \infty$.

Consider now the partial sums of the process ξ_{it}, namely $V_T(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \xi_{it}$. Then in light of Assumptions 1 and 2 and using the Beveridge-Nelson decomposition it holds that $V_T(r) \xrightarrow{d} V(r) = \beta^{-1}(1) W(r)$. In order to prove the validity of the bootstrap algorithm proposed above, it is necessary to prove a bootstrap invariance principle for the partial sums of $\xi_{it,b}^{q*}$, i.e. that $V_T^*(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \xi_{it}^{q*} \xrightarrow{d^*} V(r)$ as $(n, T) \to \infty$. This can be done noting that, using the Beveridge-Nelson decomposition

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \xi_{it}^{q*} = \hat{\beta}_q^{-1}(1) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} e_{it}^* \right) + \frac{\hat{\beta}_q^{-1}(1)}{\sqrt{T}} \left(\bar{\xi}_{i0}^{q*} - \bar{\xi}_{i\lfloor Tr \rfloor}^{q*} \right), \tag{12}$$

where $\bar{\xi}_{it}^{q*} = \sum_{j=1}^{q} \left(\sum_{i=j}^{q} \hat{\beta}_{q,i} \right) \xi_{it-j+1}^{q*}$. Then the following Lemma holds

Lemma 3 *Let Assumptions 1-5 hold. Then as* $(n, T) \to \infty$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \xi_{it}^{q*} \xrightarrow{d^*} V(r).$$

This Lemma states that the partial sums of the bootstrap process $\{\xi_{it}^{q*}\}_{t=1}^{T}$ have the same limiting distribution as the partial sums of $\{\xi_{it}\}_{t=1}^{T}$. In order for this resul to hold, two main results are needed. Firstly, the invariance principle for the partial sums of $\{e_{it}^*\}_{t=1}^{T}$ is required to hold; this follows from Lemma 1 even for the case of fixed n. Secondly, it must hold that $\hat{\beta}_q^{-1}(1) \xrightarrow{p} \beta^{-1}(1)$; as the proof of the Lemma shows, this follows from Lemma

2. Note that this result would not hold for finite n, since in that case $\hat{\beta}_{q,j}$ would not be a consistent estimator for $\beta_{q,j}$. Thus, whilst it is possible to have a valid bootstrap approximation of the partial sums of e_{it} even for finite n (when factors F_t are not estimated consistently) as shown in Lemma 1, the condition that $n \to \infty$ is necessary in order to achieve consistency of the bootstrap for ξ_{it}. This result obviously affects $\lambda^*_{i,b}$ as well, which is in apparent contradiction with the estimation theory of λ_i where only $T \to \infty$ is required.

4 Discussion

Section 3 considers the validity of an invariance principle for the partial sums of the pseudo samples $\xi^*_{it,b}$. At least in principle, Lemmas 1, 2 and 3, together with the Continuous Mapping Theorem (CMT), are sufficient in order to use sieve bootstrap to numerous applications of nonstationary large panel factor models. There remains the issue, however, as to how to generate the pseudo samples $y^*_{it,b}$, i.e. as to how to proceed with the algorithm after Step 3.2.c. The answer depends on the specific problem one wants to investigate, as there are different strategies towards:

a) the generation of $y^*_{it,b}$ itself. For example, $y^*_{it,b}$ could be computed using a "fixed regressor" approach (see also Hansen, 2000), whereby

$$y^*_{it,b} = \hat{\lambda}_i' \hat{F}_t + u^*_{it,b}; \tag{13}$$

alternatively, $y^*_{it,b}$ could be generated replacing the (generated) regressor $\hat{F}_t$ with pseudo data, so that

$$y^*_{it,b} = \hat{\lambda}_i' F^*_{t,b} + u^*_{it,b}; \tag{14}$$

b) how to obtain the bootstrap estimates λ^*_i and $\hat{F}^*_{t,b}$. One possible approach would be applying the Principal Component (PC) estimator to $y^*_{it,b}$, thereby obtaining $\lambda^{*,PC}_i$ and $\hat{F}^{*,PC}_{t,b}$; alternatively, the common factors ($\hat{F}_t$ or $F^*_{t,b}$ depending on whether one chooses a fixed regrssor approach or not to generate $y^*_{it,b}$) could be treated as observable, and thus λ^*_i could be computed by applying OLS - or equivalent techniques.

In this Section, some potential pitfalls related to points *a)* and *b)* are discussed. Particularly, two different applications are studied:

1. bias reduction for the PC estimator $\hat{\lambda}_i$;
2. bootstrap approximation for the common component $C_{it} = \lambda_i' F_t$.

Note that henceforth the convention whereby the rotation matrix H was set equal to the identity matrix (employed in Section 3) is relaxed in order to facilitate the discussion. Thus, the estimation errors for λ_i and F_t are referred to as $\hat{\lambda}_i - H^{-1}\lambda_i$ and $\hat{F}_t - HF_t$ respectively.

Bias reduction for $\hat{\lambda}_i$

The asymptotic theory developed in Bai (2004) ensures that $\hat{\lambda}_i - H^{-1}\lambda_i = O_p\left(T^{-1}\right)$. Thus, as $T \to \infty$ and irrespective of n being fixed or large, $\hat{\lambda}_i$ is an asymptotically unbiased estimator. However, considering the formula

$$\begin{aligned}
\hat{\lambda}_i - H^{-1}\lambda_i &= \left[\sum_{t=1}^{T} \hat{F}_t \hat{F}_t'\right]^{-1} \left[\sum_{t=1}^{T} \hat{F}_t \hat{u}_{it}\right] \\
&= \left[\sum_{t=1}^{T} \hat{F}_t \hat{F}_t'\right]^{-1} \left[\sum_{t=1}^{T} \hat{F}_t u_{it} + \sum_{t=1}^{T} \hat{F}_t \left(HF_t - \hat{F}_t\right)' H^{-1}\lambda_i\right],
\end{aligned}$$

it can be readily seen that small sample bias can arise from terms such as $\sum_{t=1}^{T} \hat{F}_t \left(HF_t - \hat{F}_t\right)' H^{-1}\lambda_i$ and $\sum_{t=1}^{T} \left(F_t - \hat{F}_t\right) u_{it}$, whose order of magnitude (see Bai, 2004) is $O_p\left(T^{-1}C_{nT}^{-1}\right)$. Bootstrap can be applied to correct for small sample bias as follows. The pseudo sample $y^*_{it,b}$ can be generated either according to (13) or to (14). The bootstrap estimator $\lambda^*_{i,b}$ is the OLS estimator, thereby treating the common factors as known; the estimation error is (note the absence of H)

$$\lambda_{i,b}^{(1)*} - \hat{\lambda}_i = \left[\sum_{t=1}^{T} \hat{F}_t \hat{F}_t'\right]^{-1} \left[\sum_{t=1}^{T} \hat{F}_t u_{it,b}^*\right],$$

$$\lambda_{i,b}^{(2)*} - \hat{\lambda}_i = \left[\sum_{t=1}^{T} F_{t,b}^* F_{t,b}^{*\prime}\right]^{-1} \left[\sum_{t=1}^{T} F_{t,b}^* u_{it,b}^*\right].$$

