



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
(Cass Business School, City University London)

2 (09-12)

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

September 2012

Inference on Factor Structures in Heterogeneous Panels*

Carolina Castagnetti

Eduardo Rossi

University of Pavia

University of Pavia

Lorenzo Trapani

Cass Business School, City University London

First draft: November 10, 2010

This version: September 28, 2012

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 applying Pesaran's (2006) CCE 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.

*This is a revised version of a paper previously circulated under the working title "Two-Stage Inference in Heterogeneous Panels". We 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 Horvath for providing us with valuable comments. The usual disclaimer applies.

1 Introduction

Consider the following model for stationary panel data:

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

$$x_{it} = \Lambda_i f_t + \epsilon_{it}^x, \quad (2)$$

where $i = 1, \dots, n$, $t = 1, \dots, T$, 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). Arguably, (1)-(2) is a model with a huge potential for empirical applications. For example, Kapetanios and Pesaran (2007) consider an APT model allowing for individual asset returns to be affected by common factors (both observable and unobservable); Eberhardt and Teal (2012) adopt a common factor model approach to estimate cross-country production functions for the agriculture sector; Eberhardt, Helmers and Strauss (2012) consider the impact of spillovers in the estimation of private returns to R&D within a common factor framework; Castagnetti and Rossi (2012) adopt a heterogeneous panel with a multifactor error model to study the determinants of credit spread changes in the Euro corporate bond market.

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_t s through cross sectional averages of the x_{it} and y_{it} ; such estimation techniques are referred to as the Common Correlated Effects (CCE) estimators. One of the key features of the CCE estimator is that it does not require any inference to be carried out on γ_i or f_t . Pesaran and Tosetti (2011) and Castagnetti and Rossi (2012) 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,b) 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_i s; moreover, no explicit asymptotics for γ_i or f_t is derived beyond consistency.

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,b) to the case of heterogeneous β_i s.

As a second contribution, 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 - in such case, 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. In this respect, our paper is closely 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 $y_{it} = \gamma_i' f_t + \epsilon_{it}$. The tests developed in this paper should therefore be employed before trying to estimate any factor structure, including the number of common factors, as we also discuss in Section 3.

From a methodological point of view, we use 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). From a technical point of view, in our proofs we use similar arguments to the changepoint literature (see e.g. Csorgo and Horvath, 1997): we approximate the sequences of estimated parameters with sequences of normals, and apply Extreme Value

Theory (EVT henceforth). In this respect, our paper is a first attempt to systematize the use of extrema of estimated parameters in the context of a panel regression with unobservable common factors. As far as small sample properties are concerned, we show through a Monte Carlo exercise 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.

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 Euclidean norm of a vector x is denoted as $\|x\|$; similarly, the induced Euclidean norm of a matrix A is denoted as $\|A\| = \max_{x \neq 0} \|Ax\| / \|x\|$. 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 $F = (f_1, \dots, f_T)'$; $X_i = (x_{i1}, \dots, x_{iT})'$; $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iT})'$; $y_i = (y_{i1}, \dots, y_{iT})'$; $z_{it} = (y_{it}, x'_{it})'$; $z_i = (z_{i1}, \dots, z_{iT})'$ and $\bar{H}_w = n^{-1} \sum_{i=1}^n z_i$. We also define the matrices $\bar{M}_w = I_T - \bar{H}_w (\bar{H}'_w \bar{H}_w)^{-1} \bar{H}'_w$ 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_i s in (1) can be estimated as

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

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_i s 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

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

with $M_{X_i} = I_T - X_i(X_i'X_i)^{-1}X_i'$. In (1), γ_i and f_t are not separately identifiable; as it 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 .

Consider the following assumptions.

Assumption 1. ϵ_{it} is i.i.d. across i and t with $E(\epsilon_{it}) = 0$ and $E|\epsilon_{it}|^{12} < \infty$.

Assumption 2. (i) ϵ_{it}^x is i.i.d. across i ; ϵ_{it}^x and f_t are i.i.d. across t with $E\|\epsilon_{it}^x\|^{12} < \infty$ and $E\|f_t\|^{12} < \infty$; (ii) $\{\epsilon_{it}^x, f_t\}$ and $\{\epsilon_{js}\}$ are mutually independent for all i, j, t, s ; (iii) $T^{-1}\sum_{t=1}^T f_t f_t' \rightarrow \Sigma_f$ as $T \rightarrow \infty$ with Σ_f non-singular.

Assumption 3. (i) $l_{\min}\left(\frac{X_i'\bar{M}_w X_i}{T}\right) > 0$; $l_{\min}\left(\frac{X_i' M_F X_i}{T}\right) > 0$ and $l_{\min}\left(\frac{F' M_{X_i} F}{T}\right) > 0$ a.s. for all i , where $l_{\min}(\cdot)$ denotes the smallest eigenvalue; (ii) $C \equiv n^{-1}\sum_{i=1}^n C_i$ has rank $r \leq m + 1$.

Assumption 4. (i) $\{\beta_i\}$ is independent of $\{\epsilon_{jt}, \epsilon_{jt}^x, f_t\}$ for all i, j, t ; (ii) $E\|\beta_i\|^{2+\delta} < \infty$ for some $\delta > 0$; (iii) the γ_i s are non stochastic and such that $\max_i \|\gamma_i\| < \infty$ and $n^{-1}\sum_{i=1}^n \gamma_i\gamma_i' \rightarrow \Sigma_\gamma$ as $n \rightarrow \infty$ with Σ_γ non-singular.

Assumption 1 deals with the error term ϵ_{it} , and, for example, it entails that Assumption C in Bai (2009a) holds. Assuming that the ϵ_{its} are cross sectionally independent is stronger than necessary - this is e.g. not needed in Pesaran (2006) or in Bai (2009a). In our context, we make this assumption only for the sake of simplicity; alternatively, it could be relaxed by assuming that the cross sectional covariances satisfy some summability conditions (see e.g. Assumption C in Bai, 2009a). Throughout the paper, we derive results under this assumption, and discuss,

for each result, to what extent the assumption can be relaxed. Similarly, as far as serial dependence is concerned, more general forms of dependence can be considered for ϵ_{it} ; for example, using the theory developed in Phillips and Solo (1992) (see Phillips and Moon, 1999, for the extension to the panel context), we could assume a linear process for ϵ_{it} , at the price of a more complicated algebra. Finally, note the requirement on the existence of the 12-th moment of ϵ_{it} . This assumption is stronger than what the literature normally considers - e.g. in Bai (2009a), assuming $E|\epsilon_{it}|^8 < \infty$ 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).

As far as Assumption 2 is concerned, the requirement that the ϵ_{it}^x s are cross sectionally independent, and that both ϵ_{it}^x and f_t are i.i.d. over time, are again stronger than necessary and only made for simplicity. As far as Assumption 3 is concerned, the rank condition in part (ii) is the same as equation (21) in Pesaran (2006), and it guarantees the consistency of the CCE estimator $\tilde{\beta}_i$, defined in (3). It is well known (see e.g. Remark 2 in Pesaran and Tosetti, 2011) that the CCE estimator can be consistent even if the rank condition is not met. In our context, some results in the proofs (see e.g. Lemma A.1) require this rank condition also. Finally, Assumption 4 is standard; the requirement that the γ_i s are non random in part (iii) can be relaxed in a similar fashion to Assumption 3 in Pesaran (2006).

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

Theorem 1 *Under Assumptions 1-4, it holds that, for every i*

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

As $(n, T) \rightarrow \infty$ with $\frac{\sqrt{T}}{n} \rightarrow 0$

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

where $\Sigma_{\gamma_i} = \Sigma_{fM,i}\Sigma_{fMe,i}\Sigma_{fM,i}$ and $\Sigma_{fM,i}$ and $\Sigma_{fMe,i}$ are the probability limits of $T^{-1}(F'H'M_{X_i}HF)$ and $T^{-1}(F'H'M_{X_i}\epsilon_i\epsilon_i'M_{X_i}HF)$, respectively.

Theorem 1 can be compared with Theorem 2 in Bai (2003, p.147): the rates of convergence

in (5) are exactly the same. As is typical in this literature, consistency is affected by the use of generated regressors, \hat{f}_t , when computing $\hat{\gamma}_i$ - see equation (4). This is the reason why the term $O_p\left(\frac{1}{n}\right)$ appears in (5): it is not possible to estimate the γ_i s consistently unless both n and T pass to infinity.

On the other hand, the limiting distribution of $\sqrt{T}(\hat{\gamma}_i - H^{-1}\gamma_i)$ 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} .

In the statement of the Theorem, the expression of the limiting covariance matrix is reported for the general case of serial dependence. Under independence (Assumptions 1 and 2), an estimator of Σ_{γ_i} is (see also Bai, 2003, p.150)

$$\hat{\Sigma}_{\gamma_i} = (Q'_i)^{-1} D_{0,i} (Q_i)^{-1},$$

where $Q_i = T^{-1}(\hat{F}'M_{X_i}\hat{F})$ and $D_{0,i} = \frac{1}{T} \sum_{t=j+1}^T \widehat{f}'_{x_t} \widehat{f}_{x_t} \hat{\epsilon}_{it}^2$, with \widehat{f}_{x_t} the t -th row of $M_{X_i}\hat{F}$. Under more general forms of serial dependence, Σ_{γ_i} can be estimated as $\hat{\Sigma}_{\gamma_i} = (Q'_i)^{-1} \Phi_i (Q_i)^{-1}$, where Φ_i is the HAC estimator of Newey and West (1987), viz.

$$\Phi_i = D_{0,i} + \sum_{j=1}^q \left(1 - \frac{j}{q+1}\right) (D_{j,i} + D'_{j,i}). \quad (7)$$

In (7), $D_{j,i}$ is defined as $D_{j,i} = \frac{1}{T} \sum_{t=j+1}^T \widehat{f}'_{x_t} \widehat{f}_{x_{t-j}} \hat{\epsilon}_{it} \hat{\epsilon}_{it-j}$, and the bandwidth parameter q is chosen so that $q \rightarrow \infty$ with $q/T^{1/4} \rightarrow 0$. Other HAC estimators can also be employed (see e.g. Andrews, 1991).

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

Theorem 2 *Under Assumptions 1-4, it holds that, for every t*

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

As $(n, T) \rightarrow \infty$ with $\frac{n}{T} \rightarrow 0$

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

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

Theorem 2 is the counterpart to Theorem 1 in Bai (2003, p.145). In particular, the asymptotic distribution in equation (9) is exactly the same. However, the rates of convergence given in (8) are different. In Theorem 1 in Bai (2003), $H'f_t$ is estimated at a rate $O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{T}\right)$. In our case, the fact that the x_{its} are correlated with f_t introduces a term of magnitude $O_p(T^{-1/2})$ in the estimation error $\hat{f}_t - H'f_t$. As a consequence, the restriction $\frac{n}{T} \rightarrow 0$ is needed in (9).

As in the case of Theorem 1, the asymptotic covariance matrix of $\sqrt{n}\left(\hat{f}_t - H'f_t\right)$ is expressed in its general form. We note that, since the limiting distribution is the same as in Bai (2003), 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 $\tilde{v} = (\tilde{v}_1, \dots, \tilde{v}_n)'$, and defining V_{nT} as a diagonal matrix containing the r largest eigenvalues of $\frac{1}{nT}\tilde{v}\tilde{v}'$ in descending order, the estimated Σ_{ft} is

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

Note that $\Sigma_{\Gamma\epsilon,t}$ is estimated through $n^{-1} \sum_{i=1}^n \hat{\gamma}_i \hat{\gamma}_i' \hat{\epsilon}_{it}^2$; this is consistent with Assumption 1, which postulates cross sectional independence. However, even if this assumption were relaxed, estimating $\Sigma_{\Gamma\epsilon,t}$ would still require some restrictions on the presence of cross dependence; this is because it is in general not possible 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 *Under Assumptions 1-4, as $(n, T) \rightarrow \infty$, for all i and t*

$$\hat{c}_{it} - c_{it} = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right). \quad (10)$$

As $(n, T) \rightarrow \infty$ with $\frac{n}{T} \rightarrow 0$

$$\sqrt{n}(\hat{c}_{it} - c_{it}) \xrightarrow{d} N(0, \gamma_i' \Sigma_{ft} \gamma_i), \quad (11)$$

where Σ_{f_t} is defined in Theorem 2.

The rate of convergence is the same as in Theorem 3 in Bai (2003). However, the limiting distribution is different under $\frac{T}{n} \rightarrow 0$, since now the estimation error $\hat{f}_t - H'f_t$ would also come up in the expression.

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 develop tests for the null of no factor structure in (1). Motivated by Sarafidis, Yamagata and Robertson (2009), we propose two 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_t s.

Formally, we propose two tests for the null hypotheses:

$$H_0^a : \gamma_i = \gamma \text{ for all } i; \quad (12)$$

$$H_0^b : f_t = f \text{ for all } t. \quad (13)$$

As mentioned in the Introduction, we argue that both (12) and (13) entail that there is no real factor structure in (1). Consider (12) first. When H_0^a holds, equation (1) can be rewritten as

$$y_{it} = \varphi_t + \beta_i' x_{it} + \epsilon_{it}, \quad (14)$$

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 (13), equation (1) can be rewritten as

$$y_{it} = \varphi_i + \beta_i' x_{it} + \epsilon_{it}, \quad (15)$$

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 (12) and (13) 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 (13).

Let $\widehat{\gamma} = n^{-1} \sum_{i=1}^n \widehat{\gamma}_i$ and $\widehat{f} = T^{-1} \sum_{t=1}^T \widehat{f}_t$. In order to test for (12) and (13), we propose the following max-type test statistics:

$$S_{\gamma, nT} \equiv \max_{1 \leq i \leq n} \left[T (\widehat{\gamma}_i - \widehat{\gamma})' \widehat{\Sigma}_{\gamma_i}^{-1} (\widehat{\gamma}_i - \widehat{\gamma}) \right], \quad (16)$$

$$S_{f, nT} \equiv \max_{1 \leq t \leq T} \left[n (\widehat{f}_t - \widehat{f})' \widehat{\Sigma}_{f_t}^{-1} (\widehat{f}_t - \widehat{f}) \right]. \quad (17)$$

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). 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 . Indeed, based on the same reasoning, this also entails that tests can be carried out with any, arbitrarily chosen value of r ; the asymptotics will be the same, although it can be expected that finite sample properties will differ across different values of r . From a methodological perspective, this entails that tests based on (12) and (13) 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 (12) and (13) 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 (14) or (15) 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 the information criteria in Bai and Ng (2002). The asymptotic

properties of the estimated common factors, loadings and common components are those given in Section 2.

