ISSN: 2281-1346

Department of Economics and Management

DEM Working Paper Series

Inference on Factor Structures in Heterogeneous Panels

Carolina Castagnetti
(Università di Pavia)

Eduardo Rossi
(Università di Pavia)

Lorenzo Trapani
(City University, London)

 # 88 (09-14)

Via San Felice, 5
I-27100 Pavia
http://epmq.unipv.eu/site/home.html

September 2014

Inference on Factor Structures in Heterogeneous Panels

forthcoming in

Journal of Econometrics

Carolina Castagnetti
University of Pavia

Eduardo Rossi
University of Pavia

Lorenzo Trapani
Cass Business School, City University London

September 10, 2014

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:

$$y_{it} = \beta_i' x_{it} + \gamma_i' f_t + \epsilon_{it}, \tag{1}$$

$$x_{it} = \Lambda_i f_t + \epsilon_{it}^x, \tag{2}$$

where $i = 1, ..., n$, $t = 1, ..., T$, 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), Λ_i is a matrix of coefficients of dimension $m \times r$. Model (1)-(2) is based on Pesaran (2006), 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 (2009a), 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 (2007) 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 (2013) 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 (2012) adopt a common factor model approach to estimate cross-country production functions for the agriculture sector. Similarly, Eberhardt, Helmers and Strauss (2013) 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 (Lee and Carter, 1992) has been extended to incorporate idiosyncratic explanatory variables as well as the traditional factor structure - see French and O'Hare (2013) and the references therein.

As far as conducting inference on (1) is concerned, the inferential theory on the slope coefficients β_i has been developed in various contributions. Particularly, Pesaran (2006) proposes a family of estimators for β_i based on instrumenting the f_ts 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 γ_i or f_t. Pesaran and Tosetti (2011) and Castagnetti and Rossi (2013) show that, in principle, residuals computed from (1) using CCE estimators can be used to extract γ_i and f_t using e.g. Principal Components (henceforth, PC).

However, the properties of the estimated γ_i and f_t are not discussed. In addition to the CCE estimators, Bai (2009a) develops a different estimation technique for (1)-(2) under the assumption of homogeneous slopes, i.e. $\beta_i = \beta$. Such technique is known as the Interactive Effect (henceforth IE) estimator, and it is based on iteratively computing β for given values of γ_i and f_t, and then γ_i and f_t for a given value of β. Although results are available for the estimated triple (β, λ_i, f_t), inference is developed under the assumption of homogeneous β_is; moreover, no explicit asymptotics for γ_i or f_t is derived beyond consistency. Despite this, inference on γ_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 Sarafidis, Yamagata and Robertson (2009), and Bai (2009a) 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 γ_i in (1)-(2). We estimate γ_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 (2006, p.1000), and Pesaran and Tosetti (2011), while the asymptotics of the estimated (γ_i, f_t) is studied by adapting the method of proof in Bai (2009a) to the case of heterogeneous β_is.

Secondly, we develop two tests: one for the null that $\gamma_i = \gamma$ for all i, and one for the null that $f_t = f$ for all t. The rationale for these two tests can be understood by noting that, as Pesaran (2006) points out, model (1)-(2) nests various alternative specifications. In the case of homogeneous loadings (i.e. $\gamma_i = \gamma$), 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 Sarafidis, Yamagata and Robertson (2009) to test for cross dependence in a dynamic panel context. Similarly, in the case of homogeneous factors (i.e. $f_t = f$), 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 (2012), who propose an approach based on finite sample corrections and wild bootstrap to testing for $H_0 : \gamma_i = 0$ in a standard panel factor model defined as

$y_{it} = \gamma_i' f_t + \epsilon_{it}$. We use test statistics based on extrema of the estimated γ_i and f_t, in a similar fashion to the tests for slope homogeneity developed by Kapetanios (2003) and Westerlund and Hess (2011). Monte Carlo evidence shows that the tests have correct size and satisfactory power for different levels of the signal-to-noise ratio and for several simulation designs.

The paper is organized as follows. The estimation procedure, and the asymptotics of the estimates of γ_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 "$\longrightarrow$" to denote the ordinary limit; "$\xrightarrow{d}$" and "$\xrightarrow{p}$" to denote convergence in distribution and in probability respectively; and we use "a.s." as short-hand for "almost surely". The Frobenius norm of a matrix A is denoted as $\|A\| = \sqrt{tr\,(A'A)}$, where $tr\,(A)$ denotes the trace of A. Definitional equality is denoted as "$\equiv$". 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 $F = (f_1, ..., f_T)'$; $X_i = (x_{i1}, ..., x_{iT})'$; $\epsilon_i = (\epsilon_{i1}, ..., \epsilon_{iT})'$; $y_i = (y_{i1}, ..., y_{iT})'$; $z_{it} = (y_{it}, x_{it}')'$; $z_i = (z_{i1}, ..., z_{iT})'$ and $\bar{H}_w = n^{-1}\sum_{i=1}^n z_i$. We also define the matrices $\bar{M}_w = I_T - \bar{H}_w \left(\bar{H}_w'\bar{H}_w\right)^{-1} \bar{H}_w'$ and

$$C_i = [\gamma_i|\Lambda_i'] \begin{bmatrix} 1 & 0_{1\times m} \\ \beta_i & I_m \end{bmatrix},$$

for each i. Based on this, the β_is in (1) can be estimated as

$$\tilde{\beta}_i = \left(\frac{X_i'\bar{M}_w X_i}{T}\right)^{-1} \left(\frac{X_i'\bar{M}_w y_i}{T}\right), \tag{3}$$

which is the CCE estimator of Pesaran (2006); it holds that $\tilde{\beta}_i - \beta_i = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right) + O_p\left(\frac{1}{n}\right)$.

In order to estimate γ_i and f_t, we propose the following two-step procedure.

Step 1 Estimate the β_is using the CCE estimator, and compute the residuals $\tilde{v}_i = y_i - X_i\tilde{\beta}_i$.

Step 2 Apply the PC estimator to $\tilde{v}_i$, obtaining $\hat{\gamma}_i$ and $\hat{f}_t$ under the restrictions $\hat{F}'\hat{F} = TI_r$ and $n^{-1}\sum_{i=1}^n \hat{\gamma}_i\hat{\gamma}_i'$ diagonal.

In Step 2, $\hat{F}$ is calculated as $\sqrt{T}$ times the r largest eigenvectors of $\frac{1}{nT}\sum_{i=1}^{n}\tilde{v}_i\tilde{v}_i'$. Similarly, $\hat{\gamma}_i$ is computed as

$$\hat{\gamma}_i = \left(\hat{F}' M_{Xi} \hat{F}\right)^{-1} \left(\hat{F}' M_{Xi} y_i\right), \tag{4}$$

with $M_{Xi} = I_T - X_i\left(X_i'X_i\right)^{-1} X_i'$. In (1), γ_i and f_t are not separately identifiable; as is typical in this literature, we only manage to estimate a rotation of γ_i and f_t, say $H^{-1}\gamma_i$ and $H'f_t$. However, for our purposes knowing $H^{-1}\gamma_i$ and $H'f_t$ is as good as knowing γ_i 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 β_is are estimated at a rate $O_p\left[\min\left\{T^{-1/2}, n^{-1}\right\}\right]$. Thus, the CCE is only a possible choice. Alternatives, like the Song (2013) estimator, which extends Bai (2009a) IE estimator to the case of heterogeneous slopes, may be used instead. The Song (2013) 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)* $E\left(\epsilon_{it}\right) = 0$ and $E\left|\epsilon_{it}\right|^{12} < \infty$; *(ii)* (a) $\sum_{t=1}^{T}\left|E\left(\epsilon_{it}\epsilon_{is}\right)\right| \leq M$ for all i and s, (b) $\sum_{i=1}^{n}\sum_{j=1}^{n}\left|E\left(\epsilon_{it}\epsilon_{js}\right)\right| \leq Mn$ for all t and s, (c) $\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\epsilon_{it}\epsilon_{is}\right)\right| \leq MT$ for all i, (d) $\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\epsilon_{it}\epsilon_{js}\right)\right| \leq M\left(nT\right)$; *(iii)* (a) $E\left|(nT)^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\epsilon_{it}\right|^2 \leq M$, (b) $\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{v=1}^{T}\sum_{u=1}^{T}\left|E\left(\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv}\right)\right| \leq MT^2$, (c) $\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{u=1}^{T}\left|E\left(\epsilon_{it}\epsilon_{is}\epsilon_{ju}\epsilon_{js}\right)\right| \leq M\left(nT\right)$ for all u, (d) $\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\epsilon_{it}\epsilon_{kt}\epsilon_{js}\epsilon_{ks}\right)\right| \leq M\left(nT\right)$ for all k; *(iv)* (a) $E\left|\sum_{t=1}^{T}\epsilon_{it}\right|^r \leq ME\left|\sum_{t=1}^{T}\epsilon_{it}^2\right|^{r/2}$ for all i, $r < 12$, (b) $E\left|\sum_{i=1}^{n}\epsilon_{it}\right|^r \leq ME\left|\sum_{i=1}^{n}\epsilon_{it}^2\right|^{r/2}$ for all t, $r < 12$.

Assumption 2. *[regressors and common factors] (i)* $E\left\|\epsilon_{it}^x\right\|^{12} < \infty$ and $E\left\|f_t\right\|^{12} < \infty$; *(ii)* $T^{-1}\sum_{t=1}^{T}f_tf_t' \xrightarrow{p} \Sigma_f$ as $T \to \infty$ with Σ_f non-singular; *(iii)* $\left\{\epsilon_{it}^x, f_t\right\}$ and $\left\{\epsilon_{js}\right\}$ are mutually independent for all i, j, t, s; *(iv)* $E\left|\sum_{t=1}^{T}x_{it}\epsilon_{it}\right|^r \leq ME\left|\sum_{t=1}^{T}\left(x_{it}\epsilon_{it}\right)^2\right|^{r/2}$ for all i, $r \leq 6$.

Assumption 3. *[slopes and loadings] (i)* $\left\{\beta_i\right\}$ is independent of $\left\{\epsilon_{jt}, \epsilon_{jt}^x, f_t\right\}$ for all i, j, t; *(ii)* $E\left\|\beta_i\right\|^{2+\delta} < \infty$ for some $\delta > 0$; *(iii)* the γ_is are non stochastic and such that $\max_i\left\|\gamma_i\right\| < \infty$ and $n^{-1}\sum_{i=1}^{n}\gamma_i\gamma_i' \to \Sigma_\gamma$ as $n \to \infty$ with Σ_γ non-singular.

Assumption 4. *[Step 1 estimation] (i)* $l_{\min}\left(\frac{X_i'\bar{M}_wX_i}{T}\right) > 0$; $l_{\min}\left(\frac{X_i'M_FX_i}{T}\right) > 0$ and $l_{\min}\left(\frac{F'M_{Xi}F}{T}\right) > 0$ a.s. for all i, where $l_{\min}\left(\cdot\right)$ denotes the smallest eigenvalue; *(ii)* $C \equiv n^{-1}\sum_{i=1}^{n}C_i$ has rank $r \leq m+1$.

Assumption 5. *[Central Limit Theorems] (i)* (a) there exists a nonrandom, positive definite matrix $\Sigma_{fM,i}$ such that $p\lim_{T\to\infty} T^{-1}\ F'H'M_{Xi}HF = \Sigma_{fM,i}$, (b) $T^{-1/2}F'H'M_{xi}\epsilon_i \xrightarrow{d} N\left(0, \Sigma_{fMe,i}\right)$, where $\Sigma_{fMe,i} = p\lim_{T\to\infty} T^{-1}\ F'H'M_{xi}\epsilon_i\epsilon_i'M_{xi}HF$, for all i; *(ii)* $n^{-1/2}\sum_{i=1}^{n}\gamma_i\epsilon_{it} \xrightarrow{d} N\left(0, \Phi_{\gamma\epsilon,t}\right)$, where $\Phi_{\gamma\epsilon,t}$

$= p\lim_{n\to\infty} n^{-1} \gamma_i\gamma_i'\epsilon_{it}\epsilon_{it}$, for all t.

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 ϵ_{it}, 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 (2003) and Bai (2009a), and can be shown immediately if ϵ_{it} is assumed to be independent. Part *(i)* requires the existence of the 12-th moment of ϵ_{it}, which is stronger than what the literature normally considers - e.g. in Bai (2009a), assuming $E\left|\epsilon_{it}\right|^8 < \infty$ suffices. In our context, the existence of the 12-th moment is needed in order to derive consistency of $\hat{\gamma}_i$ 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 ϵ_{it} 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 (2010, p.108).

As far as Assumption 2 is concerned, we allow for serial and cross sectional dependence in both the ϵ_{it}^xs and in the common factors f_t. The requirement in part *(ii)* is standard in the literature (see e.g. Assumption B in Bai, 2009a), and it entails that common factors are "strong" in the sense of Chudik, Pesaran and Tosetti (2011) (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 β_is, employed in Step 1. Particularly, the rank condition in part *(ii)* is the same as equation (21) in Pesaran (2006), and it guarantees the consistency of the $\tilde{\beta}_i$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 (2003).

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

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

$$\hat{\gamma}_i - H^{-1}\gamma_i = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{n}\right). \tag{5}$$

Let Assumptions 1-5 hold. As $(n,T)\to\infty$ *with* $\frac{\sqrt{T}}{n}\to 0$

$$\sqrt{T}\left(\hat{\gamma}_i - H^{-1}\gamma_i\right) \xrightarrow{d} N\left(0, \Sigma_{\gamma i}\right), \tag{6}$$

where $\Sigma_{\gamma i} = \Sigma_{fM,i}^{-1}\Sigma_{fMe,i}\Sigma_{fM,i}^{-1}$ *and* $\Sigma_{fM,i}$ *and* $\Sigma_{fMe,i}$ *are the probability limits of* $T^{-1}\left(F'H'M_{Xi}HF\right)$ *and* $T^{-1}\left(F'H'M_{Xi}\epsilon_i\epsilon_i'M_{Xi}HF\right)$*, 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 $\sqrt{T}\left(\hat{\gamma}_i - H^{-1}\gamma_i\right)$ in (6) is different from the one in Theorem 2 in Bai (2003): this is due to the presence, in our context, of the idiosyncratic regressors x_{it}.

We use the estimator of $\Sigma_{\gamma i}$ proposed in (Bai, 2003, p.150)

$$\hat{\Sigma}_{\gamma_i} = \left(Q_i'\right)^{-1}\Phi_i\left(Q_i\right)^{-1} \tag{7}$$

where $Q_i = T^{-1}(\hat{F}'M_{Xi}\hat{F})$, and $\Phi_i = D_{0,i} + \sum_{j=1}^{q}\left(1-\frac{j}{q+1}\right)\left(D_{j,i}+D_{j,i}'\right)$, with $D_{j,i} = T^{-1}\sum_{t=j+1}^{T}$ $\widehat{fx}_t'\ \widehat{fx}_{t-j}\ \hat{\epsilon}_{it}\ \hat{\epsilon}_{it-j}$, where $\widehat{fx}_t$ is the t-th row of $M_{Xi}\hat{F}$ and $\hat{\epsilon}_{it} = y_{it} - \hat{\beta}_i'x_{it} - \hat{\gamma}_i'\hat{f}_t$. The bandwidth q is chosen so that $q\to\infty$ with $q/T^{1/4}\to 0$.

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

Theorem 2 *Let Assumptions 1-4 hold; then, for every t*

$$\hat{f}_t - H'f_t = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{T}\right). \tag{8}$$

Let Assumptions 1-5 hold. As $(n,T)\to\infty$ *with* $\frac{\sqrt{n}}{T}\to 0$

$$\sqrt{n}\left(\hat{f}_t - H'f_t\right) \xrightarrow{d} N\left(0, \Sigma_{ft}\right), \tag{9}$$

where $\Sigma_{ft} = H\Sigma_f\Sigma_{\Gamma\epsilon,t}\Sigma_f H'$ *and* $\Sigma_{\Gamma\epsilon,t} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_i\gamma_j'\epsilon_{it}\epsilon_{jt}$.

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 $\sqrt{n}\left(\hat{f}_t - H'f_t\right)$ can be estimated using equation (7) in Bai (2003, p.150). Specifically, letting $\hat{\epsilon} = \left(\hat{\epsilon}_1, ..., \hat{\epsilon}_n\right)'$ with $\hat{\epsilon}_i = \left[\hat{\epsilon}_{i1}, ..., \hat{\epsilon}_{iT}\right]'$, and defining

V_{nT} as a diagonal matrix containing the r largest eigenvalues of $\frac{1}{nT}\hat{\epsilon}\hat{\epsilon}'$ in descending order, the estimated Σ_{ft} is

$$\hat{\Sigma}_{ft} = V_{nT}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_i\hat{\gamma}_i'\hat{\epsilon}_{it}^2\right)V_{nT}^{-1}. \tag{10}$$

Note that $\Sigma_{\Gamma\epsilon,t}$ is estimated through $n^{-1}\sum_{i=1}^{n}\hat{\gamma}_i\hat{\gamma}_i'\hat{\epsilon}_{it}^2$, which is valid under cross sectional independence. It is not possible, in general, to estimate $\Sigma_{\Gamma\epsilon,t}$ 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 $c_{it} = \gamma_i' f_t$, defined as $\hat{c}_{it} = \hat{\gamma}_i'\hat{f}_t$.

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

$$\hat{c}_{it} - c_{it} = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right). \tag{11}$$

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

$$\left(\frac{1}{n}\gamma_i'\Sigma_{ft}\gamma_i + \frac{1}{T}f_t'\Sigma_{\gamma i}f_t\right)^{-1/2}(\hat{c}_{it} - c_{it}) \xrightarrow{d} N(0,1), \tag{12}$$

where Σ_{ft} is defined in Theorem 2 and $\Sigma_{\gamma i}$ in Theorem 1.

After discussing the asymptotic properties of $\hat{\gamma}_i$ 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 Sarafidis, Yamagata and Robertson (2009), we study tests for, respectively: *(a)* the null of cross-sectional homogeneity of the loadings γ_is; and *(b)* the null of homogeneity, over time, of the f_ts.

Formally, we propose two tests for the null hypotheses:

$$H_0^a \quad : \quad \gamma_i = \gamma \text{ for all } i; \tag{13}$$

$$H_0^b \quad : \quad f_t = f \text{ for all } t. \tag{14}$$

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

$$y_{it} = \varphi_t + \beta_i' x_{it} + \epsilon_{it}, \tag{15}$$

where we have defined $\varphi_t = \gamma' f_t$. Thus, under H_0^a, model (1) boils down to a standard panel specification with a time effect. Similarly, under H_0^b in (14), equation (1) can be rewritten as

$$y_{it} = \varphi_i + \beta_i' x_{it} + \epsilon_{it}, \tag{16}$$

where we have defined $\varphi_i = \gamma_i' f$. Therefore, under H_0^b, 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. Sarafidis, Yamagata and Robertson (2009) propose a test for cross dependence (albeit in a different context) based on verifying the null that loadings are homogeneous, i.e. $\gamma_i = \gamma$. Our paper extends the contribution by Sarafidis, Yamagata and Robertson (2009) 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 (2009a).

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 γ_i and f_t. Specifically, we propose two max-type statistics, where the maximum is taken over the deviation of the individual estimate of γ_i (resp. of f_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 (2011), whose simulations show that the power properties are very promising, although issues may arise in presence of ties (Hall and Miller, 2010). 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 (2014b) study the use of alternative test statistics for H_0^a and H_0^b - 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 $(n, T) \to \infty$, while Hausman-type ones are inconsistent.

Define $\widehat{\overline{\gamma}} = n^{-1}\sum_{i=1}^{n}\hat{\gamma}_i$ and $\widehat{\overline{f}} = T^{-1}\sum_{t=1}^{T}\hat{f}_t$. We propose the following max-type test statistics:

$$S_{\gamma,nT} \equiv \max_{1\leq i\leq n}\left[T\left(\hat{\gamma}_i - \widehat{\overline{\gamma}}\right)'\hat{\Sigma}_{\gamma i}^{-1}\left(\hat{\gamma}_i - \widehat{\overline{\gamma}}\right)\right], \tag{17}$$

$$S_{f,nT} \equiv \max_{1\leq t\leq T}\left[n\left(\hat{f}_t - \widehat{\overline{f}}\right)'\hat{\Sigma}_{ft}^{-1}\left(\hat{f}_t - \widehat{\overline{f}}\right)\right]. \tag{18}$$

We point out that under the null hypotheses H_0^a and H_0^b, the spaces spanned by the loadings and by the factors (respectively) have rank equal to one. This fact was already noted by Sarafidis, Yamagata and Robertson (2009) who, building on it, suggest running their test setting $r = 1$. This can be applied to our context also: $S_{\gamma,nT}$ and $S_{f,nT}$ 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 (2002) and Bai (2009b). The asymptotic properties of the estimated common factors, loadings and common components are those given in Section 2.

3.1 Testing for $H_0^a : \gamma_i = \gamma$

In this section we report the asymptotics of $S_{\gamma,nT}$ under the null H_0^a, and we analyse the consistency of tests based on $S_{\gamma,nT}$. We show that, as $(n,T)\to\infty$ under some restrictions on the relative speed of divergence, $S_{\gamma,nT}$ (suitably normalised) converges to a Gumbel distribution. Further, we also show that tests based on $S_{\gamma,nT}$ have nontrivial power versus alternative hypotheses shrinking at a rate $O_p\left(\sqrt{\frac{\ln n}{T}}\right)$.

Let k_1 be the largest number for which $E\left|\epsilon_{it}\right|^{k_1}$, $E\left\|x_{it}\right\|^{k_1}$ and $E\left\|f_t\right\|^{k_1}$ are finite. In view of Assumption 1, $k_1 \geq 12$. 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 $\alpha \in (1, +\infty)$: *(i)* ϵ_{it}, f_t and x_{it} are $L_{2+\delta}$-NED (Near Epoch Dependent) of size α on a uniform mixing base $\{v_t\}_{t=-\infty}^{+\infty}$ of size $-r/(r-2)$ and $r > \frac{2\alpha-1}{\alpha-1}$; *(ii)* (a) letting $V_{iT}^{f\epsilon} \equiv T^{-1}\ E\left[\left(\sum_{t=1}^{T} f_t\epsilon_{it}\right)\left(\sum_{t=1}^{T} f_t\epsilon_{it}\right)'\right]$, $V_{iT}^{f\epsilon}$ is positive definite uniformly in T, and as $T\to\infty$, $V_{iT}^{f\epsilon} \to V_i^{f\epsilon}$ with $\left\|V_i^{f\epsilon}\right\| < \infty$, (b) the same holds for $V_{iT}^{x\epsilon} \equiv T^{-1}$

$E\left[\left(\sum_{t=1}^{T} x_{it}\epsilon_{it}\right)\left(\sum_{t=1}^{T} x_{it}\epsilon_{it}\right)'\right]$, $V_{iT}^{fx} \equiv T^{-1}E\left(\bar{w}_{iT}^{fx}\bar{w}_{iT}^{fx\prime}\right)$ with $\bar{w}_{iT}^{fx} = vec\left(\sum_{t=1}^{T} f_t x_{it}'\right) - E\left[vec\left(\sum_{t=1}^{T} f_t x_{it}'\right)\right]$, and $V_{iT}^{xx} = T^{-1}E\left(\bar{w}_{iT}^{xx}\bar{w}_{iT}^{xx\prime}\right)$ with $\bar{w}_{iT}^{xx} = vec\left(\sum_{t=1}^{T} x_{it}x_{it}'\right) - E\left[vec\left(\sum_{t=1}^{T} x_{it}x_{it}'\right)\right]$; *(iii)* (a) letting $w_{kt}^{f\epsilon}$ be the k-th element of $f_t\epsilon_{it}$ and defining $S_{kT,m}^{f\epsilon} \equiv \sum_{t=m+1}^{m+T} w_{kt}^{f\epsilon}$, there exists a positive definite matrix $\bar{\Omega}^{f\epsilon} = \left\{\varpi_{kh}^{f\epsilon}\right\}$ such that $T^{-1}\left|E\left[S_{kT,m}^{f\epsilon}S_{hT,m}^{f\epsilon}\right] - \varpi_{kh}^{f\epsilon}\right| \leq MT^{-\psi}$, for all k and h and uniformly in m, with $\psi > 0$, (b) the same holds for $x_{it}\epsilon_{it}$.