Since $u_{it,b}^*$ is obtained from bootstrapping $\hat{u}_{it}$, the distribution limit of the partial sums of the two processes are the same. Thus, both $\lambda_{i,b}^{(1)*}$ and $\lambda_{i,b}^{(2)*}$ deliver the same estimation error as the PC estimator $\hat{\lambda}_i$, and thus, using the invariance principle derived above and the CMT, they should have the same bias. Thus, the bias of $\hat{\lambda}_i$ can be reduced by calculating $\hat{\lambda}_i - \delta_B^{(1),(2)}$, with $\delta_B^{(1),(2)} = B^{-1} \sum_{b=1}^{B} \left[\lambda_{i,b}^{(1),(2)*} - \hat{\lambda}_i\right]$. Whilst the CMT and the invariance principle ensure that using (13) or to (14) should not make a difference, it is important to point out that using PC to calculate $\lambda_{i,b}^*$ would lead to the estimation error

$$\lambda_{i,b}^* - H_1^{-1}\hat{\lambda}_i = \left[\sum_{t=1}^{T} \hat{F}_{t,b}^{*,PC} \hat{F}_{t,b}^{*,PC\prime}\right]^{-1} \left\{\sum_{t=1}^{T} \hat{F}_{t,b}^{*,PC} \left[u_{it,b}^* + \hat{\lambda}_i' H_1^{-1} \left(H_1 F_{t,b}^* - \hat{F}_{t,b}^{*,PC}\right)\right]\right\},$$

where $\hat{F}_{t,b}^{*,PC}$ is the PC estimator of $F_{t,b}^*$. Thus, using PC would introduce the extra error term $\hat{\lambda}_i' H_1^{-1} \left(H_1 F_{t,b}^* - \hat{F}_{t,b}^*\right)$.

Bootstrap approximation of the limiting distribution of $\hat{C}_{it}$

Consider the estimator of the signal C_{it}, $\hat{C}_{it} = \hat{\lambda}_i' \hat{F}_t$, and let $y_{it,b}^* = \hat{\lambda}_i' F_{t,b}^* + u_{it,b}^*$. Then the bootstrap version of $\hat{C}_{it}$ is defined as $C_{it,b}^* = \hat{\lambda}_{i,b}^{*,PC\prime} \hat{F}_{t,b}^{*,PC}$, where $\lambda_{i,b}^{*,PC}$ and $\hat{F}_{t,b}^{*,PC}$ are the PC estimators. The estimation error $C_{it,b}^* - \hat{C}_{it}$ is

$$C_{it,b}^* - \hat{C}_{it} = \left(\hat{F}_{t,b}^{*,PC} - H_1 F_{t,b}^*\right)' H_1^{-1}\hat{\lambda}_i + \hat{F}_{t,b}^{*,PC\prime} \left(\lambda_{i,b}^{*,PC} - H_1^{-1}\hat{\lambda}_i\right) = I + II. \tag{15}$$

The decomposition of this error term is the bootstrap version of the decomposition of $\hat{C}_{it} - C_{it}$, and thus the CMT ensures that the limiting distributions are the same. Note that this hols for any expansion rate between n and T; particularly, as $(n, T) \to \infty$ we know that (see Bai, 2004) if $n/T \to 0$ the leading term is I; conversely, if $T/n \to 0$, it is II that dominates. Here bootstrap could prove useful since in practice one has finite n and T, and thus the law of $\hat{C}_{it} - C_{it}$, suitably normalised by $\min\left(\sqrt{n}, \sqrt{T}\right)$, depends on the contribution of both terms I and II. An alternative approach would be treating $F^*_{t,b}$ as known and estimating the loadings using OLS. Thus, $C^*_{it,b} = \hat{\lambda}^{*,OLS\prime}_{i,b} F^*_{t,b}$; this would yield

$$C^*_{it,b} - \hat{C}_{it} = \left(\hat{\lambda}^{*,OLS}_{i,b} - \hat{\lambda}_i\right)' F^*_{t,b},$$

which corresponds to only term II in (15) and which would not give an accurate approximation of the law of $\hat{C}_{it} - C_{it}$ unless T is much smaller than n.

5 Conclusions

This paper considers bootstrapping nonstationary panel factor models when possible time dependence is present in the factors dynamics. The analysis does not assume any specific DGP, and a sieve bootstrap algorithm is proposed to approximate the autocorrelation structure of the processes involved in the model. The conditions under which sieve bootstrap yields consistent estimators and test statistics are explored, and a selection rule for the order of the approximation of the AR dynamics is derived. Two main results are shown. First, an invariance principle for the partial sums of the bootstrap samples of the first differences of the estimated factors is shown to hold for large T and finite or large n. Secondly, it is proved that bootstrap estimates and test statistics are consistent only for $(n, T) \to \infty$, whilst the finite n case results in inconsistent bootstrap. Sieve bootstrap is shown to be consistent for the fixed n case only in presence of no serial correlation.

6 Appendix A: useful Lemmas

Lemma 4 *Let Assumptions hold. Then, for $r > 4$*

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\Delta\hat{F}_t-\Delta F_t\right\|^r = O_p\left(\delta_{nT}^{-r}\right), \tag{16}$$

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\Delta\hat{F}_t\right\|^r = O_p\left(1\right), \tag{17}$$

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{u}_{it}-u_{it}\right\|^r = O_p\left(\delta_{nT}^{-r}\right), \tag{18}$$

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{u}_{it}\right\|^r = O_p\left(1\right) \tag{19}$$

Proof. Letting $u_t = [u_{1t},...,u_{nt}]'$ and $\Lambda = (\lambda_1,\lambda_2,...,\lambda_n)'$, the error term $\Delta\hat{F}_t - \Delta F_t$ can be decomposed as (see e.g. Bai and Ng, 2002, p. 213)

$$\begin{aligned} & \Delta\hat{F}_t-\Delta F_t \\ = \; & T^{-1}\sum_{s=1}^{T}\Delta\hat{F}_s\gamma_{s-t}+T^{-1}\sum_{s=1}^{T}\Delta\hat{F}_s\zeta_{st}+T^{-1}\sum_{s=1}^{T}\Delta\hat{F}_s\eta_{st}+T^{-1}\sum_{s=1}^{T}\Delta\hat{F}_s\xi_{st}, \end{aligned}$$

where $\gamma_{s-t} = n^{-1}E\left(u_t'u_s\right)$, $\zeta_{st} = n^{-1}\left(u_t'u_s\right)-\gamma_{s-t}$, $\eta_{st} = n^{-1}\left(\Delta F_s'\Lambda' u_t\right)$ and $\xi_{st} = n^{-1}\left(\Delta F_t'\Lambda' u_s\right)$. Thus

