

GENERALIZED LEAST SQUARES ESTIMATION OF PANEL WITH COMMON SHOCKS

Paolo Zaffaroni*

Imperial College London

This draft: 3 February 2010

Abstract

This paper considers GLS estimation of linear panel models when the innovation and the regressors can both contain a factor structure. A novel feature of this approach is that preliminary estimation of the latent factor structure is not necessary. Under a set of regularity conditions here provided, we establish consistency and asymptotic normality of the feasible GLS estimator as both the cross-section and time series dimensions diverge to infinity. Dependence, both cross-sectional and temporal, of the idiosyncratic innovation is permitted. Our results are presented separately for time regressions with unit-specific coefficients and for cross-section regressions with time-specific coefficients. Primitive conditions of our assumptions are established for Andrews (2005) and Pesaran (2006) regression models, as particular cases of our set up. Monte Carlo experiments corroborate our results.

Key Words: panel, factor model, heterogeneous coefficients, time-varying coefficients, estimation, generalized least squares.

1 Introduction

Factor models represent one of the most popular and successful way to capture cross-sectional and temporal dependence, especially when facing a large

*I thank participants at the 2009 Cowles Foundation Econometrics Conference on "Handling Dependence: Temporal, Cross-sectional, and Spatial" and the 2009 NBER Summer Institute, Working Group on Empirical in Macroeconomics and Finance for comments. All errors are mine. Contact address: Imperial College Business School, Imperial College London, London SW7 2AZ United Kingdom, Telephone: +44 (0)20 7594 9186, email p.zaffaroni@imperial.ac.uk.

number of units (N). However, a factor structure in the innovation of a linear regression model can make the ordinary least squares (henceforth OLS) estimator invalid since it will no longer be consistent, in general, for the true regression coefficients unless some restrictions are imposed. In a linear cross-sectional regression with constant parameters Andrews (2005) shows that, when the error and the regressors both have a factor structure, consistency of the OLS estimator is preserved only with uncorrelated factor loadings. The parameters estimate has a mixed normal asymptotic distribution, as N goes to infinity. Within a linear regression across time (T) with the innovation and the regressors sharing a factor structure, when a panel of observations is available, Pesaran (2006) shows that individual-specific regression coefficients can be consistently estimated by augmenting the regressors with cross-sectional averages of the dependent variable and individual-specific regressors. Asymptotic normality is obtained, as both N, T go to infinity. Again, the essential condition is a restriction on the joint distribution of the factors loadings for the factor structure in the regressors and innovation, namely that their (population) means must be linearly independent.

This paper considers cross-sectional regressions with time-specific parameters as well as time regressions with individual-specific parameters when the innovation contains a factor structure and a panel of data is available. Both cases are of independent interest. We show that the unfeasible generalized least squares (henceforth UGLS) estimator, based on the presumption that the covariance matrix of the factor structure is known, is consistent and asymptotically normal distributed *without any* particular restriction on the factor loadings nor on the common factors, in particular even if the innovation and the regressors are mutually correlated. This is due to a form of asymptotic orthogonality between the factor loadings and the *inverse* of the factor structure covariance matrix. The difficulty arises when considering a feasible version of the UGLS estimator. A natural approach would be, exploiting the panel dimension, to consider the sample covariance matrix of the OLS residuals. Given the non-consistency of the OLS estimator, such sample covariance matrix is also non-consistent for the true covariance matrix. However, the relevant result here is that, under suitable regularity conditions, the limit of such sample covariance matrix leads to a matrix whose inverse is also asymptotically orthogonal to the factor loadings. Indeed, there is an entire class of matrices, rather than a unique matrix, that is asymptotically orthogonal to the factor loadings. As a consequence, we show that this *feasible* GLS (henceforth GLS) estimator is consistent and asymptotically normal, as both

N, T diverge to infinity, even under a set of conditions that make the OLS invalid. However, the limit covariance matrix of the OLS residuals will be in general different from the true covariance matrix of true innovations, and thus such the GLS might not be as efficient as the UGLS. Efficiency gains can be obtained, though, by an evaluating the GLS iteratively, as explained below.

The GLS estimator exhibits two main desirable properties. First, it is computationally easy to handle since it simply requires to perform a sequence of linear regressions. Second, the GLS estimator does not require knowledge of the number of factors nor of estimates of the factors themselves. Hence, we do not need to make use of the recent advances in estimation of (dynamic) factor models such as Forni, Hallin, Lippi, and Reichlin (2000), Bai and Ng (2002) and Stock and Watson (2002) which in turn would require preliminary testing of the number of factors (see Bai and Ng (2002) and Hallin and Liska (2007), Onatski (2009) for tests designed for static and dynamic factor models, respectively). In contrast to Andrews (2005), Pesaran (2006) and our approach where estimation of the factor structure is not required, Bai (2009) considers joint estimation of regression coefficients and of the innovation factor structure. Knowledge of, at least, an upper bound for the exact number of factors is required. The regression coefficients estimator is consistent although asymptotically biased, in general, but an asymptotically valid bias-correction is established. The true regression coefficients, though, have to be constant in Bai (2009) approach ruling out both unit- and time-variation.

Panel with factor structure innovations have also been considered by Holtz-Eakin, Newey, and Rosen (1988), Ahn, Hoon Lee, and Schmidt (2001), Bai and Ng (2004) Phillips and Sul (2003), Moon and Perron (2003), Phillips and Sul (2007) and Moon and Weidner (2009). Most of these papers are defined within the context of dynamic panel models. In particular, Holtz-Eakin, Newey, and Rosen (1988) note how the individual effects can be eliminated by quasi-differencing although this induces time-variation to otherwise constant regression coefficients. They consider the asymptotic properties of an instrumental variable estimator for large N where the number of instruments is of order $O(T^2)$. Ahn, Hoon Lee, and Schmidt (2001) focus on generalized method of moment estimation of cross-sectional regressions with independent and identically distributed (*i.i.d.*) regressors for fixed T , ruling out dynamic panel. For autoregressive panel models with possibly a time trend, Bai and Ng (2004) study unit root tests that permit to identify whether the non-

stationarity is associated with the factor structure part or with the idiosyncratic part. They do not treat the factor structure as a nuisance parameter but build their test on principal components estimates of common and idiosyncratic components, providing the asymptotic properties of the test for large N, T . For the same models, Phillips and Sul (2003) focus on median unbiased estimation of the autoregressive parameter, and related homogeneity and unit root tests. Their asymptotic theory holds for fixed N . Moon and Perron (2003) propose unit root testing with respect to a similar class of models, valid for both large N, T , based on de-factoring the data by means of principal components estimation of the factor structure which if ignored, would substantially reduce the power of the test. Their test has no power when a linear trend with fixed effects is allowed for. For a larger class of dynamic panels, that allows for exogenous regressors, Phillips and Sul (2007) characterize the bias of the (pooled) OLS estimator for large N , in particular showing that it converges to a random variable because of the substantial degree of cross-sectional dependence associated with the factor structure innovation. Moon and Weidner (2009) consider a similar model and estimator to Bai (2009) and established the limiting distribution of the likelihood ratio and Lagrange multiplier test for the constant regression coefficients, under weaker conditions on the observed regressors permitting dynamic panel.

This paper proceeds as follows. Section 2 illustrates the basic definitions and the general assumptions required for estimation of regressions with unit-specific parameters stating with a theorem the asymptotic properties of the OLS, UGLS and GLS estimator as T , in the first two cases, and as N, T in the last case, diverge to infinity. Section 2.3 then considers, as a special case of our set-up, the regression model with unit-specific parameters of Pesaran (2006), establishing primitive conditions for our general assumptions. In particular, we show how some, but not all, of these conditions are implied by certain of Pesaran's (2006) assumptions, summarizing the findings in a proposition. Section 3 focuses on regression models with time-specific parameters, again presenting the basic definitions and the general assumptions, summarizing the asymptotic properties of the OLS, UGLS and GLS as N and N, T , respectively, diverge to infinity. Since Andrews (2005) cross-sectional model represents a special case of this set-up, Section 3.3 investigates the extent to which Andrew's (2005) assumptions provide primitive conditions for at least some of our general assumptions. The full set of required primitive conditions is then described in a proposition. Our theoretical results are corroborated by a set of Monte Carlo experiments described in Section 4.

Section 5 concludes. The proofs of both theorems are reported in the final Appendix.

Hereafter we use the following notation: \rightarrow_p denotes convergence in probability and \rightarrow_d convergence in distribution. When $\mathbf{A} > 0$ we mean that the matrix \mathbf{A} is positive definite, $\mathbf{A} \geq 0$ that \mathbf{A} is positive semi definite, indicating by $\mathbf{A}^{\frac{1}{2}}$ its square-root matrix, $\|\mathbf{A}\| = (\text{tr}(\mathbf{A}\mathbf{A}'))^{\frac{1}{2}}$ indicates the Euclidean norm of the matrix \mathbf{A} , ι_n is a $n \times 1$ vector of ones, μ_{ja} is the limit in probability of $\iota_n' \mathbf{a}_j / n$ (as $n \rightarrow \infty$) for a sequence of $n \times 1$ random vectors \mathbf{a}_j , the matrix $\Sigma_{jA'CB}$ is the limit in probability of $\mathbf{A}_j' \mathbf{C}_j \mathbf{B}_j / n$ (as $n \rightarrow \infty$) for sequences of random matrices $\mathbf{A}_j, \mathbf{B}_j$ with n rows and a constant number of columns and for the sequence of the random $n \times n$ matrices \mathbf{C}_j , all possibly indexed by j . When \mathbf{C}_j equals the identity matrix \mathbf{I}_n , we write $\Sigma_{jA'B}$. Clearly $\Sigma_{jA'CB}' = \Sigma_{jB'C'A}$. Note that μ_{ja} and $\Sigma_{jA'CB}$ can be, and will generally be, random. A sequence of random vectors of dimension $n \times 1$, normally distributed with mean zero and identity covariance matrix, will be denoted by ζ_{nj} . We skip dependence on the index j from the above expressions when not necessary.

2 Unit-Specific Parameters Model

2.1 Definitions and assumptions

Throughout this section, the observed variables obey a linear regression model with a $k \times 1$ vector of possible unit-specific regression coefficients β_{i0} . The model for the i th unit can be expressed, in matrix form, as

$$\mathbf{y}_i = \mathbf{X}_i \beta_{i0} + \mathbf{u}_i, \quad (1)$$

for an observed $T \times 1$ vector $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})'$, an observed $T \times k$ matrix $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})'$ where either none, some or even all of the regressors can be common across units, and an unobserved $T \times 1$ vector $\mathbf{u}_i = (u_{i1}, \dots, u_{iT})'$. The innovation satisfy the factor structure

$$\mathbf{u}_i = \mathbf{F} \mathbf{b}_i + \varepsilon_i, \quad (2)$$

for an unobserved $m \times 1$ vector of factor loadings \mathbf{b}_i , an unobserved $T \times m$ matrix of common factors $\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_T)'$ and an unobserved $T \times 1$ vector of idiosyncratic innovations $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})'$. The maintained assumption

here is that k and m *do not vary* with T and N . Although model (1) is written as a regression across time for a given i , we assume that in fact a panel of observations $\{\mathbf{y}, \mathbf{X}\} = \{\mathbf{y}_1, \dots, \mathbf{y}_i, \dots, \mathbf{y}_N, \mathbf{X}_1, \dots, \mathbf{X}_i, \dots, \mathbf{X}_N\}$ is available. As pointed out in Pesaran (2006, section 2), several panel models, with either constant or unit-specific regression coefficients, are encompassed by his model, which in turn is a particular case of (1), including the traditional fixed and random effects models.

We now specify a set of general assumptions required for the estimators here considered, commenting on them through a series of remarks below. We then state, in Theorem 1, the asymptotic properties of the OLS, UGLS and GLS estimators for β_{i0} . In the subsequent section we establish a set of primitive conditions of our general assumptions for the particular case of interest of model (1) given by Pesaran (2006) model.

Assumption 1. \mathcal{H} (factor loadings)

For every i , the \mathbf{b}_i are random vector of dimension $m \times 1$ such that

$$E(\mathbf{b}_i \mathbf{b}_i' \mid \mathbf{X}_i, \mathbf{F}) = \mathcal{B}_i > 0,$$

$$N^{-1} \sum_{i=1}^N \mathcal{B}_i \xrightarrow{p} \mathcal{B} > 0 \text{ as } N \rightarrow \infty.$$

Assumption 2. \mathcal{H} (idiosyncratic innovation)

For every i , $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})'$ satisfies

$$E(\varepsilon_i \mid \mathbf{b}_i, \mathbf{X}_i, \mathbf{F}) = \mathbf{0}, \tag{3}$$

$$\mathbf{H}_i = E(\varepsilon_i \varepsilon_i' \mid \mathbf{b}_i, \mathbf{X}_i, \mathbf{F}) > 0, \tag{4}$$

$$N^{-1} \sum_{i=1}^N \mathbf{H}_i \xrightarrow{p} \mathcal{H}_T > 0 \text{ as } N \rightarrow \infty. \tag{5}$$

Assumption 3. \mathcal{H} (regressors)

For every i , the $T \times k$ matrix \mathbf{X}_i is full column rank.