Assumption 7. *[cross sectional dependence]* It holds that $T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\epsilon_{it}\epsilon_{js}\right)\right|\ln n \to 0$ as $(n,T) \to \infty$ for all $i \neq j$.

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 ϵ_{it}, f_t and 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 (1986); see also Assumption A.3 in Corradi (1999).

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 $T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\epsilon_{it}\epsilon_{js}\right)\right|$ declines (faster than $\ln n$) as n passes to infinity. This assumption is similar to the so-called "Berman condition" (Berman, 1964), which is employed in EVT for dependent time series data; we refer to Assumption 9 below for further explanations on how the Berman condition works in the case of time series data. By way of comparison, Assumption 7 can be viewed as a complement to Assumption 1*(ii)*(d), since it contains the same summation across t. As far as the amount of cross sectional dependence is concerned, the assumption is quite weak; as an example, it would be satisfied if $T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\epsilon_{it}\epsilon_{js}\right)\right| = o\left(\ln^{-1} n\right)$ for all $i \neq j$, which is a much weaker requirement than the one in Assumption 1*(ii)*(d).

Let the critical value $c_{\alpha,n}$ be defined such that $P\left(S_{\gamma,nT} \leq c_{\alpha,n}\right) = 1-\alpha$ under H_0^a, and let $\Gamma(\cdot)$ denote the Gamma function. It holds that:

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

$$\frac{\sqrt{T}n^{2/k_1}}{n} + \frac{n^{4/k_1}}{T} \to 0. \tag{19}$$

Under H_0^a, it holds that

$$P\left(A_n S_{\gamma,nT} \leq x + B_n\right) = e^{-e^{-x}}, \tag{20}$$

where $A_n = \frac{1}{2}$ *and* $B_n = \ln(n) + \left(\frac{r}{2} - 1\right) \ln\ln(n) - \ln\Gamma\left(\frac{r}{2}\right)$. *Under the alternative* $H_1^a : \gamma_i = \gamma + c_i$ *for at least one* i, *if*

$$\frac{T}{\ln n} \left\|c_i\right\|^2 \to \infty, \tag{21}$$

it holds that $P\left(S_{\gamma,nT} > c_{\alpha,n}\right) = 1$.

Theorem 3 states that $S_{\gamma,nT}$ has a Gumbel distribution. This holds in the joint limit $(n, T) \to \infty$, with the restrictions specified in (19). Since $k_1 \geq 12$, the latter condition requires $\frac{T}{n^{5/3}} \to 0$, which is marginally stricter than the condition $\frac{\sqrt{T}}{n} \to 0$ needed in for (6). Also, (19) needs that $\frac{n^{4/k_1}}{T} \to 0$; this becomes, under Assumptions 1*(i)* and 2*(i)*, $\frac{n}{T^3} \to 0$.

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

$$c_{\alpha,n} = 2B_n - \ln\left|\ln(1-\alpha)\right|^2. \tag{22}$$

Thus, for a given level α, $c_{\alpha,n}$ 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 (1987) shows that the rate of convergence for the maximum of a sequence of random variables following a Gamma distribution is $O\left(1/\ln^2 n\right)$. Unreported Monte Carlo evidence shows that tests based on using $c_{\alpha,n}$ perform quite well, although they are a bit oversized. As an alternative, one can replace B_n with $F_{\chi_r}^{-1}(1 - 1/n)$, where $F_{\chi_r}^{-1}(\cdot)$ is the inverse of the cumulative distribution function of a chi-square with r degrees of freedom, see Embrechts, Klüppelberg and Mikosch (1997).

As far as consistency of the test is concerned, equation (21) shows that nontrivial power is attained versus local alternatives shrinking at a rate $O_p\left(\sqrt{\frac{\ln n}{T}}\right)$. Thus, when using max-type statistics such as $S_{\gamma,nT}$, 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 γ_i is different from the others.

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

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

Let k_2 be the largest number such that $E\left\|f_t\right\|^{k_2}$, $E\left\|x_{it}\right\|^{k_2}$ and $E\left|\epsilon_{it}\right|^{k_2}$ are all finite. In view of Assumptions 1 and 2, $k_2 \geq 12$. 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 $\delta > 0$ and $\alpha \in (1, +\infty)$: *(i)* ϵ_{it} is $L_{2+\delta}$-NED across i, of size α on a uniform mixing base $\{v_i\}_{i=-\infty}^{+\infty}$ of size $-r/(r-2)$ and $r > \frac{2\alpha-1}{\alpha-1}$; *(ii)* letting $V_{tn}^{\epsilon\epsilon} = n^{-1} E\left[\left(\sum_{i=1}^{n} \epsilon_{it}\right)\left(\sum_{i=1}^{n} \epsilon_{it}\right)\right]$, $V_{tn}^{\epsilon\epsilon}$ is positive definite uniformly in n, and as $n \rightarrow \infty$, $V_{tn}^{\epsilon\epsilon} \rightarrow V_{t}^{\epsilon\epsilon}$ with $\left\|V_{t}^{\epsilon\epsilon}\right\| < \infty$; *(iii)* letting $S_{mt}^{\epsilon} = \sum_{i=m+1}^{m+n} \epsilon_{it}$ there exists a positive constant $\varpi^{\epsilon\epsilon}$ such that $n^{-1}\left|E\left(S_{mt}^{\epsilon 2}\right) - \varpi^{\epsilon\epsilon}\right| \leq Mn^{-\psi''}$ uniformly in m, with $\psi'' > 0$.

Assumption 9. *[serial dependence]* It holds that $\lim_{k\rightarrow\infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \left|E\left(\epsilon_{it}\epsilon_{jt-k}\right)\right| \ln k = 0$ as $(n, T) \rightarrow \infty$.

Assumption 8 is very similar, in spirit, to Assumption 6, and it requires that ϵ_{it} is NED across i. By virtue of Assumption 8, an a.s. IP holds for $\sum_{i=1}^{n} \epsilon_{it}$ and for $\sum_{i=1}^{n} \epsilon_{it}^{2}$. The definition of NED for spatial processes has been studied in Jenish and Prucha (2012), and we refer to that paper for details.

Assumption 9 is the so-called "Berman condition" (Berman, 1964): 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. Leadbetter and Rootzen, 1988). 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. Hannan and Kavalieris, 1986), 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 $c_{\alpha,T}$ be defined such that $P\left(S_{f,nT} \leq c_{\alpha,T}\right) = 1 - \alpha$ under H_0^b. It holds that:

Theorem 4 *Let Assumptions 1-4 hold and 8-9, and let* $(n, T) \rightarrow \infty$ *with*

$$\frac{\sqrt{n}T^{1/k_2}}{T} + \frac{T^{4/k_2}}{n} \rightarrow 0. \tag{23}$$

Under H_0^b, it holds that

$$P\left[A_T S_{f,nT} \leq x + B_T\right] = e^{-e^{-x}}, \tag{24}$$

where $A_T = \frac{1}{2}$ and $B_T = \ln(T) + \left(\frac{r}{2} - 1\right)\ln\ln(T) - \ln\Gamma\left(\frac{r}{2}\right)$. Under the alternative $H_1^b : f_t = f + c_t$ for at least one t, if

$$\frac{n}{\ln T}\left\|c_t\right\|^2 \to \infty, \tag{25}$$

it holds that $P\left(S_{f,nT} > c_{\alpha,T}\right) = 1$.

Theorem 4 is very similar to Theorem 3; convergence to the Gumbel distribution under the null is shown for $(n, T) \to \infty$ jointly under some restrictions between n and T, spelt out in (23). Specifically, it is required that $\frac{T^{1/k_2}\sqrt{n}}{T} \to 0$; since $k_2 \geq 12$, the former restriction is, at most, $\frac{n}{T^{11/6}} \to 0$. This is only marginally stronger than $\frac{\sqrt{n}}{T} \to 0$, which is required for (9) to hold. Similarly, requiring that $\frac{T^{4/k_2}}{n} \to 0$ entails $\frac{T}{n^3} \to 0$. 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 α can be calculated as

$$c_{\alpha,T} = 2B_T - \ln\left|\ln(1-\alpha)\right|^2; \tag{26}$$

alternatively, B_T can be approximated by $F_{\chi_r}^{-1}(1 - 1/T)$.

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

4 Small sample properties

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

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

$$y_{it} = \beta_i x_{it} + \gamma_i f_t + \epsilon_{it}, \tag{27}$$

$$x_{it} = \mu_i + \lambda_i f_t + \epsilon_{it}^x, \tag{28}$$

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 β_i and μ_i as *i.i.d.* $N(1,1)$. The common factor f_t, the loading λ_i, and both error terms ϵ_{it} and ϵ_{it}^x are all generated as *i.i.d.* $N(0,1)$ unless otherwise stated. Results are reported for $(n,T) \in \{30,50,100,200\} \times \{30,50,100,200\}$. Finally, in both exercises, simulations are carried out with 5000 iterations.

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

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

As far as $\hat{f}_t$ is concerned, we follow the same logic as in Bai (2003). We compute the correlation coefficient between $\{\hat{f}_t\}_{t=1}^T$ and $\{f_t\}_{t=1}^T$, for each Monte Carlo iteration j - say ρ_j^f. We report the average correlation coefficients, i.e. $J^{-1}\sum_{j=1}^J \rho_j^f$, in Table 1 (recall that $J = 5000$).
Table 1 illustrates that the estimated common factor $\hat{f}_t$ is highly correlated with the unobserved common factor f_t. This reinforces the results in Bai (2003), 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 ≥ 100, the estimated factors can be treated as the true ones.

n \ T	30	50	100	200
30	0.977	0.964	0.979	0.974
50	0.976	0.963	0.989	0.970
100	0.991	0.987	0.988	0.991
200	0.992	0.994	0.997	0.997

Table 1: Average correlation coefficients between $\{\hat{f}_t\}_{t=1}^T$ and $\{f_t\}_{t=1}^T$.

As far as $\hat{\gamma}_i$ is concerned, we report confidence intervals for γ_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 $(n,T) \to \infty$ with $\frac{\sqrt{T}}{n} \to 0$, the 95% confidence interval for $H^{-1}\gamma_i$ is given by $\hat{\gamma}_i \pm \frac{1.96}{\sqrt{T}} \times \hat{\Sigma}_{\gamma i}^{1/2}$. Further, let $\hat{\delta}$ be the least square estimate of δ in $\Gamma = \hat{\Gamma}\delta + error$, where $\Gamma = (\gamma_1, \ldots \gamma_n)'$ and $\hat{\Gamma} = (\hat{\gamma}_1, ..., \hat{\gamma}_n)'$. The 95% confidence interval for γ_i is therefore obtained as $\hat{\delta} \times \left(\hat{\gamma}_i \pm \frac{1.96}{\sqrt{T}} \times \hat{\Sigma}_{\gamma i}^{1/2}\right)$. By rotating $\hat{\gamma}_i$ towards γ_i, we consider the confidence interval for γ_i directly, reported in Figure 1.

Figure 1 shows that, in most cases and for all combinations of n and T, the confidence intervals

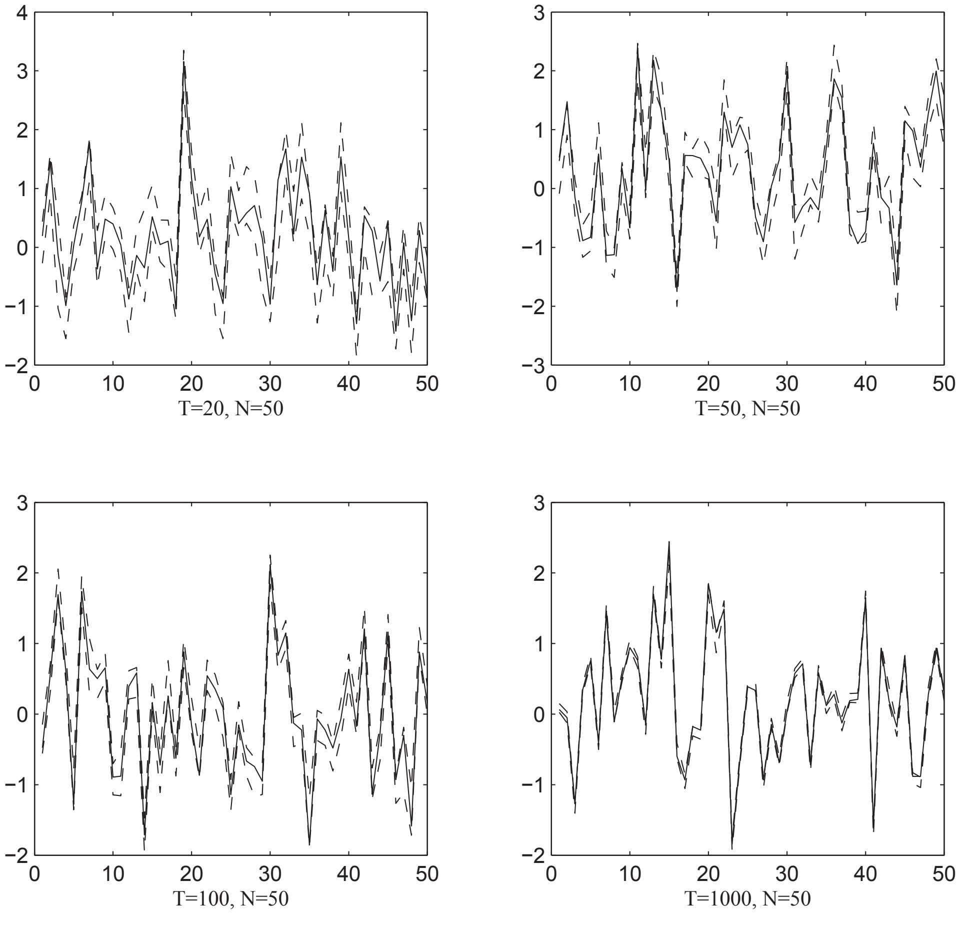

Figure 1: Confidence intervals for γ_i. For each value of $i = 1, ..., 50$ (on the horizontal axis), the solid line represents the true loading γ_i. The dashed lines are the confidence intervals at 95% confidence level for each i.

contain the true value of γ_i. This also holds true for the case $(n,T) = (50,1000)$, where the ratio $\frac{\sqrt{T}}{n}$ 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 γ_i.

4.2 Small sample properties - $S_{\gamma,nT}$ and $S_{f,nT}$

In this subsection, we report empirical rejection frequencies and power for tests based on the max-type statistics $S_{\gamma,nT}$ and $S_{f,nT}$ defined in (17) and (18) respectively.

As far as the design of the Monte Carlo is concerned, recall that the variance of the common components $c_{it} = \gamma_i f_t$ is set equal to 1 across all experiments. We conduct our simulations for different values of the signal-to-noise ratio $\frac{Var(c_{it})}{\sigma^2_\epsilon}$, where σ^2_ϵ is the variance of ϵ_{it}, equal to $\left\{\frac{1}{3}, \frac{1}{2}, 1\right\}$.

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 $H^a_0 : \gamma_i = \gamma$

When evaluating the empirical rejection frequencies for tests based on $S_{\gamma,nT}$, we run the Monte Carlo simulations under the null $\gamma_i = 1$ for all i. When evaluating power, we generate the loadings γ_i as *i.i.d.* $N\left(1, \sigma^2_\gamma\right)$, reporting results for the case of $\sigma_\gamma = 0.2$. Given that ϵ_{it} is cross sectionally uncorrelated and homoskedastic by design, $\Sigma_{\gamma i}$ is estimated as $\hat{\Sigma}_{\gamma i} = \hat{\sigma}^2_\epsilon \times T\left(\hat{F}' M_{x_i} \hat{F}\right)^{-1}$, where $\hat{\sigma}^2_\epsilon = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \hat{\epsilon}^2_{it}$.

Results for size and power when using the main DGP (52)-(53) are in Table 2.

We firstly 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 $T \geq 100$ and $n \geq 50$; even when $\sigma^2_\epsilon = 1$ (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 σ^2_ϵ increases), the tendency towards small sample oversizement worsens. This is not so when $T \geq 100$ and $n \geq 50$: the test attains the correct size even for large values of σ^2_ϵ.

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 $H^b_0 : f_t = f$

We run the Monte Carlo simulations under the null $f_t = 1$ for all t when evaluating the size of tests based on $S_{f,nT}$. When evaluating the power, we generate the common factors f_t as *i.i.d.* $N\left(1, \sigma^2_f\right)$,

[1] The simulation results are available upon request.

	Size				Power			
	T				T			
n	30	50	100	200	30	50	100	200
	$\sigma_\epsilon^2 = 1/3$				$\sigma_\epsilon^2 = 1/3$			
30	0.077	0.066	0.060	0.056	0.950	0.996	1.000	1.000
50	0.073	0.063	0.050	0.056	0.986	0.999	1.000	1.000
100	0.073	0.063	0.052	0.045	0.997	1.000	1.000	1.000
200	0.072	0.062	0.053	0.042	0.998	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1/2$				$\sigma_\epsilon^2 = 1/2$			
30	0.086	0.074	0.064	0.059	0.867	0.968	0.999	1.000
50	0.078	0.067	0.053	0.058	0.926	0.993	1.000	1.000
100	0.074	0.063	0.053	0.046	0.972	1.000	1.000	1.000
200	0.073	0.064	0.054	0.042	0.992	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1$				$\sigma_\epsilon^2 = 1$			
30	0.109	0.094	0.081	0.076	0.612	0.800	0.976	0.999
50	0.090	0.079	0.065	0.068	0.667	0.883	0.993	1.000
100	0.085	0.070	0.058	0.051	0.764	0.952	1.000	1.000
200	0.076	0.067	0.057	0.044	0.863	0.983	1.000	1.000

Table 2: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^a : \gamma_i = \gamma$, based on $S_{\gamma,nT}$. The DGP used in the simulations is (52)- (53).

reporting results for the case of $\sigma_f = 0.2$. Finally, we estimate Σ_{ft} as $\Sigma_{ft} = V_{nT}^{-1} \hat{\sigma}_\epsilon^2 \frac{1}{n} \sum_{i=1}^n \hat{\lambda}_i \hat{\lambda}_i' V_{nT}^{-1}$ where $\hat{\sigma}_\epsilon^2 = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \hat{\epsilon}_{it}^2$.

Results when using (52)-(53) are shown in Table 3.

It can be noted that the test is slightly undersized for large T, e.g. $T \geq 100$. 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 (2013). The size and power reported in Table 4, for the S_γ test, when the DGP is the one in equations (52)-(53), 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 autocor-

	Size				Power			
	T				T			
n	30	50	100	200	30	50	100	200
	$\sigma_\epsilon^2 = 1/3$				$\sigma_\epsilon^2 = 1/3$			
30	0.044	0.037	0.037	0.030	0.915	0.959	0.988	0.996
50	0.038	0.034	0.036	0.033	0.993	0.999	1.000	1.000
100	0.042	0.041	0.036	0.032	1.000	1.000	1.000	1.000
200	0.046	0.043	0.038	0.036	1.000	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1/2$				$\sigma_\epsilon^2 = 1/2$			
30	0.047	0.036	0.037	0.030	0.773	0.860	0.935	0.970
50	0.040	0.035	0.036	0.033	0.957	0.987	0.998	1.000
100	0.042	0.042	0.037	0.033	0.999	1.000	1.000	1.000
200	0.047	0.044	0.038	0.037	1.000	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1$				$\sigma_\epsilon^2 = 1$			
30	0.054	0.042	0.038	0.032	0.467	0.525	0.635	0.733
50	0.047	0.038	0.039	0.035	0.703	0.822	0.912	0.962
100	0.049	0.047	0.038	0.035	0.967	0.994	0.999	1.000
200	0.055	0.050	0.041	0.040	1.000	1.000	1.000	1.000

Table 3: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^b : f_t = f$, based on $S_{f,nT}$. The DGP used in the simulations is (52)-(53).

	Size				Power			
	T				T			
n	30	50	100	200	30	50	100	200
	$\sigma_\epsilon^2 = 1/3$				$\sigma_\epsilon^2 = 1/3$			
30	0.070	0.058	0.059	0.054	0.975	0.998	1.000	1.000
50	0.072	0.068	0.053	0.049	0.991	0.999	1.000	1.000
100	0.080	0.064	0.054	0.050	0.998	1.000	1.000	1.000
200	0.079	0.064	0.054	0.046	1.000	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1/2$				$\sigma_\epsilon^2 = 1/2$			
30	0.075	0.060	0.058	0.051	0.901	0.983	0.999	1.000
50	0.073	0.069	0.054	0.049	0.952	0.997	1.000	1.000
100	0.077	0.064	0.053	0.048	0.983	1.000	1.000	1.000
200	0.077	0.063	0.054	0.044	0.996	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1$				$\sigma_\epsilon^2 = 1$			
30	0.088	0.073	0.069	0.062	0.646	0.846	0.986	1.000
50	0.079	0.074	0.060	0.055	0.723	0.917	0.997	1.000
100	0.080	0.066	0.055	0.052	0.820	0.972	1.000	1.000
200	0.079	0.063	0.055	0.049	0.902	0.993	1.000	1.000

Table 4: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^a : \gamma_i = \gamma$, based on $S_{\gamma,nT}$. The DGP used in the simulations is (52)- (53). The first-step estimator is the one proposed Song (2013).

related and heteroskedastic, we consider the following DGP:

$$\epsilon_{it} = 0.5\epsilon_{it-1} + u_{it}$$

$$u_{it} \sim IIDN(0, \sigma^2_{ui}) \quad \sigma^2_{ui} \sim U(0.1, 0.5)$$

and we make use of the HAC estimators for Σ_γ and Σ_f given by equations (7) and (10). Apart from these features, the experiments have the same specifications as above, with $Var(\epsilon_{it}) \in (0.13, 0.67)$.

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, $(n, T) \leq 50$. 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.

	Size				Power			
	T				T			
n	30	50	100	200	30	50	100	200
30	0.103	0.087	0.088	0.100	0.838	0.905	0.966	0.994
50	0.090	0.083	0.078	0.074	0.956	0.988	0.999	1.000
100	0.081	0.071	0.063	0.065	0.999	1.000	1.000	1.000
200	0.072	0.061	0.063	0.054	1.000	1.000	1.000	1.000

Table 5: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^b : \gamma_i = \gamma$. The test is computed using the estimator of $\Sigma_{\gamma i}$ in (7).