It is worth pointing out that an alternative approach could be based on using averages of F -statistics, viz.

$$\tilde{S}_{\gamma,nT} = \sqrt{\frac{n}{2r}} \frac{1}{n} \sum_{i=1}^n \left[T (\hat{\gamma}_i - \hat{\gamma})' \hat{\Sigma}_{\gamma i}^{-1} (\hat{\gamma}_i - \hat{\gamma}) - r \right], \quad (18)$$

$$\tilde{S}_{f,nT} = \sqrt{\frac{T}{2r}} \frac{1}{T} \sum_{t=1}^T \left[n (\hat{f}_t - \hat{f})' \hat{\Sigma}_{ft}^{-1} (\hat{f}_t - \hat{f}) - r \right]; \quad (19)$$

for example, Pesaran and Yamagata (2008) propose average-type statistics for the null of slope homogeneity, in the case of a model with observable regressors. In our context, we show that average-type statistics such as $\tilde{S}_{\gamma,nT}$ and $\tilde{S}_{f,nT}$ cannot be employed. The full blown details are reported in Appendix B (Theorems B.1 and B.2), where we show that $\tilde{S}_{\gamma,nT}$ and $\tilde{S}_{f,nT}$ diverge under the null as $(n, T) \rightarrow \infty$. Thus, tests based on (18) and (19) always reject the null of no factor structure. In both cases, this is due to having the generated regressors \hat{f}_t . We also point out that Baltagi, Kao and Na (2012) propose a test for $H_0 : \gamma_i = 0$ for all i . In this case, test statistics have an asymptotic bias that makes the size equal to 1 asymptotically. Baltagi, Kao and Na (2012) propose a bias correction, and the use of wild bootstrap, to overcome this issue.

We now present results on tests based on $S_{\gamma,nT}$ (Theorem 3 in Section 3.1) and on $S_{f,nT}$ (Theorem 4 in Section 3.2). Each section contains a discussion of the main results. The technical details (including discussions as to up to which extent assumptions can be relaxed) are in the proofs of Theorems 3 and 4 in Appendix B. A heuristic preview of the main arguments used in the proofs of both theorems is as follows. Referring to (16) as a benchmark example, we approximate the sequence of the estimation errors $\sqrt{T} (\hat{\gamma}_i - H^{-1} \gamma_i)$ with a sequence of normally distributed random variables, plus an error term whose supremum taken over n is negligible. In light of this, the proofs are similar, in spirit, to the ones found in the changepoint literature (see e.g. Csorgo and Horvath, 1997).

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

In this subsection, we study 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 $(n, T) \rightarrow \infty$ 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 $E|\epsilon_{it}|^{k_1}$, $E\|x_{it}\|^{k_1}$ and $E\|f_t\|^{k_1}$ are finite. In view of Assumption 1, $k_1 \geq 12$.

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

Theorem 3 *Let Assumptions 1-4 hold, and let $(n, T) \rightarrow \infty$ with $\frac{\sqrt{T}n^{2/k_1}}{n} \rightarrow 0$ and $\frac{n^{4/k_1}}{T} \rightarrow 0$. Under H_0^a , it holds that*

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

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

$$\frac{T}{\ln n} \|c_i\|^2 \rightarrow \infty, \quad (21)$$

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

Theorem 3 states that $S_{\gamma,nT}$ has a Gumbel distribution. This holds in the joint limit $(n, T) \rightarrow \infty$, under some restrictions on the relative rate of divergence of n and T . Since $k_1 \geq 12$, we require $\frac{T}{n^{11/6}} \rightarrow 0$, which is marginally stricter than the condition $\frac{\sqrt{T}}{n} \rightarrow 0$ needed in for (6). Similarly, the restriction that $\frac{n^{4/k_1}}{T} \rightarrow 0$ becomes, under Assumptions 1 and 2, $\frac{n}{T^3} \rightarrow 0$. The two restrictions indicate that the test should be applied when n is not exceedingly larger than T , and vice versa.

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

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

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 $O(1/\ln^2 n)$. 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, Kluppelberg 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.

Finally, from a technical point of view, the assumption that the ϵ_{it} s are cross sectionally and serially independent can be relaxed. In the proof of Theorem 3, we show that (20) still holds if one assumes $|E(\epsilon_{it}\epsilon_{js})| < \sigma_{ij}$ with $\sigma_{ij} = o_p(1)$ as $n \rightarrow \infty$ for all $i \neq j$.

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

In this subsection, we study the asymptotics of $S_{f,nT}$ under H_0^b , and its consistency. Similarly to the previous subsection, we show that, as $(n, T) \rightarrow \infty$ 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 $O_p\left(\sqrt{\frac{\ln T}{n}}\right)$.

In order to present the results, we let k_2 be the largest number such that $E\|f_t\|^{k_2}$, $E\|x_{it}\|^{k_2}$ and $E|\epsilon_{it}|^{k_2}$ are all finite. In view of Assumptions 1 and 2, $k_2 \geq 12$.

Let the critical value $c_{\alpha,T}$ be defined such that $P(S_{f,nT} \leq c_{\alpha,T}) = 1 - \alpha$ under H_0^b . It holds that:

Theorem 4 *Let Assumptions 1-4 hold, and let $(n, T) \rightarrow \infty$ with $\frac{nT^{2/k_2}}{T} \rightarrow 0$ and $\frac{n^{2/k_2}T^{2/k_2}}{n} \rightarrow 0$. Under H_0^b , it holds that*

$$P[A_T S_{f,nT} \leq x + B_T] = e^{-e^{-x}}, \quad (23)$$

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

$$\frac{n}{\ln T} \|c_t\|^2 \rightarrow \infty, \quad (24)$$

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

Theorem 4 is very similar to Theorem 3; convergence to the Gumbel distribution under the null is shown for $(n, T) \rightarrow \infty$ jointly under some restrictions between n and T . Namely, we require $\frac{nT^{2/k_2}}{T} \rightarrow 0$ and $\frac{n^{2/k_2}T^{2/k_2}}{n} \rightarrow 0$. Since $k_2 \geq 12$, the former restriction is, at most, $\frac{n}{T^{5/6}} \rightarrow 0$. This is marginally stronger than $\frac{n}{T} \rightarrow 0$, which is required for (9) to hold. Similarly, requiring that $\frac{n^{2/k_2}T^{2/k_2}}{n} \rightarrow 0$ entails $\frac{T}{n^5} \rightarrow 0$. This means that T ought to be larger than n , but not massively larger (the restriction $\frac{T}{n^5} \rightarrow 0$ can be expected to always hold in practice).

Critical values for a test of level α can be calculated as

$$c_{\alpha,T} = 2B_T - \ln |\ln(1 - \alpha)|^2; \quad (25)$$

alternatively, B_T can be approximated by $F_{\chi_r}^{-1}(1 - 1/T)$. As far as power is concerned, (24) 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 .

Results are derived under the assumptions of no serial correlation in the ϵ_{it} s. In the proof, we show that this requirement can be relaxed: the Theorem still holds as long as $\lim_{s \rightarrow \infty} \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \sum_{j=1}^n \left\| \gamma_i \gamma_j' \right\| |E(\epsilon_{it} \epsilon_{jt-s})| \ln s = 0$.

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 (12) and (13) 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):

$$y_{it} = \beta_i x_{it} + \gamma_i f_t + \epsilon_{it}, \quad (26)$$

$$x_{it} = \mu_i + \lambda_i f_t + \epsilon_{it}^x, \quad (27)$$

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 1,000 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 $J = 1,000$).

[Insert Table 1 somewhere here]

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 I in Bai (2003, p.151). When n and T are ≥ 100 , the estimated factors can be treated as the true ones.

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 $(n, T) \rightarrow \infty$ with $\frac{\sqrt{T}}{n} \rightarrow 0$, 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 $\Gamma = (\gamma_1, \dots, \gamma_n)'$ and $\hat{\Gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_n)'$. 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.

[Insert Figure 1 somewhere here]

Figure 1 shows that, in most cases and for all combinations of n and T , the confidence intervals 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 (16) and (17) 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 $\frac{Var(c_{it})}{\sigma_\epsilon^2}$, where σ_ϵ^2 is the variance of ϵ_{it} , equal to $\left\{ \frac{1}{3}, \frac{1}{2}, 1 \right\}$.

In addition to conducting simulations under the DGP (26), we also consider two alternative DGPs that are nested in (26), in order to assess the robustness of the tests proposed to different specifications of (1)-(2). We firstly consider a DGP for the regressors x_{it} that modifies (27) by not containing common factors, viz.

$$x_{it} = \mu_i + \epsilon_{it}^x. \quad (28)$$

In this case, cross dependence in the y_{it} s is purely due to the presence of f_t in (26). The rank condition in Assumption 3(ii) does not hold, although the CCE estimator is still consistent. Secondly, we consider a DGP for (1) in which there are no unit specific regressors, viz.

$$y_{it} = \gamma_i f_t + \epsilon_{it}; \quad (29)$$

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).

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

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

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. $N(1, \sigma_\gamma^2)$, reporting results for the case of $\sigma_\gamma = 0.2$. Given that ϵ_{it} is cross sectionally uncorrelated and homoskedastic by design, Σ_{γ_i} is estimated as $\hat{\Sigma}_{\gamma_i} = \hat{\sigma}_\epsilon^2 \times T^{-1} (\hat{F}' M_{x_i} \hat{F})$, where $\hat{\sigma}_\epsilon^2 = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \hat{\epsilon}_{it}^2$.

Results for size and power when using the main DGP (26)-(27) are in Table 2.

[Insert Table 2 somewhere here]

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 $n \geq 50$; indeed, when $\sigma_\epsilon^2 = 1$ (high signal-to-noise ratio), the test has satisfactory size properties even for $T = 50$. The Table also shows that, as the signal-to-noise ratio decreases (i.e., as σ_ϵ^2 increases), the tendency towards small sample oversize 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_ϵ^2 .

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.

¹The simulation results are available upon request.

When considering the two alternative specifications (26)-(28) and (29), results are reported in Tables 3 and 4.

[Insert Tables 3 and 4 somewhere here]

Results do not change much with respect to the ones in Table 2, as far as both empirical rejection frequencies and power are concerned. Indeed, the size improves in both cases (especially when simulations are conducted under (29)). When the signal-to-noise ratio is sufficiently high, 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 (29) than in the other cases. 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 .

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. $N(1, \sigma_f^2)$, 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 (26)-(27) are in Table 5.

[Insert Table 5 somewhere here]

The size of the test is almost always the correct one, with few exceptions - the test is oversized for small T when σ_ϵ^2 is high. 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 .

As in the previous subsection, we also considered size and power under the alternative specifications (26)-(28) and (29); results are in Tables 6 and 7.

[Insert Tables 6 and 7 somewhere here]

Results do not differ much, when carrying out simulations under (26)-(28), from the values in Table 5. 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 (29): the size is always the correct one. The power is also very good, under all possible combinations of parameters.

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). The main finding is that results differ only marginally from the case of a stand-alone panel factor model as the one studied by Bai (2003).

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.

Appendix A: Technical Lemmas

In this Appendix and the next one, we set $H = I_r$ in the proofs (although not in the statements of the Lemmas), for the sake of notational simplicity. Inequalities are written, when possible, omitting constants.

The Lemmas in this Section extend various results in Bai (2009a,b) to our framework. All proofs rely upon the decomposition - see Proposition A.1 in Bai (2009a):

$$\begin{aligned}
 \hat{F} - F &= \frac{1}{nT} \sum_{j=1}^n X_j (\tilde{\beta}_j - \beta_j) (\tilde{\beta}_j - \beta_j)' X_j' \hat{F} \\
 &\quad - \frac{1}{nT} \sum_{j=1}^n X_j (\tilde{\beta}_j - \beta_j) \gamma_j' F' \hat{F} - \frac{1}{nT} \sum_{j=1}^n X_j (\tilde{\beta}_j - \beta_j) \epsilon_j' \hat{F} \\
 &\quad - \frac{1}{nT} \sum_{j=1}^n F \gamma_j (\tilde{\beta}_j - \beta_j)' X_j' \hat{F} - \frac{1}{nT} \sum_{j=1}^n \epsilon_j (\tilde{\beta}_j - \beta_j)' X_j' \hat{F} \\
 &\quad + \frac{1}{nT} \sum_{j=1}^n F \gamma_j \epsilon_j' \hat{F} + \frac{1}{nT} \sum_{j=1}^n \epsilon_j \gamma_j' F' \hat{F} + \frac{1}{nT} \sum_{j=1}^n \epsilon_j \epsilon_j' \hat{F}.
 \end{aligned} \tag{30}$$

In (30), 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 Appendices A and B. We define $\Upsilon_i \equiv (X_i' \bar{M}_w X_i)^{-1} (X_i' \bar{M}_w \epsilon_i)$, so that we can write

$$\begin{aligned}
 \tilde{\beta}_i - \beta_i &= \left(\frac{X_i' \bar{M}_w X_i}{T} \right)^{-1} \left(\frac{X_i' \bar{M}_w \epsilon_i}{T} \right) + \left(\frac{X_i' \bar{M}_w X_i}{T} \right)^{-1} \left(\frac{X_i' \bar{M}_w F}{T} \gamma_i \right) \\
 &= \Upsilon_i + \tilde{\Upsilon}_i,
 \end{aligned} \tag{31}$$

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

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

Proof. 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

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

where the last equality holds by symmetry. In view of Assumption 3(i), and omitting γ_i , this entails that

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

Consider I ; we have $I \leq T^{-r} E \|X_i' \epsilon_i\|^r = T^{-r} E \left\| \sum_{t=1}^T x_{it} \epsilon_{it} \right\|^r$. Using the fact that, by Assumption 1, $x_{it} \epsilon_{it}$ is an MDS,

$$\begin{aligned} T^{-r} E \|X_i' \epsilon_i\|^r &\leq T^{-r} E \left| \sum_{t=1}^T \|x_{it} \epsilon_{it}\|^2 \right|^{r/2} \leq T^{-r} E \left| T^{1-2/r} \left(\sum_{t=1}^T \|x_{it} \epsilon_{it}\|^r \right)^{2/r} \right|^{r/2} \\ &\leq T^{-r} T^{r/2} \left(\frac{1}{T} \sum_{t=1}^T E \|x_{it} \epsilon_{it}\|^r \right) \leq T^{-r/2} \left(\frac{1}{T} \sum_{t=1}^T [E \|x_{it}\|^{2r}]^{1/2} [E |\epsilon_{it}|^{2r}]^{1/2} \right) \\ &= O(T^{-r/2}), \end{aligned} \tag{32}$$

where we have used: Burkholder's inequality; 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, $(\bar{H}'_w \bar{H}_w)^{-1} = D_w^{-1} - D_w^{-1} R_w D_w^{-1}$, with $D_w = C' F' F C$ 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

$$\begin{aligned} \frac{X_i' \bar{M}_w F}{T} &= -\frac{X_i' F C D_w^{-1} \bar{\epsilon}' F}{T^2} - \frac{F' \bar{\epsilon} D_w^{-1} C' F' X_i}{T^2} \\ &\quad - \frac{X_i' \bar{\epsilon} D_w^{-1} \bar{\epsilon}' F}{T^2} - \frac{X_i' F C D_w^{-1} R_w D_w^{-1} C' F' F}{T^2} \\ &= -I - I' - II - III. \end{aligned} \tag{33}$$

Consider $E \|I\|^r$; since C has full rank by Assumption 3(i) and D_w is invertible