Assumption 4. \mathcal{H} (basic limit conditions)

All the limit matrices below, as $T \rightarrow \infty$, are *a.s.* finite:

$$\frac{\mathbf{X}'_i \mathbf{X}_i}{T} \rightarrow_p \Sigma_{iX'X} > 0, \quad \frac{\mathbf{X}'_i \mathbf{H}_i \mathbf{X}_i}{T} \rightarrow_p \Sigma_{iX'HX} > 0, \quad \frac{\mathbf{X}'_i \mathbf{F}}{T} \rightarrow_p \Sigma_{iX'F}, \quad (6)$$

$$\frac{\mathbf{X}'_i \mathbf{H}_i^{-1} \mathbf{X}_i}{T} \rightarrow_p \Sigma_{iX'H^{-1}X} > 0, \quad \frac{\mathbf{F}' \mathbf{H}_i^{-1} \mathbf{F}}{T} \rightarrow_p \Sigma_{iF'H^{-1}F} > 0, \quad \frac{\mathbf{X}'_i \mathbf{H}_i^{-1} \mathbf{F}}{T} \rightarrow_p \Sigma_{iX'H^{-1}F}, \quad (7)$$

such that *a.s.*

$$\Sigma_{iX'H^{-1}X} - \Sigma_{iX'H^{-1}F} \Sigma_{iF'H^{-1}F}^{-1} \Sigma'_{iX'H^{-1}F} > 0. \quad (8)$$

Assumption 5.H (limit conditions for GLS)

All the limit matrices below, as $N \rightarrow \infty$ and arbitrary T , are *a.s.* finite:

$$\begin{aligned} \sum_{i=1}^N \frac{\mathbf{X}_i \Sigma_{iX'X}^{-1} \Sigma_{iX'F} \mathbf{b}_i \mathbf{b}'_i \Sigma'_{iX'F} \Sigma_{iX'X}^{-1} \mathbf{X}'_i}{N} &= \mathcal{A}_{1T}(1+o_p(1)), \quad \sum_{i=1}^N \frac{\mathbf{X}_i \Sigma_{iX'X}^{-1} \Sigma_{iX'F} \mathbf{b}_i \mathbf{b}'_i \mathbf{F}'}{N} = \mathcal{A}_{2T}(1+o_p(1)), \\ \sum_{i=1}^N \frac{\mathbf{X}_i \Sigma_{iX'X}^{-1} \mathbf{X}'_i \varepsilon_i \varepsilon'_i \mathbf{X}_i \Sigma_{iX'X}^{-1} \mathbf{X}'_i}{NT} &= \mathcal{A}_{3T}(1+o_p(1)), \quad \sum_{i=1}^N \frac{\mathbf{X}_i \Sigma_{iX'X}^{-1} \mathbf{X}'_i \varepsilon_i \varepsilon'_i}{N} = \mathcal{A}_{4T}(1+o_p(1)), \\ \sum_{i=1}^N \frac{\mathbf{b}_i \varepsilon'_i}{N^{\frac{1}{2}}} &= \mathcal{C}_{1T}(1+o_p(1)), \quad \sum_{i=1}^N \frac{\mathbf{X}_i \Sigma_{iX'X}^{-1} \Sigma_{iX'F} \mathbf{b}_i \varepsilon'_i}{N^{\frac{1}{2}}} = \mathcal{C}_{2T}(1+o_p(1)), \end{aligned} \quad (9)$$

$$\sum_{i=1}^N \frac{\mathbf{b}_i \varepsilon'_i \mathbf{X}_i \Sigma_{iX'X}^{-1} \mathbf{X}'_i}{N^{\frac{1}{2}} T^{\frac{1}{2}}} = \mathcal{C}_{3T}(1+o_p(1)), \quad \sum_{i=1}^N \frac{\mathbf{X}_i \Sigma_{iX'X}^{-1} \Sigma_{iX'F} \mathbf{b}_i \varepsilon'_i \mathbf{X}_i \Sigma_{iX'X}^{-1} \mathbf{X}'_i}{N^{\frac{1}{2}} T^{\frac{1}{2}}} = \mathcal{C}_{4T}(1+o_p(1)). \quad (10)$$

Assumption 6.H (distribution conditions for OLS and UGLS)

As $T \rightarrow \infty$:

$$\frac{\mathbf{X}'_i \varepsilon_i}{T^{\frac{1}{2}}} \rightarrow_d (\Sigma_{iX'HX})^{\frac{1}{2}} \zeta_{ki}, \quad (11)$$

$$\frac{\mathbf{Z}'_i \mathbf{H}_i^{-1} \varepsilon_i}{T^{\frac{1}{2}}} \rightarrow_d (\Sigma_{iZ'H^{-1}Z})^{\frac{1}{2}} \zeta_{(k+m)i}, \quad (12)$$

setting

$$\mathbf{Z}_i = (\mathbf{X}_i, \mathbf{F}), \quad (13)$$

where $\zeta_{ki}, \zeta_{(k+m)i}$ are independent from $\Sigma_{iX'HX}, \Sigma_{iZ'H^{-1}Z}$ respectively.

Assumption 7.H (distribution and identification conditions for GLS)
Let $\mathcal{D}_{1T}, \mathcal{E}_{1T}$ be $m \times m$ matrices and $\mathcal{D}_{2T}, \mathcal{E}_{2T}$ be $T \times T$ matrices satisfying *a.s.*

$$\begin{aligned} \mathcal{A}_{1T} - (\mathcal{A}_{2T} + \mathcal{A}'_{2T}) &= \mathbf{F}\mathcal{D}_{1T}\mathbf{F}' + \mathcal{D}_{2T} \\ &\text{and} \\ \mathcal{A}_{3T} - (\mathcal{A}_{4T} + \mathcal{A}'_{4T}) &= \mathbf{F}\mathcal{E}_{1T}\mathbf{F}' + \mathcal{E}_{2T}. \end{aligned} \tag{14}$$

Then

$$\begin{aligned} \mathcal{I}_{1T} &= \mathcal{D}_{1T} + T^{-1}\mathcal{E}_{1T} + \mathcal{B}, \text{ (non-singular symmetric } m \times m \text{ matrix),} \\ \mathcal{I}_{2T} &= \mathcal{D}_{2T} + T^{-1}\mathcal{E}_{2T} + \mathcal{H}_T, \text{ (non-singular } T \times T \text{ matrix),} \end{aligned}$$

satisfy

$$\begin{aligned} \frac{\mathbf{F}'\mathcal{I}_{2T}^{-1}\mathbf{F}}{T} &\rightarrow_p \Sigma_{\mathbf{F}'\mathcal{I}_2^{-1}\mathbf{F}} \text{ (non-singular),} & \frac{\mathbf{F}'\mathcal{I}_{2T}^{-1}\mathbf{H}_i\mathcal{I}_{2T}^{-1}\mathbf{F}}{T} &\rightarrow_p \Sigma_{i\mathbf{F}'\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}\mathbf{F}}, \tag{15} \\ \frac{\mathbf{X}'_i\mathcal{I}_{2T}^{-1}\mathbf{X}_i}{T} &\rightarrow_p \Sigma_{i\mathbf{X}'\mathcal{I}_2^{-1}\mathbf{X}} \text{ (non-singular),} & \frac{\mathbf{X}'_i\mathcal{I}_{2T}^{-1}\mathbf{H}_i\mathcal{I}_{2T}^{-1}\mathbf{X}_i}{T} &\rightarrow_p \Sigma_{i\mathbf{X}'\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}\mathbf{X}}, \\ \frac{\mathbf{X}'_i\mathcal{I}_{2T}^{-1}\mathbf{F}}{T} &\rightarrow_p \Sigma_{i\mathbf{X}'\mathcal{I}_2^{-1}\mathbf{F}}, & \frac{\mathbf{X}'_i\mathcal{I}_{2T}^{-1}\mathbf{H}_i\mathcal{I}_{2T}^{-1}\mathbf{F}}{T} &\rightarrow_p \Sigma_{i\mathbf{X}'\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}\mathbf{F}}, \\ \frac{\mathbf{Z}'_i\mathcal{I}_{2T}^{-1}\varepsilon_i}{T^{\frac{1}{2}}} &\rightarrow_d \left(\Sigma_{i\mathbf{Z}'\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}\mathbf{Z}} \right)^{\frac{1}{2}} \zeta_{(k+m)i}, \tag{16} \end{aligned}$$

with $\zeta_{(k+m)i}$ independent from $\Sigma_{i\mathbf{Z}'\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}\mathbf{Z}}$, where \mathbf{Z}_i is defined in (13). All the limits above hold as $T \rightarrow \infty$ with *a.s.* finite limit matrices. Setting

$$\Sigma_T = \mathbf{F}\mathcal{I}_{1T}\mathbf{F}' + \mathcal{I}_{2T},$$

for all i and some constants $a, b, c, d > 0$:

$$\mathbf{X}'_i\Sigma_T^{-1}(\mathbf{F}\mathcal{C}_{1T} + \mathcal{C}'_{1T}\mathbf{F}' + \mathcal{C}_{2T} + \mathcal{C}'_{2T})\Sigma_T^{-1}(\mathbf{F}\mathbf{b}_i + \varepsilon_i) = O_p(T^a\iota_k), \tag{17}$$

$$\mathbf{X}'_i\Sigma_T^{-1}(\mathbf{F}\mathcal{C}_{3T} + \mathcal{C}'_{3T}\mathbf{F}' + \mathcal{C}_{4T} + \mathcal{C}'_{4T})\Sigma_T^{-1}(\mathbf{F}\mathbf{b}_i + \varepsilon_i) = O_p(T^b\iota_k), \tag{18}$$

$$\mathbf{X}'_i\Sigma_T^{-1}(\mathbf{F}\mathcal{C}_{1T} + \mathcal{C}'_{1T}\mathbf{F}' + \mathcal{C}_{2T} + \mathcal{C}'_{2T})\Sigma_T^{-1}\mathbf{X}_i = O_p(T^c\iota_k\iota'_k), \tag{19}$$

$$\mathbf{X}'_i\Sigma_T^{-1}(\mathbf{F}\mathcal{C}_{3T} + \mathcal{C}'_{3T}\mathbf{F}' + \mathcal{C}_{4T} + \mathcal{C}'_{4T})\Sigma_T^{-1}\mathbf{X}_i = O_p(T^d\iota_k\iota'_k). \tag{20}$$

Remarks 1. We are assuming that the factor loadings \mathbf{b}_i are unobserved random variables with a non-singular yet possibly heterogeneous distribution,

varying with the index i . We do not necessarily require the \mathbf{b}_i to be mutually independent from the regressors and from the factors although mutual independence is typically assumed.

2. The factors \mathbf{f}_t are assumed unobserved, whereas observed factors, if present, will be simply part of the regressors \mathbf{X}_i . Moreover, there is no restriction on the time dependence of the \mathbf{f}_t , which can be autocorrelated. One of the sufficient conditions for Assumption 7. \mathcal{H} will be, however, boundedness of $\Sigma_{F'F}$. Hence the \mathbf{f}_t can satisfy for instance a stationary vector auto-regression.

3. The idiosyncratic innovation ε_i does not need to be *i.i.d* across i , nor needs to be independent from either the factor loadings \mathbf{b}_i , the factors \mathbf{F} and the observed regressors \mathbf{X}_i . Strong exogeneity, although not assumed here, of \mathbf{X}_i and mutual independence of $\mathbf{b}_i, \mathbf{f}_t$ with ε_{it} imply (3). Note, however, that this requirement only refers to the idiosyncratic part of the regression error. Instead, \mathbf{X}_i and the common component of the regression error, $\mathbf{f}'_t \mathbf{b}_i$, are allowed to, and in general will, be cross-correlated. Moreover, \mathbf{H}_i can vary with i and does not need to be diagonal, implying a substantial degree of both heterogeneity as well as the possibility of time dependence.

4. Assuming full column rank of \mathbf{X}_i for all i is required, given that computationally the GLS estimator relies on the evaluation of a sequence of N different OLS problems.

5. When $\Sigma_{iX'X}$ and $\Sigma_{F'F}$ are finite, then the other limit matrices are also finite by the CauchySchwarz inequality when, in addition, the maximum eigenvalue of \mathbf{H}_i is bounded and its minimum eigenvalue is bounded away from zero, uniformly in T . Such assumptions are not stated explicitly since here (*a.s.*) finiteness of the limit matrices suffices. As a simple example, when $\mathbf{X}_i = \iota_T$ (intercept only), then $\mathbf{X}'_i \mathbf{H}_i \mathbf{X}_i / T = \iota'_T \mathbf{H}_i \iota_T / T$. Its limit is certainly finite (*a.s.*) when the maximum eigenvalue of \mathbf{H}_i is bounded but finiteness can also be obtained when a finite number of eigenvalues of \mathbf{H}_i grows slower than $O(T)$. Bounded-ness of the maximum eigenvalue is implied when \mathbf{u}_i satisfy an approximate factor structure (see Chamberlain (1983)).

6. Here $\Sigma_{iX'F}$ represents the cross-correlation (when $E\mathbf{F} = \mathbf{0}$) between the regressors \mathbf{X}_i and the factors \mathbf{F} and it determines the non-zero asymptotic bias of the OLS estimator, except for the trivial case of no factor structure ($\mathbf{b}_i = \mathbf{0}$). Under our assumptions, the regression innovation \mathbf{u}_i has covariance matrix

$$\mathbf{S}_i = \mathbf{F} \mathbf{B}_i \mathbf{F}' + \mathbf{H}_i$$

and, as illustrated below (see (8)), the UGLS estimator of β_{i0} requires the

limit of $T^{-1}\mathbf{X}'_i\mathbf{S}_i^{-1}\mathbf{X}_i$ to be positive definite.

