Inference on Factor Structures in Heterogeneous Panels

Carolina Castagnetti  Eduardo Rossi
University of Pavia  University of Pavia

Lorenzo Trapani
Cass Business School, City University London

October 4, 2017

Abstract

This paper develops an estimation and testing framework for a stationary large panel model with observable regressors and unobservable common factors. We allow for slope heterogeneity and for correlation between the common factors and the regressors. We propose a two stage estimation procedure for the unobservable common factors and their loadings, based on Common Correlated Effects estimator and the Principal Component estimator. We also develop two tests for the null of no factor structure: one for the null that loadings are cross sectionally homogeneous, and one for the null that common factors are homogeneous over time. Our tests are based on using extremes of the estimated loadings and common factors. The test statistics have an asymptotic Gumbel distribution under the null, and have power versus alternatives where only one loading or common factor differs from the others. Monte Carlo evidence shows that the tests have the correct size and good power.

JEL codes: C12, C33.

Keywords: Large Panels, CCE Estimator, Principal Component Estimator, Testing for Factor Structure, Extreme Value Distribution.
1 Introduction

Consider the following model for stationary panel data:

\[ y_{it} = \beta_{it} x_{it} + \delta_{it} + \epsilon_{it}; \quad (1) \]
\[ x_{it} = \beta_{i}^T f_t + \eta_{it}; \quad (2) \]

where \( i = 1; \ldots; n, t = 1; \ldots; T \), \( x_{it} \) is an \( m \)-dimensional vector of observable explanatory variables and \( f_t \) is an \( r \)-dimensional vector of unobservable common factors; in equation (2), \( \beta_{i}^T \) is a matrix of coefficients of dimension \( m \times r \). Model (1)-(2) is based on Pesaran (30), and it arguably has a huge potential for empirical applications. In the context of finance, \( y_{it} \) could represent the excess return on an asset; then, as pointed out by Bai (3), \( f_t \) could represent a vector of unobservable factor returns, which are added to the observable ones (e.g. the Book-to-Market ratio) that are typically employed. Kapetanios and Pesaran (26) consider an APT model allowing for individual asset returns to be affected by common factors (both observable and unobservable). In a similar setup, Castagnetti and Rossi (11) adopt a heterogeneous panel with a multifactor error model to study the determinants of credit spread changes in the Euro corporate bond market. Factor models are also useful in the context of estimating production functions, where \( x_{it} \) is a set of observable factor inputs, and \( f_t \) allows to consider cross sectional dependence as arising from common shocks or e.g. spillover effects determined by policy or technology shocks. For example, Eberhardt and Teal (17) adopt a common factor model approach to estimate cross-country production functions for the agriculture sector. Similarly, Eberhardt, Helmers and Strauss (16) consider the impact of spillovers in the estimation of private returns to R&D allowing for a common factor framework. Another promising field of application is the prediction of mortality rates (or their first difference), where the seminal Lee-Carter model (28) has been extended to incorporate idiosyncratic explanatory variables as well as the traditional factor structure - see French and O'Hare (21) and the references therein.

As far as conducting inference on (1) is concerned, the inferential theory on the slope coefficients is developed in various contributions. Particularly, Pesaran (30) proposes a family of estimators for \( \beta_{i}^T \) based on instrumenting the \( f_t \)'s through cross sectional averages of the \( x_{it} \) and \( y_{it} \); such estimation techniques are referred to as the Common Correlated Effects (CCE) estimators. One of the key features of the CCE estimator is that it does not require any inference to be carried out on \( \beta_{i}^T \) or \( f_t \). Pesaran and Tosetti (31) and Castagnetti and Rossi (11) show that, in principle, residuals computed from (1) using CCE estimators can be used to extract \( \beta_{i}^T \) and \( f_t \) using e.g. Principal Components (henceforth, PC). However, the properties of the estimated \( \beta_{i}^T \) and \( f_t \) are not discussed. In addition to the CCE estimators,
Bai (3) develops a different estimation technique for (1)-(2) under the assumption of homogeneous slopes, i.e. $i_i = \bar{i}$. Such technique is known as the Interactive Effect (henceforth IE) estimator, and it is based on iteratively computing $i_i$ for given values of $i$ and $f_t$, and then $i$ and $f_t$ for a given value of $i$. Although results are available for the estimated triple $(i; i_i; f_t)$, inference is developed under the assumption of homogeneous $i_i$s; moreover, no explicit asymptotics for $i$ or $f_t$ is derived beyond consistency. Despite this, inference on $i_i$ and $f_t$ is likely to be important in many settings. For instance, where a multifactor error structure is employed for the purpose of dimension reduction, or simply when explanatory variables may not be observable. In such cases, it could be relevant to know whether there is indeed a factor structure in (1), or whether common effects can be adequately represented by more parsimonious models such as a model with cross-sectional or time dummies, as also studied by Saraõdis, Yamagata and Robertson (32), and Bai (3) in the context of model (1) with homogeneous slopes. In this case, the asymptotics of the estimated common factors and loadings is obviously a first, fundamental step in order to construct tests for the presence of a multifactor error structure.

This paper makes two contributions to the literature. Firstly, we derive the inferential theory for the unobservable common factors $f_t$ and their coefficients $i_i$ in (1)-(2). We estimate $i_i$ and $f_t$ by applying PC to the residuals computed from (1) using the CCE estimator. This two-stage procedure builds on an idea of Pesaran (30, p.1000), and Pesaran and Tosetti (31), while the asymptotics of the estimated $(i_i; f_t)$ is studied by adapting the method of proof in Bai (3) to the case of heterogeneous $i_i$s.

Secondly, we develop two tests: one for the null that $i_i = \bar{i}$ for all $i$, and one for the null that $f_t = \bar{f}$ for all $t$. The rationale for these two tests can be understood by noting that, as Pesaran (30) points out, model (1)-(2) nests various alternative specifications. In the case of homogeneous loadings (i.e. $i_i = \bar{i}$), equation (1) is tantamount to a panel regression with a time effect - therefore there is no real common factor structure. This fact is used by Saraõdis, Yamagata and Robertson (32) to test for cross dependence in a dynamic panel context. Similarly, in the case of homogeneous factors (i.e. $f_t = \bar{f}$), equation (1) boils down to a heterogeneous panel with individual effects - in this case, too, there is no real common factor structure. Therefore, the two tests described above can be used to verify whether a factor structure in (1)-(2) indeed exists, or whether simpler specifications nested in (1)-(2) should be employed. Both tests should therefore be employed before trying to estimate any factor structure, including the number of common factors, as we also discuss in Section 3. In this respect, our paper is related to a recent contribution by Baltagi, Kao, and Na (5), who propose an approach based on Õnite sample corrections and wild bootstrap to testing for $H_0 : i_i = 0$ in a standard panel factor model defined as $y_t = \beta f_t + \epsilon_t$. We use test statistics based on extrema of the estimated $i_i$ and $f_t$, in a similar fashion to the tests for
The paper is organized as follows. The estimation procedure, and the asymptotics of the estimates of \( \dot{i} \) and \( f_t \) are in Section 2; Section 3 contains results about the two tests mentioned above. Section 4 contains a validation of our theory through synthetic data. Section 5 concludes. All proofs are provided either in the Appendix or in Castagnetti, Rossi and Trapani (2014).

**NOTATION.** We use \( \lim \) to denote the ordinary limit; \( \xrightarrow{d} \) and \( \xrightarrow{p} \) to denote convergence in distribution and in probability respectively; and we use \( \text{i.a.s.} \) as short-hand for \( \text{almost surely} \). The Frobenius norm of a matrix \( A \) is denoted as \( \|A\|_F = \sqrt{\text{tr}(A^T A)} \), where \( \text{tr}(A) \) denotes the trace of \( A \). Definitional equality is denoted as \( \overset{\triangle}{=} \). Other notation is defined throughout the paper and in Appendix.

### 2 Estimation

In model (1)-(2), where \( x_{it} \) is \( m \)-dimensional and \( f_t \) is \( r \)-dimensional, we consider the following notation, which we use throughout the whole paper. We define \( F = (f_1; \ldots; f_T)^0 \); \( X_i = (x_{i1}; \ldots; x_{iT})^0 \); \( \bar{\varnothing} = (\bar{\varnothing}_1; \ldots; \bar{\varnothing}_T)^0 \); \( y_i = (y_{i1}; \ldots; y_{iT})^0 \); \( z_t = (y_{it}; x_{0it})^0 \); \( z = (z_1; \ldots; z_T)^0 \) and \( \overline{H}_w = n^{-1} \sum_{i=1}^n z_i \). We also define the matrices \( \text{M}_w = I_T \); \( \text{I}_w \); \( \overline{H}_w \); \( \text{H}_w \); \( C_i = \left[ \begin{array}{ccc} 1 & 0 & 0 \\ \bar{\varnothing}_i & I_m \\ 0 \\ \end{array} \right] \).

For each \( i \), based on this, the \( \bar{\varnothing}_i \) in (1) can be estimated as

\[
\bar{\varnothing}_i = \frac{X_i \text{M}_w X_i^0}{T + 1} \frac{X_i \text{M}_w y_i}{T + 1};
\]

which is the CCE estimator of Pesaran (30); it holds that \( \bar{\varnothing}_i \overset{\text{i.a.s.}}{\rightarrow} O_p \frac{1}{T} + O_p \frac{1}{nT} + O_p \frac{1}{n} \).

In order to estimate \( \dot{i} \) and \( f_t \), we propose the following two-step procedure.

**Step 1** Estimate the \( \bar{\varnothing}_i \) using the CCE estimator, and compute the residuals \( \varnothing_i = y_i - X_i \bar{\varnothing}_i \).