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

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

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

after similar passages as in equation (32). It holds that $E \|f_t\|^{4r} < \infty$ by Assumption 2(i). Further, using the fact that the ϵ_{it} s are cross sectionally independent, we have $E \left| n^{-1} \sum_{i=1}^n \epsilon_{it} \right|^{4r} \leq n^{-2r} \left(n^{-1} \sum_{i=1}^n E |\epsilon_{it}|^{4r} \right)$ by similar passages as in equation (32). Putting all together, this yields $E \|I\|^r = O(n^{-r/2} T^{-r/2})$. The same logic yields $E \|II\|^r = O(n^{-r} T^{-r})$. Finally, consider *III*

$$\begin{aligned} E \|III\|^r &\leq \|R_w\|^r \|C\|^{2r} \|D_w^{-1}\|^{2r} E \left\| \frac{X'_i F}{T} \frac{F' F}{T} \right\|^r \\ &\leq \|R_w\|^r \left[E \left\| \frac{X'_i F}{T} \right\|^{2r} \right]^{1/2} \left[E \left\| \frac{F' F}{T} \right\|^{2r} \right]^{1/2} = O(\|R_w\|^r), \end{aligned}$$

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

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

$$\mathbf{A.2(i)} \quad T^{-1} \epsilon'_i (\hat{F} - F) = O_p(\delta_{nT}^{-2});$$

$$\mathbf{A.2(ii)} \quad n^{-1/2} T^{-1} \sum_{i=1}^n \epsilon'_i (\hat{F} - F) = O_p(n^{-1/2}) + O_p(T^{-1}).$$

Proof. 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 (30)

$$\begin{aligned}
& n^{-1/2}T^{-1} \sum_{i=1}^n \epsilon'_i (\hat{F} - F) \tag{34} \\
= & \frac{1}{n\sqrt{T}} \sum_{j=1}^n \frac{\left(\sum_{i=1}^n \epsilon_i\right)' X_j}{\sqrt{T}} (\tilde{\beta}_j - \beta_j) (\tilde{\beta}_j - \beta_j)' \frac{X'_j \hat{F}}{T} - \frac{1}{n\sqrt{T}} \sum_{j=1}^n \frac{\left(\sum_{i=1}^n \epsilon_i\right)' X_j}{\sqrt{T}} (\tilde{\beta}_j - \beta_j) \gamma'_j \frac{F' \hat{F}}{T} \\
& - \frac{1}{nT} \sum_{j=1}^n \frac{\left(\sum_{i=1}^n \epsilon_i\right)' X_j}{\sqrt{T}} (\tilde{\beta}_j - \beta_j) \frac{\epsilon'_j \hat{F}}{\sqrt{T}} - \frac{1}{n\sqrt{T}} \frac{\left(\sum_{i=1}^n \epsilon_i\right)' F}{\sqrt{T}} \sum_{j=1}^n \gamma_j (\tilde{\beta}_j - \beta_j)' \frac{X'_j \hat{F}}{T} \\
& - \frac{1}{\sqrt{n}} \frac{1}{nT} \sum_{j=1}^n \sum_{i=1}^n \epsilon'_i \epsilon_j (\tilde{\beta}_j - \beta_j)' \frac{X'_j \hat{F}}{T} + \frac{1}{\sqrt{nT}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon'_i F \frac{1}{\sqrt{T}} \sum_{j=1}^n \gamma_j \frac{\epsilon'_j \hat{F}}{T} \\
& + \frac{1}{\sqrt{nT}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon'_i \frac{1}{\sqrt{n}} \sum_{j=1}^n \epsilon_j \gamma'_j \frac{F' \hat{F}}{T} + \frac{1}{T} \frac{1}{n} \sum_{j=1}^n \left(\sum_{i=1}^n \epsilon_i\right)' \epsilon_j \frac{\epsilon'_j \hat{F}}{T} \\
= & I + II + III + IV + V + VI + VII + VIII.
\end{aligned}$$

The proof follows very similar lines to that of Lemma A.4(ii) in Bai (2009b): the only difference here is the presence of the unit-specific estimation errors, $\tilde{\beta}_j - \beta_j$. 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 $I \leq n^{-1} \sum_{j=1}^n \left\| \frac{\sum_{i=1}^n \epsilon'_i X_j}{\sqrt{nT}} \right\| \left\| \frac{X'_j \hat{F}}{T} \right\| \left\| \tilde{\beta}_j - \beta_j \right\|^2$. This is bounded by

$$\begin{aligned}
& E \left[\left\| \frac{\sum_{i=1}^n \epsilon'_i X_j}{\sqrt{nT}} \right\| \left\| \frac{X'_j \hat{F}}{T} \right\| \left\| \tilde{\beta}_j - \beta_j \right\|^2 \right] \tag{35} \\
& \leq \left[E \left(\left\| \tilde{\beta}_j - \beta_j \right\|^{2p} \right) \right]^{1/p} \left[E \left(\left\| \frac{\sum_{i=1}^n \epsilon'_i X_j}{\sqrt{nT}} \right\| \left\| \frac{X'_j \hat{F}}{T} \right\| \right)^q \right]^{1/q} \\
& \leq \left[E \left\| \tilde{\beta}_j - \beta_j \right\|^3 \right]^{2/3} \left[E \left\| \frac{\sum_{i=1}^n \epsilon'_i X_j}{\sqrt{nT}} \right\|^6 \right]^{1/6} \left[E \left\| \frac{X'_j \hat{F}}{T} \right\|^6 \right]^{1/6},
\end{aligned}$$

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 $O_p(\phi_{nT}^{-2})$ 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 $I = O_p(T^{-1/2} \phi_{nT}^{-2})$. Similar passages yield $II = O_p(T^{-1/2} \phi_{nT}^{-1})$; $III =$

$O_p(T^{-1}\phi_{nT}^{-1})$; $IV = O_p(T^{-1/2}\phi_{nT}^{-1})$; $VI = O_p(T^{-1/2}\phi_{nT}^{-1}) + O_p(\delta_{nT}^{-2})$; $VII = O_p(n^{-1/2})$ and $VIII = O_p(T^{-1}) + O_p(n^{-1/2}T^{-1/2})$.

Consider now V , whose proof is marginally different to that in Bai (2009b, p.5)

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

using the Cauchy-Schwartz inequality (first line), Holder's inequality with the same orders as in (35) (second line) and Lemma A.1. Also

$$\begin{aligned}
E \left| \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T \epsilon_{it} \epsilon_{jt} \right|^2 &= \frac{1}{nT^2} \sum_{i=1}^n \sum_{k=1}^n \sum_{t=1}^T \sum_{s=1}^T E(\epsilon_{it} \epsilon_{ks} \epsilon_{jt} \epsilon_{js}) \\
&= \begin{cases} \frac{1}{nT^2} \sum_{j \neq k} \sum_{t=1}^T E(\epsilon_{jt}^2) E(\epsilon_{kt}^2) = O\left(\frac{1}{T}\right) & \text{for } i = j = k \\ \frac{1}{nT^2} \sum_{t=1}^T \sum_{s=1}^T E(\epsilon_{it}^2 \epsilon_{is}^2) = O\left(\frac{1}{n}\right) & \text{for } i = j \neq k \end{cases},
\end{aligned}$$

and zero otherwise by Assumption 1, whence $V = O_p(\delta_{nT}^{-1}\phi_{nT}^{-1})$. This part of the proof does not require independence across the ϵ_{it} s, and it could be assumed directly. Putting all together, part A.2(ii) follows. QED

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

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

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

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

Proof. The Lemma is a refinement of Lemma A.3 in Bai (2009a). Particularly, by Lemma A.3 in Bai (2009a) we have $T^{-1} X_i' (\hat{F} - F) = o_p(1)$ and $T^{-1} F' (\hat{F} - F) = o_p(1)$.

Consider part (i). Using (30)

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

henceforth, we omit γ_i in the passages, based on Assumption 4(ii). 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 (35), this is bounded by $\left[E \left\| \frac{X'_i X_j}{T} \right\|^6 \right]^{1/6} \left[E \left\| \frac{X'_j \hat{F}}{T} \right\|^6 \right]^{1/6} \left[E \left\| \tilde{\beta}_j - \beta_j \right\|^3 \right]^{2/3}$

$= O_p(\phi_{nT}^{-2})$. Turning to II , we have $II = \frac{1}{n} \sum_{j=1}^n \frac{X'_i X_j}{T} (\tilde{\beta}_j - \beta_j) \gamma'_j \frac{F' \hat{F}}{T} + o_p(1)$, where the $o_p(1)$ term comes from $T^{-1} F' (\hat{F} - F) = o_p(1)$. By (31), this is bounded by $\left\| \frac{X_i}{\sqrt{T}} \right\| \left\| \frac{1}{n\sqrt{T}} \sum_{j=1}^n X_j \Upsilon_j \right\| + \frac{1}{n} \sum_{j=1}^n \left\| \frac{X'_i X_j}{T} \right\| \left\| \tilde{\Upsilon}_j \right\| = II_a + II_b$. Consider II_a ; since

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

where we have used Assumptions 1, 2(ii) and 3(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. Thus, $II_a = O_p(n^{-1/2} T^{-1/2})$. Turning to II_b , this is of the same order of magnitude as $E \left[\left\| \frac{X'_i X_j}{T} \right\| \left\| \tilde{\Upsilon}_j \right\| \right] \leq \left[E \left(\left\| \frac{X'_i X_j}{T} \right\|^2 \right) \right]^{1/2} \left[E \left(\left\| \tilde{\Upsilon}_j \right\|^2 \right) \right]^{1/2}$; by the proof of Lemma A.1, $II_b = O_p(n^{-1/2} \delta_{nT}^{-1})$. Thus, $II = O_p(n^{-1/2} \delta_{nT}^{-1})$. Turning to III , it can be decomposed into $\frac{1}{n\sqrt{T}} \sum_{j=1}^n \frac{X'_i X_j}{T} (\tilde{\beta}_j - \beta_j) \frac{\epsilon'_j F}{\sqrt{T}} + \frac{1}{n} \sum_{j=1}^n \frac{X'_i X_j}{T} (\tilde{\beta}_j - \beta_j)$

$\frac{\epsilon'_j(\hat{F}-F)}{T} = III_a + III_b$; 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 $III_a = O_p(T^{-1/2}\phi_{nT}^{-1})$. Similar passages and Lemma A.2(i), yield $III_b = O_p(\phi_{nT}^{-1}\delta_{nT}^{-2})$. Term IV has the same magnitude of term II . As far as V is concerned, using the fact that $T^{-1}X'_i(\hat{F}-F) = o_p(1)$, 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 $V = O_p(T^{-1/2}\phi_{nT}^{-1})$. Turning to VI , we have $VI = \frac{1}{n\sqrt{T}} \sum_{j=1}^n \frac{X'_j F}{T} \frac{\epsilon'_j F}{\sqrt{T}} + \frac{X'_j F}{T} \frac{1}{n} \sum_{j=1}^n \frac{\epsilon'_j(\hat{F}-F)}{T} = VI_a + VI_b$. Considering VI_a , similar passages as above give $VI_a = O_p(n^{-1/2}T^{-1/2})$. Turning to VI_b , this is $O_p(\delta_{nT}^{-2})$ by Lemma A.2(ii). Therefore, $VI = O_p(\delta_{nT}^{-2})$. As far as VII is concerned, it is bounded by $\left\| \frac{F'F}{T} \right\| \left\| \frac{1}{nT} \sum_{j=1}^n X'_i \epsilon_j \right\| + o_p(1)$. By using Assumption 1, $\sum_{j=1}^n X'_i \epsilon_j$ is bounded by the square root of its variance, viz. $\sum_{j=1}^n \sum_{t=1}^T E(x_{it}x'_{it}) E(\epsilon_{jt}^2) = O_p(nT)$, so that $VII = O_p(n^{-1/2}T^{-1/2})$. Finally, $VIII = \frac{1}{nT} \sum_{j=1}^n \frac{X'_i \epsilon_j}{\sqrt{T}} \frac{\epsilon'_j F}{\sqrt{T}} + \frac{1}{n\sqrt{T}} \sum_{j=1}^n \frac{X'_i \epsilon_j}{\sqrt{T}} \frac{\epsilon'_j(\hat{F}-F)}{T} = VIII_a + VIII_b$; $VIII_a$ 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 $VIII_a = O_p(T^{-1})$. Similarly, $VIII_b$ is bounded by $T^{-1/2} \left[E \left(\left\| \frac{X'_i \epsilon_j}{\sqrt{T}} \right\|^2 \right) \right]^{1/2} \left[E \left(\left\| \frac{\epsilon'_j(\hat{F}-F)}{T} \right\|^2 \right) \right]^{1/2} = O_p(T^{-1/2}\delta_{nT}^{-2})$, 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 (30), obtaining $T^{-1/2} \left\| \hat{F} - F \right\| = O_p \left(\left\| \tilde{\beta}_j - \beta_j \right\| \right) + O_p(\delta_{nT}^{-1})$, whence $T^{-1} \left\| \hat{F} - F \right\|^2 = O_p(\delta_{nT}^{-2})$. QED

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

Proof. By definition, under H_0^a

$$\sqrt{T}(\hat{\gamma}_i - \gamma) = \left(\frac{\hat{F}' M_{X_i} \hat{F}}{T} \right)^{-1} \left[\frac{\hat{F}' M_{X_i} \epsilon_i}{\sqrt{T}} - \frac{\hat{F}' M_{X_i} (\hat{F} - F) \gamma}{\sqrt{T}} \right]; \quad (36)$$

also, under H_0^a , it holds that $\hat{\gamma} - \gamma = \frac{1}{n} \sum_{i=1}^n (\hat{\gamma}_i - \gamma)$. Using (36) and neglecting higher order

terms coming from $\hat{F}'M_{X_i}\hat{F} - F'M_{X_i}F$

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

Consider I ; using cross sectional independence in Assumption 1 and Assumption 3(i), I is bounded by the square root of

$$\frac{1}{n^2} \sum_{i=1}^n E \left\| \left(\frac{F'M_{X_i}F}{T} \right)^{-1} \left(\frac{F'M_{X_i}\epsilon_i}{T} \right) \right\|^2 \leq \frac{1}{nT} E \left\| \frac{F'\epsilon_i}{\sqrt{T}} \right\|^2;$$

we know that $E \left\| \frac{F'\epsilon_i}{\sqrt{T}} \right\|^2 = O(1)$; thus, $I = O_p\left(\frac{1}{\sqrt{nT}}\right)$. Consider II ; it is bounded by

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

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

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

Proof. Consider (30) and let $\bar{F} = f \times i_T$; under H_0^b , it holds that

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

Consider I ; it is bounded by $E \left[\left\| \frac{\hat{F}' X_j}{T} \right\| \left\| \tilde{\beta}_j - \beta_j \right\|^2 \left\| \frac{1}{T} \sum_{t=1}^T x_{jt} \right\| \right] \leq E \left[\left\| \tilde{\beta}_j - \beta_j \right\|^3 \right]^{2/3} E \left[\left\| \frac{\hat{F}' X_j}{T} \right\|^6 \right]^{1/6}$
 $E \left[\left\| \frac{1}{T} \sum_{t=1}^T x_{jt} \right\|^6 \right]^{1/6} = O(\phi_{nT}^{-2})$, using a similar logic to (35) and Lemma A.1. Similar arguments yield $II = O_p(n^{-1/2}T^{-1/2}) + O_p(n^{-1})$, $III = O_p(\phi_{nT}^{-2})$, $IV = O_p(n^{-1/2}\delta_{nT}^{-1})$, and $VI = O_p(n^{-1/2}T^{-1/2})$. Consider V ; this is bounded by

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

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

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*

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

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

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

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

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

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

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

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

$$\max_{1 \leq h \leq m} |Z_h| = o_p \left(m^{1/a} \right). \quad (37)$$

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

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

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 $\max_{1 \leq i \leq n} \left\| T^{-1} X_i' \bar{M}_w F \right\|^2$. 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 $E \|x_{it} \epsilon_{it}\|^a \leq E \|x_{it}\|^a |\epsilon_{it}|^a = E \|x_{it}\|^a E |\epsilon_{it}|^a$ by independence. 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 $\max_{1 \leq i \leq n} \left\| T^{-1} X_i' \bar{M}_w F \right\|^2$ is concerned, we know from the proof of Lemma A.1 that $\left\| T^{-1} X_i' \bar{M}_w F \right\|^2$ 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 (33): $\frac{1}{\sqrt{nT}} \max_{1 \leq i \leq n} \left\| \frac{X_i' F C D_w^{-1} \sqrt{n} \bar{\epsilon}' F}{T \sqrt{T}} \right\|^2 \leq \frac{1}{\sqrt{nT}} \left(\max_{1 \leq i \leq n} \left\| \frac{X_i}{\sqrt{T}} \right\|^2 \right) \left\| \frac{F}{\sqrt{T}} \right\|^2 \left\| \frac{\bar{\epsilon}' F}{\sqrt{nT}} \right\|^2$. We have $\left\| T^{-1/2} X_i \right\|^2 = T^{-1} \sum_{t=1}^T x_{it}^2$, so that, based on (37), $\max_{1 \leq i \leq n} \left\| T^{-1/2} X_i \right\|^2 = o_p \left(n^{2/k} \right)$. 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 (33) yields $\max_{1 \leq i \leq n} \|T^{-1} X_i' \bar{M}_w F\|^2 = o_p(n^{2/k-1/2} \delta_{nT}^{-1})$. Part (i) follows putting everything together.