7. The limit matrices in Assumptions 5. \mathcal{H} and 7. \mathcal{H} arise when looking at the probability limit of the sample covariance matrix of the OLS innovations.

8. The limiting distribution results stated in Assumptions 6. \mathcal{H} and 7. \mathcal{H} , are required for OLS-UGLS and GLS respectively. Note that the possibility of a factor structure in the regressors \mathbf{X}_i implies mixed normal asymptotic distributions unless the asymptotic covariance matrix turns out to be constant. Conditional versions of the usual central limit theorems (CLT) will be required (see for instance Andrews (2005, Theorem 4)). Since we aim at providing general results, we do not specify here the primitive conditions required for such conditional CLT. These conditions can be relatively easily established when one considers particular cases of (1) such as for Pesaran (2006)'s model, examined in Section 2.3.

9. As explained below, considering the GLS requires to consider Σ_T in place of \mathbf{S}_i . Therefore, the various conditions dictated by Assumption 7. \mathcal{H} on $\mathcal{I}_{1T}, \mathcal{I}_{2T}$ make sure that Σ_T^{-1} will be asymptotically orthogonal to the matrix of latent factors \mathbf{F} . This is the essential property that guarantees that the GLS estimator will have good asymptotic properties.

10. Conditions (17)-(20) determine the speed at which N and T have to diverge to infinity, possibly at different rates, to ensure that the GLS estimator is consistent and asymptotically normal.

2.2 Estimators results

For estimation of parameters β_{i0} , the OLS estimator yield

$$\hat{\beta}_i^{OLS} = (\mathbf{X}'_i\mathbf{X}_i)^{-1}\mathbf{X}_i\mathbf{y}_i,$$

The *unfeasible* generalized least squares (UGLS) estimator is

$$\hat{\beta}_i^{UGLS} = (\mathbf{X}'_i\mathcal{S}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i\mathcal{S}_i^{-1}\mathbf{y}_i,$$

setting

$$\mathcal{S}_i = \mathbf{F}\mathcal{B}_i\mathbf{F}' + \mathbf{H}_i.$$

The *feasible* generalized least squares (GLS) estimator is

$$\hat{\beta}_i^{GLS} = (\mathbf{X}'_i\hat{\Sigma}_T^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i\hat{\Sigma}_T^{-1}\mathbf{y}_i,$$

setting

$$\hat{\Sigma}_T = N^{-1} \sum_{i=1}^N \hat{u}_i \hat{u}_i', \quad \hat{u}_i = \mathbf{y}_i - \mathbf{X}_i \hat{\beta}_i^{OLS}.$$

This requires at minimum $N \geq T$. Note, however, that if the regressors contain some observed factors, such as for instance when an intercept term is allowed for, which can be written, without loss of generality, as $\mathbf{X}_i = (\mathbf{D}, \mathbf{X}_i^*)$ for a $T \times k_1$ matrix \mathbf{D} and a $T \times k_2$ matrix \mathbf{X}_i^* , where $k = k_1 + k_2$, then $\hat{u}_i' \mathbf{D} = \mathbf{0}$ for all i . As a consequence, $\hat{\Sigma}_T$ will be at most of rank $T - k_1 < T$, no matter how large N is. Therefore, to allow non-singularity we consider instead the alternative definition

$$\hat{\Sigma}_T = N^{-1} \sum_{i=1}^N \hat{u}_i \hat{u}_i' + T^{-1} \mathbf{D} \mathbf{D}',$$

where the normalization by T^{-1} is imposed since, from our assumptions, $\sup_T \|\hat{\Sigma}_T^{-1}\| = O(1)$ a.s. is required.

Theorem 1 (*unit-specific parameters*)

(i) (OLS) When Assumptions 3.H, 4.H.(6), 6.H.(11) hold

$$T^{\frac{1}{2}}(\hat{\beta}_i^{OLS} - \beta_{i0} - \gamma_i^{OLS}) \rightarrow_d (\mathcal{V}_i^{OLS})^{\frac{1}{2}} \zeta_{ki} \text{ as } T \rightarrow \infty,$$

setting $\gamma_i^{OLS} = \Sigma_{iX'X}^{-1} \Sigma_{iX'F} \mathbf{b}_i$, $\mathcal{V}_i^{OLS} = \Sigma_{iX'X}^{-1} \Sigma_{iX'HX} \Sigma_{iX'X}^{-1}$, where ζ_{ki} and \mathcal{V}_i^{OLS} are mutually independent.

(ii) (UGLS) When Assumptions 1.H, 2.H.(4), 3.H, 4.H.(7), 6.H.(12)

$$T^{\frac{1}{2}}(\hat{\beta}_i^{UGLS} - \beta_{i0}) \rightarrow_d (\mathcal{V}_i^{UGLS})^{\frac{1}{2}} \zeta_{ki} \text{ as } T \rightarrow \infty,$$

where ζ_{ki} and \mathcal{V}_i^{UGLS} are mutually independent, setting $\mathcal{V}_i^{UGLS} = (\mathcal{M}_i^{UGLS})^{-1} \mathcal{N}_i^{UGLS} (\mathcal{M}_i^{UGLS})^{-1}$ with $\mathcal{M}_i^{UGLS} = \text{plim}_{T \rightarrow \infty} T^{-1} (\mathbf{X}_i' \mathcal{S}_i^{-1} \mathbf{X}_i)$, $\mathcal{N}_i^{UGLS} = \text{plim}_{T \rightarrow \infty} T^{-1} \mathbf{X}_i' \mathcal{S}_i^{-1} \mathbf{H}_i \mathcal{S}_i^{-1} \mathbf{X}_i$. Moreover $\mathcal{M}_i^{UGLS} = \mathcal{N}_i^{UGLS}$.

(iii) (GLS) When Assumptions 1.H, 2.H.(5), 3.H, 4.H.(6) and (8), 5.H, 7.H

$$\begin{aligned} \hat{\beta}_i^{GLS} &\rightarrow_p \beta_{i0} & \text{as } \frac{1}{T} + \frac{T^{\max(a-1, b-\frac{3}{2})}}{N^{\frac{1}{2}}} + \frac{T^{\max(c-\frac{3}{2}, d-2)}}{N^{\frac{1}{2}}} &\rightarrow 0, \\ T^{\frac{1}{2}}(\hat{\beta}_i^{GLS} - \beta_{i0}) &\rightarrow_d (\mathcal{V}_i^{GLS})^{\frac{1}{2}} \zeta_{ki} & \text{as } \frac{1}{T} + \frac{T^{\max(a-\frac{1}{2}, b-1)}}{N^{\frac{1}{2}}} + \frac{T^{\max(c-1, d-\frac{3}{2})}}{N^{\frac{1}{2}}} &\rightarrow 0, \end{aligned}$$

where ζ_{ki} and \mathcal{V}_i^{GLS} are mutually independent, setting $\mathcal{V}_i^{GLS} = (\mathcal{M}_i^{GLS})^{-1} \mathcal{N}_i^{GLS} (\mathcal{M}_i^{GLS})^{-1}$ with $\mathcal{M}_i^{GLS} = \text{plim}_{(N,T) \rightarrow \infty} T^{-1} (\mathbf{X}_i' \hat{\Sigma}_T^{-1} \mathbf{X}_i)$, $\mathcal{N}_i^{GLS} = \text{plim}_{(N,T) \rightarrow \infty} T^{-1} \mathbf{X}_i' \hat{\Sigma}_T^{-1} \mathbf{H}_i \hat{\Sigma}_T^{-1} \mathbf{X}_i$.

Remarks 1. The asymptotic bias of the OLS estimator is not simply expressed in terms an un-centered asymptotic distribution which would otherwise still ensures consistency. Instead $\beta_i^{OLS} = \beta_{i0} + \gamma_i^{OLS} + O_p(T^{-\frac{1}{2}})$ where $\gamma_i^{OLS} = \Sigma_{iX'X}^{-1} \Sigma_{iX'F} \mathbf{b}_i$ is, in general, a random variable. Consistency is achieved if either $\mathbf{b}_i = \mathbf{0}$, meaning no factor structure, or $\Sigma_{iX'F} = \mathbf{0}$, that is zero cross-correlation between the regressors and the factors (assuming the latter have mean zero).

2. It is well-known that the UGLS/GLS estimator improve efficiency with respect the the OLS estimator for non-spherical innovations. Here we find that UGLS/GLS exhibit a more profound property: they completely eliminates the factor structure's adverse effect on OLS of inducing an asymptotic bias. Moreover, when the regressors contain a factor structure, an example of which is $\mathbf{X}_i = \mathbf{D}\mathbf{A}_i + \mathbf{F}\mathbf{\Gamma}_i + \mathbf{V}_i$, as in Pesaran's (2006) model (see Section 2.3 below), then both \mathcal{V}_i^{UGLS} and \mathcal{V}_i^{GLS} will be (functionally) independent from \mathbf{D} and, in particular, from \mathbf{F} yielding $T^{\frac{1}{2}}(\hat{\beta}_i^{UGLS} - \beta_{i0}) \rightarrow_d \mathcal{N}(\mathbf{0}, \mathcal{V}_i^{UGLS})$ as $T \rightarrow \infty$ and $T^{\frac{1}{2}}(\hat{\beta}_i^{GLS} - \beta_{i0}) \rightarrow_d \mathcal{N}(\mathbf{0}, \mathcal{V}_i^{GLS})$ as $1/T + T^{\max(a-\frac{1}{2}, b-1)}/N^{\frac{1}{2}} + T^{\max(c-1, d-\frac{3}{2})}/N^{\frac{1}{2}} \rightarrow 0$. Therefore, the UGLS/GLS estimators have the conventional asymptotic normal distribution, unlike the OLS which has a mixed normal asymptotic distribution. The possibility of a different, asymptotic, behaviour of OLS and GLS has already been noted by Robinson and Hidalgo (1997) in a time series regression context with possibly long memory innovation and regressors.

3. The reason underlying this important property of the UGLS estimator here uncovered is the asymptotic orthogonality between the *inverse* of the factor structure covariance matrix \mathbf{S}_i^{-1} and the factor matrix \mathbf{F} , formalized in general terms in Lemma 1 in the Appendix. This result has been used, in the different context of financial portfolio optimization, by Pesaran and Zaffaroni (2009) who establish that mean-variance trading strategies do allow complete diversification of both idiosyncratic and common shocks to asset returns.

4. The GLS estimator here proposed does not achieve in general the same efficiency as the UGLS, as discussed below. Our estimator does, however, exhibit the desired asymptotic properties, as N, T diverge *jointly* to infinity at suitable rates, meaning that our result does not depend on the somewhat restrictive approach of taking sequential limits. When $a \leq 1, b \leq \frac{3}{2}, c \leq \frac{3}{2}$ and $d \leq 2$ then consistency is achieved without the need to specify the relative

speed at which N, T diverge to infinity. These conditions appear cumbersome due to the generality of our approach, whereas they become much simpler when looking at specific models such as Pesaran (2006), described in the next section.

The reason behind the asymptotic properties of the GLS is that, although $\hat{\Sigma}_T$ is a non-consistent estimate of the true covariance matrix \mathbf{S}_i (in the sense of element by element), the inverse of its limit $\Sigma_T = \mathbf{F}\mathcal{I}_{1T}\mathbf{F}' + \mathcal{I}_{2T}$ belongs to the space orthogonal to the factors \mathbf{F} , under suitable regularity conditions. On the other hand, the GLS estimator does not require to identify, let alone to estimate, the factor structure within the innovation so that, for instance, one does not need to know m , the true number of factors, as long as it is finite. In the case of no factor structure ($m = 0$) our method continue to work, without making use of this information which obviously suggests use of OLS.

5. Since GLS delivers consistent parameter estimates, this suggests a multi-step approach, achieving a more efficient estimator. The first step consists of getting the GLS estimator $\hat{\beta}_i^{(1)} = \hat{\beta}_i^{GLS}$ as described above. Next, one can evaluate $\hat{\Sigma}_{T(1)} = N^{-1} \sum_{i=1}^N \hat{u}_i^{(1)} \hat{u}_i^{(1)'}$ for $\hat{\mathbf{u}}_i^{(1)} = \mathbf{y}_i - \mathbf{X}_i \hat{\beta}_i^{(1)}$ in order to get the *second-step* GLS estimator $\hat{\beta}_i^{(2)} = (\mathbf{X}_i' \hat{\Sigma}_{T(1)}^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i \tilde{\Sigma}_{T(1)}^{-1} \mathbf{y}_i$. Given a further set of conditions that build on Assumptions \mathcal{H} , one can show that $\hat{\beta}_i^{(2)}$ is also consistent and asymptotically normal, as N, T diverge to infinity. Moreover, it can be shown that $\hat{\Sigma}_{T(1)} \rightarrow_p \mathbf{S}_i + T^{-1} \mathbf{R}_{i(1)}$, as $N \rightarrow \infty$, for a $T \times T$ matrix $\mathbf{R}_{i(1)}$ satisfying $\sup_T \|T^{-1} \mathbf{R}_{i(1)}\| = O(1)$ *a.s.* and where each element of $T^{-1} \mathbf{R}_{i(1)}$ goes to zero as $T \rightarrow \infty$. Hence, $\hat{\Sigma}_{T(1)}$ is closer to \mathbf{S}_i than $\hat{\Sigma}_T$, where the approximation improves the larger N and T are. This suggests that a certain efficiency improvements can already be achieved by using the two-step GLS estimator $\hat{\beta}_i^{(2)}$ and, indeed, such improvement can be substantial when N, T are both sizeable. Further gains are expected when the procedure is iterated yielding, after J steps, $\hat{\beta}_i^{(J)} = (\mathbf{X}_i' \hat{\Sigma}_{T(J-1)}^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i \tilde{\Sigma}_{T(J-1)}^{-1} \mathbf{y}_i$ where $\tilde{\Sigma}_{T(J-1)}$ is based on $\hat{\beta}_i^{(J-1)}$. Below we report some Monte Carlo results in order to quantify the possible improvements of efficiency in finite samples from using the multi-step GLS estimator with $J = 4$.