	Size				Power			
	T				T			
n	30	50	100	200	30	50	100	200
30	0.118	0.121	0.144	0.159	0.976	0.985	0.985	0.987
50	0.080	0.063	0.069	0.082	0.985	0.991	0.992	0.995
100	0.050	0.046	0.036	0.040	0.997	0.998	0.999	0.999
200	0.061	0.039	0.036	0.036	0.999	1.000	1.000	1.000

Table 6: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^b : f_t = f$. The test is computed using the estimator of Σ_{ft} 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 (2006).

We extend the framework in Pesaran (2006) 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 (2009a). In a similar vein to Sarafidis, Yamagata and Robertson (2009), 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 specific 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 β_is used in Step 1. In our paper, we focus on the CCE estimator proposed by Pesaran (2006); 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_{it}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 Everaert and Groote, 2012; and Chudik and Pesaran, 2013). 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 β_is that is consistent at a rate $O_p\left[\min\left\{T^{-1/2}, n^{-1}\right\}\right]$. A possible choice for this case is the IE estimator studied in

Song (2013), 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_{jt} for each unit i, with $i \neq j$ - indeed, both the CCE and the IE have a natural Instrumental Variable interpretation (see also Bai, 2009b). 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, December 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 Hórvath 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 (2014b).

Lemma A.1 *Under Assumptions 1-4, it holds that, for every i,* $E\left\|\tilde{\beta}_i - \beta_i\right\|^r = O\left(\phi_{nT}^{-r}\right)$*, for any* $r \leq 3$.

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

A.2*(i)* $T^{-1}\epsilon_i'\left(\hat{F} - F\right) = O_p\left(\delta_{nT}^{-2}\right)$;

A.2*(ii)* $n^{-1/2}T^{-1}\sum_{i=1}^n \epsilon_i'\left(\hat{F} - F\right) = O_p\left(n^{-1/2}\right) + O_p\left(T^{-1}\right)$.

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

A.3*(i)* $T^{-1}X_i'\left(\hat{F} - FH\right) = O_p\left(\delta_{nT}^{-2}\right)$;

A.3*(ii)* $T^{-1}F'\left(\hat{F} - FH\right) = O_p\left(\delta_{nT}^{-2}\right)$;

A.3*(iii)* $T^{-1}\left(\hat{F} - FH\right)'\left(\hat{F} - FH\right) = O_p\left(\delta_{nT}^{-2}\right)$.

Lemma A.4 *Let Assumptions 1-4 hold. Under H_0^a that $\gamma_i = \gamma$, it holds that* $\widehat{\overline{\gamma}} - \gamma = O_p\left(\delta_{nT}^{-2}\right)$ *as* $(n, T) \to \infty$.

Lemma A.5 *Let Assumptions 1-4 hold. Under H_0^b that $f_t = f$, it holds that* $\widehat{\overline{f}} - f = O_p\left(\delta_{nT}^{-2}\right)$ *as* $(n, T) \to \infty$.

Lemma A.6 *Let Assumptions 1-4 hold, and let k denote the largest finite moment of ϵ_{it}, f_t and x_{it}. It holds that*

A.6*(i)* $\max_{1\leq i\leq n}\left\|\tilde{\beta}_i - \beta_i\right\|^2 = o_p\left(n^{2/k}\phi_{nT}^{-2}\right)$;

A.6*(ii)* $\max_{1\leq t\leq T}\left\|\hat{f}_t - H'f_t\right\|^2 = o_p\left(T^{2/k}\delta_{nT}^{-2}\right)$;

A.6*(iii)* $\max_{1\leq i\leq n}\left\|\hat{\gamma}_i - H^{-1}\gamma_i\right\|^2 = o_p\left(n^{2/k}T^{-1}\right) + o_p\left(n^{2/k-2}T\right)$;

A.6*(iv)* $\max_{1\leq t\leq T}\hat{\epsilon}_{it}^2 = o_p\left(T^{2/k}\right) + o_p\left(T^{2/k}\delta_{nT}^{-2}\right)$;

A.6 ***(v)*** $\max_{1\leq t\leq T}\left|\hat{\epsilon}_{it}^2-\epsilon_{it}^2\right| = o_p\left(T^{2/k}\delta_{nT}^{-2}\right)$;

A.6 ***(vi)*** $\max_{1\leq i\leq n}\hat{\epsilon}_{it}^2 = o_p\left(n^{2/k}\right)+o_p\left(n^{4/k}\phi_{nT}^{-2}\right)+o_p\left(n^{2/k-2}T\right)$;

A.6 ***(vii)*** $\max_{1\leq i\leq n}\left|\hat{\epsilon}_{it}^2-\epsilon_{it}^2\right| = o_p\left(n^{4/k}\phi_{nT}^{-2}\right)+o_p\left(n^{2/k-2}T\right)$.

Lemma A.7 *Let Assumptions 1-4 hold, and let k denote the largest finite moment of ϵ_{it}, f_t and x_{it}:*

A.7 ***(i)*** *if, in addition, Assumption 6 holds, then* $\left\|T^{-1}\hat{F}'\hat{F}-H\Sigma_f H'\right\| = O_p\left(T^{-1/2}\right)+O_p\left(n^{-1}\right)$;

A.7 ***(ii)*** *if, in addition, Assumption 6 holds, then* $\max_{1\leq i\leq n}\left\|T^{-1}\hat{F}'M_{Xi}\hat{F}-\Sigma_{fM,i}\right\| = O_p\left(T^{-1/2}\right)+O_p\left(n^{-1}\right)$;

A.7 ***(iii)*** *if, in addition, Assumption 8 holds, then* $\max_{1\leq t\leq T}\left\|\hat{\Sigma}_{\Gamma\epsilon,t}-H^{-1}\Sigma_{\Gamma\epsilon,t}\left(H^{-1}\right)'\right\| = o_p\left(T^{2/k}\delta_{nT}^{-1}\right)$;

A.7 ***(iv)*** *if, in addition, Assumption 8 holds, then* $\max_{1\leq i\leq n}\left\|\hat{\Sigma}_{\gamma,i}-\Sigma_{\gamma,i}\right\| = o_p\left(\sqrt{T}n^{2/k}\delta_{nT}^{-2}\right)$;

A.7 ***(v)*** *if, in addition, Assumption 6 holds, then* $\max_{1\leq i\leq n}\left\|T^{-1/2}F'M_{Xi}\epsilon_i-N_i\right\| = o_p\left(n^{1/k}T^{1/k-1/2}\right)$, *where $\{N_i\}_{i=1}^n$ is a sequence of i.i.d. Gaussian random variables, with variances $\Sigma_{fMe,i}$;*

A.7 ***(vi)*** *if, in addition, Assumption 8 holds, then* $\max_{1\leq t\leq T}\left\|n^{-1/2}\sum_{i=1}^n\hat{\gamma}_i\epsilon_{it}-N_t\right\| = o_p\left(T^{1/k}n^{1/k-1/2}\right) + o_p\left(T^{1/k}\delta_{nT}^{-1}\right)+o_p\left(T^{1/k}\sqrt{n}\delta_{nT}^{-2}\right)$, *where $\{N_t\}_{t=1}^T$ is a sequence of i.i.d. Gaussian random variables, with variances $\Sigma_{\Gamma\epsilon,t}$.*

Proof of Theorem 1. By definition, we have

$$\sqrt{T}\left(\hat{\gamma}_i-\gamma_i\right)=\left(\frac{\hat{F}'M_{Xi}\hat{F}}{T}\right)^{-1}\left[\frac{\hat{F}'M_{Xi}\epsilon_i}{\sqrt{T}}-\frac{\hat{F}'M_{Xi}\left(\hat{F}-F\right)\gamma_i}{\sqrt{T}}\right]. \tag{29}$$

We start by considering the denominator of (29):

$$\begin{aligned}&\frac{\hat{F}'M_{Xi}\hat{F}}{T}-\frac{F'M_{Xi}F}{T}\\ =\ &\frac{F'M_{Xi}\left(\hat{F}-F\right)}{T}+\frac{\left(\hat{F}-F\right)'M_{Xi}F}{T}-\frac{\left(\hat{F}-F\right)'M_{Xi}\left(\hat{F}-F\right)}{T}=I+I'-II.\end{aligned}$$

Repeated application of Lemma A.3 yields $I=O_p\left(\delta_{nT}^{-2}\right)$ and $II=O_p\left(\delta_{nT}^{-4}\right)$. Thus, as $(n,T)\to\infty$, $T^{-1}\hat{F}'M_{Xi}\hat{F}=T^{-1}F'M_{Xi}F+o_p(1)$.

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

$$\frac{\hat{F}' M_{Xi}\epsilon_i}{\sqrt{T}} = \frac{F' M_{Xi}\epsilon_i}{\sqrt{T}} + \frac{\left(\hat{F} - F\right)' M_{Xi}\epsilon_i}{\sqrt{T}} = I + II.$$

By applying a similar logic as in the proof of Lemma A.4, it can be shown that $I = O_p(1)$. As far as II is concerned, note

$$II = \sqrt{T}\frac{\left(\hat{F} - F\right)' \epsilon_i}{T} + \frac{\left(\hat{F} - F\right)' X_i}{T}\left(\frac{X_i' X_i}{T}\right)^{-1}\frac{X_i'\epsilon_i}{\sqrt{T}};$$

applying Lemma A.2*(i)* (to the first term), and Lemma A.3*(i)* and Assumptions 2*(i)* and 1*(i)* (to the second term), it follows that $II = O_p\left(\sqrt{T}\delta_{nT}^{-2}\right)$. Thus, the numerator of (29) is of order $O_p(1) + O_p\left(\frac{\sqrt{T}}{n}\right)$.

Finally, as $(n, T) \to \infty$ under the restriction $\frac{\sqrt{T}}{n} \to 0$, (29) becomes

$$\sqrt{T}\left(\hat{\gamma}_i - \gamma_i\right) = \left(\frac{F' M_{Xi} F}{T}\right)^{-1}\frac{F' M_{Xi}\epsilon_i}{\sqrt{T}} + o_p(1);$$

equation (6) follows from Assumption 5*(i)*. QED

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

$$\begin{aligned}
\hat{f}_t - f_t &= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}' X_j}{T}\right)\left(\tilde{\beta}_j - \beta_j\right)\left(\tilde{\beta}_j - \beta_j\right)' x_{jt} \qquad (30) \\
&\quad -\frac{1}{n}\left(\frac{\hat{F}' F}{T}\right)\sum_{j=1}^{n}\gamma_j\left(\tilde{\beta}_j - \beta_j\right)' x_{jt} - \frac{1}{nT}\sum_{j=1}^{n}\left(\hat{F}'\epsilon_j\right)\left(\tilde{\beta}_j - \beta_j\right)' x_{jt} \\
&\quad -\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}' X_j}{T}\right)\left(\tilde{\beta}_j - \beta_j\right)\gamma_j' f_t - \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}' X_j}{T}\right)\left(\tilde{\beta}_j - \beta_j\right)\epsilon_{jt} \\
&\quad +\frac{1}{nT}\sum_{j=1}^{n}\left(\hat{F}'\epsilon_j\right)\gamma_j' f_t + \frac{1}{n}\left(\frac{\hat{F}' F}{T}\right)\sum_{j=1}^{n}\gamma_j\epsilon_{jt} + \frac{1}{nT}\sum_{j=1}^{n}\left(\hat{F}'\epsilon_j\right)\epsilon_{jt} \\
&= I - II - III - IV - V + VI + VII + VIII.
\end{aligned}$$

The order of magnitude of I follows exactly from the same passages as in the proof of Lemma A.5, with $I = O_p\left(\phi_{nT}^{-2}\right)$. Consider II; omitting γ_j in view of Assumption 3*(iii)*, we have

$$II = \frac{1}{n}\left(\frac{\hat{F}' F}{T}\right)\sum_{j=1}^{n}\Upsilon_j' x_{jt} + \frac{1}{n}\left(\frac{\hat{F}' F}{T}\right)\sum_{j=1}^{n}\bar{\Upsilon}_j' x_{jt} = II_a + II_b;$$

we have shown that $II_a = O_p\left(n^{-1/2}T^{-1/2}\right)$ and $II_b = O_p\left(n^{-1/2}T^{-1/2}\right) + O_p\left(n^{-1}\right)$ in the proof of Lemma A.3, so that $II = O_p\left(n^{-1/2}T^{-1/2}\right) + O_p\left(n^{-1}\right)$. Using Lemma A.3*(i)*, it can be shown that

$III = O_p\left(\phi_{nT}^{-2}\right)$. As far as IV is concerned, note that

$$IV = \frac{\hat{F}'}{\sqrt{T}} \frac{1}{n\sqrt{T}} \sum_{j=1}^{n} X_j \Upsilon_j \gamma_j' f_t + \frac{1}{n} \sum_{j=1}^{n} \left(\frac{\hat{F}' X_j}{T} \right) \bar{\Upsilon}_j \gamma_j' f_t = IV_a + IV_b$$

Similar passages as in the proof of the order of magnitude of II_a, and the fact that $E\left\|f_t\right\| \leq M$ entail $IV_a = O_p\left(n^{-1/2}T^{-1/2}\right)$. Similarly, IV_b is bounded by $\left\|f_t\right\| \left[E\left\|\frac{\hat{F}'X_j}{T}\right\|^2\right]^{1/2} \left[E\left\|\bar{\Upsilon}_j\right\|^2\right]^{1/2}$, which is $O_p\left(n^{-1/2}\delta_{nT}^{-1}\right)$ using Lemma A.1. Thus, $IV = O_p\left(n^{-1/2}\delta_{nT}^{-1}\right)$. Turning to V, we have

$$\begin{aligned}
V &= \frac{1}{n} \sum_{j=1}^{n} \left(\frac{\hat{F}' X_j}{T} \right) \left(\tilde{\beta}_j - \beta_j \right) \epsilon_{jt} \\
&= \frac{1}{n} \sum_{j=1}^{n} \left(\frac{\hat{F}' X_j}{T} \right) \left(\frac{X_j' \bar{M}_w X_j}{T} \right)^{-1} \left(\frac{X_j' \bar{M}_w \epsilon_j}{T} \right) \epsilon_{jt} \\
&\quad + \frac{1}{n} \sum_{j=1}^{n} \left(\frac{\hat{F}' X_j}{T} \right) \left(\frac{X_j' \bar{M}_w X_j}{T} \right)^{-1} \left(\frac{X_j' \bar{M}_w F}{T} \gamma_j \right) \epsilon_{jt} = V_a + V_b.
\end{aligned}$$

We start from $V_b \leq n^{-1} \sum_{j=1}^{n} \left\| \frac{\hat{F}'X_j}{T} \right\| \left\| \left(\frac{X_j'\bar{M}_w X_j}{T} \right)^{-1} \right\| \left\| \frac{X_j'\bar{M}_w F}{T} \right\| \left\|\gamma_j\right\| \left|\epsilon_{jt}\right|$. Using Assumptions 3*(iii)* and 4*(i)*, V_b is bounded by $E\left[\left\| \frac{\hat{F}'X_j}{T} \right\| \left\| \frac{X_j'\bar{M}_w F}{T} \right\| \left|\epsilon_{jt}\right| \right] \leq \left(E\left\| \frac{\hat{F}'X_j}{T} \right\|^6 \right)^{1/6} \left(E\left\| \frac{X_j'\bar{M}_w F}{T} \right\|^{3/2} \right)^{2/3} \left(E\left|\epsilon_{jt}\right|^6 \right)^{1/6}$ $= O\left(n^{-1}\right) + O\left(n^{-1/2}T^{-1/2}\right)$, where the passage in the middle follows from Holder's inequality. Consider now V_a:

$$\begin{aligned}
V_a &= \frac{1}{n} \sum_{j=1}^{n} \left(\frac{F' X_j}{T} \right) \left(\frac{X_j' \bar{M}_w X_j}{T} \right)^{-1} \left(\frac{X_j' \bar{M}_w \epsilon_j}{T} \right) \epsilon_{jt} \\
&\quad + \frac{1}{n} \sum_{j=1}^{n} \frac{\left(\hat{F} - F\right)' X_j}{T} \left(\frac{X_j' \bar{M}_w X_j}{T} \right)^{-1} \left(\frac{X_j' \bar{M}_w \epsilon_j}{T} \right) \epsilon_{jt} = V_{a,1} + V_{a,2}.
\end{aligned}$$

Consider $V_{a,2}$:

$$\begin{aligned}
V_{a,2} &\leq \frac{1}{n} \sum_{j=1}^{n} \left\| \frac{\left(\hat{F} - F\right)' X_j}{T} \right\| \left\| \left(\frac{X_j' \bar{M}_w X_j}{T} \right)^{-1} \right\| \left\| \frac{X_j' \bar{M}_w \epsilon_j}{T} \right\| \left|\epsilon_{jt}\right| \\
&\leq M \frac{1}{n} \sum_{j=1}^{n} \left\| \frac{\left(\hat{F} - F\right)' X_j}{T} \right\| \left\| \frac{X_j' \bar{M}_w \epsilon_j}{T} \right\| \left|\epsilon_{jt}\right|,
\end{aligned}$$

using Assumption 4*(i)*. Further, $E\left[\left\| \frac{\left(\hat{F}-F\right)'X_j}{T} \right\| \left\| \frac{X_j'\bar{M}_w \epsilon_j}{T} \right\| \left|\epsilon_{jt}\right| \right] \leq \left(E\left\| \frac{\left(\hat{F}-F\right)'X_j}{T} \right\|^{3/2} \right)^{2/3} \left(E\left\| \frac{X_j'\bar{M}_w \epsilon_j}{T} \right\|^6 \right)^{1/6}$

$\left(E\left|\epsilon_{jt}\right|^{6}\right)^{1/6}$, again by Holder's inequality. Using Lemma A.3(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 $V_{a,2}=O_p\left(T^{-1/2}\delta_{nT}^{-2}\right)$. Turning to $V_{a,1}$

$$\begin{aligned}
V_{a,1} &= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{F'X_j}{T}\right)\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\frac{X_j'\bar{M}_wE\left(\epsilon_j\epsilon_{jt}\right)}{T} \\
&\quad +\frac{1}{n}\sum_{j=1}^{n}\left(\frac{F'X_j}{T}\right)\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\frac{X_j'\bar{M}_w\left[\epsilon_j\epsilon_{jt}-E\left(\epsilon_j\epsilon_{jt}\right)\right]}{T}=V_{a,1,1}+V_{a,1,2}.
\end{aligned}$$

By virtue of Assumption 4*(i)*, $V_{a,1,1}\leq M\,n^{-1}\,T^{-2}\sum_{j=1}^{n}\left\|F'X_j\right\|\left\|X_j'\bar{M}_wE\left(\epsilon_j\epsilon_{jt}\right)\right\|$. We have $E\left[\left\|\frac{F'X_j}{T}\right\|\left\|\frac{X_j'\bar{M}_wE(\epsilon_j\epsilon_{jt})}{T}\right\|\right]$ $\leq\left(E\left\|\frac{F'X_j}{T}\right\|^2\right)^{1/2}\left(E\left\|\frac{X_j'\bar{M}_wE(\epsilon_j\epsilon_{jt})}{T}\right\|^2\right)^{1/2}$, with $E\left\|\frac{F'X_j}{T}\right\|^2\leq M$ by Assumption 2*(i)*. Further,

$$\begin{aligned}
E\left\|\frac{X_j'\bar{M}_wE\left(\epsilon_j\epsilon_{jt}\right)}{T}\right\|^2 &\leq \frac{1}{T^2}\sum_{s=1}^{T}\sum_{u=1}^{T}E\left[\left\|x_{js}\right\|\left\|x_{ju}\right\|\right]E\left(\epsilon_{js}\epsilon_{jt}\right)E\left(\epsilon_{ju}\epsilon_{jt}\right) \\
&\leq M\frac{1}{T^2}\left[\sum_{s=1}^{T}E\left(\epsilon_{js}\epsilon_{jt}\right)\right]^2=O\left(\frac{1}{T^2}\right),
\end{aligned}$$

where we have used Assumptions 4*(i)*, 2*(i)* and 1*(ii)*(a). Consider now $V_{a,1,2}$; this is bounded by the square root of

$$\begin{aligned}
&E\left\{\frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\left(\frac{F'X_j}{T}\right)\left(\frac{F'X_k}{T}\right)\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\left(\frac{X_k'\bar{M}_wX_k}{T}\right)^{-1}\right. \\
&\left.\times\frac{X_j'\bar{M}_w\left[\epsilon_j\epsilon_{jt}-E\left(\epsilon_j\epsilon_{jt}\right)\right]}{T}\frac{X_k'\bar{M}_w\left[\epsilon_k\epsilon_{kt}-E\left(\epsilon_k\epsilon_{kt}\right)\right]}{T}\right\};
\end{aligned}$$

after some algebra, this is bounded by

$$\begin{aligned}
&E\left\{\frac{1}{n^2T^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\left(\frac{F'X_j}{T}\right)\left(\frac{F'X_k}{T}\right)\sum_{s=1}^{T}\sum_{u=1}^{T}x_{js}x_{ku}\left[\epsilon_{js}\epsilon_{jt}-E\left(\epsilon_{js}\epsilon_{jt}\right)\right]\left[\epsilon_{ju}\epsilon_{jt}-E\left(\epsilon_{ju}\epsilon_{jt}\right)\right]\right\} \\
&= \frac{1}{n^2T^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{s=1}^{T}\sum_{u=1}^{T}E\left[\left(\frac{F'X_j}{T}\right)\left(\frac{F'X_k}{T}\right)x_{js}x_{ku}\right]E\left\{\left[\epsilon_{js}\epsilon_{jt}-E\left(\epsilon_{js}\epsilon_{jt}\right)\right]\left[\epsilon_{ju}\epsilon_{jt}-E\left(\epsilon_{ju}\epsilon_{jt}\right)\right]\right\} \\
&\leq \frac{1}{n^2T^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{s=1}^{T}\sum_{u=1}^{T}E\left\{\left[\epsilon_{js}\epsilon_{jt}-E\left(\epsilon_{js}\epsilon_{jt}\right)\right]\left[\epsilon_{ju}\epsilon_{jt}-E\left(\epsilon_{ju}\epsilon_{jt}\right)\right]\right\} \\
&\leq \frac{1}{nT}E\left|\frac{1}{\sqrt{nT}}\sum_{j=1}^{n}\sum_{s=1}^{T}\left[\epsilon_{js}\epsilon_{jt}-E\left(\epsilon_{js}\epsilon_{jt}\right)\right]\right|^2,
\end{aligned}$$

by using Assumption 2*(iii)* in the second line, Assumption 2*(i)* in the third line, and Assumption 1*(iii)*(c) in the final passage. Thus, $V_{a,1,2} = O_p\left(n^{-1/2}T^{-1/2}\right)$. Putting all together, $V = O_p\left(T^{-1}\right)$ $+ O_p\left(n^{-1/2}T^{-1/2}\right)$. The proofs of $VI = O_p\left(n^{-1/2}T^{-1/2}\right)$, $VII = O_p\left(n^{-1/2}\right)$ and $VIII = O_p\left(\delta_{nT}^{-2}\right)$ are based on the same arguments as in Bai (2003), since the estimation error $\tilde{\beta}_j - \beta_j$ does not appear in their expression. Putting everything together, as $(n, T) \to \infty$ with $\frac{\sqrt{n}}{T} \to 0$, the term that dominates in the expansion of $\hat{f}_t - f_t$ is VII, whose asymptotics is exactly the same as studied in Bai (2003, Theorem 1). QED

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

$$\hat{\gamma}_i - \widehat{\bar{\gamma}} = (\gamma_i - \bar{\gamma}) + (\hat{\gamma}_i - \gamma_i) - \left(\widehat{\bar{\gamma}} - \bar{\gamma}\right) = a_i + b_i - c_i.$$

Under H_0^a, $a_i = 0$; also, b_i can be rewritten as $b_i = \hat{\gamma}_i - \bar{\gamma}$. Using (29), we have

$$\begin{aligned} b_i &= \left(\hat{F}' M_{Xi} \hat{F}\right)^{-1} F' M_{Xi} \epsilon_i + \left(\hat{F}' M_{Xi} \hat{F}\right)^{-1} \left(\hat{F} - F\right)' M_{Xi} \epsilon_i \\ &\quad - \left(\hat{F}' M_{Xi} \hat{F}\right)^{-1} \hat{F}' M_{Xi} \left(\hat{F} - F\right) \gamma_i \\ &= b_{1i} + b_{2i}, \end{aligned} \tag{31}$$

where we define $b_{1i} = \left(\hat{F}' M_{Xi} \hat{F}\right)^{-1} F' M_{Xi} \epsilon_i$ and b_{2i} is the remainder. Further, we can write $\hat{\Sigma}_{\gamma i}^{-1} = \Sigma_{\gamma i}^{-1}$ $-\Sigma_{\gamma i}^{-1}\left(\hat{\Sigma}_{\gamma i} - \Sigma_{\gamma i}\right)\Sigma_{\gamma i}^{-1}$ $+ o_p\left(\left\|\hat{\Sigma}_{\gamma i} - \Sigma_{\gamma i}\right\|\right)$ for each i. Neglecting higher order terms that depend on $o_p\left(\left\|\hat{\Sigma}_{\gamma i} - \Sigma_{\gamma i}\right\|\right)$, we have

$$\begin{aligned} & T\left(\hat{\gamma}_i - \widehat{\bar{\gamma}}\right)' \hat{\Sigma}_{\gamma i}^{-1} \left(\hat{\gamma}_i - \widehat{\bar{\gamma}}\right) \\ = \; & T\left(b_{1i}' \Sigma_{\gamma i}^{-1} b_{1i}\right) + T b_{1i}' \Sigma_{\gamma i}^{-1} \left(\hat{\Sigma}_{\gamma i} - \Sigma_{\gamma i}\right) \Sigma_{\gamma i}^{-1} b_{1i} + T b_{2i}' \hat{\Sigma}_{\gamma i}^{-1} b_{2i} \\ & + 2T b_{1i}' \hat{\Sigma}_{\gamma i}^{-1} b_{2i} + T\left(\widehat{\bar{\gamma}} - \bar{\gamma}\right)' \hat{\Sigma}_{\gamma i}^{-1} \left(\widehat{\bar{\gamma}} - \bar{\gamma}\right) - 2T\left(\widehat{\bar{\gamma}} - \bar{\gamma}\right)' \hat{\Sigma}_{\gamma i}^{-1} \left(\hat{\gamma}_i - \bar{\gamma}\right) \\ = \; & T\left(b_{1i}' \Sigma_{\gamma i}^{-1} b_{1i}\right) + I_i + II_i + III_i + IV_i - V_i. \end{aligned} \tag{32}$$

After this preliminary calculations, we turn to proving (20). In order to do this, we firstly show that $\max_{1 \le i \le n} T\left(b_{1i}' \Sigma_{\gamma i}^{-1} b_{1i}\right)$ can be approximated by the maximum of a sequence of independent random variables with a χ_r^2 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 $I_i - V_i$ in (32) are also all $o_p(\ln n)$ uniformly in i.