Consider part (ii). The passages of the proof are rather repetitive. The main argument is that, based on (40), $\max_{1 \leq t \leq T} \delta_{nt}^{-2} \|\hat{f}_t - H' f_t\|^2$ is bounded by terms such as $\delta_{nt}^{-2} \left\| \frac{1}{n} \sum_{j=1}^n \left(\frac{\hat{F}' X_j}{T} \right) (\tilde{\beta}_j - \beta_j) (\tilde{\beta}_j - \beta_j)' \right\|^2 \max_{1 \leq t \leq T} \|x_{jt}\|^2$, etc. This entails that, when taking the maximum across t , the order of magnitude of the maximum is given by terms like $\max_{1 \leq t \leq T} \|x_{jt}\|^2$, $\max_{1 \leq t \leq T} \|f_t\|^2$ and $\max_{1 \leq t \leq T} \|\epsilon_{jt}\|^2$, which are of order $o_p(T^{2/k})$. This provides part (ii).

The proof of part (iii) is based on (39):

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

By Assumption 4(iii), $\max_{1 \leq i \leq n} \|T^{-1/2} \hat{F}' (\hat{F} - F)\|^2 \|\gamma_i\|^2$ has the same order of magnitude as $\|T^{-1/2} \hat{F}' (\hat{F} - F)\|^2$, i.e. $O_p(\sqrt{T}n^{-1}) + O_p(T^{-1/2})$. As far as $\max_{1 \leq i \leq n} \|T^{-1/2} \hat{F}' \epsilon_i\|^2$ is concerned

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

Consider *I*; based on the same arguments as in (32), we have $\max_{1 \leq i \leq n} \|T^{-1/2} F' \epsilon_i\|^2 \leq \max_{1 \leq i \leq n} T^{-1} \sum_{t=1}^T \|f_t \epsilon_{it}\|^2$. Also, $E \|f_t \epsilon_{it}\|^a \leq E \|f_t\|^a E |\epsilon_{it}|^a < \infty$ with the largest a being $a = 2k$, whence $\max_{1 \leq i \leq n} \|T^{-1/2} F' \epsilon_i\|^2 = o_p(n^{2/k})$. Turning to *II*, $\max_{1 \leq i \leq n} \left\| T^{-1} (\hat{F} - F)' \epsilon_i \right\|^2$ can be studied using (34). It follows that $\max_{1 \leq i \leq n} \left\| T^{-1} (\hat{F} - F)' \epsilon_i \right\|^2$ is bounded by the sum of terms

like

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

which follows from (35) (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 $E \|\epsilon_{it} x_{jt}\|^a \leq E \|x_{jt}\|^a E |\epsilon_{it}|^a$ with $a = k$ at most. We have that $\max_{1 \leq i \leq n} \left\| T^{-1/2} \epsilon'_i X_j \right\|^2 = o_p(n^{2/k})$. Applying the same logic to the squares of all the terms in (34), it follows that $II = o_p(n^{2/k} T \delta_{nT}^{-4})$. Part (iii) follows from putting everything together.

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

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

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

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

and $\max_{1 \leq i \leq n} \epsilon_{it}^2 = o_p(n^{2/k})$ by (37).

Lemma A.7 *Let Assumptions 1-4 hold, and let $\{N_i\}_{i=1}^n$ and $\{N_t\}_{t=1}^T$ be two sequences of independently distributed Gaussian random variables, with variances $\Sigma_{fMe,i}$ and $\Sigma_{\Gamma\epsilon,t}$ respectively. Let also k denote the largest finite moment of ϵ_{it} , f_t and x_{it} . It holds that*

$$\mathbf{A.7 (i)} \quad \left\| T^{-1} \hat{F}' \hat{F} - H \Sigma_f H' \right\| = O_p(T^{-1/2}) + O_p(n^{-1});$$

$$\mathbf{A.7 (ii)} \quad \max_{1 \leq i \leq n} \left\| T^{-1} \hat{F}' M_{X_i} \hat{F} - \Sigma_{fM,i} \right\| = O_p(T^{-1/2}) + O_p(n^{-1});$$

$$\mathbf{A.7 (iii)} \quad \max_{1 \leq t \leq T} \left\| \hat{\Sigma}_{\Gamma\epsilon, t} - H^{-1} \Sigma_{\Gamma\epsilon, t} (H^{-1})' \right\| = o_p \left(T^{2/k} \delta_{nT}^{-1} \right);$$

$$\mathbf{A.7 (iv)} \quad \max_{1 \leq i \leq n} \left\| \hat{\Sigma}_{\gamma, i} - \Sigma_{\gamma, i} \right\| = o_p \left(\sqrt{T} n^{2/k} \delta_{nT}^{-2} \right);$$

$$\mathbf{A.7 (v)} \quad \max_{1 \leq i \leq n} \left\| T^{-1/2} F' M_{X_i} \epsilon_i - N_i \right\| = o_p \left(n^{1/k} T^{1/k-1/2} \right);$$

$$\mathbf{A.7 (vi)} \quad \max_{1 \leq t \leq T} \left\| n^{-1/2} \sum_{i=1}^n \hat{\gamma}_i \epsilon_{it} - N_t \right\| = o_p \left(T^{1/k} n^{1/k-1/2} \right) + o_p \left(T^{1/k-1/2} n^{1/2} \right).$$

Proof. 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 $T^{-1} \hat{F}' \hat{F} - T^{-1} F' F = O_p \left(\delta_{nT}^{-2} \right)$. Also, note that the sequence $f_t f_t' - \Sigma_f$ is i.i.d. with mean zero. We have

$$E \left\| \frac{F' F}{T} - \Sigma_f \right\|^2 = E \left\| \frac{1}{T} \sum_{t=1}^T (f_t f_t' - \Sigma_f) \right\|^2 = O \left(\frac{1}{T} \right),$$

using the same passages as in the proof of Lemma A.2. Putting all together, part (i) follows. As far as part (ii) is concerned, it follows immediately from noting that $\left\| T^{-1} \hat{F}' M_{X_i} \hat{F} - \Sigma_{fM, i} \right\| \leq \left\| T^{-1} \hat{F}' \hat{F} - \Sigma_f \right\|$ for each i , by definition of M_{X_i} .

As far as showing parts (iii)-(vi) is concerned, we extensively use the fact that, for an i.i.d. sequence of random variables Z_1, \dots, Z_m such that $E(Z_1) = 0$ and $E|Z_1|^k \leq M$ for some $k > 2$, and given a Brownian motion $W(\cdot)$ with $E[W^2(1)] = E(Z_1^2)$, it holds that

$$\left\| \frac{1}{\sqrt{m}} \sum_{h=1}^m Z_h - W(1) \right\| = O_p \left(m^{1/k-1/2} \right). \quad (38)$$

Results like (38) are known as ‘‘Hungarian constructions’’; see, *inter alia*, Csorgo and Revezs (1975a,b), and Komlos, Major and Tusnady (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 $m^{-1/2} \sum_{h=1}^{\lfloor m\delta \rfloor} Z_h$ for $\delta \in [0, 1]$; in that case, (38) is stated using the sup-norm. For our purposes, we only need to consider $\delta = 1$. Note that a result like (38) does not require independence across h . Other techniques could also be employed to obtain this approximation. For example, one could use Berry-Esseen bounds (see e.g. Embrechts, Kluppelberg and Mikosch, 1997, for reference). Based on our understanding of the literature on Berry-Esseen bounds, these

are quite sensitive to the assumptions made; thus, we found the Hungarian construction an easier tool to work with.

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

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

Consider I . Note that the sequence $z_{\epsilon\gamma, it} = \gamma_i \gamma_i' \epsilon_{it}^2 - \Sigma_{\Gamma\epsilon, t}$ has mean zero and it is independent across i . Thus, $\sqrt{n} \left\| \frac{1}{n} \sum_{i=1}^n z_{\epsilon\gamma, it} \right\| = O_p(1)$. Also, by using Assumption 4(iii), the largest k for which $E \left\| \frac{1}{n} \sum_{i=1}^n z_{\epsilon\gamma, it} \right\|^k < \infty$ is $k/2$. Thus, $\max_{1 \leq t \leq T} \sqrt{n} \left\| \frac{1}{n} \sum_{i=1}^n z_{\epsilon\gamma, it} \right\| = o_p(T^{2/k})$, which entails $I = o_p(T^{2/k} n^{-1/2})$. As far as II is concerned, it has the same order of magnitude as $\max_{1 \leq t \leq T} |\hat{\epsilon}_{it}^2 - \epsilon_{it}^2|$, given in Lemma A.6(v). Turning to III , its order of magnitude is given by $O_p(\|\hat{\gamma}_i - \gamma_i\|^2) \max_{1 \leq t \leq T} \hat{\epsilon}_{it}^2$, 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 $\hat{\Sigma}_{\gamma_i} = (Q_i')^{-1} D_{0,i} (Q_i)^{-1}$; the rates for $Q_i = T^{-1} \hat{F}' M_{X_i} \hat{F}$ are given by part (ii) of this Lemma. Based on the definition of $D_{0,i}$, we have

$$\begin{aligned}
&\max_{1 \leq i \leq n} \left\| \frac{1}{T} \sum_{t=1}^T \hat{f}_t \hat{f}_t' \hat{\epsilon}_{it}^2 - H' \Sigma_f H E(\epsilon_{it}^2) \right\| \\
&\leq \max_{1 \leq i \leq n} \left\| \frac{1}{T} \sum_{t=1}^T f_t f_t' \epsilon_{it}^2 - H' \Sigma_f H E(\epsilon_{it}^2) \right\| + \max_{1 \leq i \leq n} \left\| \frac{1}{T} \sum_{t=1}^T \hat{f}_t \hat{f}_t' (\hat{\epsilon}_{it}^2 - \epsilon_{it}^2) \right\| \\
&\quad + 2 \max_{1 \leq i \leq n} \left\| \frac{1}{T} \sum_{t=1}^T f_t (\hat{f}_t - f_t)' \epsilon_{it}^2 \right\| + \max_{1 \leq i \leq n} \left\| \frac{1}{T} \sum_{t=1}^T (\hat{f}_t - f_t) (\hat{f}_t - f_t)' \hat{\epsilon}_{it}^2 \right\| \\
&= I + II + III + IV.
\end{aligned}$$

As far as I is concerned, the proof is similar to that of part (i) of this Lemma, upon noting that the largest a for which $E \|f_t f_t' \epsilon_{it}^2\|^a \leq E \|f_t\|^{2a} E |\epsilon_{it}|^{2a}$ is $a = k/2$, and that $T^{-1} \sum_{t=1}^T f_t f_t' \epsilon_{it}^2 -$

$H'\Sigma_f H E(\epsilon_{it}^2) = O_p(T^{-1/2})$; hence, $I = o_p(n^{2/k}T^{-1/2})$. Considering II , we have

$$\begin{aligned} II &\leq \frac{1}{T} \sum_{t=1}^T \left\| \hat{f}_t \hat{f}'_t \right\| \max_{1 \leq i \leq n} |\hat{\epsilon}_{it}^2 - \epsilon_{it}^2| \\ &\leq \left[\frac{1}{T} \sum_{t=1}^T \left\| \hat{f}_t \hat{f}'_t \right\|^2 \right]^{1/2} \left[E \left(\max_{1 \leq i \leq n} |\hat{\epsilon}_{it}^2 - \epsilon_{it}^2| \right)^2 \right]^{1/2}; \end{aligned}$$

applying Lemma A.6(vi), we have $II = o_p(n^{4/k}T^{-1}) + o_p(n^{2/k}T^{-1}) + O_p(\delta_{nT}^{-2})$. Considering III , a similar logic as above yields

$$\begin{aligned} III &\leq \frac{1}{T} \sum_{t=1}^T \left\| f_t (\hat{f}_t - f_t)' \right\| \max_{1 \leq i \leq n} \hat{\epsilon}_{it}^2 \\ &\leq \left[\frac{1}{T} \sum_{t=1}^T \left\| f_t (\hat{f}_t - f_t)' \right\|^2 \right]^{1/2} \left[E \left(\max_{1 \leq i \leq n} \hat{\epsilon}_{it}^2 \right)^2 \right]^{1/2}; \end{aligned}$$

also,

$$\left[\frac{1}{T} \sum_{t=1}^T \left\| f_t (\hat{f}_t - f_t)' \right\|^2 \right]^{1/2} \leq \sqrt{T} \frac{1}{T} \sum_{t=1}^T \left\| f_t (\hat{f}_t - f_t)' \right\| = O_p(\sqrt{T} \delta_{nT}^{-2}),$$

by the C_r -inequality and Lemma A.3(ii). Using Lemma A.6(vi), we have $III = o_p(n^{2/k}T^{1/2}\delta_{nT}^{-2})$.

Term IV is dominated. Putting all together, part (iv) follows.