6. Consistent estimation of the GLS asymptotic covariance matrix \mathcal{V}_i^{GLS} can be obtained using the nonparametric Newey and West (1987, Theorem 2)

approach yielding $\hat{\mathcal{V}}_i^{GLS} = (\hat{\mathcal{M}}_i^{GLS})^{-1} \hat{\mathcal{N}}_i^{GLS} (\hat{\mathcal{M}}_i^{GLS})^{-1}$ with

$$\hat{\mathcal{M}}_i^{GLS} = \frac{1}{T} (\mathbf{X}_i' \hat{\Sigma}_T^{-1} \mathbf{X}_i), \hat{\mathcal{N}}_i^{GLS} = \boldsymbol{\Omega}_{i0} + \sum_{h=1}^n (1 - \frac{h}{(n+1)}) (\boldsymbol{\Omega}_{ih} + \boldsymbol{\Omega}'_{ih}), \quad (21)$$

setting $\boldsymbol{\Omega}_{ih} = \frac{1}{T} \sum_{t=h+1}^T \hat{u}_{it} \hat{u}_{it-h} \hat{\mathbf{x}}_{it} \hat{\mathbf{x}}'_{it-h}$, $h = 0, 1, \dots, T-1$, where $\hat{X}_i = (\hat{\mathbf{x}}_{i1} \dots \hat{\mathbf{x}}_{iT})' = \hat{\Sigma}_T^{-1} \mathbf{X}_i$ and $\hat{\mathbf{u}}_t^{(1)} = (\hat{u}_{i1}^{(1)} \dots \hat{u}_{iT}^{(1)})' = \mathbf{y}_t - \mathbf{X}_i \hat{\beta}_i^{GLS}$ and the bandwidth $n = n(T, N)$ grows slowly with T, N . The same approach has been used in Pesaran (2006, eq. (51) and (52)) and Bai (2009, Remark 8) to obtain a consistent estimator of the asymptotic covariance matrix for their estimators when correlation and heterokedasticity is allowed for.

Notice that, although $\hat{\mathbf{u}}_t^{(1)}$ does contain the factor structure, since its population counterpart is $\mathbf{u}_t = \mathbf{F}\mathbf{b}_t + \varepsilon_{it}$, pre-multiplication by $\hat{\Sigma}_T^{-1}$ will make (asymptotically) negligible the contribution to $\hat{\mathbf{u}}_t^{(1)}$ of the common factors. An alternative approach consists of estimating the idiosyncratic component of $\hat{\mathbf{u}}_t^{(1)}$ directly, for example by principal components, yielding $\hat{\varepsilon}_t^{(1)}$ and then replace $\boldsymbol{\Omega}_{ih}$ by $T^{-1} \sum_{t=h+1}^T \hat{\varepsilon}_{it} \hat{\varepsilon}_{it-h} \hat{\mathbf{x}}_{it} \hat{\mathbf{x}}'_{it-h}$ into (21). Preliminary testing for the number of factors is required.

2.3 Particular model: Pesaran (2006)

The model is

$$y_{it} = \alpha'_{0i} \mathbf{d}_t + \beta'_{0i} \mathbf{x}_{it} + e_{it}, \quad (22)$$

where \mathbf{d}_t is a $n \times 1$ vector of *observed* factors, \mathbf{x}_{it} is a $k \times 1$ observed vector satisfying

$$\mathbf{x}_{it} = \mathbf{A}'_i \mathbf{d}_t + \boldsymbol{\Gamma}'_i \mathbf{f}_t + \mathbf{v}_{it} \quad (23)$$

where \mathbf{f}_t is the $m \times 1$ vector of *unobserved* factors, $\mathbf{A}_i, \boldsymbol{\Gamma}_i$ are $n \times k$ and $m \times k$ matrices of factor loadings, \mathbf{v}_{it} is the $k \times 1$ vector of specific components of the regressors \mathbf{x}_{it} . Finally

$$e_{it} = \mathbf{f}'_t \boldsymbol{\gamma}_i + \varepsilon_{it}, \quad (24)$$

with ε_{it} independent of $\mathbf{d}_t, \mathbf{x}_{it}$ and \mathbf{v}_{it} independent of $\mathbf{d}_t, \mathbf{f}_t$. With respect to our notation, (23)-(24) imply $\mathbf{F} = (\mathbf{f}_1 \dots \mathbf{f}_T)'$, $\mathbf{B} = (\boldsymbol{\gamma}_1 \dots \boldsymbol{\gamma}_T)'$, $\mathbf{X}_i = (\mathbf{D}, \mathbf{X}_i^*)$, where we set $\mathbf{X}_i^* = \mathbf{D}\mathbf{A}_i + \mathbf{F}\boldsymbol{\Gamma}_i + \mathbf{V}_i$ with $\mathbf{D} = (\mathbf{d}_1 \dots \mathbf{d}_T)'$, $\mathbf{V}_i = (\mathbf{v}_{i1} \dots \mathbf{v}_{iT})'$.

We now verify the extent to which the assumptions of Pesaran (2006) imply our Theorem 1, part (iii). It turns out that some of our conditions

are weaker and other are stronger than Pesaran (2006). Note that since the model permits common observed factors, one will need to add the term $T^{-1}\mathbf{D}\mathbf{D}'$ to $\hat{\Sigma}_T$, in particular to its \mathcal{I}_{2T} component.

Assumption 1. \mathcal{H} follows by the strong law of large numbers (LLN) and Pesaran (2006, Assumption 3) where $\mathcal{B}_i = \mathcal{B}$ equals to $\gamma\gamma' + \mathbf{\Omega}_\eta$ using Pesaran's notation. We further require $\mathcal{B} > 0$. Assumption 2. \mathcal{H} is only in part implied by Pesaran (2006, Assumption 2), in particular (4) is so, but we also require $N^{-1}\sum_{i=1}^N \mathbf{H}_i \rightarrow_p \mathcal{H}_T > 0$, not necessarily implied by Pesaran (2006, eq. (10)). Assumption 3. \mathcal{H} is implied by Pesaran (2006, Assumption 5a). Concerning Assumption 4. \mathcal{H} , (6) follows by strengthening Pesaran (2006, Assumption 1 and 2) to fourth-order covariance stationarity with absolute summable autocovariances, yielding

$$\begin{aligned} \frac{\mathbf{X}'_i \mathbf{X}_i}{T} &\rightarrow_p \Sigma_{iX'X} = \\ &\begin{pmatrix} \Sigma_{D'D} & \Sigma_{D'D}\mathbf{A}_i + \Sigma_{D'F}\mathbf{\Gamma}_i \\ \mathbf{A}'_i \Sigma_{D'D} + \mathbf{\Gamma}'_i \Sigma_{F'D} & \Sigma_{V'V} + \mathbf{A}'_i \Sigma_{D'D}\mathbf{A}_i + \mathbf{\Gamma}'_i \Sigma_{F'F}\mathbf{\Gamma}_i + \mathbf{A}'_i \Sigma_{D'F}\mathbf{\Gamma}_i + \mathbf{\Gamma}'_i \Sigma_{F'D}\mathbf{A}_i \end{pmatrix}, \\ \frac{\mathbf{X}'_i \mathbf{F}}{T} &\rightarrow_p \Sigma_{iX'F} = \begin{pmatrix} \Sigma_{D'F} \\ \mathbf{\Gamma}'_i \Sigma_{F'F} + \mathbf{A}'_i \Sigma_{D'F} \end{pmatrix}, \end{aligned}$$

since $\Sigma_{F'V}$ and $\Sigma_{D'V}$ are both matrices of zeros by Pesaran (2006, Assumption 1 and 2). By the same assumptions, $\Sigma_{iX'X}$ is bounded and, using the block matrix decomposition (Magnus and Neudecker 1988), is non-singular whenever both matrices $\Sigma_{D'D}$ and $\Sigma_{V'V} - \mathbf{\Gamma}'_i \Sigma_{(F'D)} \mathbf{\Gamma}_i$ are non-singular, where we set

$$\Sigma_{(F'D)} = \Sigma_{F'D} \Sigma_{D'D}^{-1} \Sigma_{D'F} - \Sigma_{F'F}.$$

The latter requires $\Sigma_{V'V} > 0$, implied by Pesaran (2006, Assumption 2) who defines it as Σ_i , since $-\Sigma_{(F'D)}$ is positive semi definite, in fact at most $\mathbf{0}$ for perfectly correlated $\mathbf{f}_t, \mathbf{d}_t$. However, here we require in addition $\Sigma_{D'D} > 0$. Expression for $\Sigma_{iX'HX}$ will depend on the adopted parameterization for the $h_{ts,i}$, that is on the form of the moving average coefficients a_{il} in Pesaran (2006, Assumption 2). However, under summability of the moving average coefficients a_{il} , which implies the spectral density of the ε_{it} to be finite at all frequencies, then boundedness of $\Sigma_{iX'X}$ implies, by the spectral decomposition of positive definite matrices, boundedness of $\Sigma_{iX'HX}$. Note, however, that the UGLS estimator does need $\mathbf{H}_i > 0$, as in Assumption 2. \mathcal{H} (4), which in turn requires the spectral density of the ε_{it} to be bounded away from

zero, ensuring boundedness of $\Sigma_{iX'H^{-1}X}$, $\Sigma_{iF'H^{-1}F}$, $\Sigma_{iX'H^{-1}F}$. Concerning Assumption 5. \mathcal{H} , setting $\mathbf{C}_i = (\mathbf{A}_i + \Sigma_{D'D}^{-1}\Sigma_{D'F}\Gamma_i)$, one obtains

$$\Sigma_{iX'X}^{-1} = \begin{pmatrix} \Sigma_{D'D}^{-1} + \mathbf{C}_i(\Sigma_{V'V} - \Gamma'_i\Sigma_{(F'D)}\Gamma_i)^{-1}\mathbf{C}'_i & -\mathbf{C}_i(\Sigma_{V'V} - \Gamma'_i\Sigma_{(F'D)}\Gamma_i)^{-1} \\ -(\Sigma_{V'V} - \Gamma'_i\Sigma_{(F'D)}\Gamma_i)^{-1}\mathbf{C}'_i & (\Sigma_{V'V} - \Gamma'_i\Sigma_{(F'D)}\Gamma_i)^{-1} \end{pmatrix}$$

and

$$\mathbf{X}_i\Sigma_{iX'X}^{-1}\Sigma_{iX'F} = \mathbf{D}\Sigma_{D'D}^{-1}\Sigma_{D'F} + (\mathbf{D}\Sigma_{D'D}^{-1}\Sigma_{D'F}\Gamma_i - \mathbf{F}\Gamma_i - \mathbf{V}_i)(\Sigma_{V'V} - \Gamma'_i\Sigma_{(F'D)}\Gamma_i)^{-1}\Gamma'_i\Sigma_{(F'D)}.$$

Further manipulations yield

$$\begin{aligned} \mathcal{A}_{1T} &= \mathbf{D}\Sigma_{D'D}^{-1}\Sigma_{D'F}\mathcal{B}\Sigma_{F'D}\Sigma_{D'D}^{-1}\mathbf{D}' + \mathbf{D}\Sigma_{D'D}^{-1}\Sigma_{D'F}\mathcal{B}\Sigma_{(F'D)}\mathbf{P}_{1T}\Sigma_{F'D}\Sigma_{D'D}^{-1}\mathbf{D}' \\ &\quad - \mathbf{D}\Sigma_{D'D}^{-1}\Sigma_{D'F}\mathcal{B}\Sigma_{(F'D)}\mathbf{P}_{1T}\mathbf{F}' + \mathbf{D}\Sigma_{D'D}^{-1}\Sigma_{D'F}\mathbf{P}_{1T}\Sigma_{(F'D)}\mathcal{B}\Sigma_{F'D}\Sigma_{D'D}^{-1}\mathbf{D}' \\ &\quad + \mathbf{D}\Sigma_{D'D}^{-1}\Sigma_{D'F}\mathbf{P}_{2T}\Sigma_{F'D}\Sigma_{D'D}^{-1}\mathbf{D}' - \mathbf{D}\Sigma_{D'D}^{-1}\Sigma_{D'F}\mathbf{P}_{2T}\mathbf{F}' - \mathbf{F}\mathbf{P}_{1T}\Sigma_{(F'D)}\mathcal{B}\Sigma_{F'D}\Sigma_{D'D}^{-1}\mathbf{D}' \\ &\quad - \mathbf{F}\mathbf{P}_{2T}\Sigma_{F'D}\Sigma_{D'D}^{-1}\mathbf{D}' + \mathbf{F}\mathbf{P}_{2T}\mathbf{F}' + \mathbf{P}_{3T}, \end{aligned}$$

setting

$$\begin{aligned} N^{-1} \sum_{i=1}^N \Gamma_i(\Sigma_{V'V} - \Gamma'_i\Sigma_{(F'D)}\Gamma_i)^{-1}\Gamma'_i &\rightarrow_p \mathbf{P}_{1T}, \\ N^{-1} \sum_{i=1}^N (\Gamma_i(\Sigma_{vv'} - \Gamma'_i\Sigma_{(F'D)}\Gamma_i)^{-1}\Gamma'_i\Sigma_{(F'D)}\mathcal{B}\Sigma_{(F'D)}\Gamma_i(\Sigma_{V'V} - \Gamma'_i\Sigma_{(F'D)}\Gamma_i)^{-1}\Gamma'_i) &\rightarrow_p \mathbf{P}_{2T}, \\ N^{-1} \sum_{i=1}^N (\mathbf{V}_i(\Sigma_{V'V} - \Gamma'_i\Sigma_{(F'D)}\Gamma_i)^{-1}\Gamma'_i\Sigma_{(F'D)}\mathcal{B}\Sigma_{(F'D)}\Gamma_i(\Sigma_{V'V} - \Gamma'_i\Sigma_{(F'D)}\Gamma_i)^{-1}\mathbf{V}'_i) &\rightarrow_p \mathbf{P}_{3T}, \\ N^{-1} \sum_{i=1}^N \mathbf{V}_i(\Sigma_{V'V} - \Gamma'_i\Sigma_{(F'D)}\Gamma_i)^{-1}\mathbf{V}'_i &\rightarrow_p \mathbf{P}_{4T}. \end{aligned}$$