**Step 2** Apply the PC estimator to \( \varnothing_i \), obtaining \( \hat{\gamma} \) and \( \hat{f}_t \) under the restrictions \( \text{F} \gamma = \text{T} \) and \( n^{-1} \sum_{i=1}^n \hat{\gamma} \) diagonal.
In Step 2, \(F^*\) is calculated as \(P_T\) times the \(r\) largest eigenvectors of \(\frac{1}{PT} \sum_{i=1}^n X_iY_i\). Similarly, \(\hat{\gamma}\) is computed as

\[
\hat{\gamma} = \frac{1}{PT} \sum_{i=1}^n X_iY_i \quad (4)
\]

with \(M_{X_i} = \Pi_T X_i (X_i'X_i)^{-1} X_i'\). In (1), \(\hat{\gamma}\) and \(f_t\) are not separately identifiable; as is typical in this literature, we only manage to estimate a rotation of \(\hat{\gamma}\) and \(f_t\), say \(H\hat{\gamma}\) and \(Hf_t\). However, for our purposes knowing \(H\hat{\gamma}\) and \(Hf_t\) is as good as knowing \(\hat{\gamma}\) and \(f_t\). We point out that the results in this paper do not strictly require the CCE estimator in Step 1: our results keep holding as long as the \(\hat{\gamma}\)'s are estimated at a rate \(\sqrt{\frac{\min T}{n}}\). Thus, the CCE is only a possible choice. Alternatives, like the Song (33) estimator, which extends Bai (3) IE estimator to the case of heterogeneous slopes, may be used instead. The Song (33) estimator obtains the same rate of convergence as for the CCE estimates of the individual slopes. In the remainder of the paper, we show our results based on employing the CCE in Step 1.

Consider the following assumptions.

Assumption 1. [error terms: serial and cross sectional dependence] (i) \(E(f_t) = 0\) and \(E f_t f_t' \sim \mathcal{N} \sim 1\); (ii) \(E_p f_t^T E (\Pi_t\Pi_t) f_t \sim M\) for all \(i\) and \(s\), (b) \(P_n \sum_{i=1}^n \Pi_t E (\Pi_t\Pi_t) f_t \sim M_n\) for all \(i\) and \(s\), (c) \(P_T \sum_{i=1}^n \Pi_t E (\Pi_t\Pi_t) f_t \sim M_T\) for all \(i\) and \(s\), (d) \(\sum_{i=1}^n \Pi_t E (\Pi_t\Pi_t) f_t \sim M\) for all \(u\), (e) \(\Pi_t E (\Pi_t\Pi_t) f_t \sim M\) for all \(t\), \(r < 12\), (f) \(\Pi_t E (\Pi_t\Pi_t) f_t \sim M\) for all \(t\), \(r < 12\).

Assumption 2. [regressors and common factors] (i) \(E k_t k_t' \sim \mathcal{N} k_t k_t' \sim \mathcal{N} \sim 1\); (ii) \(T^{1/2} f_t f_t^T \sim \mathcal{N} f_t f_t^T \sim \mathcal{N} \sim 1\) with \(f_t\) non-singular; (iii) \(f_t^T f_t g\) are mutually independent for all \(i\), \(j\), \(t\), \(s\); (iv) \(E f_t^T X t \sim \mathcal{N} f_t^T X t \sim \mathcal{N}\) for all \(i\), \(j\), \(t\), \(s\).

Assumption 3. [slopes and loadings] (i) \(f_t g\) is independent of \(t\), \(j\); \(f_t\) for all \(i\), \(j\), \(t\); (ii) \(E k_t k_t^T \sim \mathcal{N} k_t k_t^T \sim \mathcal{N} \sim 1\) for some \(\mathcal{N}\); (iii) \(\hat{\gamma}\) is non stochastic and such that \(\max_i k_i k_i' < 1\) and \(n^{1/2} i=1 \hat{\gamma}_i \sim \mathcal{N} \sim 0\) with \(\mathcal{N}\) non-singular.

Assumption 4. [Step 1 estimation] (i) \(\lim_{n \to \infty} \frac{1}{n^{1/2}} \sum_{i=1}^n x_t x_t' \sim \mathcal{N} x_t x_t' \sim \mathcal{N} \sim 0\) and \(\lim_{n \to \infty} \frac{1}{n^{1/2}} \sum_{i=1}^n x_t x_t' \sim \mathcal{N} x_t x_t' \sim \mathcal{N} \sim 0\) a.s. for all \(i\), where \(\lim_{n \to \infty} (\cdot)\) denotes the smallest eigenvalue; (ii) \(C_n \sim 1\) \(P_{i=1}^n C_1\) has rank \(r\) \(\sim m + 1\).

Assumption 5. [Central Limit Theorems] (i) \(a\) there exists a nonrandom, positive definite matrix \(f_M\) such that \(\lim_{n \to \infty} \sum_{i=1}^n f_M x_i H F = \sum_{i=1}^n f_M x_i H F\) for all \(i\); (b) \(T^{1/2} f t f_t^T \sim \mathcal{N} f t f_t^T \sim \mathcal{N} \sim 0\) \(a.s.\), where \(f f_M x_i H F\) for all \(i\); (ii) \(n^{1/2} i=1 \hat{\gamma}_i \sim \mathcal{N} x_t x_t' \sim \mathcal{N} \sim 0\) \(a.s.\), where \(\mathcal{N}\).
Broadly speaking, Assumptions 1-4 are needed to prove the consistency of the estimated common factors and loadings. Assumption 4 is specific to the CCE estimator, employed in Step 1. Assumption 5 is required when deriving the asymptotic distributions.

In particular, Assumption 1 deals with the error term $\epsilon_t$, and it allows for serial and cross dependence. The conditions in parts (ii) and (iii) of the assumption resemble closely (and in some cases are exactly the same as) those in Bai (2) and Bai (3), and can be shown immediately if $\epsilon_t$ is assumed to be independent. Part (i) requires the existence of the 12-th moment of $\epsilon_t$, which is stronger than what the literature normally considers - e.g. in Bai (3), assuming $E|\epsilon_t|^8 < 1$ suffices. In our context, the existence of the 12-th moment is needed in order to derive consistency of $\hat{\gamma}$ and $\hat{f}_t$ (see in particular the proof of Lemma A.1). Finally, part (iv) contains Burkholder-type inequalities: these could be shown directly under more specific assumptions on the degree of serial and cross sectional dependence. For example, part (a) holds immediately if one assumes that $\epsilon_t$ is a Martingale Difference Sequence (MDS) across $t$ (the same holds for part (b), under the MDS assumption across $i$) - see e.g. Lin and Bai (29, p.108).

As far as Assumption 2 is concerned, we allow for serial and cross sectional dependence in both the $x_{it}$s and in the common factors $f_t$. The requirement in part (ii) is standard in the literature (see e.g. Assumption B in 3), and it entails that common factors are strong in the sense of Chudik, Pesaran and Tosetti (13) (see in particular Assumption 3). Finally, according to part (iii), the $x_{it}$s are strictly exogenous. Assumption 3 is standard. Assumption 4 is specific to the CCE estimator of the $\epsilon_t$s, employed in Step 1. Particularly, the rank condition in part (ii) is the same as equation (21) in Pesaran (30), and it guarantees the consistency of the $\epsilon_t$s.

Finally, Assumption 5 contains two CLT-type results which are employed when deriving the limiting distributions of the estimated common factors and loadings: parts (i) and (ii) can be compared with Assumption F in Bai (2).

We now turn to studying the asymptotics of $\hat{\gamma}$ and $\hat{f}_t$.

**Theorem 1** Let Assumptions 1-4 hold; then, for every $i$

$$\hat{\gamma} H^1_{1,i} = O_p \frac{1}{P_T} + O_p \frac{1}{n};$$  \hspace{1cm} (5)

Let Assumptions 1-5 hold. As $(n; T)! 1$ with $\frac{P_T}{n}! 0$

$$\frac{P_T}{\hat{\gamma}} H^1_{1,i} \overset{d}{\Rightarrow} N(0; \sigma^2_{\epsilon_t});$$  \hspace{1cm} (6)
where \( \bar{M}_i = \bar{M}_i^{1/2} \bar{M}_i^{1/2} \) and \( \bar{M}_i^{1/2} \) and \( \bar{M}_i^{1/2} \) are the probability limits of \( T^1(\mathcal{F}^0 \mathcal{H}^0 \mathcal{M}_X \mathcal{H}^0 F) \) and \( T^1(\mathcal{F}^0 \mathcal{H}^0 \mathcal{M}_X \mathcal{H}^0 F) \), respectively.

Theorem 1 can be compared with Theorem 2 in Bai (2003, p.147): the rates of convergence in (5) are exactly the same. On the other hand, the limiting distribution of \( \hat{f}_t \) in (6) is different from the one in Theorem 2 in Bai (2): this is due to the presence, in our context, of the idiosyncratic regressors \( x_{it} \).

We use the estimator of \( \bar{M}_i \) proposed in (2, p.150)

\[
\hat{M}_i = (Q_i^0)^{-1/2} Q_i^{-1/2}
\]

where \( Q_i = T^1(\mathcal{F}^0 \mathcal{M}_X \mathcal{F}) \), and \( i = D_{0i} + P^0 q \sum_{j=1}^q D_{ij} + D_{0i}^{1/2} \), with \( D_{ij} = T^1 P^0 T^{tj+1} \mathcal{F}_X^0 \mathcal{F}_{Xt} \mathcal{F}_t \mathcal{F}_t \), where \( \mathcal{F}_X^0 \mathcal{F}_{Xt} \mathcal{F}_t \) is the t-th row of \( \mathcal{M}_X \mathcal{F} \) and \( \mathcal{F}_t = y_{it} \mathcal{F}_X^0 \mathcal{F}_t \mathcal{F}_t \). The bandwidth \( q \) is chosen so that \( q \to 0 \) with \( q = T^{1/4} \).

We now present the asymptotic results for \( \hat{f}_t \).

Theorem 2 Let Assumptions 1-4 hold; then, for every \( t \)

\[
\hat{f}_t \to \mathcal{H}^0_{ft} = O_p \left( \frac{1}{n} \right) + O_p \left( \frac{1}{T} \right)
\]

Let Assumptions 1-5 hold. As \( (n; T) \to 1 \) with \( \frac{p}{n} \to 0 \)

\[
\frac{P_n}{n} \hat{f}_t \to \mathcal{H}^0_{ft} \mathcal{I}^d_1 N(0; \Omega_{ft})
\]

where \( \Omega_{ft} = H_{ft} \Omega_{ft} H^0 \) and \( \Omega_{ft} = \lim_{n \to 1} n \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^P \hat{M}_i \hat{M}_j \).