Consider $\max_{1 \le i \le n} T\left(b_{1i}' \Sigma_{\gamma i}^{-1} b_{1i}\right)$, and consider in particular the sequence $\left\{\sqrt{T} b_{1i}\right\}_{i=1}^{n}$. It holds that

$\sqrt{T} b_{1i} = \left[T^{-1} \hat{F}' M_{Xi} \hat{F} \right]^{-1} \left[T^{-1/2} F' M_{Xi} \epsilon_i \right]$. As far as the numerator of this expression is concerned, by Lemma A.7*(v)* we write $T^{-1/2} F' M_{Xi} \epsilon_i = N_i + R_{Ni}$ with N_i defined in Lemma A.7 as being zero mean Gaussian with covariance matrix $\Sigma_{fMe,i}$, and $R_{Ni} = o_p \left(n^{1/k_1} T^{1/k_1 - 1/2} \right)$. As far as the denominator of $\sqrt{T} b_{1i}$ is concerned, based on Lemma A.7*(ii)* we write $\left[T^{-1} \hat{F}' M_{Xi} \hat{F} \right]^{-1} = \Sigma_{fM,i}^{-1} + R_{\Sigma fM,i}$ with $R_{\Sigma fM,i} = O_p \left(T^{-1/2} \right) + O_p \left(n^{-1} \right)$. Hence we write

$$\sqrt{T} b_{1i} = \left[\Sigma_{fM,i}^{-1} + R_{\Sigma fM,i} \right] \left[N_i + R_{Ni} \right]. \tag{33}$$

Based on (33), and on the definitions of $\Sigma_{fMe,i}$ and of $\Sigma_{fM,i}$, it holds that

$$\begin{aligned} T \left(b_{1i}' \Sigma_{\gamma i}^{-1} b_{1i} \right) &= N_i' \Sigma_{fMe,i}^{-1} N_i + 2 N_i' \Sigma_{fM,i}^{-1} \Sigma_{\gamma i}^{-1} R_{Ni} + 2 R_{Ni}' \Sigma_{fMe,i}^{-1} N_i \\ &\quad + 2 N_i' \Sigma_{fM,i}^{-1} \Sigma_{\gamma i}^{-1} R_{\Sigma fM,i} N_i + 2 N_i' \Sigma_{fM,i}^{-1} \Sigma_{\gamma i}^{-1} R_{\Sigma fM,i} R_{Ni} \\ &\quad + R_{Ni}' \Sigma_{fMe,i}^{-1} R_{Ni} + 2 R_{Ni}' \Sigma_{fM,i}^{-1} \Sigma_{\gamma i}^{-1} R_{\Sigma fM,i} R_{Ni} \\ &\quad + N_i' R_{\Sigma fM,i} \Sigma_{\gamma i}^{-1} R_{\Sigma fM,i} N_i + 2 N_i' R_{\Sigma fM,i} \Sigma_{\gamma i}^{-1} R_{\Sigma fM,i} R_{Ni} \\ &\quad + R_{Ni}' R_{\Sigma fM,i} \Sigma_{\gamma i}^{-1} R_{\Sigma fM,i} R_{Ni} \\ &= N_i' \Sigma_{fMe,i}^{-1} N_i + I_i^{b1} + II_i^{b1} + III_i^{b1} + IV_i^{b1} + V_i^{b1} + VI_i^{b1} \\ &\quad + VII_i^{b1} + VIII_i^{b1} + IX_i^{b1}. \end{aligned} \tag{34}$$

We note that the distribution of $N_i' \Sigma_{fMe,i}^{-1} N_i$ is χ_r^2. We now show that, in (34), $\max_{1 \le i \le n} I_i^{b1}, ..., \max_{1 \le i \le n} IX_i^{b1}$ are all $o_p(1)$. Consider $\max_{1 \le i \le n} I_i^{b1}$; this is bounded by $\max_{1 \le i \le n} \| N_i \| \max_{1 \le i \le n} \| R_{Ni} \| = o_p \left(n^{1/k_1} T^{1/k_1 - 1/2} \sqrt{\ln n} \right)$, in view of Lemma A.7*(v)* and the fact that $\max_{1 \le i \le n} \| N_i \| = O_p \left(\sqrt{\ln n} \right)$. The same holds for $\max_{1 \le i \le n} II_i^{b1}$. Turning to $\max_{1 \le i \le n} III_i^{b1}$, it is bounded by $\max_{1 \le i \le n} \| N_i \|^2 \max_{1 \le i \le n} \| R_{\Sigma fM,i} \| = O_p \left(T^{-1/2} \ln n \right) + O_p \left(n^{-1} \ln n \right)$ by virtue of Lemma A.7*(ii)*. As far as $\max_{1 \le i \le n} IV_i^{b1}$ is concerned, it is bounded by $\max_{1 \le i \le n} \| N_i \| \max_{1 \le i \le n} \| R_{\Sigma fM,i} \| \max_{1 \le i \le n} \| R_{Ni} \|$, and therefore it is dominated by the previously analyzed terms. Also, $\max_{1 \le i \le n} V_i^{b1}$ has the same order of magnitude as $\max_{1 \le i \le n} \| R_{Ni} \|^2$, thereby being dominated by the other terms. Similarly, $\max_{1 \le i \le n} VI_i^{b1}$ is bounded by $\max_{1 \le i \le n} \| R_{Ni} \|^2$ $\max_{1 \le i \le n} \| R_{\Sigma fM,i} \|$, and therefore it is also dominated. Turning to $\max_{1 \le i \le n} VII_i^{b1}$, it is bounded by $\max_{1 \le i \le n} \| N_i \|^2 \max_{1 \le i \le n} \| R_{\Sigma fM,i} \|^2$, so that it is smaller than $\max_{1 \le i \le n} III_i^{b1}$, and therefore negligible. Similarly, $\max_{1 \le i \le n} VIII_i^{b1}$ is bounded by $\max_{1 \le i \le n} \| N_i \| \max_{1 \le i \le n} \| R_{\Sigma fM,i} \|^2 \max_{1 \le i \le n} \| R_{Ni} \|$, which is dominated by $\max_{1 \le i \le n} IV_i^{b1}$, and thus negligible. Finally, $\max_{1 \le i \le n} IX_i^{b1}$ is bounded by

$\max_{1\leq i\leq n}\left\|R_{\Sigma fM,i}\right\|^2 \max_{1\leq i\leq n}\left\|R_{Ni}\right\|^2$, and it is dominated. Therefore

$$\max_{1\leq i\leq n} T\left(b_{1i}'\Sigma_{\gamma i}^{-1}b_{1i}\right) = \max_{1\leq i\leq n} N_i'\Sigma_{fMe,i}^{-1}N_i + o_p\left[(nT)^{1/k_1}\sqrt{\frac{\ln n}{T}}\right] + O_p\left(\frac{\ln n}{\sqrt{T}}\right) + O_p\left(\frac{\ln n}{n}\right). \quad (35)$$

After proving that $\max_{1\leq i\leq n} T\left(b_{1i}'\Sigma_{\gamma i}^{-1}b_{1i}\right)$ can be approximated by $\max_{1\leq i\leq n} N_i'\Sigma_{fMe,i}^{-1}N_i$, we turn again to equation (32). We now show that $\max_{1\leq i\leq n} I_i$, ..., $\max_{1\leq i\leq n} V_i$ are all $o_p(\ln n)$. Consider I_i; it holds that

$$\max_{1\leq i\leq n} I_i \leq \left\|\max_{1\leq i\leq n} T\left(b_{1i}'\Sigma_{\gamma i}^{-1}b_{1i}\right)\right\|\left\|\max_{1\leq i\leq n}\Sigma_{\gamma i}^{-1}\left(\hat{\Sigma}_{\gamma i}-\Sigma_{\gamma i}\right)\Sigma_{\gamma i}^{-1}\right\|.$$

Equation (35) implies that $\max_{1\leq i\leq n} T\left(b_{1i}'\Sigma_{\gamma i}^{-1}b_{1i}\right) = O_p(\ln n)$; thus, applying Lemma A.7*(iv)*, $\max_{1\leq i\leq n} I_i$ $= o_p\left(\sqrt{T}n^{2/k_1}\delta_{nT}^{-2}\ln n\right)$. Turning to $\max_{1\leq i\leq n} II_i$, note that, in equation (31), b_{2i} is defined as

$$b_{2i} = \left(\hat{F}'M_{Xi}\hat{F}\right)^{-1}\left(\hat{F}-F\right)'M_{Xi}\epsilon_i - \left(\hat{F}'M_{Xi}\hat{F}\right)^{-1}\hat{F}'M_{Xi}\left(\hat{F}-F\right)\gamma_i;$$

further, by the invertibility of $\Sigma_{\gamma i}^{-1}$ and Lemma A.7*(iv)*, $\max_{1\leq i\leq n} T\left(b_{2i}'\hat{\Sigma}_{\gamma i}^{-1}b_{2i}\right)$ has the same order of magnitude as $\max_{1\leq i\leq n}\left\|\sqrt{T}b_{2i}\right\|^2 \left\|\max_{1\leq i\leq n}\Sigma_{\gamma i}^{-1}\left(\hat{\Sigma}_{\gamma i}-\Sigma_{\gamma i}\right)\Sigma_{\gamma i}^{-1}\right\|$. Considering $\max_{1\leq i\leq n}$ $\left\|\sqrt{T}b_{2i}\right\|^2$, it can be evaluated by considering the orders of magnitude of $\max_{1\leq i\leq n}\left\|\sqrt{T}\left(\hat{F}'M_{Xi}\hat{F}\right)^{-1}\right.$ $\left.\left(\hat{F}-F\right)'M_{Xi}\epsilon_i\right\|^2$ and of $\max_{1\leq i\leq n}\left\|\sqrt{T}\left(\hat{F}'M_{Xi}\hat{F}\right)^{-1}\hat{F}'M_{Xi}\left(\hat{F}-F\right)\gamma_i\right\|^2$. The former can be shown to be $o_p\left(n^{2/k_1}T\delta_{nT}^{-4}\right)$, based on the proof of Lemma A.6*(iii)*. The latter has the same order of magnitude as $\left\|T^{-1/2}\hat{F}'\left(\hat{F}-F\right)\right\|^2$, which is $O_p\left(T\delta_{nT}^{-4}\right)$ by Lemma A.3*(iii)*. Putting all together, $\max_{1\leq i\leq n} II_i = o_p\left(T^{3/2}n^{4/k_1}\delta_{nT}^{-6}\right)$ - so, $\max_{1\leq i\leq n} II_i$ is dominated by $\max_{1\leq i\leq n} I_i$. Similar passages yield that $\max_{1\leq i\leq n} III_i$ is dominated by $\max_{1\leq i\leq n} II_i$. Turning to IV_i, it holds that $\max_{1\leq i\leq n} IV_i \leq$ $\left\|\sqrt{T}\left(\hat{\bar{\gamma}}-\bar{\gamma}\right)\right\|^2\left\|\max_{1\leq i\leq n}\Sigma_{\gamma i}^{-1}\left(\hat{\Sigma}_{\gamma i}-\Sigma_{\gamma i}\right)\Sigma_{\gamma i}^{-1}\right\|$, which is $o_p\left(Tn^{2/k_1}\delta_{nT}^{-6}\right)$ by Lemmas A.4 and A.7*(iv)*. Finally, $\max_{1\leq i\leq n} V_i$ is bounded by $\left\|\sqrt{T}\left(\hat{\bar{\gamma}}-\bar{\gamma}\right)\right\|\max_{1\leq i\leq n}\left\|\sqrt{T}b_{1i}\right\|\left\|\max_{1\leq i\leq n}\Sigma_{\gamma i}^{-1}\left(\hat{\Sigma}_{\gamma i}-\Sigma_{\gamma i}\right)\Sigma_{\gamma i}^{-1}\right\|$ $= o_p\left(Tn^{2/k_1}\delta_{nT}^{-4}\ln n\right)$. Putting all together, and using (35), it holds that

$$\begin{aligned}\max_{1\leq i\leq n} T\left(\hat{\gamma}_i-\hat{\bar{\gamma}}\right)'\hat{\Sigma}_{\gamma i}^{-1}\left(\hat{\gamma}_i-\hat{\bar{\gamma}}\right) &= \max_{1\leq i\leq n} N_i'\Sigma_{fMe,i}^{-1}N_i + o_p\left[(nT)^{1/k_1}\sqrt{\frac{\ln n}{T}}\right] \\ &+o_p\left(\frac{n^{2/k_1}}{\sqrt{T}}\ln n\right) + o_p\left(\frac{\sqrt{T}n^{2/k_1}}{n}\ln n\right) + o_p(1),\end{aligned} \quad (36)$$

where the remainders are negligible as $(n,T)\to\infty$ with $\frac{(nT)^{1/k_1}}{\sqrt{T}} + \frac{\sqrt{T}n^{2/k_1}}{n} \to 0$ and $\frac{n^{4/k_1}}{T} \to 0$, which hold in light of (19). Finally, consider the sequence $\{N_i\}_{i=1}^n$: the covariance between $\sqrt{T}b_{1i}$ and $\sqrt{T}b_{1j}$

is given by

$$E\left(\frac{F'M_{Xi}\epsilon_i\epsilon_j'M_{Xj}F}{T}\right) \leq E\left(\frac{F'\epsilon_i\epsilon_j'F}{T}\right)=\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(f_tf_s'\epsilon_{it}\epsilon_{js}\right)$$
$$\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|E\left(f_tf_s'\right)\right\|\left|E\left(\epsilon_{it}\epsilon_{js}\right)\right| \leq M\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\epsilon_{it}\epsilon_{js}\right)\right|,$$

which tends to zero as $(n,T)\to\infty$ by Assumption 7. By virtue of the asymptotic independence between N_i and N_j for all $i\neq j$, the asymptotics of $\max_{1\leq i\leq n}N_i'\Sigma_{fMe,i}^{-1}N_i$ 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 $\gamma_i\neq\bar{\gamma}$ for some (at least one) i, after some algebra it can be shown that, under the alternative, $S_{\gamma,nT}$ has non-centrality parameter given by

$$S_{\gamma,nT}^{NC}=T\max_{1\leq i\leq n}c_i'\hat{\Sigma}_{\gamma i}^{-1}c_i+2T\max_{1\leq i\leq n}c_i'\hat{\Sigma}_{\gamma i}^{-1}\left(\hat{\gamma}_i-\gamma_i\right)-2T\max_{1\leq i\leq n}c_i'\hat{\Sigma}_{\gamma i}^{-1}\left(\widehat{\bar{\gamma}}-\bar{\gamma}\right)=I+II-III,$$

with $I=O_p\left(T\left\|c_i\right\|^2\right)$ by construction. Also, II is bounded by $\sqrt{T}\left(\max_{1\leq i\leq n}\left\|c_i\right\|\right)\left(\max_{1\leq i\leq n}\sqrt{T}\left\|\hat{\gamma}_i-\gamma_i\right\|\right)=O_p\left[T\delta_{nT}^{-2}n^{1/k_1}\left\|c_i\right\|\right]$ in view of Lemma A.6*(iii)*; similarly, $III=O_p\left(\sqrt{T}\delta_{nT}^{-2}\left\|c_i\right\|\right)$ by Lemma A.4. Let $S_{nT}^{\gamma,0}$ denote the null distribution of S_{nT}^{γ}; under H_1^a it holds that

$$P\left[S_{\gamma,nT}>c_{\alpha,n}\right]=P\left[S_{nT}^{\gamma,0}>c_{\alpha,n}-S_{nT}^{\gamma,NC}\right],$$

which tends to 1 if $c_{\alpha,n}-S_{\gamma,nT}^{NC}\to-\infty$ as $(n,T)\to\infty$. In view of equation (22), we know that $c_{\alpha,n}=O\left(\ln n\right)$, whence (21) follows. QED

References

Bai, J., 2003, Inferential theory for structural models of large dimensions. Econometrica, vol. 71, 135-171.

Bai, J., 2009a, Panel data models with interactive fixed effects. Econometrica, vol. 77, 1229-1279.

Bai, J., 2009b, Supplement to Panel data models with interactive fixed effects : technical details and proofs. Econometrica, vol. 77, 1-30.

Baltagi, B., Kao, C., Na, S., 2012, Testing cross-sectional dependence in panel factor model using the wild bootstrap F-test, manuscript.

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

Berman, S.M., 1964, Limit theorems for the maximum term in stationary sequences. The Annals of Mathematical Statistics , vol. 35, 502-516.

Canto e Castro L., 1987, Uniform rate of convergence in extreme-value theory: normal and gamma models. Annales Scientifiques de l'Université de Clerment-Ferrand, 2, tome 90, Série Probabilités et Applications, vol. 6, 25-41.

Castagnetti, C., Rossi, E., Trapani L., 2014a, Supplement to "Inference on Factor Structures in Heterogeneous Panels". Technical details, proofs and further simulations. Mimeo, June 2014.

Castagnetti, C., Rossi, E., Trapani L., 2014b, Testing for no factor structures: on the use of average-type and Hausman-type statistics. Mimeo, June 2014.

Castagnetti, C., Rossi, E., 2013, Euro corporate bond risk factors. Journal of Applied Econometrics, vol. 28, 372-391.

Chudik, A., Pesaran, H., 2013, Common Correlated Effects Estimation of Heterogenous Dynamic Panel Data Models with Weakly Exogenous Regressors. CESifo Working Paper Series 4232.

Chudik, A., Pesaran, H., Tosetti, E., 2011, Weak and strong cross-section dependence and estimation of large panels. Econometrics Journal, vol. 14, 45-90.

Corradi, V., 1999, Deciding between I(0) and I(1) via flil-based bounds. Econometric Theory, vol. 15, 643-63.

Csörgö, M., Hórvath, L., 1997, Limit theorems in change-point analysis. Wiley, Chichester.

Eberhardt, M., Helmers, C., Strauss, H., 2013, Do spillovers matter when estimating private returns to R&D?. The Review of Economics and Statistics, vol. 95, 436-448.

Eberhardt, M., Teal, F., 2012, No mangos in the tundra: spatial heterogeneity in agricultural productivity analysis. Oxford Bulletin of Economics and Statistics (forthcoming).

Eberlein, E., 1986, On strong invariance principles under dependence assumptions. Annals of Probability, vol. 14, 260-270.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997, Modelling extremal events for insurance and finance. New York: Springer.

Everaert, G., Groote, T.D., 2012. Common correlated effects estimation of dynamic panels with cross-sectional dependence. Mimeo.

French, D., O'Hare, C., 2013, A dynamic factor approach to mortality modeling. Journal of Forecasting (forthcoming).

Hall, P., Miller, H., 2010, Bootstrap confidence intervals and hypothesis tests for extrema of parameters. Biometrika, vol. 97, 881-892.

Hannan, E.T., Kavalieris, L., 1986. Regression; autoregression models. Journal of Time Series Analysis, vol. 7, 27-49.

Jenish, N., Prucha, I.R., 2012. On spatial processes and asymptotic inference under Near-Epoch Dependence. Journal of Econometrics, vol. 170, 178-190.

Kapetanios, G., 2003, Determining the poolability of individual series in panel datasets. University of London Queen Mary Economics Working Paper No. 499.

Kapetanios, G., Pesaran, M.H., 2007, Alternative approaches to estimation and inference in large multifactor panels: small sample results with an application to modelling of asset returns. In Garry Phillips and Elias Tzavalis, (Eds.), The Refinenement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis. Cambridge University Press, Cambridge.

Leadbetter, M.R., Rootzen, H., 1988, Extremal theory for stochastic processes. Annals of Probability, vol. 16, 431-478.

Lee, R. D., Carter, L. R., 1992, Modeling and forecasting the time series of U.S. mortality. Journal of the American Statistical Association, vol. 87, 659-671.

Lin, Z., Bai, Z., 2010. Probability Inequalities. Berlin: Springer.

Pesaran, M. H., 2006, Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica, vol. 74, 967-1012.

Pesaran, M. H., Tosetti, E., 2011, Large panels with common factors and spatial correlation. Journal of Econometrics, vol. 161, 182-202.