Likewise $\mathcal{A}_{2T} = \mathbf{D}\Sigma_{D'D}^{-1}\Sigma_{D'F}\mathcal{B}\mathbf{F}' + \mathbf{D}\Sigma_{D'D}^{-1}\Sigma_{D'F}\mathbf{P}_{1T}\Sigma_{(f'd')}\mathcal{B}\mathbf{F}' - \mathbf{F}\mathbf{P}_{1T}\Sigma_{(F'D)}\mathcal{B}\mathbf{F}'$.

Notice how the above expression are functionally independent from \mathbf{A}_i . Hence, whether \mathbf{X}_i^* is dependent or not from \mathbf{D} , is irrelevant for the sake of the derivation of \mathcal{A}_{1T} , \mathcal{A}_{2T} whose existence is implied, using a strong LLN argument, by Pesaran (2006, Assumptions 2 and 3). No additional moment conditions on the Γ_i are required since $\sup_{\Gamma_i} \|\Gamma_i(\Sigma_{vv'} - \Gamma'_i\Sigma_{(f'd')}\Gamma_i)^{-1}\Gamma'_i\| = O(1)$ *a.s.*

and the \mathbf{V}_i have bounded fourth moment by Pesaran (2006, Assumption 2). By simple manipulations

$$\begin{aligned} \mathbf{X}_i \boldsymbol{\Sigma}_{iX'X}^{-1} \mathbf{X}'_i &= \mathbf{D} \boldsymbol{\Sigma}_{dd'}^{-1} \mathbf{D}' + [(\mathbf{D} \boldsymbol{\Sigma}_{D'D}^{-1} \boldsymbol{\Sigma}_{D'F} - \mathbf{F}) \boldsymbol{\Gamma}_i - \mathbf{V}_i] (\boldsymbol{\Sigma}_{V'V} - \boldsymbol{\Gamma}'_i \boldsymbol{\Sigma}_{(F'D)} \boldsymbol{\Gamma}'_i)^{-1} \\ &\quad \times [\boldsymbol{\Gamma}'_i (\mathbf{D} \boldsymbol{\Sigma}_{D'D}^{-1} \boldsymbol{\Sigma}_{D'F} - \mathbf{F})' - \mathbf{V}'_i], \end{aligned}$$

where all the terms involving \mathbf{A}_i drop out. Although closed-form expressions for \mathcal{A}_{3T} , \mathcal{A}_{4T} , require to specify the parameterization of the \mathbf{H}_i , existence of the limit follows by Pesaran (2006, Assumptions 2 and 3). Conditions (9) and (10) follow by direct use of the CLT which holds under suitable regularity assumptions. For instance, when $E \|\mathbf{b}_i\|^{2+\delta} < \infty$ and $E |\varepsilon_{it}|^{2+\delta} < \infty$, some $\delta > 0$, and Pesaran (2006, Assumption 2) hold with in addition *i.i.d.*-ness of the ε_{it} across i , then a conditional version of the Lyapunov condition holds and the t -th column of \mathcal{C}_{1T} satisfies $\mathcal{C}_{1tT}^{\frac{1}{2}} \zeta_{mt}$ where the standard normal distributed $m \times 1$ vector ζ_{1t} is independent of \mathcal{C}_{1tT} . The latter is obtained as the probability limit of $N^{-1} \sum_{i,j=1}^N \mathbf{b}_i \varepsilon_{it} \varepsilon_{jt} \mathbf{b}'_j$ whose existence is implied by the previously made assumptions. Cross-sectional independence of the ε_{it} can be relaxed to a limited degree of dependence of the ε_{it} such that, in particular, $\mathbf{H}_t = [h_{ij,t}]_{i,j=1}^N = E(\varepsilon_t \varepsilon'_t \mid \mathbf{f}_t, \mathbf{X}_t, \mathbf{B})$ (see (30) below) has bounded maximum eigenvalue, that is $\sup_N \|\mathbf{H}_t\| = O(1)$ *a.s.* (see Pesaran and Tosetti (2007) for a general definition cross-sectional weak dependence). Likewise, under the same conditions, for the t th column of \mathcal{C}_{2T} one gets $\mathcal{C}_{2tT}^{\frac{1}{2}} \zeta_{Tt}$, where boundedness of \mathcal{C}_{2tT} requires $E \|\mathbf{D} + \mathbf{F} + \mathbf{V}_i\|^{2+\delta} < \infty$. Similar results apply to (10) where now the Lyapunov condition require, in addition, $E \|\mathbf{D} + \mathbf{F} + \mathbf{V}_i\|^{6+\delta} < \infty$.

Assumption 6. \mathcal{H} is not required by the GLS estimator and will not be considered here. Concerning Assumption 7. \mathcal{H} , (14) follows for

$$\begin{aligned} \mathcal{D}_{1T} &= \mathbf{P}_{2T} + \mathbf{P}_{1T} \boldsymbol{\Sigma}_{(F'D)} \boldsymbol{\mathcal{B}} + \boldsymbol{\mathcal{B}} \boldsymbol{\Sigma}_{(F'D)} \mathbf{P}_{1T}, \\ \mathcal{D}_{2T} &= -\mathbf{F} (\mathcal{C}_{1T} + \boldsymbol{\mathcal{B}}) \boldsymbol{\Sigma}_{F'D} \boldsymbol{\Sigma}_{D'D}^{-1} \mathbf{D}' - \mathbf{D} \boldsymbol{\Sigma}_{D'D}^{-1} \boldsymbol{\Sigma}_{D'F} (\mathcal{C}_{1T} + \boldsymbol{\mathcal{B}}) \mathbf{F}' \\ &\quad + \mathbf{D} \boldsymbol{\Sigma}_{D'D}^{-1} \boldsymbol{\Sigma}_{D'F} (\mathcal{C}_{1T} + \boldsymbol{\mathcal{B}}) \boldsymbol{\Sigma}_{F'D} \boldsymbol{\Sigma}_{D'D}^{-1} \mathbf{D}' + \mathbf{P}_{3T}. \end{aligned}$$

For \mathcal{A}_{3T} , \mathcal{A}_{4T} , as said, closed-form expressions required to parameterize \mathbf{H}_i so, for instance, assuming for simplicity $\mathbf{H}_i = \mathbf{I}_T$ yields

$$\begin{aligned} \mathcal{E}_{1T} &= -\mathbf{P}_{1T}, \\ \mathcal{E}_{2T} &= -\mathbf{D} (\boldsymbol{\Sigma}_{D'D}^{-1} + \boldsymbol{\Sigma}_{D'D}^{-1} \boldsymbol{\Sigma}_{df'} \mathbf{P}_{1T} \boldsymbol{\Sigma}_{F'D} \boldsymbol{\Sigma}_{D'D}^{-1}) \mathbf{D}' - \mathbf{P}_{4T} \\ &\quad + \mathbf{D} \boldsymbol{\Sigma}_{D'D}^{-1} \boldsymbol{\Sigma}_{D'F} \mathbf{P}_{1T} \mathbf{F}' + \mathbf{F} \mathbf{P}_{1T} \boldsymbol{\Sigma}_{F'D} \boldsymbol{\Sigma}_{D'D}^{-1} \mathbf{D}'. \end{aligned}$$

Since $\mathcal{I}_{1T} = \mathcal{D}_{1T} + T^{-1}\mathcal{E}_{1T} + \mathcal{B}$, one requires

$$\mathcal{B} + \mathbf{P}_{2T} + \mathbf{P}_{1T}\boldsymbol{\Sigma}_{(F'D)}\mathcal{B} + \mathcal{B}\boldsymbol{\Sigma}_{(F'D)}\mathbf{P}_{1T} - T^{-1}\mathbf{P}_{1T} \text{ non-singular.} \quad (25)$$

Given $\boldsymbol{\Sigma}_{(F'D)} = -\boldsymbol{\Sigma}_{F'F}$ and, by taking into consideration the definitions of $\mathbf{P}_{1T}, \mathbf{P}_{2T}, \mathcal{I}_{1T}$ can then be expressed as the limit of

$$\begin{aligned} & N^{-1} \sum_{i=1}^N \left(\boldsymbol{\Gamma}_i (\boldsymbol{\Sigma}_{V'V} + \boldsymbol{\Gamma}'_i \boldsymbol{\Sigma}_{F'F} \boldsymbol{\Gamma}_i)^{-1} \boldsymbol{\Gamma}'_i \boldsymbol{\Sigma}_{F'F} - \mathbf{I}_m \right) \mathcal{B} \left(\boldsymbol{\Gamma}_i (\boldsymbol{\Sigma}_{V'V} + \boldsymbol{\Gamma}'_i \boldsymbol{\Sigma}_{F'F} \boldsymbol{\Gamma}_i)^{-1} \boldsymbol{\Gamma}'_i \boldsymbol{\Sigma}_{F'F} - \mathbf{I}_m \right)' \\ & - T^{-1} N^{-1} \sum_{i=1}^N \boldsymbol{\Gamma}_i (\boldsymbol{\Sigma}_{V'V} + \boldsymbol{\Gamma}'_i \boldsymbol{\Sigma}_{F'F} \boldsymbol{\Gamma}_i)^{-1} \boldsymbol{\Gamma}'_i \end{aligned} \quad (26)$$

Hence, a sufficient condition for non-singularity of the limit of (26) is non-singularity of $(\boldsymbol{\Gamma}_i (\boldsymbol{\Sigma}_{V'V} + \boldsymbol{\Gamma}'_i \boldsymbol{\Sigma}_{F'F} \boldsymbol{\Gamma}_i)^{-1} \boldsymbol{\Gamma}'_i \boldsymbol{\Sigma}_{F'F} - \mathbf{I}_m)$ for any i but in fact a milder condition might suffice. Set, as an example, $\boldsymbol{\Sigma}_{F'F} = \mathcal{B} = \mathbf{I}_m$ and $\boldsymbol{\Sigma}_{V'V} = \mathbf{I}_k$. For $m > k = 1$, non-singularity of the limit of (26) is equivalent to obtain a non-singular limit of

$$N^{-1} \sum_{i=1}^N \left(\mathbf{I}_m - \frac{(2 + \boldsymbol{\Gamma}'_i \boldsymbol{\Gamma}_i)}{(1 + \boldsymbol{\Gamma}'_i \boldsymbol{\Gamma}_i)^2} \boldsymbol{\Gamma}_i \boldsymbol{\Gamma}'_i \right) - T^{-1} N^{-1} \sum_{i=1}^N \frac{\boldsymbol{\Gamma}_i \boldsymbol{\Gamma}'_i}{(1 + \boldsymbol{\Gamma}'_i \boldsymbol{\Gamma}_i)}$$

which can be obtained under mild conditions on the $\boldsymbol{\Gamma}_i$ since each $\left(\mathbf{I}_m - \frac{(2 + \boldsymbol{\Gamma}'_i \boldsymbol{\Gamma}_i)}{(1 + \boldsymbol{\Gamma}'_i \boldsymbol{\Gamma}_i)^2} \boldsymbol{\Gamma}_i \boldsymbol{\Gamma}'_i \right)$ is non-singular for all i . Instead, when $k > m = 1$ then a non-singular limit of (26) is equivalent to a non-zero limit of

$$N^{-1} \sum_{i=1}^N \left(1 - \boldsymbol{\Gamma}_i (\mathbf{I}_k + \boldsymbol{\Gamma}'_i \boldsymbol{\Gamma}_i)^{-1} \boldsymbol{\Gamma}'_i \right)^2 - T^{-1} N^{-1} \sum_{i=1}^N \boldsymbol{\Gamma}_i (\mathbf{I}_k + \boldsymbol{\Gamma}'_i \boldsymbol{\Gamma}_i)^{-1} \boldsymbol{\Gamma}'_i,$$

where it easily follows that each of the addenda is non-zero. Similar arguments follow for the case $m = k$. Finally, notice that $\boldsymbol{\Sigma}_{V'V} > \mathbf{0}$ is strictly required, ruling out the possibility that the regressors \mathbf{x}_{it} obey a pure factor structure $\mathbf{x}_{it} = \mathbf{A}'_i \mathbf{d}_t + \boldsymbol{\Gamma}'_i \mathbf{f}_t$, otherwise (8) fails.