Theorem 2 is the counterpart to Theorem 1 in Bai (2003, p.145). Rates of convergence and limiting distribution are exactly the same: the presence of individual specific regressors does not affect inference on the common factors.

By virtue of Theorem 2, the asymptotic covariance matrix of \( \frac{P_n}{n} \hat{f}_t \to \mathcal{H}^0_{ft} \) can be estimated using equation (7) in Bai (2003, p.150). Specifically, letting \( \Omega = (\Omega_1; \cdots; \Omega_P) \) with \( \Omega = (\Omega_1; \cdots; \Omega_P) \), and defining \( V_{nt} \) as a diagonal matrix containing the r largest eigenvalues of \( \frac{1}{nt} \Omega \) in descending order, the estimated \( \Omega_{ft} \) is

\[
\hat{\Omega}_{ft} = V_{nt} \left( \sum_{i=1}^n X_i^0 \right)^{-1} \hat{M} \Omega_{ft} \hat{M} \left( \sum_{i=1}^n X_i^0 \right)^{-1} V_{nt}^{-1}
\]
Note that $\alpha_t$ is estimated through $n^{-1} \sum_{i=1}^{n} \alpha_i^0 \xi_t$, which is valid under cross sectional independence. It is not possible, in general, to estimate $\alpha_t$ consistently unless some ordering among the cross sectional units is assumed - see also Bai (2003, p.150).

Combining Theorems 1 and 2, we obtain the asymptotics for the estimated common component $c_t = \hat{\alpha}_t$, defined as $c_t = \hat{\alpha}_t^0$.

**Corollary 1** Let Assumptions 1-4 hold; then, for all $i$ and $t$

$$c_t = \hat{\alpha}_t = O_p \left( \frac{n}{T} \right) + O_p \left( \frac{1}{T^2} \right).$$

(11)

Let Assumptions 1-5 hold. As $(n;T) \to \infty$

$$\frac{1}{n} \sum_{t=1}^{T} f_t^i = \frac{1}{T} f_t^i \to f_t \quad (c_t, c_t^i) \to^d N (0; 1);$$

(12)

where $f_t$ is defined in Theorem 2 and $c_t^i$ in Theorem 1.

After discussing the asymptotic properties of $\hat{\alpha}_t$ and $\hat{f}_t$, we turn to deriving tests for the null of no factor structure.

### 3 Testing for no factor structure

In this section, we discuss and compare two approaches to testing for the null of no factor structure in (1). Motivated by Sara" etdis, Yamagata and Robertson (32), we study tests for, respectively: (a) the null of cross-sectional homogeneity of the loadings $\alpha_i$s; and (b) the null of homogeneity, over time, of the $f_t$s.

Formally, we propose two tests for the null hypotheses:

$$H_0^a : \alpha_i = \alpha \quad \text{for all } i; \quad (13)$$

$$H_0^b : f_t = f \quad \text{for all } t; \quad (14)$$

Both (13) and (14) entail that there is no real factor structure in (1). Consider (13) first. When $H_0^a$ holds, equation (1) can be rewritten as

$$y_t = \xi_t + \sum_{i=1}^{n} \alpha_i^0 \xi_t + \xi_t; \quad (15)$$
where we have defined $t = 0$ if $t$. Thus, under $H_a^0$, model (1) boils down to a standard panel specification with a time effect. Similarly, under $H_b^0$ in (14), equation (1) can be rewritten as

$$ y_{it} = \hat{\beta}_i + \hat{\mu}_t + \hat{\epsilon}_{it}; $$

(16)

where we have defined $\hat{\mu}_i = \hat{\beta}_t$. Therefore, under $H_b^0$, model (1) is tantamount to a standard panel specification with a unit specific effect.

The considerations made above also entail that testing for (13) and (14) is equivalent to testing for strong cross dependence among the $y_{it}$s. Saraòdis, Yamagata and Robertson (32) propose a test for cross dependence (albeit in a different context) based on verifying the null that loadings are homogeneous, i.e. $i = \beta$. Our paper extends the contribution by Saraòdis, Yamagata and Robertson (32) to our context, and complements it by also considering a test for (14). A similar approach to testing for factor structures versus models with individual or time dummies is also suggested in Bai (3).

In order to test for (13) and (14), we propose two tests based directly on the results in Section 2, i.e. on the estimates of $\hat{\beta}_i$ and $\hat{\mu}_t$. Specifically, we propose two max-type statistics, where the maximum is taken over the deviation of the individual estimate of $\hat{\beta}_i$ (resp. of $\hat{\mu}_t$) with respect to their cross-sectional (resp. time) average. This approach has been proposed, in the context of testing for poolability with observable regressors, by Westerlund and Hess (34), whose simulations show that the power properties are very promising, although issues may arise in presence of ties (22). In our context, we show that tests based on max-type statistics have power even versus alternatives whereby only one unit/time period has heterogeneous loadings/common factors. Castagnetti, Rossi and Trapani (10) study the use of alternative test statistics for $H_a^0$ and $H_b^0$—specifically, they consider tests based on average-type and Hausman-type statistics. Neither approach is found to be employable: average-type statistics diverge under the null as $(n; T) \to 1$, while Hausman-type ones are inconsistent.

Define $\mathbf{b} = n \sum_{i=1}^{n} \beta_i$ and $\mathbf{f} = T \sum_{t=1}^{T} \mu_t$. We propose the following max-type test statistics:

$$ S_{nT} = \max_{\beta, \mu} \left\{ \beta_i - \beta, \mu_t - \mu \right\}; $$

(17)

$$ S_{nT} = \max_{\beta, \mu} \left\{ \beta_i - \beta, \mu_t - \mu \right\}; $$

(18)

We point out that under the null hypotheses $H_a^0$ and $H_b^0$, the spaces spanned by the loadings and by the factors (respectively) have rank equal to one. This fact was already noted by Saraòdis, Yamagata and Robertson (32) who, building on it, suggest running their test setting $r = 1$. This can be applied to our
context also: \( S_{r:nT} \) and \( S_{r:nT} \) can be used setting \( r = 1 \), which avoids having to estimate \( r \).

From a methodological perspective, this entails that tests based on (17) and (18) can be implemented without prior knowledge of the number of factors: thus, testing does not require estimation of \( r \) as a preliminary step. Indeed, we note that tests for (17) and (18) are to be implemented before determining \( r \). If the null is not rejected, the conclusion can be drawn that no factor structure is needed, and either (15) or (16) is the correct specification. Conversely, if the null is rejected, then it follows that there is a genuine factor structure. Hence, the next step is determining the number of latent common factors \( r \), e.g. by applying some information criteria as discussed in Bai and Ng (6) and Bai (4). The asymptotic properties of the estimated common factors, loadings and common components are those given in Section 2.

3.1 Testing for \( H^0 \):

In this section we report the asymptotics of \( S_{r:nT} \) under the null \( H^0 \), and we analyse the consistency of tests based on \( S_{r:nT} \). We show that, as \((n;T) \rightarrow \infty \) under some restrictions on the relative speed of divergence, \( S_{r:nT} \) (suitably normalised) converges to a Gumbel distribution. Further, we also show that tests based on \( S_{r:nT} \) have nontrivial power versus alternative hypotheses shrinking at a rate \( O_p \frac{\ln n}{n} \).

Let \( k_1 \) be the largest number for which \( E \sum_{i=1}^{k_1} \|k^i \| \), \( E \sum_{i=1}^{k_1} \|k^i \| \) and \( E \sum_{i=1}^{k_1} \|k^i \| \) are \( O_p \). In view of Assumption 1, \( k_1 = 12 \). Consider the following assumptions, which complement Assumptions 1 and 2, imposing further conditions on the form of time and cross sectional dependence.

Assumption 6. [serial dependence] Let \( \delta > 0 \) and \( \delta > 2(1;+1) \): (i) \( P_t = f \) and \( x_{it} \) are L2+\( \delta \)-NED (Near Epoch Dependent) of size \( \delta \) on a uniform mixing base \( f_{0k} \) of size \( r = (r \leq 2) \) and \( r > \frac{\delta - 1}{\delta - 1} \); (ii) (a) letting \( V^{(i)}_{tT} - V^{(i)}_{tT} \) with \( V^{(i)}_{tT} \) \( < 1 \), (b) the same holds for \( V^{(i)}_{tT} - V^{(i)}_{tT} \) \( \|k^i \| \) \( \|k^i \| \) \( \|k^i \| \); (iii) (a) letting \( w_{kt} \) be the \( k \)-th element of \( f_{kt} \) and \( S_{kt} \) \( P_m+T \) \( w_{kt} \) \( \|k^i \| \) \( \|k^i \| \) \( \|k^i \| \); (b) the same holds for \( x_{it} \).

Assumption 7. [cross sectional dependence] It holds that \( T \sum_{t=1}^{T} P_{s=1}^{T} E \sum_{i=s}^{T} \|k^i \| \ln n ! 0 \) as \((n;T) \rightarrow \infty \) for all \( i \not= j \).
Assumptions 6 and 7 complement Assumptions 1 and 2, by adding further requirements on the form of serial dependence and on the amount of cross dependence respectively.

More specifically, Assumption 6 specifies the amount of memory allowed in the series \( [x_t, f_t, x_{it}] \) - these all have, by Assumptions 1 and 2, finite moments up to order 12. The assumption is needed in order to prove an a.s. version of the Invariance Principle (IP), and it is a quite general specification for the form and amount of serial dependence. Part (iii) is a bound on the growth rate of the variance of partial sums, and it is the same as equation (1.5) in Eberlein (18); see also Assumption A.3 in Corradi (14).