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T}\left\|\Delta\hat{F}_t-\Delta F_t\right\|^r \leq \; & \frac{1}{T}\sum_{t=1}^{T}\left[\left\|\frac{1}{T}\sum_{s=1}^{T}\Delta\hat{F}_s\gamma_{s-t}\right\|^2\right]^{r/2}+\frac{1}{T}\sum_{t=1}^{T}\left[\left\|\frac{1}{T}\sum_{s=1}^{T}\Delta\hat{F}_s\zeta_{st}\right\|^2\right]^{r/2} \\ & +\frac{1}{T}\sum_{t=1}^{T}\left[\left\|\frac{1}{T}\sum_{s=1}^{T}\Delta\hat{F}_s\eta_{st}\right\|^2\right]^{r/2}+\frac{1}{T}\sum_{t=1}^{T}\left[\left\|\frac{1}{T}\sum_{s=1}^{T}\Delta\hat{F}_s\xi_{st}\right\|^2\right]^{r/2} \\ = \; & I+II+III+IV. \end{aligned}$$

Consider I. Applying the Cauchy-Schwartz inequality we get

$$I \leq T^{-r/2} \left[\frac{1}{T} \sum_{s=1}^{T} \left\| \Delta \hat{F}_s \right\|^2 \right]^{r/2} \frac{1}{T} \sum_{t=1}^{T} \left[\sum_{s=1}^{T} \gamma_{s-t}^2 \right]^{r/2} .$$

Assumption 1*(iv)*-(b) ensures that $\sum_{s=1}^{T} \left| \gamma_{s-t} \right|^2 = O(1)$. Note that $T^{-1} \sum_{s=1}^{T} \left\| \Delta \hat{F}_s \right\|^2 \leq T^{-1} \sum_{s=1}^{T} \left\| \Delta F_s \right\|^2 + T^{-1} \sum_{s=1}^{T} \left\| \Delta \hat{F}_s - \Delta F_s \right\|^2$, with $T^{-1} \sum_{s=1}^{T} \left\| \Delta \hat{F}_s - \Delta F_s \right\|^2 = O_p \left(\delta_{nT}^{-2} \right)$ according to Lemma A.1 in Bai (2003, p. 159); Assumption 2*(i)* and the LLN ensure that $T^{-1} \sum_{s=1}^{T} \left\| \Delta F_s \right\|^2 = O_p(1)$. Therefore, $I = O_p \left(T^{-r/2} \right)$.

As far as II is concerned, we have

$$II \leq \left[\frac{1}{T} \sum_{s=1}^{T} \left\| \Delta \hat{F}_s \right\|^2 \right]^{r/2} \frac{1}{T} \sum_{t=1}^{T} \left[\frac{1}{T} \sum_{s=1}^{T} \zeta_{st}^2 \right]^{r/2} .$$

Since it holds that $T^{-1} \sum_{s=1}^{T} \zeta_{st}^2 = O_p \left(n^{-1} \right)$ - see Bai (2003, p. 159) - we finally have $II = O_p \left(n^{-r/2} \right)$.

Considering term III, it holds that

$$\begin{aligned} \left\| \frac{1}{T} \sum_{s=1}^{T} \Delta \hat{F}_s \eta_{st} \right\|^r &= T^{-r} \left\| \frac{1}{n} \sum_{s=1}^{T} \Delta \hat{F}_s \Delta F_s' \Lambda' u_t \right\|^r \\ &= n^{-r} \left\| \Lambda' u_t \right\|^r \left\| \frac{1}{T} \sum_{s=1}^{T} \Delta \hat{F}_s \Delta F_s' \right\|^r . \end{aligned}$$

Note that $T^{-1} \sum_{s=1}^{T} \Delta \hat{F}_s \Delta F_s' = T^{-1} \sum_{s=1}^{T} \Delta F_s \Delta F_s' + T^{-1} \sum_{s=1}^{T} \left(\Delta \hat{F}_s - \Delta F_s \right) \Delta F_s' = O_p(1) + O_p \left(\delta_{nT}^{-2} \right)$ from Lemma B.2 in Bai (2003, p. 164). Also, we have $n^{-r} \left\| \Lambda' u_t \right\|^r = n^{-r/2} \left\| n^{-1/2} \sum_{i=1}^{n} \lambda_i u_{it} \right\|^r = O_p \left(n^{-r/2} \right)$ after Assumptions 2*(i)* and 3. Thus, $III = O_p \left(n^{-r/2} \right)$.

Last, term IV can be rearranged using

$$\left\| \frac{1}{T} \sum_{s=1}^{T} \Delta \hat{F}_s \xi_{st} \right\|^r = \left[\left\| \frac{1}{nT} \sum_{s=1}^{T} \Delta \hat{F}_s u_s' \Lambda \Delta F_t \right\|^2 \right]^{r/2} ,$$

and (see Bai, 2003, p. 160) since $\left\|(nT)^{-1}\sum_{s=1}^{T}\Delta\hat{F}_s u_s'\Lambda\Delta F_t\right\|^2 = O_p\left(n^{-1/2}\delta_{nT}^{-1}\right)$ we have $IV = O_p\left(n^{-r/2}\delta_{nT}^{-r}\right)$.

Thus, we have $T^{-1}\sum_{t=1}^{T}\left\|\Delta\hat{F}_t - \Delta F_t\right\|^r = O_p\left(T^{-r/2}\right) + O_p\left(n^{-r/2}\right) + O_p\left(n^{-r/2}\right) + O_p\left(n^{-r/2}\delta_{nT}^{-r}\right) = O_p\left(\delta_{nT}^{-r}\right)$. Equation (17) can be proved noting that $T^{-1}\sum_{t=1}^{T}\left\|\Delta\hat{F}_t\right\|^r \leq T^{-1}\sum_{t=1}^{T}\left\|\Delta F_t\right\|^r + T^{-1}\sum_{t=1}^{T}\left\|\Delta\hat{F}_t - \Delta F_t\right\|^r = O_p(1) + O_p\left(\delta_{nT}^{-r}\right)$.