Consider now part (v). We provide the proof setting $M_{X_i} = I_T$ for simplicity. For an i.i.d., zero mean, Gaussian sequence $\{N_{it}^T\}_{t=1}^T$ with $E(N_{it}^T)^2 = E(f_t f_t') E(\epsilon_{it}^2)$, we can write, in view of (38)

$$T^{1/2-1/k} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T f_t \epsilon_{it} - \frac{1}{\sqrt{T}} \sum_{t=1}^T N_{it}^T \right\| = O_p(1),$$

where the largest k for which this can be written corresponds to the largest a for which $E \|f_t \epsilon_{it}\|^a < \infty$. It holds that $E \|f_t \epsilon_{it}\|^a < E \|f_t\|^a E |\epsilon_{it}|^a$, whence $a = 2k$. This also holds across i , i.e. for all i it holds that $E \left[T^{1/2-1/k} \left\| T^{-1/2} \sum_{t=1}^T f_t \epsilon_{it} - T^{-1/2} \sum_{t=1}^T N_{it}^T \right\| \right]^k < \infty$. Thus, $\max_{1 \leq i \leq n} T^{1/2-1/k} \left\| T^{-1/2} \sum_{t=1}^T f_t \epsilon_{it} - T^{-1/2} \sum_{t=1}^T N_{it}^T \right\| = o_p(n^{1/k})$. Defining $T^{-1/2} \sum_{t=1}^T N_{it}^T = N_i$, this 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 $E(N_{it}^n)^2 = \gamma_i \gamma_i' E(\epsilon_{it}^2)$, write

$$\begin{aligned} & \max_{1 \leq t \leq T} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\gamma}_i \epsilon_{it} - \frac{1}{\sqrt{n}} \sum_{i=1}^n N_{it}^2 \right\| \\ & \leq \max_{1 \leq t \leq T} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_i \epsilon_{it} - \frac{1}{\sqrt{n}} \sum_{i=1}^n N_{it}^2 \right\| + \max_{1 \leq t \leq T} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{\gamma}_i - \gamma_i) \epsilon_{it} \right\| \\ & = I + II. \end{aligned}$$

Term I satisfies (38) with

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

also, the largest existing moment over t is of order k . Thus, $\max_{1 \leq t \leq T} n^{1/2-1/k} \left\| n^{-1/2} \sum_{i=1}^n \gamma_i \epsilon_{it} - n^{-1/2} \sum_{i=1}^n N_{it}^2 \right\| = o_p(T^{1/k})$, whence $I = o_p(T^{1/k} n^{1/k-1/2})$. Also, $II \leq n^{-1/2} \sum_{i=1}^n \|\hat{\gamma}_i - \gamma_i\| \max_{1 \leq t \leq T} |\epsilon_{it}|$, which by Theorem 1 yields $II = o_p(T^{1/k-1/2} n^{1/2}) + o_p(T^{1/k} n^{-1/2})$. Putting everything together, and defining $n^{-1/2} \sum_{i=1}^n N_{it}^2 = N_t$, part (vi) follows. We point out that, in parts (v) and (vi), independence over, respectively, t and i is not required; thus, Assumptions 1 and 2 could be relaxed to accommodate for serial and cross sectional dependence, as long as the Hungarian construction (38) holds. QED

Appendix B: Proofs

Proof of Theorem 1. By definition, we have

$$\sqrt{T}(\hat{\gamma}_i - \gamma_i) = \left(\frac{\hat{F}' M_{X_i} \hat{F}}{T} \right)^{-1} \left[\frac{\hat{F}' M_{X_i} \epsilon_i}{\sqrt{T}} - \frac{\hat{F}' M_{X_i} (\hat{F} - F) \gamma_i}{\sqrt{T}} \right]. \quad (39)$$

We start by considering the denominator of (39):

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

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

We turn to the numerator of (39). It holds that

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

By Assumptions 1 and 2, it holds that $I = O_p(1)$. As far as II is concerned, note

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

applying Lemma A.1(i) (to the first term), and Lemma A.3(i) and Assumptions 1 and 2 (to the second term), it follows that $II = O_p(\sqrt{T} \delta_{nT}^{-2})$. Thus, the numerator of (39) is of order $O_p(1) + O_p\left(\frac{\sqrt{T}}{n}\right)$.

Finally, as $(n, T) \rightarrow \infty$ under the restriction $\frac{\sqrt{T}}{n} \rightarrow 0$, (39) becomes

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

Using a CLT based on Assumptions 1 and 2, $T^{-1/2} F' M_{X_i} \epsilon_i \xrightarrow{d} N[0, E(T^{-1} F' M_{X_i} \epsilon_i \epsilon_i' M_{X_i} F)]$, which proves (6). QED

Proof of Theorem 2. Using (30), we can write

$$\begin{aligned}
\hat{f}_t - f_t &= \frac{1}{n} \sum_{j=1}^n \left(\frac{\hat{F}' X_j}{T} \right) (\tilde{\beta}_j - \beta_j) (\tilde{\beta}_j - \beta_j)' x_{jt} \\
&\quad - \frac{1}{n} \left(\frac{\hat{F}' F}{T} \right) \sum_{j=1}^n \gamma_j (\tilde{\beta}_j - \beta_j)' x_{jt} - \frac{1}{nT} \sum_{j=1}^n (\hat{F}' \epsilon_j) (\tilde{\beta}_j - \beta_j)' x_{jt} \\
&\quad - \frac{1}{n} \sum_{j=1}^n \left(\frac{\hat{F}' X_j}{T} \right) (\tilde{\beta}_j - \beta_j) \gamma_j' f_t - \frac{1}{n} \sum_{j=1}^n \left(\frac{\hat{F}' X_j}{T} \right) (\tilde{\beta}_j - \beta_j) \epsilon_{jt} \\
&\quad + \frac{1}{nT} \sum_{j=1}^n (\hat{F}' \epsilon_j) \gamma_j' f_t + \frac{1}{n} \left(\frac{\hat{F}' F}{T} \right) \sum_{j=1}^n \gamma_j \epsilon_{jt} + \frac{1}{nT} \sum_{j=1}^n (\hat{F}' \epsilon_j) \epsilon_{jt} \\
&= I - II - III - IV - V + VI + VII + VIII.
\end{aligned} \tag{40}$$

The order of magnitude of I follows exactly from the same passages as in the proof of Lemma A.5, with $I = O_p(\phi_{nT}^{-2})$. Consider II ; omitting γ_j in view of Assumption 4(iii), we have

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

we have shown that $II_a = O_p\left(\frac{1}{\sqrt{nT}}\right)$ and $II_b = O_p\left(\frac{1}{\sqrt{nT}}\right) + O_p\left(\frac{1}{n}\right)$ in the proof of Lemma A.3, so that $II = O_p(n^{-1/2}T^{-1/2}) + O_p(n^{-1})$. Using Lemma A.3(i), it can be shown that $III = O_p(\phi_{nT}^{-2})$. As far as IV is concerned, note that

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

Similar passages as in the proof of the order of magnitude of II_a , and the fact that $E \|f_t\| < \infty$ entail $IV_a = O_p(n^{-1/2}T^{-1/2})$. Similarly, IV_b is bounded by $\|f_t\| \left[E \left\| \frac{\hat{F}' X_j}{T} \right\|^2 \right]^{1/2} \left[E \|\tilde{\Upsilon}_j\|^2 \right]^{1/2}$, which is $O_p(n^{-1/2}\delta_{nT}^{-1})$ using Lemma A.1. Thus, $IV = O_p(n^{-1/2}\delta_{nT}^{-1})$. Turning to V , it is bounded by $\left[E \left\| \frac{\hat{F}' X_j}{T} \right\|^4 \right]^{1/4} \left[E \|\epsilon_{jt}\|^4 \right]^{1/4} \left[E \|\tilde{\beta}_j - \beta_j\|^2 \right]^{1/2}$, whence $V = O_p(\phi_{nT}^{-1})$. The proofs of $VI = O_p\left(\frac{1}{\sqrt{nT}}\right)$, $VII = O_p\left(\frac{1}{\sqrt{n}}\right)$ and $VIII = O_p(\delta_{nT}^{-2})$ 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, it holds that $\hat{f}_t - f_t = O_p(\phi_{nT}^{-1}) + O_p(n^{-1/2})$. As $(n, T) \rightarrow \infty$ with $\frac{n}{T} \rightarrow 0$, the term that dominates is $VIII$, 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

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

Under H_0^a , $a_i = 0$; also, b_i can be rewritten as $b_i = \hat{\gamma}_i - \bar{\gamma}$. Using (39), we have

$$\begin{aligned} b_i &= \left(\hat{F}' M_{X_i} \hat{F} \right)^{-1} F' M_{X_i} \epsilon_i + \left(\hat{F}' M_{X_i} \hat{F} \right)^{-1} \left(\hat{F} - F \right)' M_{X_i} \epsilon_i \\ &\quad - \left(\hat{F}' M_{X_i} \hat{F} \right)^{-1} \hat{F}' M_{X_i} \left(\hat{F} - F \right) \gamma_i \\ &= b_{1i} + b_{2i}, \end{aligned} \tag{41}$$

where we define $b_{1i} = \left(\hat{F}' M_{X_i} \hat{F} \right)^{-1} F' M_{X_i} \epsilon_i$ 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

$$\begin{aligned} &T \left(\hat{\gamma}_i - \bar{\gamma} \right)' \hat{\Sigma}_{\gamma_i}^{-1} \left(\hat{\gamma}_i - \bar{\gamma} \right) \\ &= T \left(b'_{1i} \Sigma_{\gamma_i}^{-1} b_{1i} \right) + T b'_{1i} \Sigma_{\gamma_i}^{-1} \left(\hat{\Sigma}_{\gamma_i} - \Sigma_{\gamma_i} \right) \Sigma_{\gamma_i}^{-1} b_{1i} + T b'_{2i} \hat{\Sigma}_{\gamma_i}^{-1} b_{2i} \\ &\quad + 2T b'_{1i} \hat{\Sigma}_{\gamma_i}^{-1} b_{1i} + T \left(\hat{\gamma} - \bar{\gamma} \right)' \hat{\Sigma}_{\gamma_i}^{-1} \left(\hat{\gamma} - \bar{\gamma} \right) - 2T \left(\hat{\gamma} - \bar{\gamma} \right)' \hat{\Sigma}_{\gamma_i}^{-1} \left(\hat{\gamma}_i - \bar{\gamma} \right) \\ &= T \left(b'_{1i} \Sigma_{\gamma_i}^{-1} b_{1i} \right) + I_i + II_i + III_i + IV_i - V_i. \end{aligned} \tag{42}$$

After this preliminary calculations, we turn to proving (20). In order to do this, we firstly show that $\max_{1 \leq i \leq n} T \left(b'_{1i} \Sigma_{\gamma_i}^{-1} b_{1i} \right)$ 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 (42) are also all $o_p(\ln n)$ uniformly in i .

Consider $\max_{1 \leq i \leq n} T \left(b'_{1i} \Sigma_{\gamma_i}^{-1} b_{1i} \right)$, and consider in particular the sequence $\left\{ \sqrt{T} b_{1i} \right\}_{i=1}^n$. It holds that $\sqrt{T} b_{1i} = \left[T^{-1} \hat{F}' M_{X_i} \hat{F} \right]^{-1} \left[T^{-1/2} F' M_{X_i} \epsilon_i \right]$. As far as the numerator of this expression is concerned, by Lemma A.7(v) we write $T^{-1/2} F' M_{X_i} \epsilon_i = N_i + R_{N_i}$ with N_i de-

defined in Lemma A.7 as being zero mean Gaussian with covariance matrix $\Sigma_{fMe,i}$, and $R_{Ni} = o_p(n^{1/k}T^{1/k-1/2})$. Note also that $\sqrt{T}b_{1i}$ and $\sqrt{T}b_{1j}$ are independent for $i \neq j$ as $T \rightarrow \infty$, due to Assumption 1. Indeed, it can be shown through a Cramer-Wold device that $(\sqrt{T}b_{1i}, \sqrt{T}b_{1j})$ has, asymptotically, a normal distribution with covariance

$$\lim_{T \rightarrow \infty} E \left[\frac{F' M_{X_i} \epsilon_i \epsilon_j' M_{X_j} F}{T} \right] \leq \sigma_{ij} \lim_{T \rightarrow \infty} E \left[\frac{F' M_{X_i} M_{X_j} F}{T} \right],$$

where σ_{ij} is such that $|E(\epsilon_{it}\epsilon_{js})| \leq \sigma_{ij}$ for all (t, s) . Thus, in order to have asymptotic independence, it suffices to have $\sigma_{ij} \rightarrow 0$ as $(n, T) \rightarrow \infty$. As far as the denominator of $\sqrt{T}b_{1i}$ is concerned, based on Lemma A.7(ii) we write $[T^{-1}\hat{F}'M_{X_i}\hat{F}]^{-1} = \Sigma_{fM,i}^{-1} + R_{\Sigma fM,i}$ with $R_{\Sigma fM,i} = O_p(T^{-1/2}) + O_p(n^{-1})$. Hence we write

$$\sqrt{T}b_{1i} = [\Sigma_{fM,i}^{-1} + R_{\Sigma fM,i}] [N_i + R_{Ni}]. \quad (43)$$

Based on (43), and on the definitions of $\Sigma_{fMe,i}$ and of $\Sigma_{fM,i}$, it holds that

$$\begin{aligned} T \left(b_{1i}' \Sigma_{\gamma_i}^{-1} b_{1i} \right) &= N_i' \Sigma_{fMe,i}^{-1} N_i + 2N_i' \Sigma_{fM,i}^{-1} \Sigma_{\gamma_i}^{-1} R_{Ni} + 2R_{Ni}' \Sigma_{fMe,i}^{-1} N_i \\ &\quad + 2N_i' \Sigma_{fM,i}^{-1} \Sigma_{\gamma_i}^{-1} R_{\Sigma fM,i} N_i + 2N_i' \Sigma_{fM,i}^{-1} \Sigma_{\gamma_i}^{-1} R_{\Sigma fM,i} R_{Ni} \\ &\quad + R_{Ni}' \Sigma_{fMe,i}^{-1} R_{Ni} + 2R_{Ni}' \Sigma_{fM,i}^{-1} \Sigma_{\gamma_i}^{-1} R_{\Sigma fM,i} R_{Ni} \\ &\quad + N_i' R_{\Sigma fM,i} \Sigma_{\gamma_i}^{-1} R_{\Sigma fM,i} N_i + 2N_i' R_{\Sigma fM,i} \Sigma_{\gamma_i}^{-1} R_{\Sigma fM,i} R_{Ni} \\ &\quad + R_{Ni}' R_{\Sigma fM,i} \Sigma_{\gamma_i}^{-1} R_{\Sigma fM,i} R_{Ni} \\ &= N_i' \Sigma_{fMe,i}^{-1} N_i + I_i^{b1} + II_i^{b1} + III_i^{b1} + IV_i^{b1} + V_i^{b1} + VI_i^{b1} \\ &\quad + VII_i^{b1} + VIII_i^{b1} + IX_i^{b1}. \end{aligned} \quad (44)$$

The distribution of $N_i' \Sigma_{fMe,i}^{-1} N_i$ is χ_r^2 , and, in view of the above, this is a sequence of i.i.d. variables across i .