For (15), given

$$\begin{aligned} \mathcal{I}_{2T} = & \mathcal{H}_T + \mathbf{P}_{3T} - T^{-1} (\mathbf{D} (\boldsymbol{\Sigma}_{D'D}^{-1} - \mathbf{I}_n) \mathbf{D}' + \mathbf{P}_{4t}) \\ & + (\mathbf{F} - \mathbf{D} \boldsymbol{\Sigma}_{D'D}^{-1} \boldsymbol{\Sigma}_{D'F}) \mathcal{I}_{1T} (\mathbf{F}' - \boldsymbol{\Sigma}_{F'D} \boldsymbol{\Sigma}_{D'D}^{-1} \mathbf{D}') - \mathbf{F} \mathcal{I}_{1T} \mathbf{F}', \end{aligned}$$

one needs

$$\Sigma_{F'D} = \mathbf{0} \quad (27)$$

for otherwise $T^{-1}\mathbf{F}'\mathcal{I}_{2T}^{-1}\mathbf{F} \rightarrow_p \mathbf{0}$ by Lemma 1($\mathbf{C}, \mathbf{F}, -\mathcal{I}_{1T}, T$), setting

$$\mathbf{C} = \mathcal{H}_T + \mathbf{P}_{3T} - T^{-1}(\mathbf{D}(\Sigma_{D'D}^{-1} - \mathbf{I}_n)\mathbf{D}' + \mathbf{P}_{4t}) + (\mathbf{F} - \mathbf{D}\Sigma_{D'D}^{-1}\Sigma_{D'F})\mathcal{I}_{1T}(\mathbf{F}' - \Sigma_{F'D}\Sigma_{D'D}^{-1}\mathbf{D}').$$

Sufficient conditions for (27) are $\mu_f = \mathbf{0}$ and $\{\mathbf{f}_t, \mathbf{d}_t\}$ contemporaneously uncorrelated. Uncorrelatedness follow simply when \mathbf{d}_t is deterministic, including intercept term, trends or seasonal dummies. Hence, under (27)

$$\mathcal{I}_{2T} = \mathcal{H}_T + \mathbf{P}_{3T} - T^{-1}(\mathbf{D}(\Sigma_{D'D}^{-1} - \mathbf{I}_n)\mathbf{D}' + \mathbf{P}_{4t}) > 0.$$

Closed-form expressions for $\Sigma_{iX'\mathcal{I}_2^{-1}X}$, $\Sigma_{iX'\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}X}$, $\Sigma_{iF'\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}F}$, $\Sigma_{iX'\mathcal{I}_2^{-1}F}$, $\Sigma_{iX'\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}F}$, required for the verification of the CLT condition (16), would depend on the adopted parameterization for the $h_{ts,i}$, and thus for the a_{il} of Pesaran (2006, Assumption 2). We conclude investigating the conditions required for (17)-(20). Under the assumptions made \mathcal{C}_{1T} is a random, mean zero, matrix of dimension $m \times T$, whose rows are uncorrelated with each ε_i , \mathbf{X}_i and with each row of Σ_T^{-1} . In addition, denoting by \mathcal{C}_{1Tj} the j th row of \mathcal{C}_{1T} , we will require $\sup_T \|E\mathcal{C}_{1Tj}'\mathcal{C}_{1Tj}\| = O(1)$ for all $1 \leq j \leq m$. The same assumptions are required for all the zero mean random matrices introduced below. Hence, by standard arguments, $\mathbf{X}_i'\Sigma_T^{-1}\mathcal{C}_{1T}' = O_p(T^{\frac{1}{2}}\iota_{n+k}\iota_m')$, $\mathcal{C}_{1T}\Sigma_T^{-1}\varepsilon_i = O_p(T^{\frac{1}{2}}\iota_m)$ and, by repeated use of Lemma 2, $\mathbf{F}'\Sigma_T^{-1}\varepsilon_i = O_p(T^{-\frac{1}{2}}\iota_m)$, $\mathbf{F}'\Sigma_T^{-1}\mathcal{C}_{1T}' = O_p(T^{-\frac{1}{2}}\iota_m\iota_m')$, $\mathbf{F}'\Sigma_T^{-1}\mathbf{X}_i = O_p(\iota_{n+k}\iota_m')$, $\mathbf{F}'\Sigma_T^{-1}\mathbf{F} = O_p(\iota_m\iota_m')$ yielding

$$\begin{aligned} \mathbf{X}_i'\Sigma_T^{-1}(\mathbf{F}\mathcal{C}_{1T} + \mathcal{C}_{1T}'\mathbf{F}')\Sigma_T^{-1}\mathbf{X}_i &= O_p(T^{\frac{1}{2}}\iota_{n+k}\iota_{n+k}'), \\ \mathbf{X}_i'\Sigma_T^{-1}(\mathbf{F}\mathcal{C}_{1T} + \mathcal{C}_{1T}'\mathbf{F}')\Sigma_T^{-1}(\mathbf{F}\mathbf{b}_i + \varepsilon_i) &= O_p(T^{\frac{1}{2}}\iota_{n+k}). \end{aligned}$$

Similarly, since under (27), $\mathbf{X}_i\Sigma_{iX'X}^{-1}\Sigma_{iX'F} = (\mathbf{F}\Gamma_i + \mathbf{V}_i)(\Sigma_{V'V} + \Gamma_i'\Sigma_{F'F}\Gamma_i)^{-1}\Gamma_i'\Sigma_{F'F}$, one gets $\mathcal{C}_{2T} = \mathbf{F}\mathcal{C}_{21T} + \mathcal{C}_{22T}$ for zero mean random matrices of dimension $m \times T$ and $T \times T$ respectively. The previous bounds apply substituting \mathcal{C}_{1T} with \mathcal{C}_{21T} and when, in addition, $\mathbf{X}_i'\Sigma_T^{-1}\mathcal{C}_{22T}'\Sigma_T^{-1}\mathbf{F} = O_p(\iota_{n+k}\iota_m')$, $\mathbf{X}_i'\Sigma_T^{-1}\mathcal{C}_{22T}'\Sigma_T^{-1}\varepsilon_i = O_p(T^{\frac{1}{2}}\iota_{n+k})$ then

$$\begin{aligned} \mathbf{X}_i'\Sigma_T^{-1}(\mathcal{C}_{2T} + \mathcal{C}_{2T}')\Sigma_T^{-1}\mathbf{X}_i &= O_p(T\iota_{n+k}\iota_{n+k}'), \\ \mathbf{X}_i'\Sigma_T^{-1}(\mathcal{C}_{2T} + \mathcal{C}_{2T}')\Sigma_T^{-1}(\mathbf{F}\mathbf{b}_i + \varepsilon_i) &= O_p(T^{\frac{1}{2}}\iota_{n+k}). \end{aligned}$$

Under (27)

$$\mathbf{X}_i \boldsymbol{\Sigma}_{iX'X}^{-1} \mathbf{X}'_i = \mathbf{D} \boldsymbol{\Sigma}_{D'D}^{-1} \mathbf{D}' + (\mathbf{F} \boldsymbol{\Gamma}_i + \mathbf{V}_i) (\boldsymbol{\Sigma}_{V'V} + \boldsymbol{\Gamma}'_i \boldsymbol{\Sigma}_{F'F} \boldsymbol{\Gamma}'_i)^{-1} (\boldsymbol{\Gamma}'_i \mathbf{F}' + \mathbf{V}'_i)$$

yielding $\mathcal{C}_{3T} = T^{-\frac{1}{2}} \mathcal{C}_{1T} \mathbf{D} \boldsymbol{\Sigma}_{D'D}^{-1} \mathbf{D}' + \mathcal{C}_{31T} \mathbf{F}' + \mathcal{C}_{32T}$ for zero mean random $m \times m$ matrix \mathcal{C}_{31T} and a $m \times T$ matrix \mathcal{C}_{32T} . Again, the previous bounds apply substituting \mathcal{C}_{1T} by \mathcal{C}_{32T} and $\mathbf{X}'_i \boldsymbol{\Sigma}_T^{-1} \mathbf{D} = O_p(T \iota_{n+k} \iota'_d)$, $\mathcal{C}_{1T} \mathbf{D} = O_p(T^{\frac{1}{2}} \iota_m \iota'_d)$ yielding

$$\begin{aligned} \mathbf{X}'_i \boldsymbol{\Sigma}_T^{-1} (\mathbf{F} \mathcal{C}_{3T} + \mathcal{C}'_{3T} \mathbf{F}') \boldsymbol{\Sigma}_T^{-1} \mathbf{X}_i &= O_p(T \iota_{n+k} \iota'_{n+k}), \\ \mathbf{X}'_i \boldsymbol{\Sigma}_T^{-1} (\mathbf{F} \mathcal{C}_{3T} + \mathcal{C}'_{3T} \mathbf{F}') \boldsymbol{\Sigma}_T^{-1} (\mathbf{F} \mathbf{b}_i + \varepsilon_i) &= O_p(T \iota_{n+k}). \end{aligned}$$

Finally $\mathcal{C}_{4T} = T^{-\frac{1}{2}} \mathcal{C}_{2T} \mathbf{D} \boldsymbol{\Sigma}_{dd'}^{-1} \mathbf{D}' + \mathbf{F} \mathcal{C}_{41T} \mathbf{F}' + \mathbf{F} \mathcal{C}_{42T} + \mathcal{C}_{43T} \mathbf{F}' + \mathcal{C}_{44T}$ for zero mean random $m \times m$ matrix \mathcal{C}_{41T} , $m \times T$ matrices $\mathcal{C}_{42T}, \mathcal{C}'_{43T}$ and $T \times T$ matrix \mathcal{C}_{44T} yielding

$$\begin{aligned} \mathbf{X}'_i \boldsymbol{\Sigma}_T^{-1} (\mathcal{C}_{4T} + \mathcal{C}'_{4T}) \boldsymbol{\Sigma}_T^{-1} \mathbf{X}_i &= O_p(T^{\frac{3}{2}} \iota_{n+k} \iota'_{n+k}), \\ \mathbf{X}'_i \boldsymbol{\Sigma}_T^{-1} (\mathcal{C}_{4T} + \mathcal{C}'_{4T}) \boldsymbol{\Sigma}_T^{-1} (\mathbf{F} \mathbf{b}_i + \varepsilon_i) &= O_p(T^{\frac{3}{2}} \iota_{n+k}). \end{aligned}$$

Hence, (17),(18),(19),(20) hold with $a = 1/2, b = 1, c = 3/2, d = 3/2$. No assumption of Pesaran (2006) would imply (25), (27) nor any of the other conditions in $7.\mathcal{H}$ but, as discussed, primitive conditions can be readily found.

We summarize the result of this section as follows:

Proposition 1 *Assume that Pesaran (2006, Assumptions 1, 2, 3 and 5a) hold and, in addition, $N^{-1} \sum_{i=1}^N \mathbf{H}_i \rightarrow_p \mathcal{H}_T > 0$ as $N \rightarrow \infty$, the $(n+m) \times 1$ vector $(\mathbf{d}'_t, \mathbf{f}'_t)'$ is fourth-order covariance stationarity with absolute summable autocovariances, bounded $(6+\delta)$ th moment and $\boldsymbol{\Sigma}_{D'D} > 0$, the \mathbf{b}_i have bounded $(2+\delta)$ th moment with $\mathcal{B} > 0$, the \mathbf{v}_{it} have bounded $(6+\delta)$ th moment and the ε_{it} have bounded $(2+\delta)$ th moment and are i.i.d. across i . Finally let Assumption $7.\mathcal{H}$ hold.*

Then Theorem 1,(iii) applies to the GLS estimator for $(\alpha'_0, \beta'_0)'$ of model (22)-(23)-(24) when

$$\frac{1}{T} + \frac{1}{N} \rightarrow 0$$

for consistency and

$$\frac{1}{T} + \frac{T}{N} \rightarrow 0$$

for asymptotic normality.

No other conditions of Pesaran (2006) is required, such as in particular the $m \times (k+1)$ matrix $E(\mathbf{b}_i \boldsymbol{\Gamma}_i)$ to be full row rank m .

The bias term of the OLS for β_{i0} is, from Theorem 1 (i), $\gamma_i^{OLS} = \Sigma_{iX'X}^{-1} \Sigma_{iX'F} \mathbf{b}_i$ which is zero only if $\mathbf{b}_i = \mathbf{0}$ *a.s.* (no factor structure in the regression error) or, alternatively, if $\Sigma_{iX'F} = \mathcal{O}_i = \mathbf{0}$ *a.s.* This latter condition requires both $\Gamma_i = \mathbf{0}$ *a.s.* and $\Sigma_{F'D} = \mathbf{0}$. The GLS estimator does not require $\Gamma_i = \mathbf{0}$ *a.s.* and thus allows the unit-specific regressors \mathbf{X}_i^* to be cross-correlated with the unobserved factors \mathbf{F} .

3 Time-Specific Parameters Model

This section mirrors the previous section but we prefer to present it in full, in order to avoid a the risk of confusion in notation.