Last, consider (18). Since $\hat{u}_{it} = y_{it} - \hat{\lambda}_i'\hat{F}_t$, in light of (1) we have $\hat{u}_{it} - u_{it} = \lambda_i' F_t - \hat{\lambda}_i'\hat{F}_t$, and therefore

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\left|\hat{u}_{it} - u_{it}\right|^r &= \frac{1}{T}\sum_{t=1}^{T}\left|\left(\lambda_i - \hat{\lambda}_i\right)' F_t - \hat{\lambda}_i\left(\hat{F}_t - F_t\right)\right|^r \\
&\leq \frac{1}{T}\sum_{t=1}^{T}\left|\left(\hat{\lambda}_i - \lambda_i\right)' F_t\right|^r + \frac{1}{T}\sum_{t=1}^{T}\left|\hat{\lambda}_i'\left(\hat{F}_t - F_t\right)\right|^r \\
&= I + II.
\end{aligned}$$

Consider I; this can be rewritten as $\left\|\hat{\lambda}_i - \lambda_i\right\|^r T^{-1}\sum_{t=1}^{T}\left\|F_t\right\|^r$. Note that $\hat{\lambda}_i - \lambda_i = O_p\left(T^{-1}\right)$ - see Lemma 3 in Bai (2004, p. 148). Also, Assumptions 1*(i)*, 1*(ii)*, 2*(i)* and 2*(iii)* ensure that $\sum_{t=1}^{T}\left\|F_t\right\|^r = O_p\left(T^{1+\frac{1}{2}r}\right)$ - see Theorem 5.3 in Park and Phillips (1999). Thus, $I = O_p\left(T^{-\frac{1}{2}r}\right)$. As far as II is concerned, $\sum_{t=1}^{T}\left|\hat{\lambda}_i'\left(\hat{F}_t - F_t\right)\right|^r = \left\|\hat{\lambda}_i\right\|^r\sum_{t=1}^{T}\left\|\hat{F}_t - F_t\right\|^r$. Assumption 3*(i)* ensures $\left\|\hat{\lambda}_i\right\|^r = \left\|\lambda_i + o_p(1)\right\|^r = O(1)$, and similar calculations as before (based on the theory developed in Bai, 2004) would lead to $\sum_{t=1}^{T}\left\|\hat{F}_t - F_t\right\|^r = O_p\left(TC_{nT}^{-r}\right)$. Thus, $II = O_p\left(C_{nT}^{-r}\right)$ and therefore $T^{-1}\sum_{t=1}^{T}\left|\hat{u}_{it} - u_{it}\right|^r = O_p\left(\delta_{nT}^{-r}\right)$. Equation (19) follows from similar calculations as those derived for the proof of (17). ■

7 Appendix B: proofs and derivations

Proof of Lemma 1. Consider the $(k+1)$-dimensional vector $e_{it,b}^*$ partitioned as $\left[e_{it,b}^{F*\prime}, e_{it,b}^{u*}\right]'$, where $e_{it,b}^{F*}$ is a k-dimensional vector containing the elements corresponding to $\Delta F_{t,b}^*$ and $e_{it,b}^{u*}$ is the last element; consider also the conformed partitioning $\hat{e}_{qt} = \left[\hat{e}_{qt}^{F\prime}, \hat{e}_{qt}^{u}\right]'$. Since $\left\|e_{it,b}^{*}\right\|^{r} \leq \left\|e_{it,b}^{F*}\right\|^{r} + \left|e_{it,b}^{u*}\right|^{r}$, we shall prove (9) by showing separately

$$T^{1-\frac{1}{2}r} E^{*} \left\|e_{it,b}^{F*}\right\|^{r} \xrightarrow{p^{*}} 0, \tag{20}$$

$$T^{1-\frac{1}{2}r} E^{*} \left|e_{it,b}^{u*}\right|^{r} \xrightarrow{p^{*}} 0. \tag{21}$$

Consider (20). Recalling that $\Delta F_t = \sum_{j=1}^{q} \alpha_{q,j} \Delta F_{t-j} + e_{qt}^{F}$, the following two equations will be used henceforth:

$$\begin{aligned} \Delta \hat{F}_t &= \sum_{j=1}^{q} \hat{\alpha}_{q,j} \Delta \hat{F}_{t-j} + \hat{e}_{qt}^{F}, \\ \Delta F_t &= \sum_{j=1}^{\infty} \alpha_j \Delta F_{t-j} + e_{t}^{F}, \end{aligned}$$

where $\hat{\alpha}_{q,j}$ is the matrix containing the first k rows and columns in the estimate $\hat{\beta}_{q,j}$ derived in step 2.1 of the bootstrapping algorithm. Recalling that $\left\{e_{it,b}^{F*}\right\}_{t=1}^{T} = \left\{\hat{e}_{qt}^{F} - T^{-1} \sum_{t=1}^{T} \hat{e}_{qt}^{F}\right\}_{t=1}^{T}$, it follows that

$$\begin{aligned} E^{*} \left\|e_{it,b}^{F*}\right\|^{r} &= \frac{1}{T} \sum_{t=1}^{T} \left[\hat{e}_{qt}^{F} - \frac{1}{T} \sum_{t=1}^{T} \hat{e}_{qt}^{F}\right]^{r} \\ &\leq \frac{1}{T} \sum_{t=1}^{T} \left\|e_{t}^{F}\right\|^{r} + \frac{1}{T} \sum_{t=1}^{T} \left\|e_{qt}^{F} - e_{t}^{F}\right\|^{r} \\ &\quad + \frac{1}{T} \sum_{t=1}^{T} \left\|\hat{e}_{qt}^{F} - e_{qt}^{F}\right\|^{r} + \left\|\frac{1}{T} \sum_{t=1}^{T} \hat{e}_{qt}^{F}\right\|^{r}. \end{aligned}$$

Assumption 2*(i)* and the LLN ensure that $T^{-1} \sum_{t=1}^{T} \left\|e_{t}^{F}\right\|^{r} \xrightarrow{p} E \left\|e_{t}^{F}\right\|^{r} < \infty$ as $T \to \infty$; thus, $T^{1-\frac{1}{2}r} \left[T^{-1} \sum_{t=1}^{T} \left\|e_{t}^{F}\right\|^{r}\right] = O_p\left(T^{1-\frac{1}{2}r}\right)$.