We now show that, in (44), $\max_{1 \leq i \leq n} I_i^{b1}, \dots, \max_{1 \leq i \leq n} IX_i^{b1}$ are all $o_p(1)$. Consider $\max_{1 \leq i \leq n} I_i^{b1}$; this is bounded by $\max_{1 \leq i \leq n} \|N_i\| \max_{1 \leq i \leq n} \|R_{Ni}\| = o_p(n^{1/k_1} T^{1/k_1-1/2} \sqrt{\ln n})$, in view of Lemma A.7(ii) and the fact that $\max_{1 \leq i \leq n} \|N_i\| = O_p(\sqrt{\ln n})$. The same holds for $\max_{1 \leq i \leq n} II_i^{b1}$. Turning to $\max_{1 \leq i \leq n} III_i^{b1}$, it is bounded by $\max_{1 \leq i \leq n} \|N_i\|^2 \max_{1 \leq i \leq n} \|R_{\Sigma fM,i}\| = O_p(T^{-1/2} \ln n)$

+ $O_p(n^{-1} \ln n)$ by virtue of Lemma A.7(ii). As far as $\max_{1 \leq i \leq n} IV_i^{b1}$ is concerned, it is bounded by $\max_{1 \leq i \leq n} \|N_i\| \max_{1 \leq i \leq n} \|R_{\Sigma f M, i}\| \max_{1 \leq i \leq n} \|R_{N_i}\|$, and therefore it is dominated by the previously analysed terms. Also, $\max_{1 \leq i \leq n} V_i^{b1}$ has the same order of magnitude as $\max_{1 \leq i \leq n} \|R_{N_i}\|^2$, thereby being dominated by the other terms. Similarly, $\max_{1 \leq i \leq n} VI_i^{b1}$ is bounded by $\max_{1 \leq i \leq n} \|R_{N_i}\|^2 \max_{1 \leq i \leq n} \|R_{\Sigma f M, i}\|$, and therefore it is also dominated. Turning to $\max_{1 \leq i \leq n} VII_i^{b1}$, it is bounded by $\max_{1 \leq i \leq n} \|N_i\|^2 \max_{1 \leq i \leq n} \|R_{\Sigma f M, i}\|^2$, so that it is smaller than $\max_{1 \leq i \leq n} III_i^{b1}$, and therefore negligible. Similarly, $\max_{1 \leq i \leq n} VIII_i^{b1}$ is bounded by $\max_{1 \leq i \leq n} \|N_i\| \max_{1 \leq i \leq n} \|R_{\Sigma f M, i}\|^2 \max_{1 \leq i \leq n} \|R_{N_i}\|$, which is dominated by $\max_{1 \leq i \leq n} IV_i^{b1}$, and thus negligible. Finally, $\max_{1 \leq i \leq n} IX_i^{b1}$ is bounded by $\max_{1 \leq i \leq n} \|R_{\Sigma f M, i}\|^2 \max_{1 \leq i \leq n} \|R_{N_i}\|^2$, and it is dominated. Therefore

$$\max_{1 \leq i \leq n} T \left(b'_{1i} \Sigma_{\gamma_i}^{-1} b_{1i} \right) = \max_{1 \leq i \leq n} N'_i \Sigma_{f M e, i}^{-1} N_i + o_p \left[(nT)^{1/k_1} \sqrt{\frac{\ln n}{T}} \right] + O_p \left(\frac{\ln n}{\sqrt{T}} \right) + O_p \left(\frac{\ln n}{n} \right), \quad (45)$$

and the approximation error is negligible as long as $\frac{(nT)^{2/k_1}}{T} \rightarrow 0$ as $(n, T) \rightarrow \infty$. Under Assumption 1 ($k_1 \geq 12$), and under $\frac{n^{4/k_1}}{T} \rightarrow 0$, this is always verified.

After proving that $\max_{1 \leq i \leq n} T \left(b'_{1i} \Sigma_{\gamma_i}^{-1} b_{1i} \right)$ can be approximated by $\max_{1 \leq i \leq n} N'_i \Sigma_{f M e, i}^{-1} N_i$, we turn again to equation (42). We now show that $\max_{1 \leq i \leq n} I_i, \dots, \max_{1 \leq i \leq n} V_i$ 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 (45) 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 (41), b_{2i} is defined as

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

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 $\max_{1 \leq i \leq n} \left\| \sqrt{T} \left(\hat{F}' M_{X_i} \hat{F} \right)^{-1} \right\|$

$\left(\hat{F} - F\right)' M_{X_i \epsilon_i} \left\| \right\|^2$ and of $\max_{1 \leq i \leq n} \left\| \sqrt{T} \left(\hat{F}' M_{X_i \hat{F}}\right)^{-1} \hat{F}' M_{X_i} \left(\hat{F} - F\right) \gamma_i \right\|^2$. 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 $\max_{1 \leq i \leq n} IV_i \leq \left\| \sqrt{T} \left(\hat{\gamma} - \bar{\gamma}\right) \right\|^2 \left\| \max_{1 \leq i \leq n} \Sigma_{\gamma_i}^{-1} \left(\hat{\Sigma}_{\gamma_i} - \Sigma_{\gamma_i}\right) \Sigma_{\gamma_i}^{-1} \right\|$, which is $o_p \left(T n^{2/k_1} \delta_{nT}^{-6}\right)$ by Lemmas A.4 and A.7(iv). Finally, $\max_{1 \leq i \leq n} V_i$ is bounded by $\left\| \sqrt{T} \left(\hat{\gamma} - \bar{\gamma}\right) \right\| \max_{1 \leq i \leq n} \left\| \sqrt{T} b_{1i} \right\| \left\| \max_{1 \leq i \leq n} \Sigma_{\gamma_i}^{-1} \left(\hat{\Sigma}_{\gamma_i} - \Sigma_{\gamma_i}\right) \Sigma_{\gamma_i}^{-1} \right\| = o_p \left(T n^{2/k_1} \delta_{nT}^{-4} \ln n\right)$. Putting all together, and using (45), it holds that

$$\begin{aligned}
\max_{1 \leq i \leq n} T \left(\hat{\gamma}_i - \hat{\gamma}\right)' \hat{\Sigma}_{\gamma_i}^{-1} \left(\hat{\gamma}_i - \hat{\gamma}\right) &= \max_{1 \leq i \leq n} N_i' \Sigma_{f_{Me,i}}^{-1} N_i + o_p \left[\left(nT\right)^{1/k_1} \sqrt{\frac{\ln n}{T}} \right] \\
&+ o_p \left(\frac{n^{2/k_1}}{\sqrt{T}} \ln n \right) + o_p \left(\frac{\sqrt{T} n^{2/k_1}}{n} \ln n \right) + o_p(1),
\end{aligned} \tag{46}$$

where the remainders are negligible as $(n, T) \rightarrow \infty$ with $\frac{\sqrt{T} n^{2/k_1}}{n} \rightarrow 0$ and $\frac{n^{4/k_1}}{T} \rightarrow 0$. Equation (20) follows from (46); the asymptotics of $\max_{1 \leq i \leq n} N_i' \Sigma_{f_{Me,i}}^{-1} N_i$ is studied e.g. in Embrechts, Kluppelberg and Mikosch (1997, Table 3.4.4, p.156)

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

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

with $I = O_p \left(T \|c_i\|^2\right)$ by construction. Also, II is bounded by $\sqrt{T} \left(\max_{1 \leq i \leq n} \|c_i\|\right) \left(\max_{1 \leq i \leq n} \sqrt{T} \|\hat{\gamma}_i - \gamma_i\|\right) = O_p \left[T \delta_{nT}^{-2} n^{1/k_1} \|c_i\|\right]$ in view of Lemma A.6(iii); similarly, $III = O_p \left(\sqrt{T} \delta_{nT}^{-2} \|c_i\|\right)$ 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 $c_{\alpha,n} - S_{\gamma,nT}^{NC} \rightarrow -\infty$ as $(n, T) \rightarrow \infty$. In view of equation (22), we know that $c_{\alpha,n} = O(\ln n)$, whence (21) 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

$$\hat{f}_t - \hat{\bar{f}} = (f_t - f) + (\hat{f}_t - f_t) - (\hat{\bar{f}} - f) = a_t + b_t - c_t.$$

Under H_0^b , $a_t = 0$ and $b_t = \hat{f}_t - f$; using (40), we can write

$$b_t = \left(\frac{\hat{F}'F}{T} \right) \frac{1}{n} \sum_{i=1}^n \gamma_i \epsilon_{it} + b_{2t} = b_{1t} + b_{2t}, \quad (47)$$

where b_{2t} contains terms $I-VI$ and $VIII$ in (40). Also, for each t , $\hat{\Sigma}_{ft}^{-1} = \Sigma_{ft}^{-1} + \Sigma_{ft}^{-1} (\hat{\Sigma}_{ft} - \Sigma_{ft}) \Sigma_{ft}^{-1} + o_p \left(\left\| \hat{\Sigma}_{ft} - \Sigma_{ft} \right\| \right)$.

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

$$\begin{aligned} & n \left(\hat{f}_t - \hat{\bar{f}} \right)' \hat{\Sigma}_{ft}^{-1} \left(\hat{f}_t - \hat{\bar{f}} \right) \\ &= n \left(b_{1t}' \Sigma_{ft}^{-1} b_{1t} \right) + n b_{1t}' \Sigma_{ft}^{-1} \left(\hat{\Sigma}_{ft} - \Sigma_{ft}^{-1} \right) \Sigma_{ft}^{-1} b_{1t} + n b_{2t}' \Sigma_{ft}^{-1} b_{2t} \\ & \quad + 2n b_{1t}' \Sigma_{ft}^{-1} b_{2t} + n \left(\hat{\bar{f}} - f \right)' \hat{\Sigma}_{ft}^{-1} \left(\hat{\bar{f}} - f \right) - 2n \left(\hat{\bar{f}} - f \right)' \hat{\Sigma}_{ft}^{-1} \left(\hat{f}_t - f \right) \\ &= n \left(b_{1t}' \Sigma_{ft}^{-1} b_{1t} \right) + I_t + II_t + III_t + IV_t - V_t. \end{aligned} \quad (48)$$

After this preliminary calculations, we now turn to proving (23). 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 stationary sequence of random variables with a χ_r^2 distribution, up to a negligible error. Secondly, we show that, in (48), $\max_{1 \leq t \leq T} I_t, \dots, \max_{1 \leq t \leq T} V_t$ 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}

$= o_p(T^{1/k_2}n^{1/k_2-1/2}) + o_p(T^{1/k_2-1/2}n^{1/2})$. The N_t s are independent across t by Assumption 1. More generally, the covariance matrix between N_t and N_s for $s \neq t$ is given by $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \sum_{j=1}^n \gamma_i \gamma_j' E(\epsilon_{it} \epsilon_{js})$. We point out that all the results derived here are still valid if there is serial dependence, as long as it holds that

$$\lim_{s \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \|\gamma_i \gamma_j'\| |E(\epsilon_{it} \epsilon_{js})| \ln s = 0, \quad (49)$$

which is known as Berman condition, Berman (1964). Note also that

$$\frac{\hat{F}'F}{T} = \Sigma_f + \left(\frac{F'F}{T} - \Sigma_f \right) + \frac{(\hat{F} - F)'F}{T} = \Sigma_f + R_f,$$

where $R_f = O_p(T^{-1/2}) + O_p(n^{-1})$ by Lemmas A.7(i) and A.3(ii). Hence, we write

$$\sqrt{nb_{1t}} = (\Sigma_f + R_f)(N_t + R_{Nt}). \quad (50)$$

Thence

$$\begin{aligned} n \left(b'_{1t} \hat{\Sigma}_{ft}^{-1} b_{1t} \right) &= N_t' \Sigma_{\Gamma\epsilon,t}^{-1} N_t + 2N_t' \Sigma_f^{-1} \Sigma_{ft}^{-1} R_{Nt} + 2R_{Nt}' \Sigma_{\Gamma\epsilon,t}^{-1} N_t \\ &\quad + 2N_t' \Sigma_f^{-1} \Sigma_{ft}^{-1} R_f N_t + 2N_t' \Sigma_f^{-1} \Sigma_{ft}^{-1} R_f R_{Nt} \\ &\quad + R_{Nt}' \Sigma_{\Gamma\epsilon,t}^{-1} R_{Nt} + 2R_{Nt}' \Sigma_f^{-1} \Sigma_{ft}^{-1} R_f R_{Nt} \\ &\quad + N_t' R_f \Sigma_{ft}^{-1} R_f N_t + 2N_t' R_f \Sigma_{ft}^{-1} R_f R_{Nt} \\ &\quad + R_{Nt} R_f \Sigma_{ft}^{-1} R_f R_{Nt} \\ &= N_t' \Sigma_{\Gamma\epsilon,t}^{-1} N_t + I_t^{b1} + II_t^{b1} + III_t^{b1} + IV_t^{b1} + V_t^{b1} + VI_t^{b1} \\ &\quad + VII_t^{b1} + VIII_t^{b1} + IX_t^{b1}. \end{aligned} \quad (51)$$

Passages are very similar to those after (44) 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 $o_p\left(T^{1/k_2}n^{1/k_2-1/2}\sqrt{\ln T}\right) + o_p\left(T^{1/k_2-1/2}n^{1/2}\sqrt{\ln T}\right)$; $\max_{1 \leq t \leq T} III_t^{b1} = O_p(T^{-1/2} \ln T) + O_p(n^{-1} \ln T)$; and that $\max_{1 \leq t \leq T} IV_t^{b1}, \dots,$

$\max_{1 \leq t \leq T} IX_t^{b1}$ are all dominated and therefore negligible. Thus

$$\max_{1 \leq t \leq T} n \left(b'_{1t} \Sigma_{ft}^{-1} b_{1t} \right) = \max_{1 \leq t \leq T} N'_t \Sigma_{\Gamma \epsilon, t}^{-1} N_t + o_p \left[(nT)^{1/k_2} \sqrt{\frac{\ln T}{n}} \right] + o_p \left(T^{1/k_2} \sqrt{\frac{n \ln T}{T}} \right) + o_p(1), \quad (52)$$

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

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 (48). We show that $\max_{1 \leq t \leq T} I_t, \dots, \max_{1 \leq t \leq T} V_t$ in (48) are all $o_p(\ln T)$. We have that $\max_{1 \leq t \leq T} I_t \leq \max_{1 \leq t \leq T} \|\sqrt{n} b_{1t}\|^2 \max_{1 \leq t \leq T} \left\| \Sigma_{ft}^{-1} \left(\hat{\Sigma}_{ft} - \Sigma_{ft}^{-1} \right) \Sigma_{ft}^{-1} \right\| = o_p(T^{2/k_2} \delta_{nT}^{-1} \ln T)$ by using Lemma A.7(iii). Also, combining Lemmas A.5 and Lemma A.7(iii), we have $\max_{1 \leq t \leq T} IV_t = O_p(n \delta_{nT}^{-4}) + o_p(n T^{2/k_2} \delta_{nT}^{-5})$ and $\max_{1 \leq t \leq T} V_t = O_p(n \delta_{nT}^{-3}) + o_p(n T^{2/k_2} \delta_{nT}^{-4})$. 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 $\max_{1 \leq t \leq T} \|b_{2t}\|$ and $\max_{1 \leq t \leq T} \|b_{2t}\|^2$. Recall that

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

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

$$\begin{aligned} \max_{1 \leq t \leq T} n \left(\hat{f}_t - \hat{\tilde{f}} \right)' \hat{\Sigma}_{ft}^{-1} \left(\hat{f}_t - \hat{\tilde{f}} \right) &= \max_{1 \leq t \leq T} N'_t \Sigma_{\Gamma \epsilon, t}^{-1} N_t + o_p \left[(nT)^{1/k_2} \sqrt{\frac{\ln T}{n}} \right] \\ &\quad + o_p \left(T^{1/k_2} \sqrt{\frac{n \ln T}{T}} \right) + o_p(1). \end{aligned} \quad (53)$$

Equation (23) follows from standard EVT (Embrechts, Kluppelberg and Mikosch, 1997) in case

of serial dependence, if the Berman condition (49) holds, equation (23) can be shown e.g. by using 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

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

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

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

$P[S_{f,nT} > c_{\alpha,T}]$ tends to 1 if $c_{\alpha,T} - S_{f,nT}^{NC} \rightarrow -\infty$ as $(n, T) \rightarrow \infty$; since we know from (25) that $c_{\alpha,T} = O(\ln T)$, (21) follows from (24). QED

In addition to the proofs of the main results in the paper, we now report two negative results concerning the use of average-type statistics to test for the null of no factor structure, as discussed in Section 3 - see equations (18) and (19).