Sarafidis, V., Yamagata, T., Robertson, D., 2009, A test of cross section dependence for a linear dynamic panel model with regressors. Journal of Econometrics, vol. 148, 149-461.

Song, M., 2013. Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. Mimeo, January 2013.

Westerlund, J., Hess, W., 2011, A new poolability test for cointegrated panels. Journal of Applied Econometrics, vol. 26, 56-88.

Supplement to: "Inference on Factor Structures in Heterogeneous Panels".

Technical details, proofs and further simulations

Carolina Castagnetti Eduardo Rossi Lorenzo Trapani

September 10, 2014

1 Proofs

In this Supplement we report the proofs of Lemmas A.1-A.7, and of Theorem 4 in Castagnetti, Rossi and Trapani (2014). Henceforth, we set $H = I_r$, for the sake of notational simplicity. Inequalities are written, when possible, omitting constants.

The proofs of all Lemmas rely upon the decomposition - see Proposition A.1 in Bai (2009a):

$$\begin{aligned}
\hat{F} - F &= \frac{1}{nT}\sum_{j=1}^{n} X_j \left(\tilde{\beta}_j - \beta_j\right)\left(\tilde{\beta}_j - \beta_j\right)' X_j' \hat{F} \\
&\quad - \frac{1}{nT}\sum_{j=1}^{n} X_j \left(\tilde{\beta}_j - \beta_j\right) \gamma_j' F' \hat{F} - \frac{1}{nT}\sum_{j=1}^{n} X_j \left(\tilde{\beta}_j - \beta_j\right) \epsilon_j' \hat{F} \\
&\quad - \frac{1}{nT}\sum_{j=1}^{n} F\gamma_j \left(\tilde{\beta}_j - \beta_j\right)' X_j' \hat{F} - \frac{1}{nT}\sum_{j=1}^{n} \epsilon_j \left(\tilde{\beta}_j - \beta_j\right)' X_j' \hat{F} \\
&\quad + \frac{1}{nT}\sum_{j=1}^{n} F\gamma_j \epsilon_j' \hat{F} + \frac{1}{nT}\sum_{j=1}^{n} \epsilon_j \gamma_j' F' \hat{F} + \frac{1}{nT}\sum_{j=1}^{n} \epsilon_j \epsilon_j' \hat{F}.
\end{aligned} \tag{1}$$

In (1), the main difference with Bai (2009a) is the presence of the unit specific estimates, $\tilde{\beta}_j$. Consider also the following notation, which we use henceforth throughout the Appendices. We define $\Upsilon_i \equiv \left(X_i' \bar{M}_w X_i\right)^{-1} \left(X_i' \bar{M}_w \epsilon_i\right)$, so that we can write

$$\begin{aligned}
\tilde{\beta}_i - \beta_i &= \left(\frac{X_i' \bar{M}_w X_i}{T}\right)^{-1} \left(\frac{X_i' \bar{M}_w \epsilon_i}{T}\right) + \left(\frac{X_i' \bar{M}_w X_i}{T}\right)^{-1} \left(\frac{X_i' \bar{M}_w F}{T}\gamma_i\right) \\
&= \Upsilon_i + \bar{\Upsilon}_i,
\end{aligned} \tag{2}$$

for every i; by construction, $\bar{\Upsilon}_i = O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right)$. We extensively use the notation $\delta_{nT} = \min\left\{\sqrt{n}, \sqrt{T}\right\}$ and $\phi_{nT} = \min\left\{n, \sqrt{T}\right\}$.

Proof of Lemma A.1. Let $\|A\|_1$ denote the L_1-norm of a matrix A, i.e. $\|A\|_1 = \max_{x \neq 0} \|Ax\|_1 / \|x\|_1$. By a well known norm inequality (see e.g. Strang, 1988, p. 369, exercise 7.2.3), it holds that

$$\begin{aligned}\left\|\tilde{\beta}_i - \beta_i\right\|^r &\leq \left\|\left(\frac{X_i'\bar{M}_w X_i}{T}\right)^{-1}\right\|_1^r \left\|\frac{X_i'\bar{M}_w \epsilon_i}{T} + \frac{X_i'\bar{M}_w F}{T}\gamma_i\right\|^r \\ &= \left[l_{\min}^{-1}\left(\frac{X_i'\bar{M}_w X_i}{T}\right)\right]^r \left\|\frac{X_i'\bar{M}_w \epsilon_i}{T} + \frac{X_i'\bar{M}_w F}{T}\gamma_i\right\|^r,\end{aligned}$$

where the last equality holds by symmetry. In view of Assumption 4*(i)*, and omitting γ_i by virtue of Assumption 3*(iii)*

$$E\left\|\tilde{\beta}_i - \beta_i\right\|^r \leq E\left\|\frac{X_i'\bar{M}_w \epsilon_i}{T}\right\|^r + E\left\|\frac{X_i'\bar{M}_w F}{T}\right\|^r = I + II.$$

Consider I; we have $I \leq T^{-r}E\|X_i'\epsilon_i\|^r = T^{-r}E\left\|\sum_{t=1}^T x_{it}\epsilon_{it}\right\|^r$. It holds that

$$\begin{aligned}T^{-r}E\|X_i'\epsilon_i\|^r &\leq T^{-r}E\left|\sum_{t=1}^T \|x_{it}\epsilon_{it}\|^2\right|^{r/2} \leq T^{-r}E\left|T^{1-2/r}\left(\sum_{t=1}^T \|x_{it}\epsilon_{it}\|^r\right)^{2/r}\right|^{r/2} \\ &\leq T^{-r}T^{r/2}\left(\frac{1}{T}\sum_{t=1}^T E\|x_{it}\epsilon_{it}\|^r\right) \leq T^{-r/2}\left(\frac{1}{T}\sum_{t=1}^T \left[E\|x_{it}\|^{2r}\right]^{1/2}\left[E|\epsilon_{it}|^{2r}\right]^{1/2}\right) \\ &= O\left(T^{-r/2}\right),\end{aligned} \tag{3}$$

where we have used: Assumption 2*(iv)*; Holder's inequality; the C_r-inequality and Jensen's inequality; the Cauchy-Schwartz inequality; and the fact that, by Assumptions 1 and 2*(i)*, $E|\epsilon_{it}|^{2r} < \infty$ and $E\|x_{it}\|^{2r} < \infty$ respectively. Using the Cauchy-Schwartz inequality in this context is more than what is necessary, since x_{it} and ϵ_{it} are independent. Turning to II, note that, for sufficiently large n and omitting higher order terms, $\left(\bar{H}_w'\bar{H}_w\right)^{-1} = D_w^{-1} - D_w^{-1}R_w D_w^{-1}$, with $D_w = C'F'FC$ and $\|R_w\| = O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right)$ - see e.g. equation (29) in Pesaran (2006). Therefore, letting $\bar{\epsilon} = n^{-1}\sum_{i=1}^n \epsilon_i$ and omitting higher order terms

$$\begin{aligned}\frac{X_i'\bar{M}_w F}{T} &= -\frac{X_i'FCD_w^{-1}\bar{\epsilon}'F}{T^2} - \frac{F'\bar{\epsilon}D_w^{-1}C'F'X_i}{T^2} \\ &\quad -\frac{X_i'\bar{\epsilon}D_w^{-1}\bar{\epsilon}'F}{T^2} - \frac{X_i'FCD_w^{-1}R_w D_w^{-1}C'F'F}{T^2} \\ &= -I - I' - II - III.\end{aligned} \tag{4}$$

Consider $E\left\|I\right\|^{r}$; since C has full rank by Assumption 4*(ii)* and D_w is invertible

$$E\left\|I\right\|^{r}\leq E\left\|\frac{X_i'F}{T}\frac{\bar{\epsilon}'F}{T}\right\|^{r}\leq M\left[E\left\|\frac{X_i'F}{T}\right\|^{2r}\right]^{1/2}\left[E\left\|\frac{\bar{\epsilon}'F}{T}\right\|^{2r}\right]^{1/2}.$$

Consider the first term; we have $T^{-1}\sum_{t=1}^{T}E\left\|x_{it}f_t'\right\|^{2r}\leq T^{-1}\sum_{t=1}^{T}\left[E\left\|x_{it}\right\|^{4r}\right]^{1/2}\left[E\left\|f_t\right\|^{4r}\right]^{1/2}$, which is finite by Assumption 2*(i)*. As far as the second term is concerned, note

$$E\left\|\frac{\bar{\epsilon}'F}{T}\right\|^{2r}\leq T^{-2r}\sum_{t=1}^{T}\left[E\left\|f_t\right\|^{4r}\right]^{1/2}\left[E\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_{it}\right|^{4r}\right]^{1/2},$$

after similar passages as in equation (3). It holds that $E\left\|f_t\right\|^{4r}<\infty$ by Assumption 2*(i)*. By using Assumption 1*(iv)*(b) and following thereafter a similar logic as in the proof of (3), we have $E\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_{it}\right|^{4r}=O\left(n^{-r/2}\right)$, so that $E\left\|I\right\|^{r}=O\left(n^{-r/2}T^{-r/2}\right)$. The same logic yields $E\left\|II\right\|^{r}=O\left(n^{-r}T^{-r}\right)$. Finally, consider III; after some passages

$$E\left\|III\right\|^{r}\leq\left\|R_w\right\|^{r}\left[E\left\|\frac{X_i'F}{T}\right\|^{2r}\right]^{1/2}\left[E\left\|\frac{F'F}{T}\right\|^{2r}\right]^{1/2}=O\left(\left\|R_w\right\|^{r}\right),$$

again by similar passages as above. Therefore, $E\left\|\frac{X_i'\bar{M}_wF}{T}\right\|^{r}=O\left(\left\|R_w\right\|^{r}\right)$. Putting everything together, the Lemma follows. QED

Proof of Lemma A.2. The proof of A.2*(i)* is very similar, and in fact simpler, than that of A.2*(ii)*; thus we focus on the latter only. Using (1)

$$\begin{aligned}
& n^{-1/2}T^{-1}\sum_{i=1}^{n}\epsilon_i'\left(\hat{F}-F\right) \qquad (5)\\
= & \frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\frac{\left(\frac{\sum_{i=1}^{n}\epsilon_i}{\sqrt{n}}\right)'X_j}{\sqrt{T}}\left(\tilde{\beta}_j-\beta_j\right)\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T}-\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\frac{\left(\frac{\sum_{i=1}^{n}\epsilon_i}{\sqrt{n}}\right)'X_j}{\sqrt{T}}\left(\tilde{\beta}_j-\beta_j\right)\gamma_j'\frac{F'\hat{F}}{T}\\
& -\frac{1}{nT}\sum_{j=1}^{n}\frac{\left(\frac{\sum_{i=1}^{n}\epsilon_i}{\sqrt{n}}\right)'X_j}{\sqrt{T}}\left(\tilde{\beta}_j-\beta_j\right)\frac{\epsilon_j'\hat{F}}{\sqrt{T}}-\frac{1}{n\sqrt{T}}\frac{\left(\frac{\sum_{i=1}^{n}\epsilon_i}{\sqrt{n}}\right)'F}{\sqrt{T}}\sum_{j=1}^{n}\gamma_j\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T}\\
& -\frac{1}{\sqrt{n}}\frac{1}{nT}\sum_{j=1}^{n}\sum_{i=1}^{n}\epsilon_i'\epsilon_j\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T}+\frac{1}{\sqrt{nT}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\epsilon_i'F}{\sqrt{T}}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\gamma_j\frac{\epsilon_j'\hat{F}}{T}\\
& +\frac{1}{\sqrt{nT}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i'\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\epsilon_j\gamma_j'\frac{F'\hat{F}}{T}+\frac{1}{T}\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\sum_{i=1}^{n}\epsilon_i}{\sqrt{n}}\right)'\epsilon_j\frac{\epsilon_j'\hat{F}}{T}\\
= & I+II+III+IV+V+VI+VII+VIII.
\end{aligned}$$

The proof follows very similar lines to that of Lemma A.8 in Song (2013): the only difference is the different expansion of the estimation error $\tilde{\beta}_j - \beta_j$ when using the CCE. Thus, we report only the complete passages to determine the order of magnitude of I; the same logic applies to all the other terms in the expansion. The only term for which passages slightly differ is V, and we report the full blown proof for it.

Consider I; it holds that $I \leq n^{-1} \sum_{j=1}^{n} \left\| \frac{\sum_{i=1}^{n} \epsilon_i' X_j}{\sqrt{n}T} \right\| \left\| \frac{X_j' \hat{F}}{T} \right\| \left\| \tilde{\beta}_j - \beta_j \right\|^2$. This is bounded by

$$\begin{aligned} & E\left[\left\| \frac{\sum_{i=1}^{n} \epsilon_i' X_j}{\sqrt{n}T} \right\| \left\| \frac{X_j' \hat{F}}{T} \right\| \left\| \tilde{\beta}_j - \beta_j \right\|^2 \right] \quad (6) \\ \leq & \left[E\left(\left\| \tilde{\beta}_j - \beta_j \right\|^{2p} \right) \right]^{1/p} \left[E\left(\left\| \frac{\sum_{i=1}^{n} \epsilon_i' X_j}{\sqrt{n}T} \right\| \left\| \frac{X_j' \hat{F}}{T} \right\| \right)^q \right]^{1/q} \\ \leq & \left[E\left\| \tilde{\beta}_j - \beta_j \right\|^3 \right]^{2/3} \left[E\left\| \frac{\sum_{i=1}^{n} \epsilon_i' X_j}{\sqrt{n}T} \right\|^6 \right]^{1/6} \left[E\left\| \frac{X_j' \hat{F}}{T} \right\|^6 \right]^{1/6}, \end{aligned}$$

using Holder's inequality in the first line (with $p = \frac{3}{2}$ and $q = 3$), and the Cauchy-Schwartz inequality in the second line. The first term is of order $O\left(\phi_{nT}^{-2}\right)$ in light of Lemma A.1. Similar passages as in the proof of Lemma A.1 yield that both the second and third terms are of order $O(1)$. This entails that $I = O_p\left(T^{-1/2}\phi_{nT}^{-2}\right)$. Similar passages yield $II = O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$; $III = O_p\left(T^{-1}\phi_{nT}^{-1}\right)$; $IV = O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$; $VI = O_p\left(T^{-1/2}\phi_{nT}^{-1}\right) + O_p\left(\delta_{nT}^{-2}\right)$; $VII = O_p\left(n^{-1/2}\right)$ and $VIII = O_p\left(T^{-1}\right) + O_p\left(n^{-1/2}T^{-1/2}\right)$.

Consider now V, whose proof is marginally different to that of Song (2013)

$$\begin{aligned} V \leq & \left[\frac{1}{n} \sum_{j=1}^{n} E\left| \frac{1}{\sqrt{n}T} \sum_{i=1}^{n} \sum_{t=1}^{T} \epsilon_{it}\epsilon_{jt} \right|^2 \right]^{1/2} \left[\frac{1}{n} \sum_{j=1}^{n} E\left(\left\| \frac{X_j' \hat{F}}{T} \right\| \left\| \tilde{\beta}_j - \beta_j \right\| \right)^2 \right]^{1/2} \\ \leq & \left[\frac{1}{n} \sum_{j=1}^{n} E\left| \frac{1}{\sqrt{n}T} \sum_{i=1}^{n} \sum_{t=1}^{T} \epsilon_{it}\epsilon_{jt} \right|^2 \right]^{1/2} \left[\frac{1}{n} \sum_{j=1}^{n} E\left\| \tilde{\beta}_j - \beta_j \right\|^3 \right]^{1/3} \left[\frac{1}{n} \sum_{j=1}^{n} E\left\| \frac{X_j' \hat{F}}{T} \right\|^6 \right]^{1/6} \\ = & \left[\frac{1}{nT} \sum_{j=1}^{n} E\left| \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \epsilon_{it}\epsilon_{jt} \right|^2 \right]^{1/2} O_p\left(\phi_{nT}^{-1}\right), \end{aligned}$$

using the Cauchy-Schwartz inequality (first line), Holder's inequality with the same orders as in (6) (second line) and Lemma A.1. Also, $E\left| \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \epsilon_{it}\epsilon_{jt} \right|^2 \leq (nT)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} \left| E\left(\epsilon_{it}\epsilon_{kt}\epsilon_{js}\epsilon_{ks}\right) \right| \leq M$, by Assumption 1*(iii)*(d), so that $V = O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$. Putting all together, part A.2*(ii)* follows. QED

Proof of Lemma A.3. The Lemma is a refinement of Lemma A.3 in Bai (2009a). Particularly, by

Lemma A.3 in Bai (2009a) we have $T^{-1}X_i'\left(\hat{F}-F\right)=o_p(1)$ and $T^{-1}F'\left(\hat{F}-F\right)=o_p(1)$.

Consider part *(i)*. Using (1)

$$\begin{aligned}
\frac{X_i'\left(\hat{F}-F\right)}{T} &= \frac{1}{n}\sum_{j=1}^{n}\frac{X_i'X_j}{T}\left(\tilde{\beta}_j-\beta_j\right)\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T} \\
&-\frac{1}{n}\sum_{j=1}^{n}\frac{X_i'X_j}{T}\left(\tilde{\beta}_j-\beta_j\right)\gamma_j'\frac{F'\hat{F}}{T}-\frac{1}{nT}\sum_{j=1}^{n}\frac{X_i'X_j}{T}\left(\tilde{\beta}_j-\beta_j\right)\epsilon_j'\hat{F} \\
&-\frac{1}{n}\sum_{j=1}^{n}\frac{X_i'F}{T}\gamma_j\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T}-\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\frac{X_i'\epsilon_j}{\sqrt{T}}\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T} \\
&+\frac{1}{nT}\sum_{j=1}^{n}\frac{X_i'F}{T}\gamma_j\epsilon_j'\hat{F}+\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\frac{X_i'\epsilon_j}{\sqrt{T}}\gamma_j'\frac{F'\hat{F}}{T}+\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\frac{X_i'\epsilon_j}{\sqrt{T}}\frac{\epsilon_j'\hat{F}}{T} \\
&= I-II-III-IV-V+VI+VII+VIII;
\end{aligned}$$

henceforth, we omit γ_i in the passages, based on Assumption 3*(iii)*. Consider I; it is bounded by $E\left(\left\|\frac{X_i'X_j}{T}\right\|\left\|\frac{X_j'\hat{F}}{T}\right\|\left\|\tilde{\beta}_j-\beta_j\right\|^2\right)$. Using the Holder's inequality and the Cauchy-Schwartz inequality in a similar way to (6), this is bounded by $\left[E\left\|\frac{X_i'X_j}{T}\right\|^6\right]^{1/6}\left[E\left\|\frac{X_j'\hat{F}}{T}\right\|^6\right]^{1/6}\left[E\left\|\tilde{\beta}_j-\beta_j\right\|^3\right]^{2/3}=O_p\left(\phi_{nT}^{-2}\right)$. Turning to II, we have $II=\frac{1}{n}\sum_{j=1}^{n}\frac{X_i'X_j}{T}\left(\tilde{\beta}_j-\beta_j\right)\gamma_j'\frac{F'F}{T}+o_p(1)$, where the $o_p(1)$ term comes from $T^{-1}F'\left(\hat{F}-F\right)=o_p(1)$. By (2), this is bounded by $\left\|\frac{X_i}{\sqrt{T}}\right\|\left\|\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}X_j\Upsilon_j\right\|+\frac{1}{n}\sum_{j=1}^{n}\left\|\frac{X_i'X_j}{T}\right\|\left\|\bar{\Upsilon}_j\right\|$ $=II_a+II_b$. Consider II_a; since

$$\begin{aligned}
E\left\|\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}X_j\Upsilon_j\right\|^2 &= E\left\|\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}X_j\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\left(\frac{X_j'\bar{M}_w\epsilon_j}{T}\right)\right\|^2 \\
&= \frac{1}{n^2T}\sum_{j=1}^{n}E\left\|X_j\left(\frac{X_j'\bar{M}_wX_j}{T}\right)^{-1}\left(\frac{X_j'\bar{M}_w\epsilon_j}{T}\right)\right\|^2 \\
&\leq \frac{1}{n^2T}\sum_{j=1}^{n}\left[E\left\|\frac{X_j}{\sqrt{T}}\right\|^2\right]^{1/2}\left[E\left\|\frac{X_j'\epsilon_j}{\sqrt{T}}\right\|^2\right]^{1/2}=O\left(\frac{1}{nT}\right),
\end{aligned}$$

where we have used Assumptions 4*(i)*, 1, 2*(i)*; the Cauchy-Schwartz inequality; and the facts that both $E\left\|\frac{X_j}{\sqrt{T}}\right\|^2$ and $E\left\|\frac{X_j'\epsilon_j}{\sqrt{T}}\right\|^2$ are finite. The latter statement can be shown as follows

$$E\left\|\frac{X_j'\epsilon_j}{\sqrt{T}}\right\|^2=E\left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}x_{jt}x_{js}'\epsilon_{jt}\epsilon_{js}\right\|\leq\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left\|x_{jt}x_{js}'\right\|E\left|\epsilon_{jt}\epsilon_{js}\right|\leq M\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left|\epsilon_{jt}\epsilon_{js}\right|\leq M',$$

using, respectively, Assumptions 2*(iii)*, 2*(i)* and 1*(ii)*(c). Thus, $II_a=O_p\left(n^{-1/2}T^{-1/2}\right)$. Turning to

II_b, this is of the same order of magnitude as $E\left[\left\|\frac{X_i'X_j}{T}\right\|\left\|\bar{\Upsilon}_j\right\|\right] \leq \left[E\left(\left\|\frac{X_i'X_j}{T}\right\|^2\right)\right]^{1/2}\left[E\left(\left\|\bar{\Upsilon}_j\right\|^2\right)\right]^{1/2}$; by the proof of Lemma A.1, $II_b = O_p\left(n^{-1/2}\delta_{nT}^{-1}\right)$. Thus, $II = O_p\left(n^{-1/2}\delta_{nT}^{-1}\right)$. Turning to III, it can be decomposed into $\frac{1}{n\sqrt{T}}\sum_{j=1}^n \frac{X_i'X_j}{T}\left(\tilde{\beta}_j - \beta_j\right)\frac{\epsilon_j'F}{\sqrt{T}} + \frac{1}{n}\sum_{j=1}^n \frac{X_i'X_j}{T}\left(\tilde{\beta}_j - \beta_j\right)\frac{\epsilon_j'\left(\hat{F}-F\right)}{T} = III_a + III_b$; III_a is bounded by $\frac{1}{\sqrt{T}}E\left[\left\|\frac{X_i'X_j}{T}\right\|\left\|\tilde{\beta}_j - \beta_j\right\|\left\|\frac{\epsilon_j'F}{\sqrt{T}}\right\|\right]$. Using the same logic as above, this entails $III_a = O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$. Similar passages and Lemma A.2*(i)*, yield $III_b = O_p\left(\phi_{nT}^{-1}\delta_{nT}^{-2}\right)$. Term IV has the same magnitude of term II. As far as V is concerned, using the fact that $T^{-1}X_i'\left(\hat{F}-F\right) = o_p(1)$, V is bounded by $T^{-1/2}E\left[\left\|\frac{X_i'\epsilon_j}{\sqrt{T}}\right\|\left\|\frac{X_j'F}{T}\right\|\left\|\tilde{\beta}_j - \beta_j\right\|\right]$; a similar logic to the proof of I yields $V = O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$. Turning to VI, we have $VI = \frac{1}{n\sqrt{T}}\sum_{j=1}^n \frac{X_i'F}{T}\frac{\epsilon_j'F}{\sqrt{T}} + \frac{X_i'F}{T}\frac{1}{n}\sum_{j=1}^n \frac{\epsilon_j'\left(\hat{F}-F\right)}{T} = VI_a + VI_b$. Considering VI_a, similar passages as above give $VI_a = O_p\left(n^{-1/2}T^{-1/2}\right)$. Turning to VI_b, this is $O_p\left(\delta_{nT}^{-2}\right)$ by Lemma A.2*(ii)*. Therefore, $VI = O_p\left(\delta_{nT}^{-2}\right)$. As far as VII is concerned, it is bounded by $\left\|\frac{F'F}{T}\right\|\left\|\frac{1}{nT}\sum_{j=1}^n X_i'\epsilon_j\right\| + o_p(1)$. The term $\sum_{j=1}^n X_i'\epsilon_j$ is bounded by the square root of its variance, viz.