3.1 Definitions and assumptions

Consider linear regression models with possibly time-specific parameters, such that for the t th time period

$$\mathbf{y}_t = \mathbf{X}_t \beta_{t0} + \mathbf{u}_t, \quad (28)$$

for an observed $N \times 1$ vector $\mathbf{y}_t = (y_{1t}, \dots, y_{it}, \dots, y_{Nt})'$ and an observed $N \times k$ matrix $\mathbf{X}_t = (\mathbf{x}_{1t}, \dots, \mathbf{x}_{it}, \dots, \mathbf{x}_{Nt})'$ related by a $k \times 1$ vector of possibly time-specific regression coefficients β_{t0} . The unobserved $N \times 1$ vector $\mathbf{u}_t = (u_{1t}, \dots, u_{it}, \dots, u_{Nt})'$ obeys the same factor structure described previously which, staking the u_{it} across units i , can be expressed as

$$\mathbf{u}_t = \mathbf{B} \mathbf{f}_t + \varepsilon_t.$$

As before, \mathbf{f}_t denotes an unobserved $m \times 1$ vector of factors, $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_N)'$ is an unobserved $N \times m$ matrix of factor loadings and $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$ is the unobserved $N \times 1$ vector of idiosyncratic innovations. Cross-sectional regressions with constant regression coefficients, such as Andrews (2005), or time-specific coefficients, are particular cases of (57).

A set of general assumptions required for the estimators here considered are introduced below, and commented subsequently. Theorem 2 states the asymptotic properties of the OLS, UGLS and GLS estimators for β_{t0} . Section 3.3 discusses a set of primitive conditions for our general assumptions with respect to a particular case of interest of model (57), namely Andrews (2005)'s model.

Assumption 1. \mathcal{T} (common factors)

For every t , the \mathbf{f}_t are random vector of dimension $m \times 1$ such that

$$E(\mathbf{f}_t \mathbf{f}_t' | \mathbf{X}_t, \mathbf{B}) = \mathcal{F}_t > 0,$$

$$T^{-1} \sum_{t=1}^T \mathcal{F}_t \rightarrow_p \mathcal{F} > 0 \text{ as } T \rightarrow \infty.$$

Assumption 2. \mathcal{T} (idiosyncratic innovation)

For every t , $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{it}, \dots, \varepsilon_{Nt})'$ satisfies

$$E(\varepsilon_t | \mathbf{f}_t, \mathbf{X}_t, \mathbf{B}) = \mathbf{0}, \quad (29)$$

$$\mathbf{H}_t = E(\varepsilon_t \varepsilon_t' | \mathbf{f}_t, \mathbf{X}_t, \mathbf{B}) > 0, \quad (30)$$

$$T^{-1} \sum_{t=1}^T \mathbf{H}_t \rightarrow_p \mathcal{H}_N > 0 \text{ as } T \rightarrow \infty. \quad (31)$$

Assumption 3. \mathcal{T} (regressors)

For every t , the $N \times k$ matrix \mathbf{X}_t is full column rank.

Assumption 4. \mathcal{T} (basic limit conditions)

All the limit matrices below, as $N \rightarrow \infty$, are *a.s.* finite:

$$\frac{\mathbf{X}_t' \mathbf{X}_t}{N} \rightarrow_p \Sigma_{tX'X} > 0, \quad \frac{\mathbf{X}_t' \mathbf{H}_t \mathbf{X}_t}{N} \rightarrow_p \Sigma_{tX'HX} > 0, \quad \frac{\mathbf{X}_t' \mathbf{B}}{N} \rightarrow_p \Sigma_{tX'B}, \quad (32)$$

$$\frac{\mathbf{X}_t' \mathbf{H}_t^{-1} \mathbf{X}_t}{N} \rightarrow_p \Sigma_{tX'H^{-1}X} > 0, \quad \frac{\mathbf{B}' \mathbf{H}_t^{-1} \mathbf{B}}{N} \rightarrow_p \Sigma_{tB'H^{-1}B} > 0, \quad \frac{\mathbf{X}_t' \mathbf{H}_t^{-1} \mathbf{B}}{N} \rightarrow_p \Sigma_{tX'H^{-1}B} \quad (33)$$

such that *a.s.*

$$\Sigma_{tX'H^{-1}X} - \Sigma_{tX'H^{-1}B} \Sigma_{tB'H^{-1}B}^{-1} \Sigma_{tX'H^{-1}B}' > 0. \quad (34)$$

Assumption 5. \mathcal{T} (limit conditions for GLS)

All the limit matrices below, as $T \rightarrow \infty$ and arbitrary N , are *a.s.* finite:

$$\begin{aligned} \sum_{t=1}^T \frac{\mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \boldsymbol{\Sigma}_{tX'B} \mathbf{f}_t \mathbf{f}_t' \boldsymbol{\Sigma}'_{tX'B} \boldsymbol{\Sigma}_{tX'X}^{-1} \mathbf{X}_t'}{T} &= \mathcal{A}_{1N}(1+o_p(1)), \quad \sum_{t=1}^T \frac{\mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \boldsymbol{\Sigma}_{tX'B} \mathbf{f}_t \mathbf{f}_t' \mathbf{B}'}{T} = \mathcal{A}_{2N}(1+o_p(1)), \\ \sum_{t=1}^T \frac{\mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \mathbf{X}_t' \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \mathbf{X}_t'}{NT} &= \mathcal{A}_{3N}(1+o_p(1)), \quad \sum_{t=1}^T \frac{\mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \mathbf{X}_t' \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'}{T} = \mathcal{A}_{4N}(1+o_p(1)), \\ \sum_{t=1}^T \frac{\mathbf{f}_t \boldsymbol{\varepsilon}_t'}{T^{\frac{1}{2}}} &= \mathcal{C}_{1N}(1+o_p(1)), \quad \sum_{t=1}^T \frac{\mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \boldsymbol{\Sigma}_{tX'B} \mathbf{f}_t \boldsymbol{\varepsilon}_t'}{T^{\frac{1}{2}}} = \mathcal{C}_{2N}(1+o_p(1)), \end{aligned} \quad (35)$$

$$\sum_{t=1}^T \frac{\mathbf{f}_t \boldsymbol{\varepsilon}_t' \mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \mathbf{X}_t'}{T^{\frac{1}{2}} N^{\frac{1}{2}}} = \mathcal{C}_{3N}(1+o_p(1)), \quad \sum_{t=1}^T \frac{\mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \boldsymbol{\Sigma}_{tX'B} \mathbf{f}_t \boldsymbol{\varepsilon}_t' \mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \mathbf{X}_t'}{T^{\frac{1}{2}} N^{\frac{1}{2}}} = \mathcal{C}_{4N}(1+o_p(1)). \quad (36)$$

Assumption 6.T (distribution conditions for OLS and UGLS)

As $N \rightarrow \infty$:

$$\frac{\mathbf{X}_t' \boldsymbol{\varepsilon}_t}{N^{\frac{1}{2}}} \rightarrow_d (\boldsymbol{\Sigma}_{tX'HX})^{\frac{1}{2}} \zeta_{kt}, \quad (37)$$

$$\frac{\mathbf{Z}_t' \mathbf{H}_t^{-1} \boldsymbol{\varepsilon}_t}{N^{\frac{1}{2}}} \rightarrow_d (\boldsymbol{\Sigma}_{tZ'H^{-1}Z})^{\frac{1}{2}} \zeta_{(k+m)t}, \quad (38)$$

setting

$$\mathbf{Z}_t = (\mathbf{X}_t, \mathbf{B}), \quad (39)$$

where $\zeta_{kt}, \zeta_{(k+m)t}$ are independent from $\boldsymbol{\Sigma}_{tX'HX}, \boldsymbol{\Sigma}_{tZ'H^{-1}Z}$ respectively.

Assumption 7.T (distribution and identification conditions for GLS)

Let $\mathcal{D}_{1N}, \mathcal{E}_{1N}$ be $m \times m$ matrices and $\mathcal{D}_{2N}, \mathcal{E}_{2N}$ be $N \times N$ matrices satisfying

$$\begin{aligned} \mathcal{A}_{1N} - (\mathcal{A}_{2N} + \mathcal{A}'_{2N}) &= \mathbf{B} \mathcal{D}_{1N} \mathbf{B}' + \mathcal{D}_{2N} \\ &\text{and} \\ \mathcal{A}_{3N} - (\mathcal{A}_{4N} + \mathcal{A}'_{4N}) &= \mathbf{B} \mathcal{E}_{1N} \mathbf{B}' + \mathcal{E}_{2N}. \end{aligned} \quad (40)$$

Then

$$\begin{aligned} \mathcal{I}_{1N} &= \mathcal{D}_{1N} + N^{-1} \mathcal{E}_{1N} + \mathcal{F}(\text{non-singular symmetric } m \times m \text{ matrix}), \\ \mathcal{I}_{2N} &= \mathcal{D}_{2N} + N^{-1} \mathcal{E}_{2N} + \mathcal{H}_N(\text{non-singular } T \times T \text{ matrix}), \end{aligned}$$

satisfy

$$\begin{aligned}
\frac{\mathbf{B}'\mathcal{I}_{2N}^{-1}\mathbf{B}}{N} &\rightarrow_p \Sigma_{\mathbf{B}'\mathcal{I}_2^{-1}\mathbf{B}}(\text{non-singular}), \quad \frac{\mathbf{B}'\mathcal{I}_{2N}^{-1}\mathbf{H}_t\mathcal{I}_{2N}^{-1}\mathbf{B}}{N} \rightarrow_p \Sigma_{t\mathbf{B}'\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}\mathbf{B}}, \quad (41) \\
\frac{\mathbf{X}'_t\mathcal{I}_{2N}^{-1}\mathbf{X}_t}{N} &\rightarrow_p \Sigma_{t\mathbf{X}'\mathcal{I}_2^{-1}\mathbf{X}}(\text{non-singular}), \quad \frac{\mathbf{X}'_t\mathcal{I}_{2N}^{-1}\mathbf{H}_t\mathcal{I}_{2N}^{-1}\mathbf{X}_t}{N} \rightarrow_p \Sigma_{t\mathbf{X}'\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}\mathbf{X}}, \\
\frac{\mathbf{X}'_t\mathcal{I}_{2N}^{-1}\mathbf{B}}{N} &\rightarrow_p \Sigma_{t\mathbf{X}'\mathcal{I}_2^{-1}\mathbf{B}}, \quad \frac{\mathbf{X}'_t\mathcal{I}_{2N}^{-1}\mathbf{H}_t\mathcal{I}_{2N}^{-1}\mathbf{B}}{N} \rightarrow_p \Sigma_{t\mathbf{X}'\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}\mathbf{B}}, \\
\frac{\mathbf{Z}'_t\mathcal{I}_{2N}^{-1}\varepsilon_t}{N^{\frac{1}{2}}} &\rightarrow_d (\Sigma_{t\mathbf{Z}'\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}\mathbf{Z}})^{\frac{1}{2}}\zeta_{(k+m)t}, \quad (42)
\end{aligned}$$

with $\zeta_{(k+m)t}$ independent from $\Sigma_{t\mathbf{Z}'\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}\mathbf{Z}}$, where \mathbf{Z}_t is defined in (39). All the limits above hold as $N \rightarrow \infty$ with *a.s.* finite limit matrices. Setting

$$\Sigma_N = \mathbf{B}\mathcal{I}_{1N}\mathbf{B}' + \mathcal{I}_{2N},$$

for all t , and some constants $a, b, c, d > 0$:

$$\mathbf{X}'_t\Sigma_N^{-1}(\mathbf{B}\mathcal{C}_{1N} + \mathcal{C}'_{1N}\mathbf{B}' + \mathcal{C}_{2N} + \mathcal{C}'_{2N})\Sigma_N^{-1}(\mathbf{B}\mathbf{f}_t + \varepsilon_t) = O_p(N^a l_k), \quad (43)$$

$$\mathbf{X}'_t\Sigma_N^{-1}(\mathbf{B}\mathcal{C}_{3N} + \mathcal{C}'_{3N}\mathbf{B}' + \mathcal{C}_{4N} + \mathcal{C}'_{4N})\Sigma_N^{-1}(\mathbf{B}\mathbf{f}_t + \varepsilon_t) = O_p(N^b l_k), \quad (44)$$

$$\mathbf{X}'_t\Sigma_N^{-1}(\mathbf{B}\mathcal{C}_{1N} + \mathcal{C}'_{1N}\mathbf{B}' + \mathcal{C}_{2N} + \mathcal{C}'_{2N})\Sigma_N^{-1}\mathbf{X}_t = O_p(N^c l_k l'_k), \quad (45)$$

$$\mathbf{X}'_t\Sigma_N^{-1}(\mathbf{B}\mathcal{C}_{3N} + \mathcal{C}'_{3N}\mathbf{B}' + \mathcal{C}_{4N} + \mathcal{C}'_{4N})\Sigma_N^{-1}\mathbf{X}_t = O_p(N^d l_k l'_k). \quad (46)$$

Remark: The comments made to Assumptions 1. \mathcal{H} -7. \mathcal{H} apply now but replacing $T, \mathbf{F}, \mathbf{b}_i, \mathbf{X}_i, \varepsilon_i, \mathbf{H}_i, \mathbf{u}_i$ with $N, \mathbf{B}, \mathbf{f}_t, \mathbf{X}_t, \varepsilon_t, \mathbf{H}_t, \mathbf{u}_t$, respectively.