Theorem B.1 *Let Assumptions 1-4 hold, and assume that, as $(n, T) \rightarrow \infty$*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\epsilon_i' M_{xi} F \Sigma_{fMe,i}^{-1} F' M_{Xi} \epsilon_i - r \right] = O_p(1).$$

Then, under H_0^a

$$\tilde{S}_{\gamma,nT} = O_p(1) + O_p\left(\sqrt{\frac{T}{n}}\right) + O_p\left(\sqrt{\frac{n}{T}}\right).$$

Proof of Theorem B.1 Under H_0^a we have

$$\begin{aligned}\sqrt{2r}\tilde{S}_{\gamma,nT} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[T(\hat{\gamma}_i - \gamma)' \hat{\Sigma}_{\gamma^i}^{-1} (\hat{\gamma}_i - \gamma) - r \right] + \frac{1}{\sqrt{n}} \sum_{i=1}^n T(\hat{\gamma} - \gamma)' \hat{\Sigma}_{\gamma^i}^{-1} (\hat{\gamma} - \gamma) \quad (54) \\ &\quad - \frac{2}{\sqrt{n}} \sum_{i=1}^n T(\hat{\gamma}_i - \gamma)' \hat{\Sigma}_{\gamma^i}^{-1} (\hat{\gamma} - \gamma) \\ &= I + II - III.\end{aligned}$$

Consider I ; using (36), we can write

$$\begin{aligned}I &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\frac{\epsilon'_i M_{X_i} F}{\sqrt{T}} \left(\frac{F' M_{X_i} F}{T} \right)^{-1} \hat{\Sigma}_{\gamma^i}^{-1} \left(\frac{F' M_{X_i} F}{T} \right)^{-1} \frac{F' M_{X_i} \epsilon_i}{\sqrt{T}} - r \right] \\ &\quad + \frac{1}{\sqrt{n}} T \sum_{i=1}^n \frac{\epsilon'_i M_{X_i} (\hat{F} - F)}{T} \left(\frac{F' M_{X_i} F}{T} \right)^{-1} \hat{\Sigma}_{\gamma^i}^{-1} \left(\frac{F' M_{X_i} F}{T} \right)^{-1} \frac{(\hat{F} - F)' M_{X_i} \epsilon_i}{T} \\ &\quad + \frac{1}{\sqrt{n}} T \sum_{i=1}^n \frac{\gamma' (\hat{F} - F)' M_{X_i} \hat{F}}{T} \left(\frac{F' M_{X_i} F}{T} \right)^{-1} \hat{\Sigma}_{\gamma^i}^{-1} \left(\frac{F' M_{X_i} F}{T} \right)^{-1} \frac{\hat{F}' M_{X_i} (\hat{F} - F) \gamma}{T} \\ &\quad + \frac{2}{\sqrt{n}} \sqrt{T} \sum_{i=1}^n \frac{\epsilon'_i M_{X_i} (\hat{F} - F)}{T} \left(\frac{F' M_{X_i} F}{T} \right)^{-1} \hat{\Sigma}_{\gamma^i}^{-1} \left(\frac{F' M_{X_i} F}{T} \right)^{-1} \frac{F' M_{X_i} \epsilon_i}{\sqrt{T}} \\ &\quad + \frac{2}{\sqrt{n}} T \sum_{i=1}^n \frac{\epsilon'_i M_{X_i} (\hat{F} - F)}{T} \left(\frac{F' M_{X_i} F}{T} \right)^{-1} \hat{\Sigma}_{\gamma^i}^{-1} \left(\frac{F' M_{X_i} F}{T} \right)^{-1} \frac{\hat{F}' M_{X_i} (\hat{F} - F) \gamma}{T} \\ &\quad + \frac{2}{\sqrt{n}} \sqrt{T} \sum_{i=1}^n \frac{\gamma' (\hat{F} - F)' M_{X_i} \hat{F}}{T} \left(\frac{F' M_{X_i} F}{T} \right)^{-1} \hat{\Sigma}_{\gamma^i}^{-1} \left(\frac{F' M_{X_i} F}{T} \right)^{-1} \frac{F' M_{X_i} \epsilon_i}{\sqrt{T}} \\ &= I_a + I_b + I_c + I_d + I_e + I_f.\end{aligned}$$

By assumption, I_a is $O_p(1)$. The joint limit assumed in the statement of the Theorem could be shown under more primitive conditions, but this suffices for our purposes. Turning to I_b , it is bounded by

$$\begin{aligned}\sqrt{n}TE \left[\frac{\epsilon'_i M_{X_i} (\hat{F} - F)}{T} \left(\frac{F' M_{X_i} F}{T} \right)^{-1} \hat{\Sigma}_{\gamma^i}^{-1} \left(\frac{F' M_{X_i} F}{T} \right)^{-1} \frac{(\hat{F} - F)' M_{X_i} \epsilon_i}{T} \right] \\ \leq M\sqrt{n}TE \left\| \frac{\epsilon'_i (\hat{F} - F)}{T} \right\|^2,\end{aligned}$$

where we have used the consistency of $\hat{\Sigma}_{\gamma_i}$ and Assumptions 3(i) and 4(iii). Applying Lemma A.2(i), we have $I_b = O_p(\sqrt{nT}\delta_{nT}^{-4})$. By a similar logic, it can be shown that I_c is bounded by $\sqrt{nTE} \left\| T^{-1} (\hat{F} - F) F' \right\|^2$, which is $O_p(\sqrt{nT}\delta_{nT}^{-4})$ by virtue of Lemma A.3(ii). Turning to I_d , similar passages as above entails that it is bounded by

$$\begin{aligned} \sqrt{nTE} \left[\frac{\epsilon'_i M_{Xi} (\hat{F} - F)}{T} \frac{F' M_{Xi} \epsilon_i}{\sqrt{T}} \right] \\ \leq \sqrt{nT} \left[E \left(\left\| \frac{\epsilon'_i (\hat{F} - F)}{T} \right\|^2 \right) \right]^{1/2} \left[E \left(\left\| \frac{F' \epsilon_i}{\sqrt{T}} \right\|^2 \right) \right]^{1/2} \\ = \sqrt{nT} O_p(\delta_{nT}^{-2}) O_p(1). \end{aligned}$$

Similarly, I_e is bounded by

$$\begin{aligned} \sqrt{nTE} \left[\frac{\epsilon'_i M_{Xi} (\hat{F} - F)}{T} \frac{\hat{F}' M_{Xi} (\hat{F} - F) \gamma}{T} \right] \\ \leq \sqrt{nT} \left[E \left(\left\| \frac{\epsilon'_i (\hat{F} - F)}{T} \right\|^2 \right) \right]^{1/2} \left[E \left(\left\| \frac{\hat{F}' (\hat{F} - F)}{T} \right\|^2 \right) \right]^{1/2} \\ = \sqrt{nT} O_p(\delta_{nT}^{-2}) O_p(\delta_{nT}^{-2}); \end{aligned}$$

using a similar logic, it can be shown that $I_f = O_p(\sqrt{nT}\delta_{nT}^{-2})$. Putting all together, $I = O_p(1) + O_p(\sqrt{nT}\delta_{nT}^{-2}) + O_p(\sqrt{nT}\delta_{nT}^{-4})$. Finally, consider II and III in (54). As far as II is concerned, note that $II = \sqrt{nT} (\hat{\gamma} - \bar{\gamma})' \Sigma_{\gamma_i}^{-1} (\hat{\gamma} - \bar{\gamma}) + o_p(1)$ by consistency of $\hat{\Sigma}_{\gamma_i}$. Thus, Lemma A.4 entails that $II = O_p(\sqrt{nT}\delta_{nT}^{-4})$. Turning to III , this is bounded by $\sqrt{nT} \max_i \Sigma_{\gamma_i}^{-1} \|\hat{\gamma} - \bar{\gamma}\| \left\| \frac{1}{n} \sum_{i=1}^n (\hat{\gamma}_i - \bar{\gamma}) \right\|$, which has the same order of magnitude as II .

Putting all together, it holds that $\sqrt{2r} \tilde{S}_{\gamma, nT} = O_p(1) + O_p(\sqrt{nT}\delta_{nT}^{-2}) + O_p(\sqrt{nT}\delta_{nT}^{-4}) = O_p(1) + O_p(\sqrt{\frac{n}{T}}) + O_p\left(\sqrt{\frac{T}{n}}\right)$. QED

Theorem B.2 *Let Assumptions 1-4 hold, and assume that, as $(n, T) \rightarrow \infty$*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \left[\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \gamma_i' \left(\frac{\hat{F}' F}{T} \right) \left(\frac{F' \hat{F}}{T} \right) \gamma_j \epsilon_{it} \epsilon_{jt} - r \right] = O_p(1).$$

Then, under H_0^b

$$\tilde{S}_{F,nT} = O_p(1) + O_p(\sqrt{n}) + O_p\left(\frac{n}{\sqrt{T}}\right) + O_p\left(\frac{\sqrt{T}}{n}\right).$$

Proof of Theorem B.2 Under H_0^b , we have

$$\begin{aligned} \sqrt{2r} \tilde{S}_{f,nT} &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \left[n (\hat{f}_t - f)' \Sigma_{ft} (\hat{f}_t - f) - r \right] \\ &\quad + \frac{1}{\sqrt{T}} \sum_{t=1}^T n (\hat{f} - f)' \Sigma_{ft} (\hat{f} - f) - \frac{2}{\sqrt{T}} \sum_{t=1}^T n (\hat{f}_t - f)' \Sigma_{ft} (\hat{f} - f) \\ &= I + II - III. \end{aligned} \tag{55}$$

Consider I ; using (47) we may write

$$I = \frac{1}{\sqrt{T}} \sum_{t=1}^T (nb'_{1t} \Sigma_{ft} b_{1t} - r) + \frac{1}{\sqrt{T}} \sum_{t=1}^T nb'_{2t} \Sigma_{ft} b_{2t} + \frac{2}{\sqrt{T}} \sum_{t=1}^T nb'_{1t} \Sigma_{ft} b_{2t} = I_a + I_b + I_c.$$

By assumption, $T^{-1/2} \sum_{t=1}^T (na'_t \Sigma_{ft} a_t - r) = O_p(1)$; this joint limit could be shown under more primitive assumptions. As far as I_b is concerned, by virtue of the consistency of Σ_{ft} , it is bounded by $n\sqrt{T}E \|b_t\|^2 = n\sqrt{T} \min\{T^{-1}, n^{-2}\} = O_p\left(\frac{n}{\sqrt{T}}\right) + O_p\left(\frac{\sqrt{T}}{n}\right)$, which follows from

the proof of Theorem 2. Finally, turning to I_c and setting $\hat{\Sigma}_{Ft}^{-1} = I_r$ for simplicity, we may write

$$\begin{aligned}
\frac{1}{2}I_c &= n \left(\frac{F' \hat{F}}{T} \right) \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma'_i \left(\frac{\hat{F}' X_j}{T} \right) (\tilde{\beta}_j - \beta_j) (\tilde{\beta}_j - \beta_j)' \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T x_{jt} \epsilon_{it} \right) \\
&\quad - n \left(\frac{F' \hat{F}}{T} \right) \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma'_i \left(\frac{\hat{F}' F}{T} \right) \gamma_j (\tilde{\beta}_j - \beta_j)' \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T x_{jt} \epsilon_{it} \right) \\
&\quad - n \left(\frac{F' \hat{F}}{T} \right) \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma'_i \frac{1}{T} (\hat{F}' \epsilon_j) (\tilde{\beta}_j - \beta_j)' \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T x_{jt} \epsilon_{it} \right) \\
&\quad - n \left(\frac{F' \hat{F}}{T} \right) \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma'_i \left(\frac{\hat{F}' X_j}{T} \right) (\tilde{\beta}_j - \beta_j) \gamma'_j \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T f_{t} \epsilon_{it} \right) \\
&\quad - n \left(\frac{F' \hat{F}}{T} \right) \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma'_i \left(\frac{\hat{F}' X_j}{T} \right) (\tilde{\beta}_j - \beta_j) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \epsilon_{jt} \epsilon_{it} \right) \\
&\quad + n \left(\frac{F' \hat{F}}{T} \right) \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma'_i \frac{\hat{F}' \epsilon_j}{T} \gamma'_j \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T f_{t} \epsilon_{it} \right) \\
&\quad + n \left(\frac{F' \hat{F}}{T} \right) \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma'_i \frac{\hat{F}' \epsilon_j}{T} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \epsilon_{jt} \epsilon_{it} \right) \\
&= I_{c,1} - I_{c,2} - I_{c,3} - I_{c,4} - I_{c,5} + I_{c,6} + I_{c,7}.
\end{aligned}$$

Studying the order of magnitude of each term is based on similar passages to the ones in the proof of Theorem 2. The only differences are: the summation across t ; the normalisation by $T^{-1/2}$; and the multiplication by n . The effect of summing across t is washed out by the normalisation by $T^{-1/2}$ for all terms $I_{c,1} - I_{c,4}$ and $I_{c,6}$, which can be shown by the same arguments as in (32). We have $I_{c,1} = O_p(\sqrt{n}T^{-1})$; $I_{c,2} = O_p(\delta_{nT}^{-1})$; $I_{c,3} = O_p(\sqrt{n}T^{-1})$, $I_{c,4} = O_p(\delta_{nT}^{-1})$ and $I_{c,6} = O_p(T^{-1/2})$. As far as $I_{c,5}$ and $I_{c,7}$ are concerned, the contribution of $T^{-1/2} \sum_{t=1}^T \epsilon_{jt} \epsilon_{it}$ is at most $O_p(\sqrt{T})$; thus, $I_{c,5} = O_p(\sqrt{n})$ and $I_{c,7} = O_p(\sqrt{T} \delta_{nT}^{-2})$. We now turn to analysing *II* and *III* in (55). By Lemma A.5, *II* = $n\sqrt{T} O_p(\delta_{nT}^{-4})$. As far as *III* is concerned, using the consistency of Σ_{ft} , it is bounded by $n\sqrt{T} \max_t \|\Sigma_{ft}\| \left\| \hat{f} - f \right\| \left\| \frac{1}{T} \sum_{t=1}^T (\hat{f}_t - f) \right\| = n\sqrt{T} O_p(\delta_{nT}^{-4})$, again by Lemma A.5. Putting all together, the result follows. QED

References

- Andrews, D.W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817-858.
- Bai, J., 2003, Inferential theory for structural models of large dimensions. *Econometrica* 71, 135-171.
- Bai, J., 2009a, Panel data models with interactive fixed effects. *Econometrica* 77, 1229-1279.
- Bai, J., 2009b, Supplement to Panel data models with interactive fixed effects : technical details and proofs. *Econometrica* 77, 1229-1279.
- 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* 70, 191-221.
- Berman, S.M., 1964, Limit theorems for the maximum term in stationary sequences. *The Annals of Mathematical Statistics* 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 Clermont-Ferrand*, 2, tome 90, Série Probabilités et Applications 6, 25-41.
- Castagnetti, C., Rossi, E., 2012, Euro corporate bond risk factors. *Journal of Applied Econometrics* (forthcoming).
- Csorgo, M., Horvath, L., 1997, *Limit theorems in change-point analysis*. Wiley, Chichester.
- Csorgo, M., Revesz, P., 1975a, A new method to prove Strassen-type laws of invariance principle, I. *Probability Theory and Related Fields* 31, 255-260.
- Csorgo, M., Revesz, P., 1975b, A new method to prove Strassen-type laws of invariance principle, II. *Probability Theory and Related Fields* 31, 261-269.