As far as Assumption 7 is concerned, it complements the summability conditions in Assumption 1 by allowing for some cross dependence. In essence, it requires that \( \sum_{t=1}^{T} \sum_{s=1}^{T} \frac{1}{T} \ln \left( \frac{1}{T} \right) \Gamma \left( \frac{1}{T} \right) = o \left( \ln \frac{1}{T} \right) \) for all \( i \neq j \), which is a much weaker requirement than the one in Assumption 1(ii)(d).

Let the critical value \( c_{j,n} \) be defined such that \( P \left( S_{j,T} \leq c_{j,n} \right) = 1 \) under \( H_0 \), and let \( (\cdot) \) denote the Gamma function. It holds that:

**Theorem 3** Let Assumptions 1-4 and 6-7 hold, and let \( (n;T) \to 1 \) with

\[
T^{-1} \sum_{t=1}^{T} x_t^2 + n^{-1} \to 0.
\]

Under \( H_0 \), it holds that

\[
P \left( A_n S_{n,T} \leq x + B_n \right) = e^{-e^x};
\]

where \( A_n = \frac{1}{2} \) and \( B_n = \ln (n) + \frac{1}{2} \ln \ln (n) \ln \frac{1}{T} \). Under the alternative \( H_1 \), for at least one \( i \), if

\[
\frac{T}{\ln n} k_i \ln k_i^2 \to 1;
\]

it holds that \( P \left( S_{i,n,T} > c_{j,n} \right) = 1 \).

Theorem 3 states that \( S_{n,T} \) has a Gumbel distribution. This holds in the joint limit \( (n;T) \to 1 \), with the restrictions specified in (19). Since \( k_1 \leq 12 \), the latter condition requires \( \frac{T}{\ln n} \to 0 \), which is
marginally stricter than the condition $\frac{p}{r}! 0$ needed in for (6). Also, (19) needs that $\frac{n^{1+\epsilon}}{\rho_1}! 0$; this becomes, under Assumptions 1(i) and 2(i), $\frac{n^{1+\epsilon}}{\rho_1}! 0$.

Equation (20) also provides a rule to calculate asymptotic critical values $c_{i,n}$, which are given by

$$c_{i,n} = 2B_n \ln j \ln (1 - \ln (1 - j^{2}))$$

Thus, for a given level $\alpha$, $c_{i,n}$ is nuisance free, and it depends only on the cross-sectional sample size, $n$. A well known issue in EVT is that convergence to Extreme Value distributions is in general rather slow. Canto e Castro (8) shows that the rate of convergence for the maximum of a sequence of random variables following a Gamma distribution is $O(\frac{1}{\ln n})$. Unreported Monte Carlo evidence shows that tests based on using $c_{i,n}$ perform quite well, although they are a bit oversized. As an alternative, one can replace $B_n$ with $F_{1r}(1 - 1/n)$, where $F_{1r}(\cdot)$ is the inverse of the cumulative distribution function of a chi-square with $r$ degrees of freedom, see Embrechts, Klüppelberg and Mikosch (19).

As far as consistency of the test is concerned, equation (21) shows that nontrivial power is attained versus local alternatives shrinking at a rate $O(\frac{p \ln n}{T})$. Thus, when using max-type statistics such as $S_{i,nT}$, $n$ does not play a role in enhancing the power of the test. On the other hand, the test is powerful as long as just one $i$ is different from the others.

### 3.2 Testing for $H_0^b : f_t = f$

We report the asymptotics of $S_{r,nT}$ under $H_0^b$, and its consistency. Similarly to the previous subsection, we show that, as $(n;T) \to 1$ under some restrictions on the relative speed of divergence, $S_{r,nT}$ (suitably normalised) converges to a Gumbel distribution. Further, we also show that tests based on $S_{r,nT}$ have nontrivial power versus alternative shrinking at a rate $O(\frac{p \ln n}{T})$.

Let $k_2$ be the largest number such that $E(k_t k^2) = E(k_1 k^2) = E(\|x\|^2)$ are all finite. In view of Assumptions 1 and 2, $k_2 \in [2, \cdot]$. Consider also the following assumption, which, as in the previous section, complement Assumptions 1 and 2 by adding further structure to the serial and cross sectional dependence of the series.

**Assumption 8.** [cross sectional dependence] Let $\alpha > 0$ and $\beta \in (1; +1)$: (i) $\{x\}$ is $L_{2+\alpha}$-NED across $i$, of size $\alpha$ on a uniform mixing base $\mathcal{F}_i$, $\alpha = 1$ of size $\alpha = 2$ and $\alpha > \frac{1}{\beta + 1}$; (ii) letting $V_{\alpha}^{\|x\|} = n^{-1} E(|\mathcal{F}_i|)^{\alpha}$, $\mathcal{F}_i$ is positive definite uniformly in $\alpha$, and as $n \to 1$, $V_{\alpha}^{\|x\|} \to V_{\alpha}^{\|x\|}$ with $kV_{\alpha}^{\|x\|} < 1$; (iii) letting $S_{\alpha}^{\|x\|} = \mathbb{P}_{i=1}^{n+1} |\mathcal{F}_i|$ there exists a positive constant $\alpha$ such that $n^{-1} \mathbb{E} S_{\alpha}^{\|x\|} \to \alpha$ uniformly in $\alpha$, with $\alpha > 0$.  

12
Assumption 9. [serial dependence] It holds that \( \lim_{k \to 1} n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} E(\mathbb{I}[t \leq k]) \ln k = 0 \) as \((n; T) \to 1\).

Assumption 8 is very similar, in spirit, to Assumption 6, and it requires that \( \mathbb{I}[\mathbb{I}] \) is NED across \( i \). By virtue of Assumption 8, an a.s. IP holds for \( \sum_{i=1}^{n} \mathbb{I} \) and for \( \sum_{i=1}^{n} \mathbb{I} \). The definition of NED for spatial processes has been studied in Jenish and Prucha (24), and we refer to that paper for details.

Assumption 9 is the so-called Berman condition (7): as mentioned when discussing Assumption 7, standard EVT, which holds for i.i.d. data, can be applied under such condition, yielding the same results as in the case of independence. Berman condition holds as long as serial correlations have at least a logarithmic rate of decay, and it is a sufficient condition used to verify more general mixing conditions which are typical of EVT (and more difficult to verify; see e.g. 27). Assumption 9 is a very mild requirement: for example in the case of ARMA processes, typically the autocovariances have an exponential rate of decay (see e.g. 23), which is more than enough to ensure that Assumption 9 holds. Further, Assumption 9 can be shown to hold in contexts where the autocorrelation function is not absolutely summable, as e.g. fractional ARIMA processes. In our context, Assumption 9 can be compared to Assumption 1(ii)(d), and it contains the same summation across \( i \).

Let the critical value \( c_{[i; T]} \) be defined such that \( \mathbb{P}(S_{f;nT} > c_{[i; T]}) = 1 \) under \( H_{0} \). It holds that:

**Theorem 4** Let Assumptions 1-4 hold and 8-9, and let \((n; T) \to 1\) with

\[
\frac{P}{nT^{\frac{1}{2}} + T^{2}} + \frac{4k_{2}}{n} \to 0; \tag{23}
\]

Under \( H_{0} \), it holds that

\[
\mathbb{P}[A_{T}S_{f;nT} \times + B_{T}] = e^{e^{-x}}; \tag{24}
\]

where \( A_{T} = \frac{1}{2} \) and \( B_{T} = \ln(T) + \frac{1}{2} - \ln \ln(T) \ln \frac{1}{2} \). Under the alternative \( H_{1}^{P} : f_{t} = f + c_{t} \) for at least one \( t \), if

\[
\frac{n}{\ln T} \ln k_{G} k^{2} \to 1; \tag{25}
\]

it holds that \( \mathbb{P}(S_{f;nT} > c_{[i; T]}) = 1 \).

Theorem 4 is very similar to Theorem 3; convergence to the Gumbel distribution under the null is shown for \((n; T) \to 1\) jointly under some restrictions between \( n \) and \( T \), spelt out in (23). Specifically, it is required that \( \frac{T^{\frac{1}{2}}}{n} \to 0 \); since \( k_{2} \to 12 \), the former restriction is, at most, \( \frac{n}{\ln} \to 0 \). This is only marginally stronger than \( \frac{P}{n} \to 0 \), which is required for (9) to hold. Similarly, requiring that \( \frac{T^{4k_{2}}}{n} \to 0 \)
entails $\frac{1}{nT} \neq 0$. As in the case of Theorem 3, the test should be applied when $n$ is not exceedingly larger than $T$, and vice versa.

Critical values for a test of level $\alpha$ can be calculated as

$$c_{\alpha;T} = 2B_T \ln j_n \ln (1 - \alpha)^2;$$

(26)

alternatively, $B_T$ can be approximated by $F_{\text{tails}}^{-1}(1 - \alpha = T)$.

As far as power is concerned, (25) stipulates that the test is consistent versus alternatives shrinking as $O \left( \frac{\ln n}{n} \right)$. Similarly to Theorem 3, it suffices that $f_t$ differs from $f$ in just one period $t$ for the test to reject $H_0$.

### 4 Small sample properties

In this section, we evaluate, through synthetic data, the small sample properties of estimators of $\gamma$ and $\hat{f}_t$ (discussed in Section 2), and the power and size of tests for (13) and (14) based on $S_{1:nT}$ and $S_{f_1:nT}$ (discussed in Section 3).

The Monte Carlo settings are as follows. Based on model (1)-(2), we consider the following data generating process (DGP):

$$y_{it} = \beta x_{it} + \gamma f_{it} + \epsilon_{it};$$

(27)

$$x_{it} = \beta_1 + \beta_2 f_{it} + \epsilon_{it};$$

(28)

i.e. we consider model (1)-(2) with $m = r = 1$ - only one individual specific regressor, $x_{it}$, and only one common factor, $f_t$. Unreported simulations show that increasing either $r$ or $m$ does not alter the results. In the simulations, we generate the parameters $\beta_1$ and $\beta_2$ as i.i.d. $N(1; 1)$. The common factor $f_t$, the loading $\beta_1$, and both error terms $\epsilon_{it}$ and $\epsilon_{it}^T$ are all generated as i.i.d. $N(0; 1)$ unless otherwise stated. Results are reported for $(n; T) = (30; 50; 100; 200)$ unless otherwise stated. Results are reported for $(n; T) = (30; 50; 100; 200)$. Finally, in both exercises, simulations are carried out with 5000 iterations.