As far as $T^{-1} \sum_{t=1}^{T} \left\|e_{qt}^{F} - e_{t}^{F}\right\|^{r}$ is concerned, note that $e_{qt}^{F} - e_{t}^{F} = \sum_{j=q+1}^{\infty} \alpha_j \Delta F_{t-j}$

and therefore Minkowski's inequality and the stationarity of ΔF_t lead to

$$\begin{aligned} &\frac{1}{T}\sum_{t=1}^{T}\left\|e_{qt}^{F}-e_{t}^{F}\right\|^{r} \\ = &\frac{1}{T}\sum_{t=1}^{T}\left\|\sum_{j=q+1}^{\infty}\alpha_j\Delta F_{t-j}\right\|^{r}\leq\frac{1}{T}\sum_{t=1}^{T}\left\|\Delta F_t\right\|^{r}\left(\sum_{j=q+1}^{\infty}\left|\alpha_j\right|\right)^{r}. \end{aligned}$$

The term $T^{-1}\sum_{t=1}^{T}\left\|\Delta F_t\right\|^{r}$ is finite in light of Assumption 2*(i)* and the LLN, and Assumption 1*(ii)* ensures that $\sum_{j=q+1}^{\infty}\left|\alpha_j\right| = o\left(q^{-s}\right)$. This entails $T^{1-\frac{1}{2}r}\left[T^{-1}\sum_{t=1}^{T}\left\|e_{qt}^{F}-e_{t}^{F}\right\|^{r}\right] = O_p\left(T^{1-\frac{1}{2}r}q^{-rs}\right)$.

The term $T^{-1}\sum_{t=1}^{T}\left\|\hat{e}_{qt}^{F}-e_{qt}^{F}\right\|^{r}$ can be rewritten as

$$\hat{e}_{qt}^{F}-e_{qt}^{F}=\sum_{j=0}^{q}\alpha_{q,j}\left(\Delta\hat{F}_{t-j}-\Delta F_{t-j}\right)-\sum_{j=1}^{q}\left(\hat{\alpha}_{q,j}-\alpha_{q,j}\right)\Delta\hat{F}_{t-j},$$

where $\alpha_{q,0}=1$. Hence

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T}\left\|\hat{e}_{qt}^{F}-e_{qt}^{F}\right\|^{r} &\leq \frac{1}{T}\sum_{t=1}^{T}\left\|\sum_{j=0}^{q}\alpha_{q,j}\left(\Delta\hat{F}_{t-j}-\Delta F_{t-j}\right)\right\|^{r}+\frac{1}{T}\sum_{t=1}^{T}\left\|\sum_{j=1}^{q}\left(\hat{\alpha}_{q,j}-\alpha_{q,j}\right)\Delta\hat{F}_{t-j}\right\|^{r} \\ &= I+II. \end{aligned}$$

After Minkowsi's inequality we have

$$I\leq\frac{1}{T}\sum_{t=1}^{T}\left\|\Delta\hat{F}_t-\Delta F_t\right\|^{r}\left(\sum_{j=0}^{q}\left|\alpha_{q,j}\right|\right)^{r},$$

and it holds that $\sum_{j=0}^{q}\left|\alpha_{q,j}\right|\leq\sum_{j=0}^{\infty}\left|\alpha_j\right|=O\left(1\right)$. Also, $T^{-1}\sum_{t=1}^{T}\left\|\Delta\hat{F}_t-\Delta F_t\right\|^{r}=O_p\left(\delta_{nT}^{-r}\right)$ according to (16) in Lemma 4. Thus, $I=O_p\left(\delta_{nT}^{-r}\right)$. As far as II is concerned, we have

$$II\leq\frac{1}{T}\sum_{t=1}^{T}\left\|\Delta\hat{F}_t\right\|^{r}\left(\sum_{j=0}^{q}\left|\hat{\alpha}_{q,j}-\alpha_{q,j}\right|\right)^{r}.$$

Equation (17) in Lemma 4 ensures that $T^{-1}\sum_{t=1}^{T}\left\|\Delta\hat{F}_t\right\|^{r}=O_p\left(1\right)$. Also,

$\sum_{j=0}^{q}|\hat{\alpha}_{q,j}-\alpha_{q,j}| \leq q\max_{1\leq j\leq q}|\hat{\alpha}_{q,j}-\alpha_{q,j}|$, and Lemma 2 leads to $[q\max_{1\leq j\leq q}|\hat{\alpha}_{q,j}-\alpha_{q,j}|]^r = O_p\left[q^r T^{-r/2}(\log T)^{r/2} + q^r \delta_{nT}^{-2r}\right]$. Thus, it holds that $T^{1-\frac{1}{2}r}\left[T^{-1}\sum_{t=1}^{T}\left\|\hat{e}_{qt}^F - e_{qt}^F\right\|^r\right] = O_p\left(T^{1-\frac{1}{2}r}\delta_{nT}^{-r}\right) + O_p\left[T^{1-r}q^r(\log T)^{r/2}\right] + O_p\left(T^{1-\frac{1}{2}r}q^r\delta_{nT}^{-2r}\right)$.

Last, consider $\left\|T^{-1}\sum_{t=1}^{T}\hat{e}_{qt}^F\right\|^r$; we have $\hat{e}_{qt}^F = -\sum_{j=0}^{q}\hat{\alpha}_{q,j}\Delta\hat{F}_{t-j}$ with $\hat{\alpha}_{q,0}=-1$. Thus

$$
\begin{aligned}
-\frac{1}{T}\sum_{t=1}^{T}\hat{e}_{qt}^F &= \sum_{j=0}^{q}\alpha_{q,j}\left(\frac{1}{T}\sum_{t=1}^{T}\Delta\hat{F}_{t-j}\right) + \sum_{j=0}^{q}(\hat{\alpha}_{q,j}-\alpha_{q,j})\left(\frac{1}{T}\sum_{t=1}^{T}\Delta\hat{F}_{t-j}\right) \\
&= I + II.
\end{aligned}
$$

Noting that $T^{-1}\sum_{t=1}^{T}\Delta\hat{F}_{t-j} = O_p\left(T^{-1/2}\right)$ for all js, it holds that

$$
\begin{aligned}
I &\leq \left(\sum_{j=0}^{q}|\alpha_{q,j}|^2\right)^{1/2}\left(\sum_{j=0}^{q}\left|\frac{1}{T}\sum_{t=1}^{T}\Delta\hat{F}_{t-j}\right|^2\right)^{1/2} \\
&\leq O(1)\left[q\max_{1\leq j\leq q}\left|\frac{1}{T}\sum_{t=1}^{T}\Delta\hat{F}_{t-j}\right|^2\right]^{1/2} = O_p\left(\sqrt{\frac{q}{T}}\right),
\end{aligned}
$$

and also

$$
\begin{aligned}
II &\leq \left(\sum_{j=0}^{q}|\hat{\alpha}_{q,j}-\alpha_{q,j}|^2\right)^{1/2}\left(\sum_{j=0}^{q}\left|\frac{1}{T}\sum_{t=1}^{T}\Delta\hat{F}_{t-j}\right|^2\right)^{1/2} \\
&\leq \left(q\max_{1\leq j\leq q}|\hat{\alpha}_{q,j}-\alpha_{q,j}|^2\right)^{1/2} O_p\left(\sqrt{\frac{q}{T}}\right).
\end{aligned}
$$

Since, in light of Lemma 2, $\max_{1\leq j\leq q}|\hat{\alpha}_{q,j}-\alpha_{q,j}|^2 = O_p\left(\varphi_{nT}^{-2}\right)$, we have $\left\|T^{-1}\sum_{t=1}^{T}\hat{e}_{qt}^F\right\|^r = O_p\left(q^{r/2}T^{-r/2}\right) = o_p(1)$.