- Eberhardt, M., Helmers, C., Strauss, H., 2012, Do spillovers matter when estimating private returns to R&D?. *The Review of Economics and Statistics* (forthcoming).
- Eberhardt, M., Teal, F., 2012, No mangos in the tundra: spatial heterogeneity in agricultural productivity analysis. *Oxford Bulletin of Economics and Statistics* (forthcoming).
- Embrechts, P., Kluppelberg, C., Mikosch, T., 1997, *Modelling extremal events for insurance and finance*. New York: Springer.
- Hall, P., Miller, H., 2010, Bootstrap confidence intervals and hypothesis tests for extrema of parameters. *Biometrika* 97, 881-892.
- 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 Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*. Cambridge University Press, Cambridge.
- Komlos, J., Major, P., Tusnady, G., 1975, An approximation of partial sums of independent rv's and the sample distribution function, I. *Probability Theory and Related Fields* 32, 111-131.
- Komlos, J., Major, P., Tusnady, G., 1976, An approximation of partial sums of independent rv's and the sample distribution function, II. *Probability Theory and Related Fields* 34, 33-58.
- Leadbetter, M.R., Rootzen, H., 1988, Extremal theory for stochastic processes. *Annals of Probability* 16, 431-478.
- Newey, W.K., West, K.D., 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703-708.
- Pesaran, M. H., 2006, Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74, 967-1012.

- Pesaran, M. H., Tosetti, E., 2011, Large panels with common factors and spatial correlation. *Journal of Econometrics* 161, 182-202.
- Pesaran, M.H., Yamagata, T., 2008, Testing slope homogeneity in large panels. *Journal of Econometrics* 142, 50-93.
- Phillips, P.C.B., Moon, H. R., 1999, Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057-1112.
- Phillips, P.C.B., Solo, V., 1992, Asymptotics for linear processes. *Annals of Statistics* 20, 971-1001.
- Sarafidis, V., Yamagata, T., Robertson, D., 2009, A test of cross section dependence for a linear dynamic panel model with regressors. *Journal of Econometrics* 148, 149-461.
- Shorack, G.R., Wellner, J.A., 1986, *Empirical processes with applications to statistics*. Wiley, New York.
- Strang, G., 1988, *Linear algebra and its applications*. Third Edition, Harcourt, Orlando.
- Westerlund, J., Hess, W., 2011, A new poolability test for cointegrated panels. *Journal of Applied Econometrics* 26, 56-88.

n	T	20	30	50	100	200
20		0.900	0.932	0.949	0.959	0.881
30		0.947	0.971	0.958	0.965	0.894
50		0.867	0.986	0.979	0.979	0.980
100		0.971	0.990	0.970	0.993	0.993
200		0.991	0.993	0.994	0.993	0.992

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

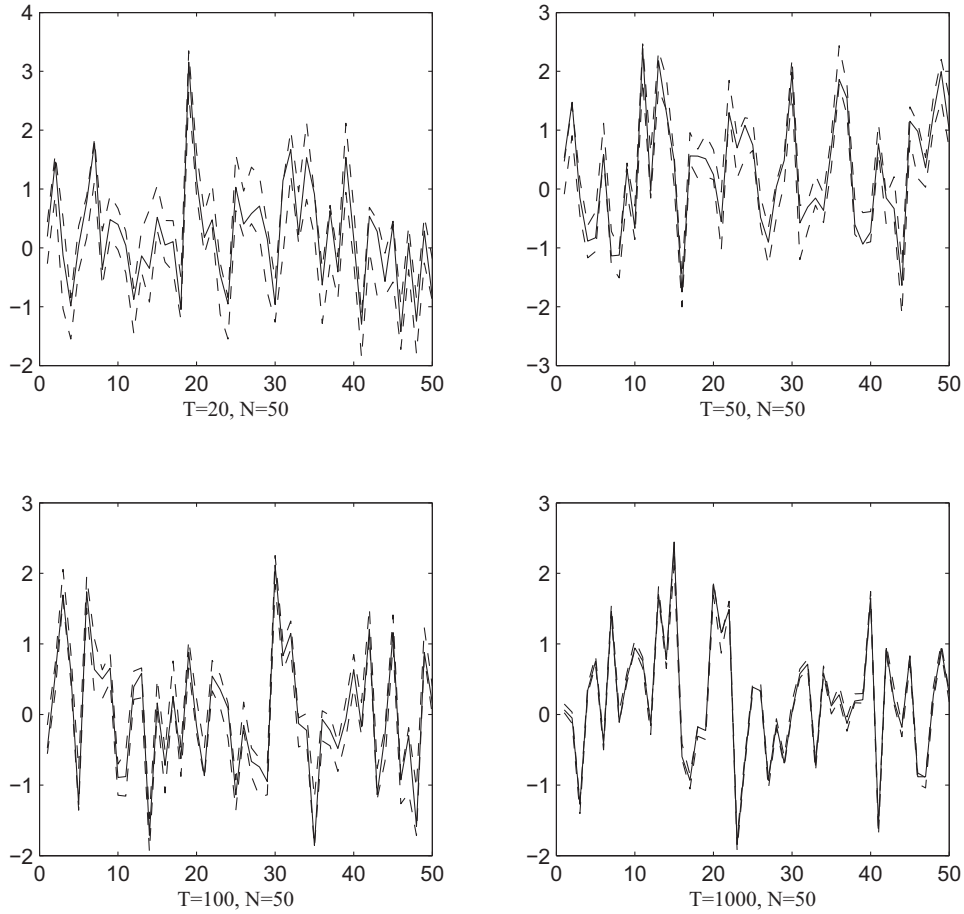


Figure 1: Confidence intervals for γ_i . For each value of $i = 1, \dots, 50$ (on the horizontal axis), the solid line represents the true loading γ_i . The dashed lines are the confidence intervals at 95% confidence level for each i .

n	Size				Power			
	T				T			
	30	50	100	200	30	50	100	200
$\sigma_\epsilon^2 = 1/3$								
30	0.086	0.068	0.057	0.053	0.917	1.000	1.000	1.000
50	0.065	0.058	0.050	0.049	0.978	1.000	1.000	1.000
100	0.077	0.061	0.062	0.041	0.999	1.000	1.000	1.000
200	0.080	0.049	0.045	0.039	1.000	1.000	1.000	1.000
$\sigma_\epsilon^2 = 1/2$								
30	0.093	0.071	0.065	0.056	0.764	0.968	1.000	1.000
50	0.069	0.058	0.054	0.053	0.905	0.999	1.000	1.000
100	0.079	0.064	0.063	0.041	0.977	0.999	1.000	1.000
200	0.080	0.050	0.048	0.042	0.993	1.000	1.000	1.000
$\sigma_\epsilon^2 = 1$								
30	0.122	0.095	0.082	0.070	0.470	0.770	1.000	1.000
50	0.096	0.073	0.061	0.063	0.615	0.947	0.999	1.000
100	0.085	0.066	0.070	0.043	0.773	0.964	1.000	1.000
200	0.088	0.053	0.050	0.043	0.858	0.995	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 (26)-(27).

n	Size				Power				
	T				T				
	30	50	100	200	30	50	100	200	
$\sigma_\epsilon^2 = 1/3$					$\sigma_\epsilon^2 = 1/3$				
30	0.061	0.050	0.046	0.048	0.994	1.000	1.000	1.000	
50	0.058	0.059	0.036	0.055	1.000	1.000	1.000	1.000	
100	0.080	0.052	0.051	0.040	1.000	1.000	1.000	1.000	
200	0.061	0.056	0.050	0.044	1.000	1.000	1.000	1.000	
$\sigma_\epsilon^2 = 1/2$					$\sigma_\epsilon^2 = 1/2$				
30	0.062	0.058	0.049	0.053	0.950	1.000	1.000	1.000	
50	0.059	0.063	0.036	0.056	0.997	1.000	1.000	1.000	
100	0.078	0.054	0.050	0.039	0.996	1.000	1.000	1.000	
200	0.065	0.056	0.050	0.045	1.000	1.000	1.000	1.000	
$\sigma_\epsilon^2 = 1$					$\sigma_\epsilon^2 = 1$				
30	0.069	0.069	0.059	0.060	0.721	0.994	1.000	1.000	
50	0.069	0.071	0.045	0.059	0.900	0.987	1.000	1.000	
100	0.081	0.062	0.055	0.042	0.831	0.982	1.000	1.000	
200	0.070	0.055	0.050	0.048	0.954	0.996	1.000	1.000	

Table 3: 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 (26)-(28), i.e. the case of no common factor structure in the regressors.

n	Size				Power			
	T				T			
	30	50	100	200	30	50	100	200
	$\sigma_\epsilon^2 = 1/3$				$\sigma_\epsilon^2 = 1/3$			
30	0.067	0.047	0.038	0.038	1.000	0.999	1.000	1.000
50	0.067	0.049	0.055	0.038	0.982	1.000	1.000	1.000
100	0.064	0.061	0.048	0.039	1.000	1.000	1.000	1.000
200	0.068	0.067	0.050	0.061	1.000	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1/2$				$\sigma_\epsilon^2 = 1/2$			
30	0.065	0.047	0.038	0.039	0.993	0.992	1.000	1.000
50	0.068	0.047	0.050	0.037	0.924	1.000	1.000	1.000
100	0.063	0.057	0.049	0.039	0.994	1.000	1.000	1.000
200	0.066	0.064	0.049	0.057	0.994	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1$				$\sigma_\epsilon^2 = 1$			
30	0.078	0.056	0.053	0.052	0.862	0.853	0.987	1.000
50	0.072	0.053	0.054	0.050	0.668	0.986	1.000	1.000
100	0.064	0.060	0.049	0.041	0.899	0.994	1.000	1.000
200	0.067	0.064	0.052	0.060	0.883	1.000	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 (29), i.e. the case of a pure factor model for y_{it} .

n	Size				Power			
	T				T			
	30	50	100	200	30	50	100	200
$\sigma_\epsilon^2 = 1/3$								
30	0.049	0.052	0.038	0.041	0.922	0.961	0.981	0.998
50	0.047	0.028	0.042	0.027	0.992	0.998	1.000	1.000
100	0.039	0.039	0.038	0.046	1.000	1.000	1.000	1.000
200	0.048	0.051	0.036	0.037	1.000	1.000	1.000	1.000
$\sigma_\epsilon^2 = 1/2$								
30	0.050	0.046	0.036	0.038	0.781	0.858	0.934	0.973
50	0.044	0.026	0.039	0.027	0.956	0.989	1.000	1.000
100	0.037	0.038	0.037	0.047	0.998	1.000	1.000	1.000
200	0.055	0.047	0.039	0.036	1.000	1.000	1.000	1.000
$\sigma_\epsilon^2 = 1$								
30	0.061	0.053	0.038	0.041	0.454	0.554	0.632	0.729
50	0.050	0.027	0.042	0.027	0.730	0.807	0.919	0.970
100	0.039	0.041	0.038	0.050	0.968	0.994	1.000	1.000
200	0.058	0.045	0.040	0.038	1.000	1.000	1.000	1.000

Table 5: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^b : f_t = f$, based on $S_{f,nT}$. The DGP used in the simulations is (26)-(27).

n	Size				Power			
	T				T			
	30	50	100	200	30	50	100	200
	$\sigma_\epsilon^2 = 1/3$				$\sigma_\epsilon^2 = 1/3$			
30	0.053	0.046	0.045	0.042	0.957	0.981	0.997	1.000
50	0.038	0.036	0.029	0.037	0.997	0.999	1.000	1.000
100	0.042	0.047	0.044	0.032	0.999	1.000	1.000	1.000
200	0.044	0.040	0.039	0.044	1.000	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1/2$				$\sigma_\epsilon^2 = 1/2$			
30	0.054	0.047	0.045	0.042	0.856	0.923	0.976	0.994
50	0.040	0.040	0.029	0.037	0.980	0.995	1.000	1.000
100	0.046	0.049	0.046	0.032	0.999	1.000	1.000	1.000
200	0.045	0.043	0.040	0.046	1.000	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1$				$\sigma_\epsilon^2 = 1$			
30	0.070	0.051	0.051	0.043	0.576	0.665	0.778	0.866
50	0.056	0.046	0.032	0.041	0.817	0.910	0.966	0.990
100	0.051	0.054	0.047	0.034	0.986	0.996	1.000	1.000
200	0.055	0.048	0.041	0.048	1.000	1.000	1.000	1.000

Table 6: Empirical rejection frequencies (for a nominal size of 5%) and power for tests for $H_0^b : f_t = f$, based on $S_{f,nT}$. The DGP used in the simulations is (26)-(28), i.e. the case of no common factor structure in the regressors x_{it} .

n	Size				Power			
	T				T			
	30	50	100	200	30	50	100	200
	$\sigma_\epsilon^2 = 1/3$				$\sigma_\epsilon^2 = 1/3$			
30	0.045	0.041	0.054	0.053	0.972	0.991	1.000	1.000
50	0.040	0.044	0.054	0.048	0.998	1.000	1.000	1.000
100	0.041	0.038	0.043	0.038	1.000	1.000	1.000	1.000
200	0.045	0.043	0.031	0.044	1.000	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1/2$				$\sigma_\epsilon^2 = 1/2$			
30	0.047	0.043	0.055	0.053	0.904	0.959	0.993	0.993
50	0.042	0.044	0.058	0.048	0.987	0.997	0.999	1.000
100	0.041	0.042	0.043	0.040	1.000	1.000	1.000	1.000
200	0.046	0.041	0.031	0.044	1.000	1.000	1.000	1.000
	$\sigma_\epsilon^2 = 1$				$\sigma_\epsilon^2 = 1$			
30	0.061	0.049	0.059	0.053	0.657	0.699	0.842	0.897
50	0.050	0.046	0.060	0.050	0.869	0.919	0.980	0.991
100	0.055	0.050	0.043	0.041	0.989	1.000	1.000	1.000
200	0.052	0.048	0.041	0.047	1.000	1.000	1.000	1.000

Table 7: 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 (29), i.e. the case of a pure factor model for y_{it} .