#### 4.1 Small sample properties - $\hat{\gamma}$ and $\hat{f}_t$

We evaluate the small sample properties of the estimators $\hat{\gamma}$ and $\hat{f}_t$.

As far as $\hat{f}_t$ is concerned, we follow the same logic as in Bai (2). We compute the correlation coefficient between $f f_t^\top g_{1:t}^\top$ and $f f_t^\top g_{1:t}^\top$, for each Monte Carlo iteration $j$ - say $j^\top f$. We report the average correlation...
coefficients, i.e. $J = 1 \sum_{j=1}^{P} f_{j}$, in Table 1 (recall that $J = 5000$).

Table 1 illustrates that the estimated common factor $f_t^*$ is highly correlated with the unobserved common factor $f_t$. This reinforces the results in Bai (2), albeit obtained in a different context, that the estimated factors are quite good at tracking the true ones; indeed, numerical values are very similar to those in Table 1 in Bai (2003, p.151). When $n$ and $T$ are both 100, the estimated factors can be treated as the true ones.

<table>
<thead>
<tr>
<th>$n$</th>
<th>30</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.977</td>
<td>0.964</td>
<td>0.979</td>
<td>0.974</td>
</tr>
<tr>
<td>50</td>
<td>0.976</td>
<td>0.963</td>
<td>0.989</td>
<td>0.970</td>
</tr>
<tr>
<td>100</td>
<td>0.991</td>
<td>0.987</td>
<td>0.988</td>
<td>0.991</td>
</tr>
<tr>
<td>200</td>
<td>0.992</td>
<td>0.994</td>
<td>0.997</td>
<td>0.997</td>
</tr>
</tbody>
</table>

Table 1: Average correlation coefficients between $f_t^*$ and $f_t$.

As far as $\gamma_i$ is concerned, we report confidence intervals for $i$. In order to illustrate how confidence intervals shrink as $T$ expands, we set $n = 50$ and $T = 20; 50; 100; 1000$.

According to equation (6) in Theorem 1, as $(n; T) \to (0, \infty)$, with $\frac{p}{n} \to 0$, the 95% confidence interval for $H \gamma_i$ is given by $\gamma_i \pm \frac{1.96}{\sqrt{T}} \gamma_i^{1/2}$. Further, let $\gamma_i$ be the least square estimate of $\gamma_i$ in $\gamma_i = \gamma_i + \text{error}$, where $\gamma_i = (\gamma_i; \cdots; \gamma_i)^T$ and $\gamma_i = (\gamma_i; \cdots; \gamma_i)^T$. The 95% confidence interval for $\gamma_i$ is therefore obtained as $\gamma_i \pm \frac{1.96}{\sqrt{T}} \gamma_i^{1/2}$. By rotating $\gamma_i$ towards $i$, we consider the confidence interval for $i$ directly, reported in Figure 1.

Figure 1: Confidence intervals for $\gamma_i$. For each value of $i = 1; \cdots; 50$ (on the horizontal axis), the solid line represents the true loading $\gamma_i$. The dashed lines are the confidence intervals at 95% confidence level for each $i$.

Figure 1 shows that, in most cases and for all combinations of $n$ and $T$, the confidence intervals contain the true value of $\gamma_i$. This also holds true for the case $(n; T) = (50; 1000)$, where the ratio $\frac{p}{n}$ is not negligible, as the theory would require. As predicted by the theory, as $T$ grows, the confidence intervals collapse to the true value of $\gamma_i$.

4.2 Small sample properties - $S_{nT}$ and $S_{nT}$

In this subsection, we report empirical rejection frequencies and power for tests based on the max-type statistics $S_{nT}$ and $S_{nT}$ defined in (17) and (18) respectively.
Table 2: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^i : \beta = 1$. Based on $S_{nT}$. The DGP used in the simulations is (27)-(28).

As far as the design of the Monte Carlo is concerned, recall that the variance of the common components $c_{it} = \beta_f t$ is set equal to 1 across all experiments. We conduct our simulations for different values of the signal-to-noise ratio $\frac{\text{Var}(c_{it})}{\sigma_{it}^2}$, where $\sigma_{it}^2$ is the variance of $c_{it}$, equal to $\frac{1}{3}; \frac{1}{2}; 1$.

Critical values have been computed by approximating $B_n$ and $B_T$ as discussed in Section 3. Unreported simulations show that results worsen only slightly when using the asymptotic critical values.¹

Testing for $H_0^i : \beta = 1$

When evaluating the empirical rejection frequencies for tests based on $S_{nT}$, we run the Monte Carlo simulations under the null $\beta = 1$ for all $i$. When evaluating power, we generate the loadings $\beta_i$ as i.i.d. $N(1; 1^2)$, reporting results for the case of $\beta = 0.2$. Given that $\beta_i$ is cross-sectionally uncorrelated and homoskedastic by design, $\beta_i$ is estimated as $\hat{\beta}_i = \frac{1}{n} \sum_{t=1}^{T} \hat{\sigma}_{it}^2 T^\top F^\top M_{xt} F \hat{\sigma}_{it}^2$, where $\hat{\sigma}_{it}^2 = \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{\sigma}_{it}^2$.

Results for size and power when using the main DGP (27)-(28) are in Table 2.

We Örstly consider the empirical rejection frequencies (left panel in the table). The test has a tendency to be oversized in small samples; as a general rule, the correct size is attained when $T = 100$ and $n = 50$; even when $\hat{\sigma}_{it}^2 = 1$ (low signal-to-noise ratio), the test has satisfactory size properties for $T = 50$. The Table also shows that, as the signal-to-noise ratio decreases (i.e., as $\hat{\sigma}_{it}^2$ increases), the tendency towards

¹The simulation results are available upon request.
Table 3: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^b : f_t = f$, based on $S_{ft; nT}$. The DGP used in the simulations is (27)-(28).

<table>
<thead>
<tr>
<th>n</th>
<th>30</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>30</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\overline{\omega}_n^2 = 1 \Rightarrow \overline{\omega}_T^2 = 1 \Rightarrow$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.044</td>
<td>0.037</td>
<td>0.037</td>
<td>0.030</td>
<td>0.915</td>
<td>0.959</td>
<td>0.988</td>
<td>0.996</td>
</tr>
<tr>
<td>50</td>
<td>0.038</td>
<td>0.034</td>
<td>0.036</td>
<td>0.033</td>
<td>0.993</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>100</td>
<td>0.042</td>
<td>0.041</td>
<td>0.036</td>
<td>0.032</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>200</td>
<td>0.046</td>
<td>0.043</td>
<td>0.038</td>
<td>0.036</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

As far as the power is concerned (right panel in the Table), the test has good power properties in all cases: the power is above 50% for almost all cases. We note that, similarly to the size, the power deteriorates as the signal-to-noise ratio decreases; when $n$ and $T$ are sufficiently large, this disappears.

Testing for $H_0^b : f_t = f$

We run the Monte Carlo simulations under the null $f_t = 1$ for all $t$ when evaluating the size of tests based on $S_{ft; nT}$. When evaluating the power, we generate the common factors $f_t$ as i.i.d. $N(1; \overline{\omega}^2)$, reporting results for the case of $\overline{\omega} = 0.2$. Finally, we estimate $\overline{\omega}_{ft}$ as $\overline{\omega}_{rt} = V_{nT} \overline{\omega}_n^2 \sum_{i=1}^{n} \sum_{t=1}^{T} V_{nT}^{-1}$ where $\overline{\omega}_n^2 = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \overline{\omega}_{rt}$.

Results when using (27)-(28) are shown in Table 3.

It can be noted that the test is slightly undersized for large $T$, e.g. $T = 100$. However, both $n$ and $T$ have a quite limited impact on the results. The test has very good power properties, especially when the signal-to-noise ratio is high. We note that the power increases with both $n$ and $T$, in a more pronounced way with $n$.

For the sake of completeness, we run both tests using as a first step estimator the IE proposed by Song (33). The size and power reported in Table 4, for the $S_{ft; nT}$ test, when the DGP is the one in equations...
Table 4: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^* : \lambda = 1$, based on $S_{nT}$. The DGP used in the simulations is (27)-(28). The first-step estimator is the one proposed Song (33).

(27)-(28), show that the test procedure is unaffected by the choice of the first-step estimator when this is a consistent one. Finally, we point out that in Castagnetti, Rossi and Trapani (2014), we provide further Monte Carlo evidence based on alternative DGPs. The Monte Carlo results confirm for both tests good properties in terms of size and power.

Autocorrelated and heteroskedastic idiosyncratic errors

In order to assess the finite sample properties of the two test procedures when the errors are autocorrelated and heteroskedastic, we consider the following DGP:

$$\eta_t = 0.5\eta_{t-1} + u_t$$

$$u_t \sim \text{IIDN}(0; \sigma^2_{u_t}) \quad \eta_t \sim \text{U}(0.1; 0.5)$$

and we make use of the HAC estimators for $\eta$ and $\eta_t$ given by equations (7) and (10). Apart from these features, the experiments have the same specifications as above, with $\text{Var}(\eta_t) \sim N(0.13; 0.67)$.

The results in Tables 5 and 6 can be compared with the i.i.d. cases in Tables 2 and 3 respectively. In the case of non i.i.d. errors, both tests have a tendency to be oversized in small samples, $(n; T) < 50$. However, as both dimensions are larger than 50, the empirical rejection frequencies become almost
undistinguishable from the ones computed with i.i.d. errors. As far as, the power is concerned, both tests have good properties and are very close to the i.i.d. case.

<table>
<thead>
<tr>
<th>Size</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>n</td>
<td>30</td>
</tr>
<tr>
<td>30</td>
<td>0.103</td>
</tr>
<tr>
<td>50</td>
<td>0.090</td>
</tr>
<tr>
<td>100</td>
<td>0.081</td>
</tr>
<tr>
<td>200</td>
<td>0.072</td>
</tr>
</tbody>
</table>

Table 5: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0: \theta_i = \theta$. The test is computed using the estimator of $\theta_i$ in (7).