Combining all these results together, it follows that

$$
\begin{aligned}
E^*\left\|e_{it,b}^{F*}\right\|^r &= E\left\|e_t^F\right\|^r + O_p\left(q^{-rs}\right) + O_p\left(q^r\delta_{nT}^{-2r}\right) + O_p\left[q^r T^{-r/2}(\log T)^{r/2}\right] + O_p\left(q^{r/2}T^{-r/2}\right) \\
&= E\left\|e_t^F\right\|^r + o_p(1),
\end{aligned}
$$

which proves (10), and hence also

$$\begin{aligned}
T^{1-\frac{1}{2}r}E^{*}\left\|e_{it,b}^{F*}\right\|^{r} &= O_{p}\left(T^{1-\frac{1}{2}r}\right)+O_{p}\left(T^{1-\frac{1}{2}r}q^{-rs}\right)+O_{p}\left(T^{1-\frac{1}{2}r}q^{r}\varphi_{nT}^{-r}\right) \\
&\quad +O_{p}\left(T^{1-r}q^{\frac{1}{2}r}\right)+o_{p}\left(1\right) \\
&= O_{p}\left(T^{1-\frac{1}{2}r}\right)+O_{p}\left(T^{1-\frac{1}{2}r}q^{r}\varphi_{nT}^{-r}\right)+o_{p}\left(1\right).
\end{aligned}$$

Then $T^{1-\frac{1}{2}r}E^{*}\left\|e_{it,b}^{F*}\right\|^{r}=o_{p}\left(1\right)$ for any $r>2$.

As far as (21) is concerned, recall that $u_{it}=\sum_{j=1}^{q}\gamma_{q,j}^{(i)}u_{it-j}+e_{qt}^{u(i)}$, consider the notation

$$\begin{aligned}
\hat{u}_{it} &= \sum_{j=1}^{q}\hat{\gamma}_{q,j}^{(i)}\hat{u}_{it-j}+\hat{e}_{qt}^{u(i)}, \\
u_{it} &= \sum_{j=1}^{\infty}\gamma_{j}^{(i)}u_{it-j}+e_{t}^{u(i)},
\end{aligned}$$

where $\hat{\gamma}_{q,j}^{(i)}$ is the element in position $(k+1,k+1)$ in the matrix $\hat{\beta}_{q,j}$ derived in step 2.1 of the bootstrapping algorithm. Suppressing the dependence on i, Then we can write

$$\begin{aligned}
E^{*}\left|e_{it,b}^{u*}\right|^{r} &= \frac{1}{T}\sum_{t=1}^{T}\left[\hat{e}_{qt}^{u}-\frac{1}{T}\sum_{t=1}^{T}\hat{e}_{qt}^{u}\right]^{r} \\
&\leq \frac{1}{T}\sum_{t=1}^{T}\left|e_{t}^{u}\right|^{r}+\frac{1}{T}\sum_{t=1}^{T}\left|e_{qt}^{u}-e_{t}^{u}\right|^{r} \\
&\quad +\frac{1}{T}\sum_{t=1}^{T}\left|\hat{e}_{qt}^{u}-e_{qt}^{u}\right|^{r}+\left|\frac{1}{T}\sum_{t=1}^{T}\hat{e}_{qt}^{u}\right|^{r}.
\end{aligned}$$

Using Assumption 2*(i)* and similar arguments as in Park (2002), it can be shown that $T^{-1}\sum_{t=1}^{T}\left|e_{t}^{u}\right|^{r}=O_{p}\left(1\right)$ and $T^{-1}\sum_{t=1}^{T}\left|e_{qt}^{u}-e_{t}^{u}\right|^{r}=O_{p}\left(q^{-rs}\right)$. Note that

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\left|\hat{e}_{qt}^{u}-e_{qt}^{u}\right|^{r} &\leq \frac{1}{T}\sum_{t=1}^{T}\left|\sum_{j=0}^{q}\gamma_{q,j}^{(i)}\left(\hat{u}_{it-j}-u_{it-j}\right)\right|^{r}+\frac{1}{T}\sum_{t=1}^{T}\left|\sum_{j=1}^{q}\left(\hat{\gamma}_{q,j}^{(i)}-\gamma_{q,j}^{(i)}\right)\hat{u}_{it-j}\right|^{r} \\
&= I+II,
\end{aligned}$$

with $\beta^u_{q,0} = 1$. Then it holds that

$$I \leq \frac{1}{T} \sum_{t=1}^{T} \left| \hat{u}_{it} - u_{it} \right|^r \left(\sum_{j=0}^{q} \left| \gamma^{(i)}_{q,j} \right| \right)^r,$$

and (18) in Lemma 4 entails $I = O_p\left(q^r \delta^{-r}_{nT}\right)$. Also,

$$II \leq \frac{1}{T} \sum_{t=1}^{T} \left| \hat{u}_{it} \right|^r \left(\sum_{j=0}^{q} \left| \hat{\gamma}^{(i)}_{q,j} - \gamma^{(i)}_{q,j} \right| \right)^r.$$

Equation (19) in Lemma 4 ensures that $T^{-1} \sum_{t=1}^{T} \left| \hat{u}_{it} \right|^r = O_p(1)$. Also, $\sum_{j=0}^{q} \left| \hat{\gamma}^{(i)}_{q,j} - \gamma^{(i)}_{q,j} \right| \leq q \max_{1 \leq j \leq q} \left| \hat{\gamma}^{(i)}_{q,j} - \gamma^{(i)}_{q,j} \right|$, and Lemma 2 leads to $\left[q \max_{1 \leq j \leq q} \left\| \hat{\gamma}^{(i)}_{q,j} - \gamma^{(i)}_{q,j} \right\| \right]^r$ $= O_p\left(q^r \varphi^{-r}_{nT}\right)$. Last, it can be shown straightforwardly that

$$\left| \frac{1}{T} \sum_{t=1}^{T} \hat{e}^u_{qt} \right|^r \leq \frac{1}{T} \sum_{t=1}^{T} \left| \hat{u}_{it} \right|^r \left(\sum_{j=0}^{q} \left| \hat{\gamma}^{(i)}_{q,j} \right| \right)^r = O_p\left[\left(\sqrt{\frac{q}{T}} \right)^r \right],$$