$$
\begin{aligned}
E\left\|\sum_{t=1}^T\sum_{s=1}^T x_{it}x_{is}'\left(\sum_{j=1}^n\sum_{k=1}^n \epsilon_{jt}\epsilon_{ks}\right)\right\| &\leq \sum_{j=1}^n\sum_{k=1}^n\sum_{t=1}^T\sum_{s=1}^T E\left\|x_{it}x_{is}'\right\| E\left|\epsilon_{jt}\epsilon_{ks}\right| \\
&\leq M\sum_{j=1}^n\sum_{k=1}^n\sum_{t=1}^T\sum_{s=1}^T E\left|\epsilon_{jt}\epsilon_{ks}\right| \leq M'(nT).
\end{aligned}
$$

Thus, $VII = O_p\left(n^{-1/2}T^{-1/2}\right)$. Finally, $VIII = \frac{1}{nT}\sum_{j=1}^n \frac{X_i'\epsilon_j}{\sqrt{T}}\frac{\epsilon_j'F}{\sqrt{T}} + \frac{1}{n\sqrt{T}}\sum_{j=1}^n \frac{X_i'\epsilon_j}{\sqrt{T}}\frac{\epsilon_j'\left(\hat{F}-F\right)}{T} = VIII_a + VIII_b$; $VIII_a$ is bounded by $\left[E\left\|\frac{X_i'\epsilon_j}{\sqrt{T}}\right\|^2\right]^{1/2}\left[E\left\|\frac{\epsilon_j'F}{\sqrt{T}}\right\|^2\right]^{1/2}$, which is $O(1)$, so that $VIII_a = O_p\left(T^{-1}\right)$. Similarly, $VIII_b$ is bounded by $T^{-1/2}\left[E\left(\left\|\frac{X_i'\epsilon_j}{\sqrt{T}}\right\|^2\right)\right]^{1/2}\left[E\left(\left\|\frac{\epsilon_j'\left(\hat{F}-F\right)}{T}\right\|^2\right)\right]^{1/2} = O_p\left(T^{-1/2}\delta_{nT}^{-2}\right)$, using Lemma A.2*(i)*. Putting all together, part *(i)* of the Lemma follows. The proof of part *(ii)* follows essentially the same passages, and is therefore omitted. As far as part *(iii)* is concerned, the same logic as above can be applied directly to (1), obtaining $T^{-1/2}\left\|\hat{F}-F\right\| = O_p\left(\left\|\tilde{\beta}_j - \beta_j\right\|\right) + O_p\left(\delta_{nT}^{-1}\right)$, whence $T^{-1}\left\|\hat{F}-F\right\|^2 = O_p\left(\delta_{nT}^{-2}\right)$. QED

Proof of Lemma A.4. By definition, under H_0^a

$$
\sqrt{T}\left(\hat{\gamma}_i - \gamma\right) = \left(\frac{\hat{F}'M_{Xi}\hat{F}}{T}\right)^{-1}\left[\frac{\hat{F}'M_{Xi}\epsilon_i}{\sqrt{T}} - \frac{\hat{F}'M_{Xi}\left(\hat{F}-F\right)\gamma}{\sqrt{T}}\right]; \tag{7}
$$

also, under H_0^a, it holds that $\widehat{\overline{\gamma}} - \gamma = \frac{1}{n}\sum_{i=1}^n\left(\hat{\gamma}_i - \gamma\right)$. Using (7) and neglecting higher order terms coming from $\hat{F}'M_{Xi}\hat{F} - F'M_{Xi}F$

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(\hat{\gamma}_i-\gamma) &= \frac{1}{n}\sum_{i=1}^{n}\left(F'M_{Xi}F\right)^{-1}F'M_{Xi}\epsilon_i+\frac{1}{n}\sum_{i=1}^{n}\left(F'M_{Xi}F\right)^{-1}\left(\hat{F}-F\right)'M_{Xi}\epsilon_i \\ &\quad +\frac{1}{n}\sum_{i=1}^{n}\left(F'M_{Xi}F\right)^{-1}\hat{F}'M_{Xi}\left(\hat{F}-F\right)\gamma \\ &= I+II+III.\end{aligned}$$

Term I is bounded by the square root of

$$\begin{aligned}&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left\|\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\left(\frac{F'M_{Xi}\epsilon_i}{T}\right)\left(\frac{F'M_{Xj}F}{T}\right)^{-1}\left(\frac{F'M_{Xj}\epsilon_j}{T}\right)\right\| \\ \leq\ & M\frac{1}{n^2T^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left\|\left(F'\epsilon_i\right)\left(F'\epsilon_j\right)\right\|\leq M\frac{1}{n^2T^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left\|f_tf_s'\right\|E\left|\epsilon_{it}\epsilon_{js}\right| \\ \leq\ & M'\frac{1}{n^2T^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left|\epsilon_{it}\epsilon_{js}\right|\leq M''\frac{1}{nT},\end{aligned}$$

using, respectively, Assumptions 4*(i)*, 2*(iii)*, 2*(i)* and 1*(ii)*(c). Thus, $I = O_p\left(\frac{1}{\sqrt{nT}}\right)$. Consider II; it is bounded by

$$E\left\|\left(\frac{F'M_{Xi}F}{T}\right)^{-1}\frac{\left(\hat{F}-F\right)'M_{Xi}\epsilon_i}{T}\right\|\leq E\left\|\frac{\epsilon_i'\left(\hat{F}-F\right)}{T}\right\|=O_p\left(\delta_{nT}^{-2}\right),$$

by Assumption 3*(i)* and Lemma A.2*(i)*. Similarly, III is bounded by $E\left\|\hat{F}'\left(\hat{F}-F\right)\right\| = O_p\left(\delta_{nT}^{-2}\right)$, by Lemma A.3*(ii)*. The bounds for II and III are not necessarily the sharpest ones, but are sufficient for our purpose. Putting all together, $\hat{\bar{\gamma}} = \gamma + O_p\left(\delta_{nT}^{-2}\right)$. QED

Proof of Lemma A.5. Consider (1) and let $\bar{F} = f\times i_T$; under H_0^b, it holds that

$$\begin{aligned}\hat{\bar{F}}-\bar{F} &= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\left(\tilde{\beta}_j-\beta_j\right)'\frac{1}{T}\sum_{t=1}^{T}x_{jt} \\ &\quad -\frac{1}{n}\left(\frac{\hat{F}'F}{T}\right)\sum_{j=1}^{n}\gamma_j\left(\tilde{\beta}_j-\beta_j\right)'\frac{1}{T}\sum_{t=1}^{T}x_{jt}-\frac{1}{nT}\sum_{j=1}^{n}\left(\hat{F}'\epsilon_j\right)\left(\tilde{\beta}_j-\beta_j\right)'\frac{1}{T}\sum_{t=1}^{T}x_{jt} \\ &\quad -\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\gamma_j'\frac{1}{T}\sum_{t=1}^{T}f_t-\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_{jt} \\ &\quad +\frac{1}{nT}\sum_{j=1}^{n}\left(\hat{F}'\epsilon_j\right)\gamma_j'\frac{1}{T}\sum_{t=1}^{T}f_t+\frac{1}{n\sqrt{T}}\left(\frac{\hat{F}'F}{T}\right)\sum_{j=1}^{n}\gamma_j\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_{jt}+\frac{1}{nT^{3/2}}\sum_{j=1}^{n}\left(\hat{F}'\epsilon_j\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_{jt} \\ &= I-II-III-IV-V+VI+VII+VIII.\end{aligned}$$

Consider I; it is bounded by $E\left[\left\|\frac{\hat{F}'X_j}{T}\right\|\left\|\tilde{\beta}_j-\beta_j\right\|^2\left\|\frac{1}{T}\sum_{t=1}^T x_{jt}\right\|\right]\leq E\left[\left\|\tilde{\beta}_j-\beta_j\right\|^3\right]^{2/3} E\left[\left\|\frac{\hat{F}'X_j}{T}\right\|^6\right]^{1/6}$ $E\left[\left\|\frac{1}{T}\sum_{t=1}^T x_{jt}\right\|^6\right]^{1/6}=O\left(\phi_{nT}^{-2}\right)$, using a similar logic to (6) and Lemma A.1. Similar arguments yield $II=O_p\left(n^{-1/2}T^{-1/2}\right)+O_p\left(n^{-1}\right)$, $III=O_p\left(\phi_{nT}^{-2}\right)$, $IV=O_p\left(n^{-1/2}\delta_{nT}^{-1}\right)$, and $VI=O_p\left(n^{-1/2}T^{-1/2}\right)$. Consider V; this is bounded by

$$\begin{aligned}&\frac{1}{\sqrt{T}}E\left[\left\|\frac{\hat{F}'X_j}{T}\right\|\left\|\tilde{\beta}_j-\beta_j\right\|\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T\epsilon_{jt}\right|\right]\\ \leq\;&\frac{1}{\sqrt{T}}\left[E\left\|\frac{\hat{F}'X_j}{T}\right\|^4\right]^{1/4}\left[E\left\|\tilde{\beta}_j-\beta_j\right\|^2\right]^{1/2}\left[E\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T\epsilon_{jt}\right|^4\right]^{1/4}=\frac{1}{\sqrt{T}}O\left(1\right)O\left(\phi_{nT}^{-1}\right)O\left(1\right),\end{aligned}$$

using again Lemma A.1 and the fact that $E\left\|T^{-1/2}\sum_{t=1}^T\epsilon_{jt}\right\|^4=O\left(1\right)$ - this can be shown using Assumption 1*(iv)*(b) and similar passages as in the proof of Lemma A.1. Hence, $V=O_p\left(T^{-1/2}\phi_{nT}^{-1}\right)$; similarly, $VII=O_p\left(n^{-1/2}T^{-1/2}\right)$ and $VIII=O_p\left(T^{-1/2}\delta_{nT}^{-2}\right)$. Putting all together, this yields $\widehat{\bar{F}}-\bar{F}=O_p\left(\delta_{nT}^{-2}\right)$. QED

Proof of Lemma A.6. In the proof, we extensively use the fact that, for an arbitrary sequence of random variables $Z_1,...,Z_m$ such that $\max_{1\leq h\leq m}E\left|Z_h\right|^a\leq M$ for some $a>0$, it holds that

$$\max_{1\leq h\leq m}\left|Z_h\right|=o_p\left(m^{1/a}\right). \tag{8}$$

The proofs are rather repetitive, and where possible we only provide an intuition of the main argument, omitting passages.

Consider part *(i)*. We know, from the proof of Lemma A.1, that

$$\left\|\tilde{\beta}_i-\beta_i\right\|^2\leq\left\|\left(\frac{X_i'\bar{M}_wX_i}{T}\right)^{-1}\right\|_1^2\left\|\frac{X_i'\bar{M}_w\epsilon_i}{T}+\frac{X_i'\bar{M}_wF}{T}\gamma_i\right\|^2\leq\left\|\frac{X_i'\epsilon_i}{T}\right\|^2+\left\|\frac{X_i'\bar{M}_wF}{T}\right\|^2,$$

so that the order of magnitude of $\max_{1\leq i\leq n}\left\|\tilde{\beta}_i-\beta_i\right\|^2$ can be derived by studying $T^{-1}\max_{1\leq i\leq n}\left\|T^{-1/2}X_i'\epsilon_i\right\|^2$ and $\max_{1\leq i\leq n}\left\|T^{-1}X_i'\bar{M}_wF\right\|^2$. Consider the former. By the proof of Lemma A.1, we know that $E\left\|T^{-1/2}X_i'\epsilon_i\right\|^a$ is bounded by $E\left\|x_{it}\epsilon_{it}\right\|^a\leq E\left\|x_{it}\right\|^a\left|\epsilon_{it}\right|^a = E\left\|x_{it}\right\|^a E\left|\epsilon_{it}\right|^a$ by using Assumption 2*(iii)*. The largest a for which this moment exists is $a=k/2$, whence $\max_{1\leq i\leq n}\left\|T^{-1/2}X_i'\epsilon_i\right\|^2=o_p\left(n^{2/k}\right)$. This entails $T^{-1}\max_{1\leq i\leq n}\left\|T^{-1/2}X_i'\epsilon_i\right\|^2=o_p\left(n^{2/k}T^{-1}\right)$. As far as $\max_{1\leq i\leq n}\left\|T^{-1}X_i'\bar{M}_wF\right\|^2$ is concerned, we know from the proof of Lemma A.1 that $\left\|T^{-1}X_i'\bar{M}_wF\right\|^2$ has magnitude $O_p\left(n^{-1/2}\delta_{nT}^{-1}\right)$. When applying $\max_{1\leq i\leq n}$, this only affects the x_{it}s. To illustrate this, consider term I in (4): $\frac{1}{\sqrt{nT}}\max_{1\leq i\leq n}\left\|\frac{X_i'FCD_w^{-1}\sqrt{n}\bar{\epsilon}'F}{T\sqrt{T}}\right\|^2$

$\leq \frac{1}{\sqrt{nT}} \left(\max_{1\leq i\leq n} \left\| \frac{X_i}{\sqrt{T}} \right\|^2 \right) \left\| \frac{F}{\sqrt{T}} \right\|^2 \left\| \frac{\bar{\epsilon}' F}{\sqrt{nT}} \right\|^2$. We have $\left\| T^{-1/2} X_i \right\|^2 = T^{-1} \sum_{t=1}^{T} x_{it}^2$, so that, based on (8), $\max_{1\leq i\leq n} \left\| T^{-1/2} X_i \right\|^2 = o_p \left(n^{2/k} \right)$. Therefore, the whole expression is of order $o_p \left(n^{-1/2} T^{-1/2} n^{2/k} \right)$. Applying similar passages to terms II and III in (4) yields $\max_{1\leq i\leq n} \left\| T^{-1} X_i' \bar{M}_w F \right\|^2 = o_p \left(n^{2/k-1/2} \delta_{nT}^{-1} \right)$. Part *(i)* follows putting everything together.

Consider part *(ii)*. The passages of the proof are rather repetitive. The main argument is that, based on (32), $\max_{1\leq t\leq T} \delta_{nt}^{-2} \left\| \hat{f}_t - H' f_t \right\|^2$ is bounded by terms such as $\delta_{nt}^{-2} \left\| \frac{1}{n} \sum_{j=1}^{n} \left(\frac{\hat{F}' X_j}{T} \right) \left(\tilde{\beta}_j - \beta_j \right) \left(\tilde{\beta}_j - \beta_j \right)' \right\|^2 \max_{1\leq t\leq T} \left\| x_{jt} \right\|^2$, etc. This entails that, when taking the maximum across t, the order of magnitude of the maximum is given by terms like $\max_{1\leq t\leq T} \left\| x_{jt} \right\|^2$, $\max_{1\leq t\leq T} \left\| f_t \right\|^2$ and $\max_{1\leq t\leq T} \left\| \epsilon_{jt} \right\|^2$, which are of order $o_p \left(T^{2/k} \right)$. This provides part *(ii)*.

The proof of part *(iii)* is based on (31):

$$
\begin{aligned}
T \left\| \hat{\gamma}_i - \gamma_i \right\|^2 &\leq \left\| \left(\frac{\hat{F}' M_{Xi} \hat{F}}{T} \right)^{-1} \right\|_1^2 \left\| \frac{\hat{F}' M_{Xi} \epsilon_i}{\sqrt{T}} - \frac{\hat{F}' M_{Xi} \left(\hat{F} - F \right) \gamma_i}{\sqrt{T}} \right\|^2 \\
&\leq \left\| \frac{\hat{F}' \epsilon_i}{\sqrt{T}} \right\|^2 + \left\| \frac{\hat{F}' \left(\hat{F} - F \right)}{\sqrt{T}} \right\|^2 \left\| \gamma_i \right\|^2 .
\end{aligned}
$$

By Assumption 3*(iii)*, $\max_{1\leq i\leq n} \left\| T^{-1/2} \hat{F}' \left(\hat{F} - F \right) \right\|^2 \left\| \gamma_i \right\|^2$ has the same order of magnitude as $\left\| T^{-1/2} \hat{F}' \left(\hat{F} - F \right) \right\|^2$, i.e. $O_p \left(\sqrt{T} n^{-1} \right) + O_p \left(T^{-1/2} \right)$. As far as $\max_{1\leq i\leq n} \left\| T^{-1/2} \hat{F}' \epsilon_i \right\|^2$ is concerned

$$
\max_{1\leq i\leq n} \left\| \frac{\hat{F}' \epsilon_i}{\sqrt{T}} \right\|^2 \leq \max_{1\leq i\leq n} \left\| \frac{F' \epsilon_i}{\sqrt{T}} \right\|^2 + T \max_{1\leq i\leq n} \left\| \frac{\left(\hat{F} - F \right)' \epsilon_i}{T} \right\|^2 = I + II.
$$

Consider I; based on the same arguments as in (3), we have $\max_{1\leq i\leq n} \left\| T^{-1/2} F' \epsilon_i \right\|^2 \leq \max_{1\leq i\leq n} T^{-1} \sum_{t=1}^{T} \left\| f_t \epsilon_{it} \right\|^2$. Also, $E \left\| f_t \epsilon_{it} \right\|^a \leq E \left\| f_t \right\|^a E \left| \epsilon_{it} \right|^a < \infty$ with the largest a being $a = 2k$, whence $\max_{1\leq i\leq n} \left\| T^{-1/2} F' \epsilon_i \right\|^2 = o_p \left(n^{2/k} \right)$. Turning to II, $\max_{1\leq i\leq n} \left\| T^{-1} \left(\hat{F} - F \right)' \epsilon_i \right\|^2$ can be studied using

(5). It follows that $\max_{1\leq i\leq n}\left\|T^{-1}\left(\hat{F}-F\right)'\epsilon_i\right\|^2$ is bounded by the sum of terms like

$$\begin{aligned} & \max_{1\leq i\leq n}\left\|\frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\left(\frac{\epsilon_i'X_j}{\sqrt{T}}\right)\left(\tilde{\beta}_j-\beta_j\right)\left(\tilde{\beta}_j-\beta_j\right)'\frac{X_j'\hat{F}}{T}\right\|^2 \\ \leq\ & \frac{1}{T}\max_{1\leq i\leq n}\left[\frac{1}{n}\sum_{j=1}^{n}\left\|\frac{\epsilon_i'X_j}{\sqrt{T}}\right\|^6\right]^{1/3}\left[\frac{1}{n}\sum_{j=1}^{n}\left\|\tilde{\beta}_j-\beta_j\right\|^3\right]^{4/3}\left[\frac{1}{n}\sum_{j=1}^{n}\left\|\frac{X_j'\hat{F}}{T}\right\|^6\right]^{1/3} \\ \leq\ & \left[\frac{1}{nT}\sum_{j=1}^{n}\max_{1\leq i\leq n}\left\|\frac{\epsilon_i'X_j}{\sqrt{T}}\right\|^2\right]\left[\frac{1}{n}\sum_{j=1}^{n}\left\|\tilde{\beta}_j-\beta_j\right\|^3\right]^{4/3}\left[\frac{1}{n}\sum_{j=1}^{n}\left\|\frac{X_j'\hat{F}}{T}\right\|^6\right]^{1/3}, \end{aligned}$$

which follows from (6) (first line) and from the C_r-inequality (second line). Note that $E\left\|T^{-1/2}\epsilon_i'X_j\right\|^a$ is bounded by $E\left\|\epsilon_{it}x_{jt}\right\|^a\leq E\left\|x_{jt}\right\|^a E\left|\epsilon_{it}\right|^a$ with $a=k$ at most. We have that $\max_{1\leq i\leq n}\left\|T^{-1/2}\epsilon_i'X_j\right\|^2=o_p\left(n^{2/k}\right)$. Applying the same logic to the squares of all the terms in (5), it follows that $II=o_p\left(n^{2/k}T\delta_{nT}^{-4}\right)$. Part *(iii)* follows from putting everything together.

Consider parts *(iv)* and *(v)*. Using the definition of $\hat{\epsilon}_{it}$:

$$\begin{aligned} \max_{1\leq t\leq T}\hat{\epsilon}_{it}^2 \leq\ & \max_{1\leq t\leq T}\epsilon_{it}^2+\left\|\tilde{\beta}_i-\beta_i\right\|^2\max_{1\leq t\leq T}\left\|x_{it}\right\|^2+\left\|\hat{\gamma}_i-\gamma_i\right\|^2\max_{1\leq t\leq T}\left\|f_t\right\|^2 \\ & +\left\|\hat{\gamma}_i\right\|^2\max_{1\leq t\leq T}\left\|\hat{f}_t-f_t\right\|^2=I+II+III+IV. \end{aligned}$$

Parts *(iv)* and *(v)* follow immediately from Assumptions 1 and 2. Explicit rates are derived using the other parts of this Lemma. Parts *(vi)* and *(vii)* can be proved similarly, using

$$\max_{1\leq i\leq n}\hat{\epsilon}_{it}^2\leq\max_{1\leq i\leq n}\epsilon_{it}^2+\max_{1\leq i\leq n}\left\|\tilde{\beta}_i-\beta_i\right\|^2\left\|x_{it}\right\|^2+\left\|f_t\right\|^2\max_{1\leq i\leq n}\left\|\hat{\gamma}_i-\gamma_i\right\|^2+\left\|\hat{f}_t-f_t\right\|^2\max_{1\leq i\leq n}\left\|\hat{\gamma}_i\right\|^2,$$

and $\max_{1\leq i\leq n}\epsilon_{it}^2=o_p\left(n^{2/k}\right)$ by (8).

Proof of Lemma A.7. As a preliminary result, note that Assumptions 1*(i)*, 2*(i)* and 6*(i)* entail that ϵ_{it}^2, $f_t\epsilon_{it}$, $x_{it}\epsilon_{it}$, $\mathbf{vec}\left(f_tf_t'\right)$, $\mathbf{vec}\left(f_tx_{it}'\right)$ and $\mathbf{vec}\left(f_tf_t'\epsilon_{it}^2\right)$ are all $L_{2+\delta}$-NED of size $\alpha'>\frac{1}{2}$ on $\{v_t\}_{t=-\infty}^{+\infty}$, for each i. These results are applications of Example 17.17 in Davidson (1994, p. 273), and are explicitly reported in Kao, Trapani and Urga (2012; see in particular Lemmas 8 and 9 therein). Similarly, Assumption 8 entails that ϵ_{it}^2 is $L_{2+\delta}$-NED of size $\alpha'>\frac{1}{2}$ on $\{v_i\}_{i=-\infty}^{+\infty}$, for each t.