<table>
<thead>
<tr>
<th>Size</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>n</td>
<td>30</td>
</tr>
<tr>
<td>30</td>
<td>0.118</td>
</tr>
<tr>
<td>50</td>
<td>0.080</td>
</tr>
<tr>
<td>100</td>
<td>0.050</td>
</tr>
<tr>
<td>200</td>
<td>0.061</td>
</tr>
</tbody>
</table>

Table 6: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0: \theta_t = \theta$. The test is computed using the estimator of $\theta_t$ in (10).

5 Conclusions

In this contribution, we develop an inferential theory for the unobservable common factors and their loadings in a large, stationary panel model with observable regressors. Our framework allows for slope heterogeneity; we also allow for correlation between common factors and observable regressors, by modelling the DGP of the observable regressors as containing the common factors, in a similar spirit as in Pesaran (30).

We extend the framework in Pesaran (30) by providing a two stage estimator for the unobserved common factors and their loading. We derive rates of convergence and limiting distribution of both the estimated factors and loadings, using a similar method of proof to Bai (3). In a similar vein to Saraidis, Yamagata and Robertson (32), we also develop two tests for the null of no factor structure, based on the null that factor loadings are homogeneous, and that common factors are homogeneous over time, respectively. In either case, the assumed factor model boils down to a model with (time speci...
or unit specific common effects, so that common features in the panel can be captured by inserting time dummies or unit specific dummies. The proposed test procedures simplify the specification analysis of heterogeneous panel data models with unobserved factors. From a methodological perspective, this entails that the tests can be implemented without prior knowledge of the number of factors. The only thing which is needed is a consistent preliminary estimation of the slope parameters. Building on this, we propose statistics based on extrema of the estimated loadings and common factors. Under the null, the test statistics converge to an Extreme Value distribution. As far as power is concerned, from a theoretical point of view our tests are consistent even under alternatives where only one loading or common factor differs from the average. Monte Carlo evidence shows that both tests have the correct size and good power properties.

Building on the theory developed in this paper, there are several interesting avenues for further developments. An important case is the estimator of the $\beta_i$'s used in Step 1. In our paper, we focus on the CCE estimator proposed by Pesaran (30); this estimator is easy to treat analytically, but it is only a possible choice. In particular, our setup requires strict exogeneity, thereby ruling out e.g. the possibility of having lagged values of the $y_{i,t}$'s among the regressors. This requirement is due to the estimation method employed in Step 1, rather than to the inference on factors and loadings per se. Indeed, the CCE is known not to work in presence of weakly exogenous regressors (see 20; and 12). However, the assumption of strict exogeneity can be readily relaxed (accommodating e.g. for dynamic models), upon employing, in Step 1, an estimator of the $\beta_i$'s that is consistent at a rate $O_p(\min T, n^{1/2})$. A possible choice for this case is the IE estimator studied in Song (33), which has the desired convergence rate, even in presence of dynamic models. Alternatively, a different approach, based on unit specific estimators can be used, by instrumenting the unobservable common factors $f_t$ using the regressors $x_{j,t}$ for each unit $i$, with $i \neq j$ - indeed, both the CCE and the IE have a natural Instrumental Variable interpretation (see also 4). Such extensions are currently under investigation of the authors.

Acknowledgement

This is a revised version of a paper previously circulated under the working title "Two-Stage Inference in Heterogeneous Panels." We are very grateful to the Editor (Cheng Hsiao), one anonymous Associate Editor and two anonymous Referees for very constructive feedback which has greatly improved the generality of the paper. We also wish to thank the participants to the Faculty of Finance Workshops at Cass Business School; to the New York Camp Econometrics V (Syracuse University, NY, October 2010); to the 4th CSDA International Conference on Computational and Financial Econometrics (London, De-
cember 2010); to the 18th International Conference on Panel Data (Banque de France, July 2012), in particular Chihwa Kao, Jean-Pierre Urbain and Takashi Yamagata. Special thanks go to Lajos Hûrvath for providing us with valuable comments. The usual disclaimer applies.
Appendix: technical results and proofs

In this section we set the rotation matrix $H = I_r$ whenever possible in order to simplify the notation. The proofs of the Lemmas, and the proof of Theorem 4, are given in Castagnetti, Rossi and Trapani (10).

Lemma A.1 Under Assumptions 1-4, it holds that, for every $i, E \left( \delta_i r \right) = O_{nT}$, for any $r \geq 3$.

Lemma A.2 Under Assumptions 1-4, it holds that, for every $i$

A.2(i) $T^{-1} \delta_i' F^\delta = O_p \left( n^{-2} \right)$;
A.2(ii) $n^{-1/2} T^{-1} \delta_i' F_i = O_p \left( n^{-1} + O_p T^{-1} \right)$.

Lemma A.3. It holds that, for every $i$

A.3(i) $T^{-1} X_i' F^\delta H = O_p \left( h^{-2} \right)$;
A.3(ii) $T^{-1} F_0^\delta F_i = O_p \left( h^{-2} \right)$;
A.3(iii) $T^{-1} F_i^\delta F_i = O_p \left( h^{-2} \right)$.

Lemma A.4 Let Assumptions 1-4 hold. Under $H_0^a$ that $a = 1$, it holds that $\delta = O_p \left( n^{-2} \right)$ as $(n; T) \rightarrow 1$.

Lemma A.5 Let Assumptions 1-4 hold. Under $H_0^b$ that $f_t = f$, it holds that $\delta f = O_p \left( h^{-2} \right)$ as $(n; T) \rightarrow 1$.

Lemma A.6 Let Assumptions 1-4 hold, and let $k$ denote the largest finite moment of $\Omega_t, f_t$ and $x_t$. It holds that

A.6(i) $\max_{t \in [1]} \left( \delta_t \right)^2 = O_p \left( n^{2-k} h^{-2} \right)$;
A.6(ii) $\max_{t \in [1]} \left( \delta_t \right)^2 = O_p \left( T_n^{2-k} h^{-2} \right)$;
A.6(iii) $\max_{t \in [1]} \left( \delta_t \right)^2 = O_p \left( n^{2-k} T_n + n^{2-k} h^{-2} \right)$;
A.6(iv) $\max_{t \in [1]} \left( \delta_t \right)^2 = O_p \left( T_n^{2-k} + T_n^{2-k} h^{-2} \right)$;
A.6(v) $\max_{t \in [1]} \left( \delta_t \right)^2 = O_p \left( T_n^{2-k} h^{-2} \right)$;
Proof of Theorem 1. By definition, we have

\[ p \bar{T}(\hat{N}_t^i) = \frac{\hat{F}^\flat M_{X_i}F^\flat}{\hat{F}^\flat M_{X_i}F^\flat} \frac{1}{\hat{F}^\flat M_{X_i}F^\flat} \frac{1}{\hat{F}^\flat M_{X_i}F^\flat} \frac{3}{\hat{F}^\flat M_{X_i}F^\flat} \]  \quad \text{(29)}

We start by considering the denominator of (29):

\[ \frac{\hat{F}^\flat M_{X_i}F^\flat}{\hat{F}^\flat M_{X_i}F^\flat} \frac{1}{\hat{F}^\flat M_{X_i}F^\flat} \frac{1}{\hat{F}^\flat M_{X_i}F^\flat} \frac{3}{\hat{F}^\flat M_{X_i}F^\flat} = I + I^0 \quad \text{II:}\]

Repeated application of Lemma A.3 yields I = O_p \bar{n}_T^2 and II = O_p \bar{n}_T^4. Thus, as \(n;T) \to T \hat{F}^\flat M_{X_i}F^\flat = T \hat{F}^\flat M_{X_i}F^\flat + O_p(1).

We turn to the numerator of (29). It holds that

\[ \frac{\hat{F}^\flat M_{X_i}F^\flat}{\hat{F}^\flat M_{X_i}F^\flat} \frac{1}{\hat{F}^\flat M_{X_i}F^\flat} \frac{3}{\hat{F}^\flat M_{X_i}F^\flat} = I + II:\]
By applying a similar logic as in the proof of Lemma A.4, it can be shown that \( I = \mathbb{O}_p(1) \). As far as \( I \) is concerned, note

\[
II = \frac{p}{T} \left( \sum_{i=1}^{\infty} F \sum_{j=1}^{F} X_i \sum_{k=1}^{F} X_j \sum_{l=1}^{F} X_{kl} \frac{1}{T} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{T} \right);
\]

applying Lemma A.2(i) (to the \( O \)rst term), and Lemma A.3(i) and Assumptions 2(i) and 1(i) (to the second term), it follows that \( II = \mathbb{O}_p \left( \frac{p}{T} \right)^2 \). Thus, the numerator of (29) is of order \( \mathbb{O}_p(1)+\mathbb{O}_p \left( \frac{p}{T} \right)^2 \).