in light of (19). Thus

$$E^* \left| e^{u*}_{it,b} \right|^r = E \left| e^u_{it} \right|^r + O_p\left(q^{-rs}\right) + O_p\left(q^r \delta^{-r}_{nT}\right) + O_p\left(q^r \varphi^{-r}_{nT}\right) + o_p(1),$$

which proves (10). Equation (21) also follows. ■

Proof of Lemma 2. Note first that $\max_{1 \leq j \leq q} \left\| \hat{\beta}_{q,j} - \beta_j \right\| \leq \max_{1 \leq j \leq q} \left\| \hat{\beta}_{q,j} - \beta_{q,j} \right\| +$ $\max_{1 \leq j \leq q} \left\| \beta_{q,j} - \beta_j \right\|$, and note that, in light of Assumptions 1(ii) and 2(iii) we have $\max_{1 \leq j \leq q} \left\| \beta_{q,j} - \beta_j \right\| \leq \sum_{j=1}^{q} \left\| \beta_{q,j} - \beta_j \right\| = o\left(q^{-s}\right)$ - see e.g. Theorem 2.1 in Hannan and Kavalieris (1986).

In order to find a bound for $\max_{1 \leq j \leq q} \left\| \hat{\beta}_{q,j} - \beta_{q,j} \right\|$, assume first, for the sake of the notation and without loss of generality, that $k = 1$, so that ΔF_t and related quantities are scalars. Then the $\alpha_{q,j}$'s are scalars as well as the $\gamma^{(i)}_{q,j}$'s; letting $\beta_{q,j} = \left[\alpha_{q,j}, \gamma^{(i)}_{q,j} \right]'$, it can be shown that $\max_{1 \leq j \leq q} \left| \hat{\alpha}_{q,j} - \alpha_{q,j} \right| =$ $O_p\left(\sqrt{\log T / T} \right) + O_p\left(\delta^{-2}_{nT} \right)$ and $\max_{1 \leq j \leq q} \left| \hat{\gamma}^{(i)}_{q,j} - \gamma^{(i)}_{q,j} \right| = O_p\left(\sqrt{\log T / T} \right) +$ $O_p\left(\delta^{-2}_{nT} \right)$. Consider first $\hat{\alpha}_{q,j}$. Letting $\alpha_q = \left[\alpha_{q,1}, ..., \alpha_{q,q} \right]'$, OLS regression of

$\Delta\hat{F}_t$ against the vector $\Delta\hat{F}_{q,t} = \left[\Delta\hat{F}_{t-1}, ..., \Delta\hat{F}_{t-q}\right]'$ leads to

$$\hat{\alpha}_q = \left[\sum_{t=q+1}^{T} \Delta\hat{F}_{q,t}\Delta\hat{F}'_{q,t}\right]^{-1} \left[\sum_{t=q+1}^{T} \Delta\hat{F}_{q,t}\Delta\hat{F}_t\right].$$

Consider $\sum_{t=q+1}^{T} \Delta\hat{F}_{q,t}\Delta\hat{F}'_{q,t}$; application of Lemma A.1 and Lemma B.2 in Bai (2003) entails $T^{-1}\sum_{t=q+1}^{T} \Delta\hat{F}_{q,t}\Delta\hat{F}'_{q,t} = T^{-1}\sum_{t=q+1}^{T} \Delta F_{q,t}\Delta F'_{q,t} + O_p\left(\delta_{nT}^{-2}\right)$. Letting $\sum_{t=q+1}^{T} \Delta F_{q,t}\Delta F'_{q,t} = d = O_p(T)$, after some algebra we have

$$\begin{aligned}
\hat{\alpha}_q - \alpha_q &= d^{-1}\left\{\sum_{t=q+1}^{T} \Delta F_{q,t} e_{qt}^{F} + \sum_{t=q+1}^{T} \Delta F_{q,t}\left(\Delta\hat{F}_t - \Delta F_t\right)\right. \\
&\quad \left. + \sum_{t=q+1}^{T}\left(\Delta\hat{F}_{q,t} - \Delta F_{q,t}\right)\Delta F_t + \sum_{t=q+1}^{T}\left(\Delta\hat{F}_{q,t} - \Delta F_{q,t}\right)\left(\Delta\hat{F}_t - \Delta F_t\right)\right\} \\
&= I + II + III + IV,
\end{aligned}$$

and therefore

$$\max_{1\leq j\leq q}|\hat{\alpha}_{q,j} - \alpha_{q,j}| \leq \max_{1\leq j\leq q}|I| + \max_{1\leq j\leq q}|II| + \max_{1\leq j\leq q}|III| + \max_{1\leq j\leq q}|IV|.$$

From Theorem 2.1 in Hannan and Kavalieris (1986) we know that $\max_{1\leq j\leq q}|I| = O_p\left(\sqrt{\log T/T}\right)$. Also, using Lemma B.2 in Bai (2003) it can be proved that $II = O_p\left(\delta_{nT}^{-2}\right)$ and $III = O_p\left(\delta_{nT}^{-2}\right)$, and Lemma A.1 in Bai (2003) entails $IV = O_p\left(\delta_{nT}^{-2}\right)$; note that these results hold for all q, and thus $\max_{1\leq j\leq q}|a| = O_p\left(\delta_{nT}^{-2}\right)$ for $a = II$, III and IV.

The proof for $\hat{\gamma}_{q,j}^{(i)}$ follows similar lines. Defining $\gamma_q = \left[\gamma_{q,1}, ..., \gamma_{q,q}\right]'$ (and suppressing the dependence on i for the sake of notation) and $\hat{u}_{it,q} = \left[\hat{u}_{it-1,q}, ..., \hat{u}_{it-q,q}\right]'$ we have $\hat{\gamma}_q = \left[\sum_{t=q+1}^{T}\hat{u}_{it,q}\hat{u}'_{it,q}\right]^{-1}\left[\sum_{t=q+1}^{T}\hat{u}_{it,q}\hat{u}_{it}\right]$. Consider $\sum_{t=q+1}^{T}\hat{u}_{it,q}\hat{u}'_{it,q} = \sum_{t=q+1}^{T}u_{it,q}u'_{it,q} + \lambda_i\sum_{t=q+1}^{T}\hat{u}_{it,q}\left(\hat{F}_{t,q} - F_{t,q}\right)' + \lambda_i\sum_{t=q+1}^{T}\left(\hat{F}_{t,q} - F_{t,q}\right)\hat{u}'_{it}$ $+\lambda_i^2\sum_{t=q+1}^{T}\left(\hat{F}_{t,q} - F_{t,q}\right)\left(\hat{F}_{t,q} - F_{t,q}\right)'$. Lemma A.1 in Bai (2004) ensures that $\sum_{t=q+1}^{T}\left(\hat{F}_{t,q}-F_{t,q}\right)\left(\hat{F}_{t,q}-F_{t,q}\right)=O_p\left(TC_{nT}^{-2}\right)$. Also,

$$\begin{aligned}\left|\sum_{t=q+1}^{T}\left(\hat{F}_{t,q}-F_{t,q}\right)\hat{u}_{it,q}'\right| &\leq \left(\sum_{t=q+1}^{T}\left\|\hat{F}_{t,q}-F_{t,q}\right\|^2\right)^{1/2}\left(\sum_{t=q+1}^{T}\left\|\hat{u}_{it,q}\right\|^2\right)^{1/2} \\ &= O_p\left(\sqrt{T}C_{nT}^{-1}\right)O_p\left(\sqrt{T}\right)=O_p\left(TC_{nT}^{-1}\right). \qquad (22)\end{aligned}$$