Consider part *(i)*, writing $T^{-1}\hat{F}'\hat{F}-\Sigma_f=\left(T^{-1}\hat{F}'\hat{F}-T^{-1}F'F\right)+\left(T^{-1}F'F-\Sigma_f\right)$. Lemma A.3 (parts *(ii)* and *(iii)*) entails that $T^{-1}\hat{F}'\hat{F}-T^{-1}F'F=O_p\left(\delta_{nT}^{-2}\right)$. The CLT for NED sequences can be applied (Theorem 24.6 and Corollary 24.7 in Davidson , 1994, p. 386-387), so that $E\left\|T^{-1}F'F-\Sigma_f\right\|^2$

$= E\left\|T^{-1}\sum_{t=1}^{T}\left(f_t f_t' - \Sigma_f\right)\right\|^2 = O\left(T^{-1}\right)$. Putting all together, part *(i)* follows. As far as part *(ii)* is concerned, it follows immediately from noting that $\left\|T^{-1}\hat{F}'M_{Xi}\hat{F} - \Sigma_{fM,i}\right\| \leq \left\|T^{-1}\hat{F}'\hat{F} - \Sigma_f\right\|$ for each i, by definition of M_{Xi}.

As far as showing parts *(iii)*-*(vi)* is concerned, we extensively use the following result, which is an application of Theorem 2.1 in Berkes, Liu and Wu (2013; see also Theorem 2.2 in Ling, 2007). Given an $L_{2+\delta}$-NED zero mean sequence $Z_1, ..., Z_m$ of size (equal to or greater than) $\frac{1}{2}$, such that $E\left|Z_1\right|^k \leq M$ for some $k > 2$, and that the conditions spelt out in Assumptions 6*(ii)* and 6*(iii)* hold; and given a Brownian motion $W(\cdot)$ with $E\left[W^2(1)\right] = \lim_{m\to\infty} E\left[\left(m^{-1/2}\sum_{h=1}^{m} Z_h\right)^2\right]$, it holds that, redefining Z_h in a richer probability space

$$\left\|\frac{1}{\sqrt{m}}\sum_{h=1}^{m} Z_h - W(1)\right\| = O_p\left(m^{1/k-1/2}\right). \tag{9}$$

Results like (9) are known as "Hungarian constructions"; see, *inter alia*, Csörgö and Révész (1975a,b), and Komlós, Major and Tusnády (1975, 1976); we also refer to Shorack and Wellner (1986) for a review. Hungarian constructions are usually stated in terms of the partial sum process $m^{-1/2}\sum_{h=1}^{\lfloor m\tau\rfloor} Z_h$ for $\tau \in [0, 1]$; in that case, (9) is stated using the sup-norm. For our purposes, we only need to consider $\tau = 1$. The rate in (9) is sharp, and this result was shown, for the case of dependent data, only very recently (Berkes, Liu and Wu, 2013). By Assumptions 6 and 8, and by the fact that, as stated above, ϵ_{it}^2, $f_t\epsilon_{it}$, $x_{it}\epsilon_{it}$, $\mathbf{vec}\left(f_t f_t'\right)$, $\mathbf{vec}\left(f_t x_{it}'\right)$ and $\mathbf{vec}\left(f_t f_t' \epsilon_{it}^2\right)$ are all $L_{2+\delta}$-NED of size $\alpha' > \frac{1}{2}$, equation (9) can be applied to the normalized sums of all these sequences - in the case of ϵ_{it}^2, to both sums across t and across i. As a final remark, we point out that if all moments of Z_h exist (e.g. if Z_h is Gaussian), the rate in (9) becomes exponential, i.e. (9) holds with a rate $O_p\left(\frac{\ln m}{\sqrt{m}}\right)$.

We turn to the proof of part *(iii)* of the Lemma. We have

$$\begin{aligned}
\max_{1\leq t\leq T}\left\|\hat{\Sigma}_{\Gamma\epsilon,t} - \Sigma_{\Gamma\epsilon,t}\right\| &\leq \max_{1\leq t\leq T}\left\|\frac{1}{n}\sum_{i=1}^{n}\gamma_i\gamma_i'\epsilon_{it}^2 - \Sigma_{\Gamma\epsilon,t}\right\| + \max_{1\leq t\leq T}\left\|\frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_i\hat{\gamma}_i'\left(\hat{\epsilon}_{it}^2 - \epsilon_{it}^2\right)\right\| \\
&\quad +2\max_{1\leq t\leq T}\left\|\frac{1}{n}\sum_{i=1}^{n}\gamma_i\left(\hat{\gamma}_i - \gamma_i\right)'\hat{\epsilon}_{it}^2\right\| + \max_{1\leq t\leq T}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\gamma}_i - \gamma_i\right)\left(\hat{\gamma}_i - \gamma_i\right)'\hat{\epsilon}_{it}^2\right\| \\
&= I + II + III + IV.
\end{aligned}$$

Consider I. Recall that the sequence $z_{\epsilon\gamma,it} = \gamma_i\gamma_i'\epsilon_{it}^2 - \Sigma_{\Gamma\epsilon,t}$ is, as stated above, $L_{2+\delta}$-NED of size $\alpha' > \frac{1}{2}$. Therefore, by Theorem 17.5(b) in Davidson (1994, p. 264), $z_{\epsilon\gamma,it}$ is an $L_{2+\delta}$-mixingale of size $\min\left\{\alpha', \frac{k-1}{k-2}\right\} > \frac{1}{2}$. Using Assumption 3*(iii)* and Corollary 1 in Peligrad, Utev, and Wu (2007), it follows that $E\left\|n^{-1/2}\sum_{i=1}^{n} z_{\epsilon\gamma,it}\right\|^a \leq ME\left|\epsilon_{it}^2\right|^a < \infty$. By Assumption 1*(i)*, the largest a for which

$E\left\|n^{-1/2}\sum_{i=1}^{n}z_{\epsilon\gamma,it}\right\|^{k}<\infty$ is $k/2$. Thus, $\max_{1\leq t\leq T}\left\|n^{-1/2}\sum_{i=1}^{n}z_{\epsilon\gamma,it}\right\|=o_{p}\left(T^{2/k}\right)$, which entails $I=o_{p}\left(T^{2/k}n^{-1/2}\right)$. As far as II is concerned, it has the same order of magnitude as $\max_{1\leq t\leq T}\left|\hat{\epsilon}_{it}^{2}-\epsilon_{it}^{2}\right|$, given in Lemma A.6*(v)*. Turning to III, its order of magnitude is given by $O_{p}\left(\left\|\hat{\gamma}_{i}-\gamma_{i}\right\|^{2}\right)\max_{1\leq t\leq T}\hat{\epsilon}_{it}^{2}$, which comes from Lemma A.6*(iv)*. Term IV is dominated. Putting all together, part *(iii)* of the Lemma follows.

As far as part *(iv)* is concerned, the proof is similar, in spirit, to that of part *(iii)*. Recall that $\hat{\Sigma}_{\gamma i}=\left(Q_{i}^{\prime}\right)^{-1}D_{0,i}\left(Q_{i}\right)^{-1}$; the rates for $Q_{i}=T^{-1}\hat{F}^{\prime}M_{Xi}\hat{F}$ are given by part *(ii)* of this Lemma. Based on the definition of $D_{0,i}$, we have

$$
\begin{aligned}
& \max_{1\leq i\leq n}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{f}_{t}\hat{f}_{t}^{\prime}\hat{\epsilon}_{it}^{2}-H^{\prime}\Sigma_{f}HE\left(\epsilon_{it}^{2}\right)\right\| \\
\leq & \max_{1\leq i\leq n}\left\|\frac{1}{T}\sum_{t=1}^{T}f_{t}f_{t}^{\prime}\epsilon_{it}^{2}-H^{\prime}\Sigma_{f}HE\left(\epsilon_{it}^{2}\right)\right\|+\max_{1\leq i\leq n}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{f}_{t}\hat{f}_{t}^{\prime}\left(\hat{\epsilon}_{it}^{2}-\epsilon_{it}^{2}\right)\right\| \\
& +2\max_{1\leq i\leq n}\left\|\frac{1}{T}\sum_{t=1}^{T}f_{t}\left(\hat{f}_{t}-f_{t}\right)^{\prime}\epsilon_{it}^{2}\right\|+\max_{1\leq i\leq n}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\hat{f}_{t}-\hat{f}_{t}\right)\left(\hat{f}_{t}-\hat{f}_{t}\right)^{\prime}\epsilon_{it}^{2}\right\| \\
= & I+II+III+IV.
\end{aligned}
$$

As far as I is concerned, the proof is similar to that of part *(iii)* of this Lemma, upon recalling that $f_{t}f_{t}^{\prime}\epsilon_{it}^{2}$ is, across t, $L_{2+\delta}$-NED of size $\alpha^{\prime}>\frac{1}{2}$. Indeed, the largest a for which $E\left\|f_{t}f_{t}^{\prime}\epsilon_{it}^{2}\right\|^{a}\leq E\left\|f_{t}\right\|^{2a}E\left|\epsilon_{it}\right|^{2a}$ is $a=k/2$, and that $T^{-1}\sum_{t=1}^{T}f_{t}f_{t}^{\prime}\epsilon_{it}^{2}-H^{\prime}\Sigma_{f}HE\left(\epsilon_{it}^{2}\right)=O_{p}\left(T^{-1/2}\right)$; hence, $I=o_{p}\left(n^{2/k}T^{-1/2}\right)$. Considering II, we have

$$
\begin{aligned}
II & \leq \frac{1}{T}\sum_{t=1}^{T}\left\|\hat{f}_{t}\hat{f}_{t}^{\prime}\right\|\max_{1\leq i\leq n}\left|\hat{\epsilon}_{it}^{2}-\epsilon_{it}^{2}\right| \\
& \leq \left[\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{f}_{t}\hat{f}_{t}^{\prime}\right\|^{2}\right]^{1/2}\left[E\left(\max_{1\leq i\leq n}\left|\hat{\epsilon}_{it}^{2}-\epsilon_{it}^{2}\right|\right)^{2}\right]^{1/2};
\end{aligned}
$$

applying Lemma A.6*(v)*, we have $II=o_{p}\left(T^{-2/k}\delta_{nT}^{-2}\right)$. Considering III, a similar logic as above yields

$$
\begin{aligned}
III & \leq \frac{1}{T}\sum_{t=1}^{T}\left\|f_{t}\left(\hat{f}_{t}-f_{t}\right)^{\prime}\right\|\max_{1\leq i\leq n}\hat{\epsilon}_{it}^{2} \\
& \leq \left[\frac{1}{T}\sum_{t=1}^{T}\left\|f_{t}\left(\hat{f}_{t}-f_{t}\right)^{\prime}\right\|^{2}\right]^{1/2}\left[E\left(\max_{1\leq i\leq n}\hat{\epsilon}_{it}^{2}\right)^{2}\right]^{1/2};
\end{aligned}
$$

also,

$$
\left[\frac{1}{T}\sum_{t=1}^{T}\left\|f_{t}\left(\hat{f}_{t}-f_{t}\right)^{\prime}\right\|^{2}\right]^{1/2}\leq\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\left\|f_{t}\left(\hat{f}_{t}-f_{t}\right)^{\prime}\right\|=O_{p}\left(\sqrt{T}\delta_{nT}^{-2}\right),
$$

by the C_r-inequality and Lemma A.3*(ii)*. Using Lemma A.6*(vi)*, we have $III = o_p\left(n^{2/k}T^{1/2}\delta_{nT}^{-2}\right) + o_p\left(n^{4/k}T^{1/2}\phi_{nT}^{-2}\delta_{nT}^{-2}\right) + o_p\left(n^{2/k-2}T^{3/2}\delta_{nT}^{-2}\right)$. Term IV is dominated. Putting all together, part *(iv)* follows.

Consider now part *(v)*. We have $T^{-1/2}F'M_{Xi}\epsilon_i = T^{-1/2}\sum_{t=1}^T f_t\epsilon_{it} - \left(T^{-1}\sum_{t=1}^T f_t x_{it}'\right)\left(T^{-1}\sum_{t=1}^T x_{it}x_{it}'\right)^{-1} T^{-1/2}\sum_{t=1}^T x_{it}\epsilon_{it}$. Since $f_t\epsilon_{it}$ and $x_{it}\epsilon_{it}$ are $L_{2+\delta}$-NED of size $\alpha' > \frac{1}{2}$, equation (9) holds with $k^* = 4$: there are two sequences of *i.i.d.* Gaussian, zero mean random variables, say $\left\{N_{it}^{f\epsilon}\right\}_{t=1}^T$ and $\{N_{it}^{x\epsilon}\}_{t=1}^T$, such that $E\left[\left(N_{it}^{f\epsilon}\right)^2\right] = \Sigma_{f\epsilon,i}$ and $E\left[(N_{it}^{x\epsilon})^2\right] = \Sigma_{x\epsilon,i}$ and $T^{1/k-1/2}\left\|T^{-1/2}\sum_{t=1}^T f_t\epsilon_{it} - T^{-1/2}\sum_{t=1}^T N_{it}^{f\epsilon}\right\| = O_p(1)$ and $T^{1/k-1/2}\left\|T^{-1/2}\sum_{t=1}^T x_{it}\epsilon_{it} - T^{-1/2}\sum_{t=1}^T N_{it}^{x\epsilon}\right\| = O_p(1)$. Further, by the CLT for NED processes (see e.g. Theorem 24.6 in Davidson , 1994, p. 386), $T^{-1}\sum_{t=1}^T f_t x_{it}' = \Sigma_{fx,i} + O_p\left(T^{-1/2}\right)$ and $T^{-1}\sum_{t=1}^T x_{it}x_{it}' = \Sigma_{xx,i} + O_p\left(T^{-1/2}\right)$. Putting all together, and defining the *i.i.d.* Gaussian sequence $N_i \equiv T^{-1/2}\sum_{t=1}^T N_{it}^{f\epsilon} - \Sigma_{fx,i}\Sigma_{xx,i}^{-1}T^{-1/2}\sum_{t=1}^T N_{it}^{x\epsilon}$, we have that $T^{1/k-1/2}\left\|T^{-1/2}F'M_{Xi}\epsilon_i - N_i\right\| = O_p(1)$. Note further that, $E\left\|T^{-1/2}F'M_{Xi}\epsilon_i\right\|^r \leq E\left\|T^{-1/2}\sum_{t=1}^T f_t\epsilon_{it}\right\|^r$; note that $f_t\epsilon_{it}$ is $L_{2+\delta}$-NED of size $\alpha' > \frac{1}{2}$ on $\{v_t\}_{t=-\infty}^{+\infty}$, which entails that it is an $L_{2+\delta}$-mixingale of size $\min\left\{\alpha', \frac{k-1}{k-2}\right\} > \frac{1}{2}$. Therefore, using again Corollary 1 in Peligrad, Utev, and Wu (2007), $E\left\|T^{-1/2}\sum_{t=1}^T f_t\epsilon_{it}\right\|^r \leq M\ E\left\|f_t\epsilon_{it}\right\|^r$; by Assumptions 1*(i)*, 2*(i)* and 2*(iii)*, the largest r for which $E\left\|T^{-1/2}F'M_{Xi}\epsilon_i\right\|^r < \infty$ is $r = k$. Thus, $\max_{1\leq i\leq n} T^{1/k-1/2}\left\|T^{-1/2}F'M_{Xi}\epsilon_i - N_i\right\| = o_p\left(n^{1/k}\right)$, which proves part *(v)*.

The proof of part *(vi)* is similar. Indeed, given an independent, zero mean, Gaussian sequence $\{N_{it}^n\}_{i=1}^n$ with $E\left(N_{it}^n\right)^2 = \gamma_i\gamma_i'\ E\left(\epsilon_{it}^2\right)$, write

$$
\begin{aligned}
& \max_{1\leq t\leq T}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\gamma}_i\epsilon_{it} - \frac{1}{\sqrt{n}}\sum_{i=1}^n N_{it}^n\right\| \\
\leq & \max_{1\leq t\leq T}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \gamma_i\epsilon_{it} - \frac{1}{\sqrt{n}}\sum_{i=1}^n N_{it}^n\right\| + \max_{1\leq t\leq T}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\hat{\gamma}_i - \gamma_i\right)\epsilon_{it}\right\| \\
= & I + II.
\end{aligned}
$$

By virtue of Assumption 8, and using similar considerations as for the proof of part *(v)*, term I satisfies (9) with

$$
n^{1/2-1/k}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \gamma_i\epsilon_{it} - \frac{1}{\sqrt{n}}\sum_{i=1}^n N_{it}^n\right\| = O_p(1);
$$

also, the largest existing moment over t is of order k. Thus, $\max_{1\leq t\leq T} n^{1/k-1/2}\left\|n^{-1/2}\sum_{i=1}^n \gamma_i\epsilon_{it} - n^{-1/2}\sum_{i=1}^n N_{it}^n\right\| = o_p\left(T^{1/k}\right)$, whence $I = o_p\left(T^{1/k}n^{1/k-1/2}\right)$. Consider now II. By (31), we can write

$$
II = \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(F'M_{Xi}F\right)^{-1}\left(F'M_{Xi}\epsilon_i\right)\epsilon_{it} + II_b = II_a + II_b,
$$

where II_b contains the remainder of $(\hat{\gamma}_i - \gamma_i)$. By the results in the proof of Theorem 1, and by using the Cauchy-Schwartz inequality, it follows immediately that $II_b = o_p\left(T^{1/k}\sqrt{n}\delta_{nT}^{-2}\right) + o_p\left(T^{1/k}\sqrt{n}\delta_{nT}^{-2}\right)$. Also, $II_a \leq M\ n^{-1/2}\ T^{-1}\ \left\|\sum_{i=1}^n \sum_{s=1}^T f_s \epsilon_{is}\epsilon_{it}\right\|$, so that

$$II_a \leq M\frac{1}{\sqrt{n}T}\left\|\sum_{i=1}^n \sum_{s=1}^T f_s\left[\epsilon_{is}\epsilon_{it} - E\left(\epsilon_{is}\epsilon_{it}\right)\right]\right\| + M\frac{1}{\sqrt{n}T}\left\|\sum_{i=1}^n \sum_{s=1}^T f_s E\left(\epsilon_{is}\epsilon_{it}\right)\right\| = II_{a,1} + II_{a,2}.$$

We have that $II_{a,1}$ is bounded by the square root of

$$\begin{aligned} & E\left\|\frac{1}{nT^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{u=1}^T\sum_{s=1}^T f_s f_u' \left[\epsilon_{is}\epsilon_{it} - E\left(\epsilon_{is}\epsilon_{it}\right)\right]\left[\epsilon_{ju}\epsilon_{jt} - E\left(\epsilon_{ju}\epsilon_{jt}\right)\right]\right\| \\ \leq\ & \frac{1}{nT^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{u=1}^T\sum_{s=1}^T E\left\|f_s f_u'\right\| E\left|\left[\epsilon_{is}\epsilon_{it} - E\left(\epsilon_{is}\epsilon_{it}\right)\right]\left[\epsilon_{ju}\epsilon_{jt} - E\left(\epsilon_{ju}\epsilon_{jt}\right)\right]\right| \\ \leq\ & M\frac{1}{nT^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{u=1}^T\sum_{s=1}^T E\left|\left[\epsilon_{is}\epsilon_{it} - E\left(\epsilon_{is}\epsilon_{it}\right)\right]\left[\epsilon_{ju}\epsilon_{jt} - E\left(\epsilon_{ju}\epsilon_{jt}\right)\right]\right| \\ \leq\ & M'\frac{1}{nT^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{u=1}^T\sum_{s=1}^T E\left|\epsilon_{is}\epsilon_{it}\epsilon_{ju}\epsilon_{jt}\right| \leq M''\frac{1}{T}, \end{aligned}$$

on account of Assumptions 2*(iii)* and 1*(iii)*(c); hence, $II_{a,1} = O_p\left(T^{-1/2}\right)$. The same logic entails $II_{a,2} = O_p\left(T^{-1/2}\right)$ also. Putting all together, $II = o_p\left(T^{1/k}\delta_{nT}^{-1}\right) + o_p\left(T^{1/k}\sqrt{n}\delta_{nT}^{-2}\right)$. Defining $n^{-1/2}\sum_{i=1}^n N_{it}^n = N_t$, part *(vi)* follows. QED

Proof of Theorem 4. The proof is very similar, in spirit, to the proof of Theorem 3, and therefore some passages are omitted to save space. Consider the following preliminary notation and derivations. We write

$$\hat{f}_t - \widehat{\overline{f}} = \left(f_t - f\right) + \left(\hat{f}_t - f_t\right) - \left(\widehat{\overline{f}} - f\right) = a_t + b_t - c_t.$$

Under H_0^b, $a_t = 0$ and $b_t = \hat{f}_t - f$; using (30),we can write

$$b_t = \left(\frac{\hat{F}'F}{T}\right)\frac{1}{n}\sum_{i=1}^n \gamma_i\epsilon_{it} + b_{2t} = b_{1t} + b_{2t}, \tag{10}$$

where b_{2t} contains terms $I - VI$ and $VIII$ in (30). Also, for each t, $\hat{\Sigma}_{ft}^{-1} = \Sigma_{ft}^{-1} - \Sigma_{ft}^{-1}\left(\hat{\Sigma}_{ft} - \Sigma_{ft}\right)\Sigma_{ft}^{-1}$ $+ o_p\left(\left\|\hat{\Sigma}_{ft} - \Sigma_{ft}\right\|\right)$.

Neglecting higher order terms containing $o_p\left(\left\|\hat{\Sigma}_{ft}-\Sigma_{ft}\right\|\right)$, we have

$$\begin{aligned}
& n\left(\hat{f}_t-\widehat{\overline{f}}\right)'\hat{\Sigma}_{ft}^{-1}\left(\hat{f}_t-\widehat{\overline{f}}\right) \qquad (11)\\
= \; & n\left(b_{1t}'\Sigma_{ft}^{-1}b_{1t}\right)+nb_{1t}'\Sigma_{ft}^{-1}\left(\hat{\Sigma}_{ft}-\Sigma_{ft}^{-1}\right)\Sigma_{ft}^{-1}b_{1t}+nb_{2t}'\Sigma_{ft}^{-1}b_{2t}\\
& +2nb_{1t}'\Sigma_{ft}^{-1}b_{2t}+n\left(\widehat{\overline{f}}-f\right)'\hat{\Sigma}_{ft}^{-1}\left(\widehat{\overline{f}}-f\right)-2n\left(\widehat{\overline{f}}-f\right)'\hat{\Sigma}_{ft}^{-1}\left(\hat{f}_t-f\right)\\
= \; & n\left(b_{1t}'\Sigma_{ft}^{-1}b_{1t}\right)+I_t+II_t+III_t+IV_t-V_t.
\end{aligned}$$

After this preliminary calculations, we now turn to proving (24). Similarly to the proof of Theorem 3, we firstly prove that $\max_{1\leq t\leq T} n\left(b_{1t}'\Sigma_{ft}^{-1}b_{1t}\right)$ can be approximated by the maximum of a sequence of random variables with a χ_r^2 distribution, up to a negligible error. Secondly, we show that, in (11), $\max_{1\leq t\leq T} I_t$, ..., $\max_{1\leq t\leq T} V_t$ are all $o_p(\ln T)$ uniformly in t.