Finally, as \( \frac{p}{T} \to 1 \) under the restriction \( \frac{p}{T} \to 0 \), (29) becomes

\[
\frac{p}{T} \left( \sum_{i=1}^{\infty} F \sum_{j=1}^{F} X_i \sum_{k=1}^{F} X_j \sum_{l=1}^{F} X_{kl} \frac{1}{T} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{T} \right) + \mathbb{O}_p(1);
\]
equation (6) follows from Assumption 5(i). QED

Proof of Theorem 2. Using (1) in Castagnetti, Rossi and Trapani (2014), we can write

\[
f_t = \frac{1}{n} \sum_{j=1}^{\infty} F \sum_{i=1}^{F} X_i \sum_{j=1}^{F} X_j \sum_{k=1}^{F} X_{kl} \frac{1}{T} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{T} + \mathbb{O}_p(1);
\]

The order of magnitude of \( I \) follows exactly from the same passages as in the proof of Lemma A.5, with \( I = \mathbb{O}_p \left( \frac{p}{T} \right)^2 \). Consider \( III \); omitting \( j \) in view of Assumption 3(iii), we have

\[
III = \frac{1}{n} \sum_{j=1}^{\infty} F \sum_{i=1}^{F} X_i \sum_{j=1}^{F} X_j \sum_{k=1}^{F} X_{kl} \frac{1}{T} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{T} + \mathbb{O}_p(1);
\]

we have shown that \( III_a = \mathbb{O}_p \left( \frac{p}{T} \right)^2 \) and \( III_b = \mathbb{O}_p \left( \frac{p}{T} \right)^2 \) in the proof of Lemma A.3, so that \( III = \mathbb{O}_p \left( \frac{p}{T} \right)^2 \). Using Lemma A.3(i), it can be shown that \( III = \mathbb{O}_p \left( \frac{p}{T} \right)^2 \). As far as \( IV \) is concerned, note that

\[
IV = \frac{1}{n} \sum_{j=1}^{\infty} F \sum_{i=1}^{F} X_i \sum_{j=1}^{F} X_j \sum_{k=1}^{F} X_{kl} \frac{1}{T} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{T} + \mathbb{O}_p(1);
\]
Similar passages as in the proof of the order of magnitude of $I_{\frac{1}{2}}$, and the fact that $E k_f \mathbb{1} M$ entail $IV_a = O_p n^{1/2} T^{1/2}$. Similarly, $IV_b$ is bounded by $k_f k E \sum_{j=1}^{1/2} T^{1/2}$, which is $O_p n^{1/2} T^{1/2}$ using Lemma A.1. Thus, $IV = O_p n^{1/2} T^{1/2}$. Turning to $V$, we have

$$V = \frac{1}{n} \sum_{j=1}^{n} \frac{f_{X_j}}{T} \quad \text{and} \quad \frac{1}{n} \sum_{j=1}^{n} \frac{X_j^\ast \mathbb{1} X_j}{T} \quad \text{are bounded by} \quad \lambda \frac{1}{n} \sum_{j=1}^{n} \frac{X_j^\ast \mathbb{1} X_j}{T}$$

We start from $V_b$ and $V_a$ in the middle follows from Holder's inequality. Consider now $V_b$:

$$V_a = \frac{1}{n} \sum_{j=1}^{n} \frac{f_{X_j}}{T} \quad \text{and} \quad \frac{1}{n} \sum_{j=1}^{n} \frac{X_j^\ast \mathbb{1} X_j}{T} \quad \text{are bounded by} \quad \lambda \frac{1}{n} \sum_{j=1}^{n} \frac{X_j^\ast \mathbb{1} X_j}{T}$$

Consider $V_{a;2}$:

$$V_{a;2} = \frac{1}{n} \sum_{j=1}^{n} \frac{f_{X_j}}{T} \quad \text{and} \quad \frac{1}{n} \sum_{j=1}^{n} \frac{X_j^\ast \mathbb{1} X_j}{T} \quad \text{are bounded by} \quad \lambda \frac{1}{n} \sum_{j=1}^{n} \frac{X_j^\ast \mathbb{1} X_j}{T}$$

Using Assumption 4(i). Further, $E \frac{f_{j\mathbb{1} X_j}}{T} \quad \text{and} \quad E \frac{f_{j\mathbb{1} X_j}}{T} \quad \text{are bounded by} \quad \lambda \frac{1}{n} \sum_{j=1}^{n} \frac{X_j^\ast \mathbb{1} X_j}{T}$. Again by Holder's inequality. Using Lemma A.2(i), Assumption 2(iv) and similar passages as in the proof of (3) in Castagnetti, Rossi and Trapani (2014), and Assumption 1(i), we have
By virtue of Assumption 4(i), \( V_{a:1;1} \) in the \( O \) of 2(i) and \( 1(ii)(a) \). Consider now \( V_{a;2} \); this is bounded by the square root of

\[
E \left[ \frac{X_j^T \hat{\mathbf{W}}_w E (\{q \}_{t})}{\hat{\mathbf{W}}_w (\{q \}_{t})} \right]^2 \leq \frac{1}{n^2} \frac{X_j^T \hat{\mathbf{W}}_w E (\{q \}_{t})}{\hat{\mathbf{W}}_w (\{q \}_{t})} + \frac{1}{n^2} \frac{X_k^T \hat{\mathbf{W}}_w E (\{q \}_{t})}{\hat{\mathbf{W}}_w (\{q \}_{t})} \]

after some algebra, this is bounded by

\[
E \left[ \frac{X_j^T \hat{\mathbf{W}}_w E (\{q \}_{t})}{\hat{\mathbf{W}}_w (\{q \}_{t})} \right]^2 \leq \frac{1}{n^2} \frac{X_j^T \hat{\mathbf{W}}_w E (\{q \}_{t})}{\hat{\mathbf{W}}_w (\{q \}_{t})} + \frac{1}{n^2} \frac{X_k^T \hat{\mathbf{W}}_w E (\{q \}_{t})}{\hat{\mathbf{W}}_w (\{q \}_{t})} \]

by using Assumption 2(iii) in the second line, Assumption 2(ii) in the third line, and Assumption 1(ii)(c) in the final passage. Thus, \( V_{a;1;2} = O_p \) \( n \) \( 1 \). Putting all together, \( V = O_p \) \( T \).
Lemma A.7

expression. Putting everything together, as based on the same arguments as in Bai (2), since the estimation error \( \frac{p}{T} \) does not appear in their expression. Putting everything together, as \((n; T)! 1\) with \( \frac{p}{T} \) ! 0, the term that dominates in the expansion of \( f_t \) is \( VII \), whose asymptotics is exactly the same as studied in Bai (2, Theorem 1).

\[ \text{QED} \]

Proof of Theorem 3. Prior to proving the Theorem, we lay out some preliminary results and notation. We write

\[ \bigwedge b = (\bigwedge a) + (\bigwedge i) \quad \bigwedge a = a_1 + b \quad a_i \]  

Under \( H_0 \), \( a_i = 0 \); also, \( b \) can be rewritten as \( b = \bigwedge i \). Using (29), we have

\[ b = F^\omega X_1 F^1 F^i M X_1 i + F^\omega X_1 F^1 F^i M X_1 H \]

\[ = b_1 + b_2; \]

where we de\(\hat{\text{One}} b_1 = F^\omega X_1 F^1 F^i M X_1 i \) and \( b_2 \) is the remainder. Further, we can write \( \bigwedge i = \bigwedge i + a_p \bigwedge i \) for each \( i \). Neglecting higher order terms that depend on \( a_p \), we have

\[ \bigwedge b = (\bigwedge b_1) + (\bigwedge i \bigwedge i) \]

After this preliminary calculations, we turn to proving (20). In order to do this, \( \hat{\text{O}} \)rstly show that \( \max_{\bigwedge i \bigwedge i \bigwedge i} T \bigwedge b_1 \bigwedge b_2 \) can be approximated by the maximum of a sequence of independent random variables with a \( \bigwedge i \) distribution, up to a negligible error. Given that the maximum of a sequence of chi-squares is of order \( O_p (\ln n) \), the approximation error should be \( O_p (\ln n) \) at most. Secondly, we show that \( I + V_i \) in (32) are also all \( O_p (\ln n) \) uniformly in \( i \).

Consider \( \max_{\bigwedge i \bigwedge i \bigwedge i} T \bigwedge b_1 \bigwedge b_2 \), and consider in particular the sequence \( n_p T_{b_1} \). It holds that \( \frac{p}{T} T_{b_1} = T^1 F^\omega X_1 F^1 i + \bigwedge i \bigwedge 1 = \omega M X_1 i \). As far as the numerator of this expression is concerned, by Lemma A.7(v) we write \( T^1 = \omega M X_1 i = N_i + R_{N_i} \) with \( N_i \) de\(\hat{\text{Oned}} \) in Lemma A.7 as being zero mean.
Gaussian with covariance matrix \( \| f_{M,e} \|^2 \) and \( R_{N,i} = O_p n^{1/s_k} T^{1/s_k} 1/s_k^2 \). As far as the denominator of \( p_{T_{Di}} \) is concerned, based on Lemma A.7(ii) we write \( T 1^{2} f_{M,X} f_{M,X} = \| f_{M,i} \|^2 + R_{M,i} \) with \( R_{M,i} = O_p T^{1+2/s_k} + O_p n^{1/s_k} \). Hence we write

\[
p_{T_{Di}} = h \| f_{M,i} \|^2 + R_{M,i} [N_i + R_{N,i}];
\]

(33)

Based on (33), and on the definitions of \( \| f_{M,e} \|^2 \) and of \( \| f_{M,i} \|^2 \), it holds that

\[
T \| b_{Di} \|^2 = N_i^0 \| f_{M,e} \|^2 N_i + 2N_i^0 \| f_{M,i} \|^2 \| R_{N,i} \|^2 + 2R_{N,i} \| f_{M,e} \|^2 N_i
\]

\[
+ 2N_i^0 \| f_{M,i} \|^2 \| R_{M,i} \|^2 N_i + 2N_i^0 \| f_{M,i} \|^2 \| R_{M,i} \|^2 R_{N,i}
\]

\[
+ R_{N,i} \| f_{M,i} \|^2 R_{N,i} + 2R_{N,i} \| f_{M,i} \|^2 \| R_{M,i} \|^2 R_{N,i}
\]

\[
+ N_i^0 \| f_{M,i} \|^2 \| R_{M,i} \|^2 R_{N,i} + 2N_i^0 \| f_{M,i} \|^2 \| R_{M,i} \|^2 R_{N,i}
\]

\[
+ R_{N,i} \| f_{M,i} \|^2 \| R_{M,i} \|^2 R_{N,i}
\]

\[
= N_i^0 \| f_{M,e} \|^2 N_i + T_{\| b_{Di} \|^2} + T_{\| b_{Di} \|^2} + T_{\| b_{Di} \|^2} + T_{\| b_{Di} \|^2} + T_{\| b_{Di} \|^2};
\]

(34)

We note that the distribution of \( N_i^0 \| f_{M,e} \|^2 N_i \) is \( \| N \|^2 \). We now show that, in (34), \( \max_{1 \leq i \leq n} \| b_{Di} \|^2 \) are all \( O_p(1) \). Consider \( \max_{1 \leq i \leq n} \| b_{Di} \|^2 \); this is bounded by \( \max_{1 \leq i \leq n} k_{N_i} k_{R_{N,i}} k = O_p \| N \|^2 \ln n \), in view of Lemma A.7(v) and the fact that \( \max_{1 \leq i \leq n} k_{N_i} k = O_p \| N \|^2 \ln n \). The same holds for \( \max_{1 \leq i \leq n} \| b_{Di} \|^2 \).