Thus, $T^{-1}\sum_{t=q+1}^{T}\hat{u}_{it,q}\hat{u}_{it,q}'=T^{-1}\sum_{t=q+1}^{T}u_{it,q}u_{it,q}'+o_p(1)$. As far as $\sum_{t=q+1}^{T}\hat{u}_{it,q}\hat{u}_{it}$ is concerned, note that $\sum_{t=q+1}^{T}\hat{u}_{it,q}\hat{u}_{it}=\sum_{t=q+1}^{T}u_{it,q}u_{it}+\lambda_i\sum_{t=q+1}^{T}u_{it}\left(F_{t,q}-\hat{F}_{t,q}\right)'$ $+\lambda_i\sum_{t=q+1}^{T}\left(F_t-\hat{F}_t\right)\hat{u}_{it,q}+\lambda_i^2\sum_{t=q+1}^{T}\left(\hat{F}_{t,q}-F_{t,q}\right)'\left(\hat{F}_t-F_t\right)$. Similar arguments as for the proof of the consistency of $\hat{\alpha}_{q,j}$, and (22) lead to $\max_{1\leq j\leq q}\left|\hat{\gamma}_{q,j}-\gamma_{q,j}\right|=O_p\left(\sqrt{\log T/T}\right)+O_p\left(C_{nT}^{-1}\right)$. ■

Proof of Lemma 3. Assumptions 1 and 2 ensure that $V(r)=\beta^{-1}(1)W(r)$, with $\beta(1)=1-\sum_{j=1}^{\infty}\beta_j$ and $T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}e_{it}\xrightarrow{d}W(r)$. Consider (12) and note that $\sum_{j=1}^{q}\hat{\beta}_{q,j}=\sum_{j=1}^{\infty}\beta_j-\sum_{j=q+1}^{\infty}\beta_j+\sum_{j=1}^{q}\left(\hat{\beta}_{q,j}-\beta_{q,j}\right)$. After Assumption 1*(ii)* and 2*(iii)* it holds that $\sum_{j=q+1}^{\infty}\beta_j=o\left(q^{-s}\right)$, and after Lemma 2 we have $\sum_{j=1}^{q}\left(\hat{\beta}_{q,j}-\beta_{q,j}\right)\leq q\max_{1\leq j\leq q}\left|\hat{\beta}_{q,j}-\beta_{q,j}\right|=O_p\left(q\varphi_{nT}^{-1}\right)$, which is negligible under Assumption 5. Thus, $\hat{\beta}_q^{-1}(1)\xrightarrow{p}\beta^{-1}(1)$. The bootstrap invariance principle in Lemma 1 ensures that $T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}e_{it,b}^{*}\xrightarrow{d}W(r)$. Also, following the same lines as Park (2002, proof of Theorem 3.3), we may show that $T^{-1/2}\sup_{1\leq t\leq T}\left|\bar{\xi}_{it}^{q*}\right|=o_p(1)$. Therefore the CMT entails

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}\xi_{it}^{q*} &= \hat{\beta}_q^{-1}(1)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}e_{it,b}^{*}\right)+o_p(1) \\ &\xrightarrow{d^*}\beta^{-1}(1)\,W(r).\end{aligned}$$

■

References

[1] Bai, J. (2003), "Inferential Theory for Structural Models of Large Dimensions", *Econometrica,* 71, 135-171.

[2] Bai, J. (2004), "Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data", *Journal of Econometrics*, 122, 137-183.

[3] Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models", *Econometrica*, 70, 191-221.

[4] Bai, J., and Ng, S. (2004), "A Panic Attack on Unit Roots and Cointegration", *Econometrica,* 72, 1127-1178.

[5] Bühlmann, P. (1997), "Sieve Bootstrap for Time Series", *Bernoulli*, 3, 123-148.

[6] Chamberlain, G., and Rothschild, M. (1983), "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica*, 51, 1305-1324.

[7] Chang, Y., Park, J.Y., Song, K. (2006), "Bootstrapping Cointegrating Regressions", *Journal of Econometrics*, 133, 703-739.

[8] Cragg, J., Donald, S. (1997), "Inferring the Rank of a Matrix", *Journal of Econometrics*, 76, 223-250.

[9] Donald, S.G. (1997), "Inference Concerning the Number of Factors in a Multivariate Nonparametric Relationship", *Econometrica*, 65(1), 103-131.

[10] Hannan, E.T., L. Kavalieris, L. (1986), "Regression; Autoregression Models", *Journal of Time Series Analysis*, 7, 27-49.

[11] Horowitz

[12] Johansen, S. (1991), "Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregression models", *Econometrica*, 59, 1551-1580.

[13] Kao, C., Trapani, L., Urga, G. (2007a), "Asymptotics for Panel Models with Common Shicks", mimeo.

[14] Kao, C., Trapani, L., Urga, G. (2007b), "Modelling and Testing for Structural Breaks in Panel Cointegration Models with Common and Idiosyncratic Stochastic Trends", mimeo.

[15] Lee, R.D., Carter L. (1992), "Modeling and Forecasting the Time Series of U.S. Mortality", *Journal of the American Statistical Association*, 87, 659-71

[16] Lewbel, A. (1991), "The Rank of Demand Systems: Theory and Nonparametric Estimation", *Econometrica*, 59(3), 711-730.

[17] Onatski, A. (2006), "Determining the Number of Factors form Empirical Distribution of Eigenvalues", unpublished manuscript.

[18] Park, J.Y. (2002), "An Invariance Principle for Sieve Bootstrap in Time Series", *Econometric Theory,* 18, 469–490.

[19] Park, J.Y., Phillips, P. C. B. (1999), "Asymptotics for Nonlinear Transformations of Integrated Time Series", *Econometric Theory,* 15, 269, 298.

[20] Phillips, P. C. B., Moon, H. R. (1999), "Linear Regression Limit Theory for Nonstationary Panel Data", *Econometrica*, 67, 1057-1112.