We start from $\max_{1\leq t\leq T} n\left(b_{1t}'\Sigma_{ft}^{-1}b_{1t}\right)$. We show that the sequence $\{\sqrt{n}b_{1t}\}_{t=1}^{T}$ can be approximated by a sequence of *i.i.d.* Gaussian random variables with covariance matrix Σ_{ft}. To show this, recall that by Lemma A.7*(vi)*, we can write $n^{-1/2}\sum_{i=1}^{n}\gamma_i\epsilon_{it}=N_t+R_{Nt}$, with N_t defined in Lemma A.7 as being zero mean Gaussian with covariance matrix $\Sigma_{\Gamma\epsilon,t}$, and $R_{Nt}=o_p\left(T^{1/k_2}n^{1/2-1/k_2}\right)+o_p\left(T^{1/k_2}\delta_{nT}^{-1}\sqrt{\ln T}\right)$ $+\ o_p\left(T^{1/k}\sqrt{n}\delta_{nT}^{-2}\sqrt{\ln T}\right)$. Further, $T^{-1}\hat{F}'F=\Sigma_f+\left(T^{-1}F'F-\Sigma_f\right)+T^{-1}\left(\hat{F}-F\right)'F=\Sigma_f+R_f$, with $R_f=O_p\left(T^{-1/2}\right)+O_p\left(n^{-1}\right)$ by Lemmas A.7*(i)* and A.3*(ii)*. Hence

$$\sqrt{n}b_{1t}=\left(\Sigma_f+R_f\right)\left(N_t+R_{Nt}\right), \qquad (12)$$

and

$$\begin{aligned}
n\left(b_{1t}'\hat{\Sigma}_{ft}^{-1}b_{1t}\right) \; = \; & N_t'\Sigma_{\Gamma\epsilon,t}^{-1}N_t+2N_t'\Sigma_f^{-1}\Sigma_{ft}^{-1}R_{Nt}+2R_{Nt}'\Sigma_{\Gamma\epsilon,t}^{-1}N_t \qquad (13)\\
& +2N_t'\Sigma_f^{-1}\Sigma_{ft}^{-1}R_fN_t+2N_t'\Sigma_f^{-1}\Sigma_{ft}^{-1}R_fR_{Nt}\\
& +R_{Nt}'\Sigma_{\Gamma\epsilon,t}^{-1}R_{Nt}+2R_{Nt}'\Sigma_f^{-1}\Sigma_{ft}^{-1}R_fR_{Nt}\\
& +N_t'R_f\Sigma_{ft}^{-1}R_fN_t+2N_t'R_f\Sigma_{ft}^{-1}R_fR_{Nt}\\
& +R_{Nt}R_f\Sigma_{ft}^{-1}R_fR_{Nt}\\
= \; & N_t'\Sigma_{\Gamma\epsilon,t}^{-1}N_t+I_t^{b1}+II_t^{b1}+III_t^{b1}+IV_t^{b1}+V_t^{b1}+VI_t^{b1}\\
& +VII_t^{b1}+VIII_t^{b1}+IX_t^{b1}.
\end{aligned}$$

Passages are very similar to those after (34) in the proof of Theorem 3. In particular, it can be shown using Lemma A.7 that: $\max_{1\leq t\leq T} I_t^{b1}$ and $\max_{1\leq t\leq T} II_t^{b1}$ are both $o_p\left(T^{1/k_2}n^{1/2-1/k_2}\sqrt{\ln T}\right)$

$+\ o_p\left(T^{1/k_2}\delta_{nT}^{-1}\sqrt{\ln T}\right)\ +\ o_p\left(T^{1/k}\sqrt{n}\delta_{nT}^{-2}\sqrt{\ln T}\right)$; $\max_{1\leq t\leq T} III_t^{b1}\ =\ O_p\left(T^{-1/2}\ln T\right)\ +\ O_p\left(n^{-1}\ln T\right)$; and that $\max_{1\leq t\leq T} IV_t^{b1}$,..., $\max_{1\leq t\leq T} IX_t^{b1}$ are all dominated and therefore negligible. Thus

$$\max_{1\leq t\leq T} n\left(b_{1t}'\Sigma_{ft}^{-1}b_{1t}\right)=\max_{1\leq t\leq T} N_t'\Sigma_{\Gamma\epsilon,t}^{-1}N_t+o_p\left[\sqrt{\frac{\ln T}{n}}\left(nT\right)^{1/k_2}\right]+o_p\left(T^{1/k_2}\frac{\sqrt{n\ln T}}{T}\right)+o_p\left(1\right), \quad (14)$$

where the approximation errors are negligible as long as $(n,T)\to\infty$ with $\frac{T^{4/k_2}}{n}\to 0$ and $T^{1/k_2}\frac{\sqrt{n}}{T}\to 0$.

After showing that $\max_{1\leq t\leq T} n\left(b_{1t}'\Sigma_{ft}^{-1}b_{1t}\right)$ can be approximated by $\max_{1\leq t\leq T} N_t'\Sigma_{\Gamma\epsilon,t}^{-1}N_t$, we turn back to equation (11). We show that $\max_{1\leq t\leq T} I_t$, ..., $\max_{1\leq t\leq T} V_t$ in (11) are all $o_p\left(\ln T\right)$. We have that $\max_{1\leq t\leq T} I_t\leq\max_{1\leq t\leq T}\left\|\sqrt{n}b_{1t}\right\|^2\max_{1\leq t\leq T}\left\|\Sigma_{ft}^{-1}\left(\hat{\Sigma}_{ft}-\Sigma_{ft}^{-1}\right)\Sigma_{ft}^{-1}\right\|=o_p\left(T^{2/k_2}\delta_{nT}^{-1}\ln T\right)$ by using Lemma A.7*(iii)*. Also, combining Lemmas A.5 and Lemma A.7*(iii)*, we have $\max_{1\leq t\leq T} IV_t=O_p\left(n\delta_{nT}^{-4}\right)+o_p\left(nT^{2/k_2}\delta_{nT}^{-5}\right)$ and $\max_{1\leq t\leq T} V_t=O_p\left(n\delta_{nT}^{-3}\right)+o_p\left(nT^{2/k_2}\delta_{nT}^{-4}\right)$. As far as $\max_{1\leq t\leq T} II_t$ and $\max_{1\leq t\leq T} III_t$ are concerned, studying their order of magnitude involves finding a bound for $\max_{1\leq t\leq T}\left\|b_{2t}\right\|$ and $\max_{1\leq t\leq T}\left\|b_{2t}\right\|^2$. Recall that

$$\begin{aligned} b_{2t} &= \frac{1}{n}\sum_{j=1}^n\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\left(\tilde{\beta}_j-\beta_j\right)'x_{jt}-\frac{1}{n}\left(\frac{\hat{F}'F}{T}\right)\sum_{j=1}^n\gamma_j\left(\tilde{\beta}_j-\beta_j\right)'x_{jt} \\ &-\frac{1}{nT}\sum_{j=1}^n\left(\hat{F}'\epsilon_j\right)\left(\tilde{\beta}_j-\beta_j\right)'x_{jt}-\frac{1}{n}\sum_{j=1}^n\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\gamma_j'f_t \\ &-\frac{1}{n}\sum_{j=1}^n\left(\frac{\hat{F}'X_j}{T}\right)\left(\tilde{\beta}_j-\beta_j\right)\epsilon_{jt}+\frac{1}{nT}\sum_{j=1}^n\left(\hat{F}'\epsilon_j\right)\gamma_j'f_t+\frac{1}{nT}\sum_{j=1}^n\left(\hat{F}'\epsilon_j\right)\epsilon_{jt}. \end{aligned}$$

Similar passages as in the proof of Lemma A.6*(ii)* yield $\max_{1\leq t\leq T}\left\|b_{2t}\right\|=o_p\left(T^{1/k_2}\delta_{nT}^{-2}\right)$ and $\max_{1\leq t\leq T}\left\|b_{2t}\right\|^2$ $=o_p\left(T^{2/k_2}\delta_{nT}^{-4}\right)$. We now turn to analyzing $\max_{1\leq t\leq T} II_t$ and $\max_{1\leq t\leq T} III_t$. As far as the former is concerned, $\max_{1\leq t\leq T} II_t\leq n\max_{1\leq t\leq T}\left\|b_{2t}\right\|^2=o_p\left(T^{2/k_2}\phi_{nT}^{-2}\right)$. Also, $\max_{1\leq t\leq T} III_t\leq\sqrt{n}$ $\max_{1\leq t\leq T}\left\|\sqrt{n}b_{1t}\right\|\max_{1\leq t\leq T}\left\|b_{2t}\right\|=\sqrt{n}\ o_p\left(T^{1/k_2}\delta_{nT}^{-2}\sqrt{\ln T}\right)$. Putting all together, we have

$$\begin{aligned} \max_{1\leq t\leq T} n\left(\hat{f}_t-\widehat{\overline{f}}\right)'\hat{\Sigma}_{ft}^{-1}\left(\hat{f}_t-\widehat{\overline{f}}\right) &= \max_{1\leq t\leq T} N_t'\Sigma_{\Gamma\epsilon,t}^{-1}N_t+o_p\left[\sqrt{\frac{\ln T}{n}}\left(nT\right)^{1/k_2}\right] \\ &+o_p\left(\frac{T^{2/k_2}}{\sqrt{n}}\right)+o_p\left(T^{1/k_2}\frac{\sqrt{n\ln T}}{T}\right)+o_p\left(1\right); \end{aligned} \quad (15)$$

under (23), the error term is negligible. Consider the sequence $\left\{N_t\right\}_{t=1}^T$. The covariance between N_t and

N_{t-k} is proportional to, for $(n,T)\to\infty$

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left(\gamma_i\gamma_j'\epsilon_{it}\epsilon_{jt-k}\right) \\
\leq\; & \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|E\left(\gamma_i\gamma_j'\epsilon_{it}\epsilon_{jt-k}\right)\right\| \leq \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|\gamma_i\gamma_j'\right\|\left|E\left(\epsilon_{it}\epsilon_{jt-k}\right)\right| \\
\leq\; & M\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|E\left(\epsilon_{it}\epsilon_{jt-k}\right)\right|,
\end{aligned}$$

so that, under Assumption 9, $\lim_{k,n\to\infty}E\left(N_tN_{t-k}\right)\ln k=0$. By virtue of such Berman condition, equation (24)holds - see e.g. Theorem 3.5.1 in Leadbetter and Rootzen (1988, p.470).

We now complete the proof of the Theorem by studying the power versus local alternatives. Under H_1^b, it can be shown that $S_{f,nT}$ has non-centrality parameter given by

$$\begin{aligned}
S_{f,nT}^{NC} &= n\max_{1\leq t\leq T}c_t'\hat{\Sigma}_{ft}^{-1}c_t+2n\max_{1\leq t\leq T}c_t'\hat{\Sigma}_{ft}^{-1}\left(\hat{f}_t-f_t\right)-2n\max_{1\leq t\leq T}c_t'\hat{\Sigma}_{ft}^{-1}\left(\hat{\overline{f}}-f\right) \\
&= I+II+III,
\end{aligned}$$

with $I=O_p\left(n\left\|c_t\right\|^2\right)$ by construction. Also, II is bounded by $n\left(\max_{1\leq t\leq T}\left\|c_t\right\|\right)\max_{1\leq t\leq T}\left\|\left(\hat{f}_t-f_t\right)\right\|=O_p\left(n\left\|c_t\right\|T^{2/k_2}\delta_{nT}^{-2}\right)$ by Lemma A.6*(ii)*; similarly, $III=O_p\left(n\delta_{nT}^{-2}\left\|c_t\right\|\right)$. Let $S_{nT}^{f,0}$ denote the null distribution of $S_{f,nT}$. Then, under H_1^b we have

$$P\left[S_{f,nT}>c_{\alpha,T}\right]=P\left[S_{nT}^{f,0}>c_{\alpha,T}-S_{f,nT}^{NC}\right];$$

$P\left[S_{f,nT}>c_{\alpha,T}\right]$ tends to 1 if $c_{\alpha,T}-S_{f,nT}^{NC}\to-\infty$ as $(n,T)\to\infty$; this holds because, by (26), $c_{\alpha,T}=O\left(\ln T\right)$. QED

2 Further simulations

In addition to conducting simulations under the DGP

$$y_{it}=\beta_ix_{it}+\gamma_if_t+\epsilon_{it}, \tag{16}$$

$$x_{it}=\mu_i+\lambda_if_t+\epsilon_{it}^x, \tag{17}$$

we also consider two alternative DGPs that are nested in (16), in order to assess the robustness of the tests proposed to different specifications.

We firstly consider a DGP for the regressors x_{it} that modifies (17) by not containing common factors, viz.

$$x_{it} = \mu_i + \epsilon_{it}^x. \tag{18}$$

In this case, cross dependence in the y_{it}s is purely due to the presence of f_t in (16). The rank condition in Assumption 3*(ii)* does not hold, although the CCE estimator is still consistent. Secondly, we consider a DGP in which there are no unit specific regressors, viz.

$$y_{it} = \gamma_i f_t + \epsilon_{it}; \tag{19}$$

this is a pure factor model, that fits in the class of models considered by Bai (2003). In this case, it can be argued that testing for no factor structure (either by using $S_{\gamma,nT}$ or $S_{f,nT}$) complements the information criteria in Bai and Ng (2002), by being a test for $r = 0$. This is can also be compared with the framework in Baltagi, Kao, and Na (2012).

Testing for $H_0^a : \gamma_i = \gamma$

When evaluating the empirical rejection frequencies for tests based on $S_{\gamma,nT}$, we run the Monte Carlo simulations under the null $\gamma_i = 1$ for all i. When evaluating power, we generate the loadings γ_i as *i.i.d.* $N\left(1, \sigma_\gamma^2\right)$, reporting results for the case of $\sigma_\gamma = 0.2$. Given that ϵ_{it} is cross sectionally uncorrelated and homoskedastic by design, $\Sigma_{\gamma i}$ is estimated as $\hat{\Sigma}_{\gamma i} = \hat{\sigma}_\epsilon^2 \times T\left(\hat{F}' M_{x_i} \hat{F}\right)^{-1}$, where $\hat{\sigma}_\epsilon^2 = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \hat{\epsilon}_{it}^2$.

When considering the two alternative specifications, the first given by (16) and (18) whereas the second is simply the one in (19), the results are reported in Tables 1 and 2, respectively.

As far as empirical rejection frequencies and power are concerned, results for both specifications do not change much with respect to the ones obtained under (16)-(17) and presented in the paper (Table 2 in the paper). Indeed, the size improves in both cases. For all signal-to-noise cases considered, the test attains its nominal size for all values of n, as long as $T \geq 100$.

It is interesting to note that both size and power become much better under (19) than in the other cases, see Table 2. The correct size is attained as long as $n \geq 30$ and $T \geq 50$; moreover, the power is always above 90% for all combinations of n and T considered.

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 $S_{f,nT}$. When evaluating the power, we generate the common factors f_t as *i.i.d.* $N\left(1, \sigma_f^2\right)$,

	Size				Power			
	T				T			
n	30	50	100	200	30	50	100	200
	$\sigma_\epsilon^2 = 1/3$				$\sigma_\epsilon^2 = 1/3$			
30	0.067	0.053	0.046	0.043	0.999	1.000	1.000	1.000
50	0.067	0.055	0.048	0.047	1.000	1.000	1.000	1.000
100	0.066	0.064	0.052	0.040	1.000	1.000	1.000	1.000
200	0.069	0.063	0.050	0.041	1.000	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1/2$				$\sigma_\epsilon^2 = 1/2$			
30	0.069	0.055	0.051	0.046	0.979	0.998	1.000	1.000
50	0.069	0.057	0.049	0.049	0.996	1.000	1.000	1.000
100	0.067	0.064	0.054	0.041	0.999	1.000	1.000	1.000
200	0.069	0.064	0.05	0.041	1.000	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1$				$\sigma_\epsilon^2 = 1$			
30	0.081	0.066	0.060	0.052	0.921	0.989	1.000	1.000
50	0.077	0.063	0.054	0.057	0.965	0.999	1.000	1.000
100	0.072	0.068	0.057	0.044	0.992	1.000	1.000	1.000
200	0.073	0.065	0.052	0.043	0.998	1.000	1.000	1.000

Table 1: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^a : \gamma_i = \gamma$, based on $S_{\gamma,nT}$. The DGP used in the simulations is (16)- (18), i.e. the case of no common factor structure in the regressors.

	Size				Power			
	T				T			
n	30	50	100	200	30	50	100	200
	$\sigma_\epsilon^2 = 1/3$				$\sigma_\epsilon^2 = 1/3$			
30	0.058	0.050	0.044	0.045	0.980	0.999	1.000	1.000
50	0.059	0.049	0.046	0.044	0.998	1.000	1.000	1.000
100	0.062	0.046	0.048	0.040	0.999	1.000	1.000	1.000
200	0.071	0.050	0.046	0.046	1.000	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1/2$				$\sigma_\epsilon^2 = 1/2$			
30	0.061	0.052	0.047	0.047	0.904	0.976	1.000	1.000
50	0.062	0.051	0.049	0.047	0.978	1.000	1.000	1.000
100	0.064	0.048	0.049	0.041	0.986	1.000	1.000	1.000
200	0.073	0.051	0.046	0.047	0.998	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1$				$\sigma_\epsilon^2 = 1$			
30	0.070	0.064	0.056	0.055	0.611	0.739	0.991	1.000
50	0.070	0.057	0.055	0.052	0.778	0.966	1.000	1.000
100	0.067	0.050	0.053	0.044	0.791	0.980	1.000	1.000
200	0.074	0.054	0.047	0.048	0.938	0.999	1.000	1.000

Table 2: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^a : \gamma_i = \gamma$, based on $S_{\gamma,nT}$. The DGP used in the simulations is (19), i.e. the case of a pure factor model for y_{it}.

reporting results for the case of $\sigma_f = 0.2$. Finally, we estimate Σ_{ft} as $\Sigma_{ft} = V_{nT}^{-1}\hat{\sigma}_\epsilon^2 \frac{1}{n}\sum_{i=1}^n \hat{\lambda}_i\hat{\lambda}_i' V_{nT}^{-1}$ where $\hat{\sigma}_\epsilon^2 = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \hat{\epsilon}_{it}^2$.

When considering the two alternative specifications, the first given by (16) and (18) whereas the second is simply the one in (19), the results are reported in Tables 3 and 4, respectively.

	Size				Power			
	T				T			
n	30	50	100	200	30	50	100	200
	$\sigma_\epsilon^2 = 1/3$				$\sigma_\epsilon^2 = 1/3$			
30	0.044	0.039	0.045	0.039	0.954	0.989	0.998	0.987
50	0.044	0.042	0.038	0.036	0.998	1.000	1.000	0.999
100	0.042	0.038	0.040	0.038	1.000	1.000	1.000	1.000
200	0.043	0.047	0.041	0.036	1.000	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1/2$				$\sigma_\epsilon^2 = 1/2$			
30	0.045	0.041	0.046	0.041	0.863	0.933	0.979	0.995
50	0.045	0.043	0.039	0.037	0.978	0.996	1.000	1.000
100	0.044	0.039	0.040	0.038	0.999	1.000	1.000	1.000
200	0.048	0.048	0.043	0.037	1.000	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1$				$\sigma_\epsilon^2 = 1$			
30	0.052	0.049	0.050	0.043	0.561	0.646	0.780	0.859
50	0.047	0.049	0.042	0.039	0.809	0.894	0.960	0.991
100	0.052	0.042	0.043	0.040	0.978	0.997	1.000	1.000
200	0.058	0.052	0.044	0.039	1.000	1.000	1.000	1.000

Table 3: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^b : f_t = f$, based on $S_{f,nT}$. The DGP used in the simulations is (16)-(18), i.e. the case of no common factor structure in the regressors x_{it}.

Results do not differ much, when carrying out simulations under (16) and (18), from the values of Table 3 in the paper. Actually, as it was noted for the case of $S_{\gamma,nT}$, results improve slightly, in particular the power. Similar considerations hold for the empirical rejection frequencies computed under (19) (Table 4): the size is close to the correct one. The power is also very good, under all possible combinations of parameters.

	Size				Power			
	T				T			
n	30	50	100	200	30	50	100	200
	$\sigma_\epsilon^2 = 1/3$				$\sigma_\epsilon^2 = 1/3$			
30	0.048	0.054	0.053	0.056	0.973	0.994	0.998	1.000
50	0.042	0.041	0.046	0.049	0.999	1.000	1.000	1.000
100	0.039	0.043	0.044	0.041	1.000	1.000	1.000	1.000
200	0.043	0.040	0.039	0.040	1.000	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1/2$				$\sigma_\epsilon^2 = 1/2$			
30	0.049	0.056	0.054	0.057	0.904	0.955	0.988	0.997
50	0.046	0.042	0.047	0.050	0.989	0.997	1.000	1.000
100	0.042	0.044	0.046	0.041	1.000	1.000	1.000	1.000
200	0.045	0.041	0.040	0.041	1.000	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1$				$\sigma_\epsilon^2 = 1$			
30	0.060	0.064	0.057	0.058	0.646	0.729	0.830	0.905
50	0.053	0.047	0.050	0.051	0.869	0.934	0.977	0.995
100	0.049	0.052	0.049	0.043	0.991	0.999	1.000	1.000
200	0.052	0.046	0.044	0.043	1.000	1.000	1.000	1.000

Table 4: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^b : f_t = f$, based on $S_{f,nT}$. The DGP used in the simulations is (19), i.e. the case of a pure factor model for y_{it}.

References

Bai, J., 2003, Inferential theory for structural models of large dimensions. Econometrica, vol. 71, 135-171.

Bai, J., 2009a, Panel data models with interactive fixed effects. Econometrica, vol. 77, 1229-1279.

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

Baltagi, B., Kao, C., Na, S., 2012, Testing cross-sectional dependence in panel factor model using the wild bootstrap F-test, manuscript.

Berkes, I., Liu, W., Wu, W.B., 2013, Komlos-Major-Tusnady approximation under dependence. Annals of Probability (forthcoming).

Castagnetti, C., Rossi, E., Trapani L., 2014, Inference on Factor Structures in Heterogeneous Panels, Mimeo.

Csörgö, M., Révész, P., 1975a, A new method to prove Strassen-type laws of invariance principle. I. Probability Theory and Related Fields, vol. 31, 255-260.

Csörgö, M., Révész, P., 1975b, A new method to prove Strassen-type laws of invariance principle. II. Probability Theory and Related Fields, vol. 31, 261-269.

Davidson J., 1994, Stochastic Limit Theory. Oxford University Press, Oxford.

Kao, C., Trapani, L., Urga, G., 2012, Testing for Instability in Covariance Structures. Center for Policy Research Working Papers No. 131.

Komlós, J., Major, P., Tusnády, G., 1975, An approximation of partial sums of independent rv's and the sample df. I. Probability Theory and Related Fields, vol. 32, 111-131.

Komlós, J., Major, P., Tusnády, G., 1976, An approximation of partial sums of independent rv's and the sample df. II. Probability Theory and Related Fields, vol. 34, 33-58.

Leadbetter, M.R., Rootzen, H., 1988, Extremal theory for stochastic processes. Annals of Probability, vol. 16, 431-478.

Ling, S., 2007. Testing for change points in time series models and limiting theorems for NED sequences. Annals of Statistics, vol. 35, 1213-1237.

Peligrad, M., Utev, S., Wu, W.B., 2007. A maximal L_p-inequality for stationary sequences and application. Proceedings of the American Statistical Association, vol. 135, 541-550.

Pesaran, M. H., 2006, Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica, vol. 74, 967-1012.

Shorack, G.R., Wellner, J.A., 1986, Empirical processes with applications to statistics. Wiley, New York.

Song, M., 2013. Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. Mimeo, January 2013.

Strang, G., 1988, Linear algebra and its applications. Third Edition, Harcourt, Orlando.