Turning to \( \max_{1 \leq i \leq n} \| b_{Di} \|^2 \), it is bounded by \( \max_{1 \leq i \leq n} k_{N_i} k^2 \max_{1 \leq i \leq n} k_{R_{M,i}} k = O_p T^{1+2/s_k} \ln n + O_p n^{1+1/s_k} \ln n \) by virtue of Lemma A.7(ii). As far as \( \max_{1 \leq i \leq n} \| V_{b,i} \|^2 \) is concerned, it is bounded by \( \max_{1 \leq i \leq n} k_{N_i} k^2 \max_{1 \leq i \leq n} k_{R_{M,i}} k \max_{1 \leq i \leq n} k_{R_{N,i}} k \), and therefore it is dominated by the previously analyzed terms. Also, \( \max_{1 \leq i \leq n} \| V_{b,i} \|^2 \) has the same order of magnitude as \( \max_{1 \leq i \leq n} k_{R_{N,i}} k^2 \), thereby being dominated by the other terms. Similarly, \( \max_{1 \leq i \leq n} \| V_{l,i} \|^2 \) is bounded by \( \max_{1 \leq i \leq n} k_{R_{N,i}} k^2 \max_{1 \leq i \leq n} k_{R_{M,i}} k \), and therefore it is also dominated. Turning to \( \max_{1 \leq i \leq n} \| V_{l,i} \|^2 \), it is bounded by \( \max_{1 \leq i \leq n} k_{N_i} k^2 \max_{1 \leq i \leq n} k_{R_{M,i}} k^2 \), so that it is smaller than \( \max_{1 \leq i \leq n} \| b_{Di} \|^2 \), and therefore negligible. Similarly, \( \max_{1 \leq i \leq n} \| V_{l,i} \|^2 \) is bounded by \( \max_{1 \leq i \leq n} k_{N_i} k \max_{1 \leq i \leq n} k_{R_{M,i}} k^2 \max_{1 \leq i \leq n} k_{R_{N,i}} k \), which is dominated by \( \max_{1 \leq i \leq n} \| b_{Di} \|^2 \), and thus negligible. Finally, \( \max_{1 \leq i \leq n} \| X_{b,i} \|^2 \) is bounded by \( \max_{1 \leq i \leq n} k_{R_{M,i}} k^2 \max_{1 \leq i \leq n} k_{R_{N,i}} k^2 \), and it is dominated. Therefore

\[
\max_{1 \leq i \leq n} T \| b_{Di} \|^2 = \max_{1 \leq i \leq n} N_i^0 \| f_{M,e} \|^2 N_i + O_p (nT)^{1+1/s_k} \ln n \# + O_p \| n \|^2 + O_p \| n \|^2:
\]

(35)
After proving that $\max_{i \in I_0} T \ b_{0i} = 1 \ b_{0i}$ can be approximated by $\max_{i \in I_0} N_{fM} \ f^{-1} \ N_i$, we turn again to equation (32). We now show that $\max_{i \in I_0} I_1, \ldots, \max_{i \in I_0} V_i$ are all $o_p (\ln n)$. Consider $I_i$; it holds that

$$\max_{i \in I_0} I_i \leq \max_{i \in I_0} T \ b_{0i} = 1 \ b_{0i} \leq \max_{i \in I_0} I_i \leq 1 :$$

Equation (35) implies that $\max_{i \in I_0} T \ b_{0i}^2 = o_p (\ln n)$; thus, applying Lemma A.7(iv), $\max_{i \in I_0} I_i = o_p (\ln n)$ as $\ln n \to \infty$. Turning to $\max_{i \in I_0} II_i$, note that, in equation (31), $b_{0i}$ is defined as

$$b_{0i} = F^0 M_{X_i} F^0 F^0 I_i F^0 F^0 M_{X_i} F^0 F^0 I_i F^0 F^0 I_i :$$

further, by the invertibility of $I_i$ and Lemma A.7(iv), $\max_{i \in I_0} T \ b_{0i}^2 = o_p (\ln n)$ has the same order of magnitude as $\max_{i \in I_0} \max_{i \in I_0} P \ T b_{0i}^2 = o_p (\ln n)$. Considering $\max_{i \in I_0} P \ T b_{0i}^2$, it can be evaluated by considering the orders of magnitude of $\max_{i \in I_0} P \ T F^0 M_{X_i} F^0 F^0 I_i^2$, and of $\max_{i \in I_0} P \ T F^0 M_{X_i} F^0 F^0 I_i^2$. The former can be shown to be $o_p (\ln n)$ based on the proof of Lemma A.6(iii). The latter has the same order of magnitude as

$$T^{1/2} F^0 F^0 F^0 F^0 I_i^2$$

which is $o_p (\ln n)$ by Lemma A.3(iii). Putting all together, $\max_{i \in I_0} II_i = o_p (\ln n)$ so, $\max_{i \in I_0} II_i$ is dominated by $\max_{i \in I_0} I_i$. Similar passages yield that $\max_{i \in I_0} III_i$ is dominated by $\max_{i \in I_0} II_i$. Turning to $IV_i$, it holds that $\max_{i \in I_0} IV_i = o_p (\ln n)$, which is $o_p (\ln n)$ by Lemmas A.4 and A.7(iv). Finally, $\max_{i \in I_0} V_i$ is bounded by $\max_{i \in I_0} IV_i = o_p (\ln n)$, so, $\max_{i \in I_0} V_i$ is bounded by $\max_{i \in I_0} IV_i = o_p (\ln n)$. Putting all together, and using (35), it holds that

$$\max_{i \in I_0} T \ b_{0i}^2 = \max_{i \in I_0} I_i F^0 M_{X_i} F^0 F^0 I_i F^0 F^0 M_{X_i} F^0 F^0 I_i F^0 F^0 I_i \leq o_p (\ln n)$$

where the remainders are negligible as $(n; T) \to 1$ with $n \to \infty$ and $n \to \infty$, which hold in light of (19). Finally, consider the sequence $f \ N_c g_{-1}^0$: the covariance between $P \ T b_{0i}$ and $P \ T b_{0j}$ is given by

$$E \ F^0 M_{X_i} F^0 F^0 M_{X_i} F^0 T = \begin{cases} 1 & T = 1 \ s = 1 \ E (f_{t} f_{s}^0 I_{t} I_{s}) \\ 0 & T = 1 \ s = 1 \ k E (f_{t} f_{s}^0 I_{t} I_{s}) \end{cases}$$

$k E (f_{t} f_{s}^0 I_{t} I_{s})$ is defined as

$$k E (f_{t} f_{s}^0 I_{t} I_{s}) := \sum_{t \leq s \leq 1} E (f_{t} f_{s}^0 I_{t} I_{s})$$

$k E (f_{t} f_{s}^0 I_{t} I_{s})$ is defined as

$$k E (f_{t} f_{s}^0 I_{t} I_{s}) := \sum_{t \leq s \leq 1} E (f_{t} f_{s}^0 I_{t} I_{s})$$

$k E (f_{t} f_{s}^0 I_{t} I_{s})$ is defined as

$$k E (f_{t} f_{s}^0 I_{t} I_{s}) := \sum_{t \leq s \leq 1} E (f_{t} f_{s}^0 I_{t} I_{s})$$

$k E (f_{t} f_{s}^0 I_{t} I_{s})$ is defined as

$$k E (f_{t} f_{s}^0 I_{t} I_{s}) := \sum_{t \leq s \leq 1} E (f_{t} f_{s}^0 I_{t} I_{s})$$

$k E (f_{t} f_{s}^0 I_{t} I_{s})$ is defined as

$$k E (f_{t} f_{s}^0 I_{t} I_{s}) := \sum_{t \leq s \leq 1} E (f_{t} f_{s}^0 I_{t} I_{s})$$
which tends to zero as \((n; T) \to 1\) by Assumption 7. By virtue of the asymptotic independence between \(N_i\) and \(N_j\) for all \(i \neq j\), the asymptotics of \(\max_{1 \leq i, j \leq n} N_i \cdot f_{ei}; N_i\) is studied e.g. in Embrechts, Klüppelberg and Mikosch (1997, Table 3.4.4, p.156). Thus, equation (20) follows from (36).

We now finish the proof of the Theorem, analysing the power properties of the test. In order to evaluate the presence of power when \(i \neq j\) for some (at least one) \(i\), after some algebra it can be shown that, under the alternative, \(S_{n; nT}\) has non-centrality parameter given by

\[
S^{NC}_{n; nT} = T \max_{1 \leq i \leq n} c_i + 2T \max_{1 \leq i < j \leq n} c_i c_j (\wedge_{1 \leq i < j \leq n}) 2T \max_{1 \leq i \leq n} c_i \|_{i < j}^{\|_{i < j}} = I + II + III;
\]

with \(I = O_p(n)\) by construction. Also, \(II\) is bounded by

\[
P_T(\max_{1 \leq i \leq n} c_i k) \leq O_p(T)\max_{1 \leq i \leq n} P_{k_i} k_i \leq O_p(T) n^{1/k_1} k_1\]

in view of Lemma A.6(iii); similarly, \(III = O_p(T) n^{2/k_2} k_2\) by Lemma A.4. Let \(S_{n; nT}\) denote the null distribution of \(S_{n; nT}\); under \(H_1\) it holds that

\[
P[S_{n; nT} > C_{n; n}] = P[S_{n; nT} > C_{n; n}] = P[S^{NC}_{n; nT} > C_{n; n}] = 1.
\]

which tends to 1 if \(C_{n; n} = O(\ln n)\), whence (21) follows. QED
References

[1]