3.2 Estimators results

The ordinary least squares (OLS) estimator is

$$\hat{\beta}_t^{OLS} = (\mathbf{X}'_t\mathbf{X}_t)^{-1}\mathbf{X}'_t\mathbf{y}_t,$$

The *unfeasible* generalized least squares estimator (UGLS) is

$$\hat{\beta}_t^{UGLS} = (\mathbf{X}'_t\mathcal{S}_t^{-1}\mathbf{X}_t)^{-1}\mathbf{X}'_t\mathcal{S}_t^{-1}\mathbf{y}_t,$$

setting

$$\mathcal{S}_t = \mathbf{B}\mathcal{F}_t\mathbf{B}' + \mathbf{H}_t.$$

The *feasible* generalized least squares estimator (GLS) estimator estimator is

$$\hat{\beta}_t^{GLS} = (\mathbf{X}_t' \hat{\Sigma}_N^{-1} \mathbf{X}_t)^{-1} \mathbf{X}_t' \hat{\Sigma}_N^{-1} \mathbf{y}_t,$$

setting

$$\hat{\Sigma}_N = T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}_t', \quad \hat{u}_t = \mathbf{y}_t - \mathbf{X}_t \hat{\beta}_t^{OLS},$$

which requires $T \geq N$. Again, if $\mathbf{X}_t = (\mathbf{D}, \mathbf{X}_t^*)$ for a $N \times k_1$ matrix \mathbf{D} and a $N \times k_2$ matrix \mathbf{X}_t^* , where $k = k_1 + k_2$, such as when an intercept term is allowed for, then in order to allow non-singularity we consider

$$\hat{\Sigma}_N = T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}_t' + N^{-1} \mathbf{D} \mathbf{D}'.$$

Theorem 2 (*time-specific parameters*)

(i) (OLS) When Assumptions 3.T, 4.T.(32), 6.T.(37) hold

$$N^{\frac{1}{2}}(\hat{\beta}_t^{OLS} - \beta_{t0} - \gamma_t^{OLS}) \rightarrow_d (\mathcal{V}_t^{OLS})^{\frac{1}{2}} \zeta_{kt} \text{ as } N \rightarrow \infty,$$

setting $\gamma_t^{OLS} = \Sigma_{tX'X}^{-1} \Sigma_{tX'B} \mathbf{f}_t$, $\mathcal{V}_t^{OLS} = \Sigma_{tX'X}^{-1} \Sigma_{tX'HX}^{-1} \Sigma_{tX'X}^{-1}$, where ζ_{kt} and \mathcal{V}_t^{OLS} are mutually independent.

(ii) (UGLS) When Assumptions 1.T, 2.T.(30), 3.T, 4.T.(33), 6.T.(38)

$$N^{\frac{1}{2}}(\hat{\beta}_t^{UGLS} - \beta_{t0}) \rightarrow_d \mathcal{N}_k(\mathcal{V}_t^{UGLS})^{\frac{1}{2}} \zeta_{kt} \text{ as } N \rightarrow \infty,$$

where ζ_{kt} and \mathcal{V}_t^{UGLS} are mutually independent, setting $\mathcal{V}_t^{UGLS} = (\mathcal{M}_t^{UGLS})^{-1} \mathcal{N}_t^{UGLS} (\mathcal{M}_t^{UGLS})^{-1}$ with $\mathcal{M}_t^{UGLS} = \text{plim}_{N \rightarrow \infty} N^{-1} (\mathbf{X}_t' \mathcal{S}_t^{-1} \mathbf{X}_t)$, $\mathcal{N}_t^{UGLS} = \text{plim}_{N \rightarrow \infty} N^{-1} \mathbf{X}_t' \mathcal{S}_t^{-1} \mathbf{H}_t \mathcal{S}_t^{-1} \mathbf{X}_t$. Moreover $(\mathcal{M}_t^{UGLS}) = \mathcal{N}_t^{UGLS}$.

(iii) (GLS)

When Assumptions 1.T, 2.T.(31), 3.T, 4.T.(32) and (34), 5.T, 7.T

$$\begin{aligned} \hat{\beta}_t^{GLS} \rightarrow_p \beta_{t0} & \quad \text{as } \frac{1}{N} + \frac{N^{\max(a-1, b-\frac{3}{2})}}{T^{\frac{1}{2}}} + \frac{N^{\max(c-\frac{3}{2}, d-2)}}{T^{\frac{1}{2}}} \rightarrow 0, \\ N^{\frac{1}{2}}(\hat{\beta}_t^{GLS} - \beta_{t0}) \rightarrow_d (\mathcal{V}_t^{GLS})^{\frac{1}{2}} \zeta_{kt} & \quad \text{as } \frac{1}{N} + \frac{N^{\max(a-\frac{1}{2}, b-1)}}{T^{\frac{1}{2}}} + \frac{N^{\max(c-1, d-\frac{3}{2})}}{T^{\frac{1}{2}}} \rightarrow 0, \end{aligned}$$

where ζ_{kt} and \mathcal{V}_t^{GLS} are mutually independent, setting $\mathcal{V}_t^{GLS} = (\mathcal{M}_t^{GLS})^{-1} \mathcal{N}_t^{GLS} (\mathcal{M}_t^{GLS})^{-1}$ with $\mathcal{M}_t^{GLS} = \text{plim}_{(N,T) \rightarrow \infty} N^{-1} (\mathbf{X}_t' \hat{\Sigma}_N^{-1} \mathbf{X}_t)$, $\mathcal{N}_t^{GLS} = \text{plim}_{(N,T) \rightarrow \infty} N^{-1} \mathbf{X}_t' \hat{\Sigma}_N^{-1} \mathbf{H}_t \hat{\Sigma}_N^{-1} \mathbf{X}_t$.

Remarks 1. Most of the comments to Theorem 1 apply here and will not be repeated. Now $\beta_t^{OLS} = \beta_{t0} + \gamma_t^{OLS} + O_p(N^{-\frac{1}{2}})$ where $\gamma_t^{OLS} = \Sigma_{tX'X}^{-1} \Sigma_{tX'B}^{-1} \mathbf{f}_t$ which is a random variable. Consistency is achieved if either $\mathbf{f}_t = \mathbf{0}$ (no factors) or $\Sigma_{tX'B}^{-1} = \mathbf{0}$, implied by no cross-correlation between the regressors and the factors (assuming the latter have mean zero or, alternatively, when \mathbf{X}_t contains an vector of ones).

2. The GLS uses $\hat{\Sigma}_N$ which, although a non-consistent estimate of the true covariance matrix $\mathbf{S}_t = \mathbf{B}\mathcal{F}_t\mathbf{B}' + \mathbf{H}_t$ (in the sense of element by element), has limit $\Sigma_N = \mathbf{B}\mathcal{I}_{1N}\mathbf{B}' + \mathcal{I}_{2N}$ which, once taking the inverse, belongs to the space orthogonal to the factor loadings \mathbf{B} under suitable regularity conditions.

3.3 Particular model: Andrews (2005)

The model is

$$y_{it} = \beta'_{0t}(\mathbf{1} \mathbf{x}'_{it})' + u_{it}, \quad (47)$$

where $(y_{it}, \mathbf{x}_{it})$ are assumed *i.i.d.* across units conditional on $\mathbf{c}_{1t}, \mathbf{C}_{2t}$ by Andrews (2005, Assumption 1), with

$$u_{it} = \mathbf{c}'_{1t} \mathbf{u}_i^* + \varepsilon_{it}, \quad (48)$$

$$\mathbf{x}_{it} = \mathbf{C}_{2t} \mathbf{x}_i^* + \mathbf{v}_{it}, \quad (49)$$

with $\mathbf{c}_{1t}, \mathbf{u}_i^*$ are $d_1 \times 1$ random vectors and $\mathbf{C}_{2t}, \mathbf{x}_i^*$ respectively a random matrix of dimension $k \times d_2$, with $d_2 \geq k$, and a random vector of dimension $d_2 \times 1$ and ε_{it} and \mathbf{v}_{it} are *i.i.d.* innovations across i and t , respectively scalar and $k \times 1$, with zero mean and variances $h_{ii,t}$ and $\Sigma_{tV'V}$. We focus on Andrews (2005)'s *standard factor* structure, spelled out in his Assumption SF1, here slightly extended to allow for an idiosyncratic component in both the regression error u_{it} and the regressors \mathbf{x}_{it} as well as time-variation in parameters, common factors and covariance matrices. The first extension is compulsory since when $\varepsilon_{it} = 0$ *a.s.* our theory does not apply. Let us start assuming $\Sigma_{tV'V} = \mathbf{0}$ implying $\mathbf{v}_{it} = 0$ *a.s.* as in Andrews (2005), although this is not necessary for our arguments to go through. (48)-(49) imply $\mathbf{F} = (\mathbf{c}_{11} \dots \mathbf{c}_{1t} \dots \mathbf{c}_{1T})'$, $\mathbf{B} = (\mathbf{u}_1^* \dots \mathbf{u}_N^*)'$, $\mathbf{X}_t = (\iota_N, \mathbf{X}^* \mathbf{C}'_{2t})$, where we set $\mathbf{X}^* = (\mathbf{x}_1^* \dots \mathbf{x}_i^* \dots \mathbf{x}_N^*)'$. We do not consider here Andrew's other, more general, forms cross-sectional dependence named heterogeneous and functional factor structures.

We now verify the extent to which the assumptions of Andrews (2005) imply our Theorem 2, part (iii). We are able to relax some of his assumptions,

although here more conditions need to be made here with respect to the time-series properties of $\mathbf{c}_{1t}, \mathbf{C}_{2t}, \varepsilon_t$, unnecessary to Andrews (2005) since all his results hold *conditional* on the σ -field \mathcal{C} induced by $\{\mathbf{c}_{1t}, \mathbf{C}_{2t}\}$.

Assumptions 1. \mathcal{T} and 2. \mathcal{T} do not follow from any of Andrews (2005)'s assumptions although, when imposing *iidness* conditional on \mathcal{C} , $\mathbf{H}_t = \sigma_t^2 \mathbf{I}_N$. We also require $\text{plim} T^{-1} \sum_{t=1}^T \sigma_t^2 = \sigma^2 > 0$ yielding $\mathcal{H}_N = \sigma \mathbf{I}_N$. Assumption 3. \mathcal{T} is implied by Andrews (2005, Assumption 2d). Concerning Assumption 4. \mathcal{T}

$$\begin{aligned} \mathbf{X}'_t \mathbf{X}_t &= \begin{pmatrix} N & \sum_{i=1}^N \mathbf{x}_i^{*'} \mathbf{C}'_{2t} \\ \mathbf{C}_{2t} \sum_{i=1}^N \mathbf{x}_i^* & \mathbf{C}_{2t} \sum_{i=1}^N \mathbf{x}_i^* \mathbf{x}_i^{*'} \mathbf{C}'_{2t} \end{pmatrix}, \quad \mathbf{X}'_t \mathbf{H}_t \mathbf{X}_t = \sigma_t^2 \mathbf{X}'_t \mathbf{X}_t, \\ \mathbf{X}'_t \mathbf{H}_t^{-1} \mathbf{X}_t &= \sigma_t^{-2} \mathbf{X}'_t \mathbf{X}_t, \quad \mathbf{B}' \mathbf{H}_t^{-1} \mathbf{B} = \sigma_t^{-2} \sum_{i=1}^N \mathbf{u}_i^* \mathbf{u}_i^{*'}, \\ \mathbf{X}'_t \mathbf{B} &= \begin{pmatrix} \sum_{i=1}^N \mathbf{u}_i^{*'} \\ \mathbf{C}_{2t} \sum_{i=1}^N \mathbf{x}_i^* \mathbf{u}_i^{*'} \end{pmatrix}, \quad \mathbf{X}'_t \mathbf{H}_t^{-1} \mathbf{B} = \sigma_t^{-2} \mathbf{X}'_t \mathbf{B}. \end{aligned}$$

and the limits are well defined by Andrews (2005, Assumptions 1, 2 and 3(a)). Then

$$\Sigma_{tX'X} = \begin{pmatrix} 1 & \mu_x' \mathbf{C}_{2t}' \\ \mathbf{C}_{2t} \mu_x & \mathbf{C}_{2t} \Sigma_{xx'} \mathbf{C}_{2t}' \end{pmatrix} > 0, \quad \Sigma_{tX'B} = \begin{pmatrix} \mu_u' \\ \mathbf{C}_{2t} \Sigma_{xu'} \end{pmatrix},$$

with non-singularity ensured by Andrews (2005, Assumption 2(d)), where throughout this section $\mu_x = E \mathbf{x}_i^*$, $\Sigma_{xx'} = E \mathbf{x}_i^* \mathbf{x}_i^{*'}$, $\Sigma_{uu'} = E \mathbf{u}_i^* \mathbf{u}_i^{*'}$, $\mu_u = E \mathbf{u}_i^*$, $\Sigma_{xu'} = E \mathbf{x}_i^* \mathbf{u}_i^{*'}$. Using the block matrix decomposition (Magnus and Neudecker 1988) one gets

$$\Sigma_{tX'X}^{-1} = \begin{pmatrix} q_t^{-1} & -q_t^{-1} \mu_x' \mathbf{C}_{2t}' (\mathbf{C}_{2t} \Sigma_{xx'} \mathbf{C}_{2t}')^{-1} \\ -q_t^{-1} (\mathbf{C}_{2t} \Sigma_{xx'} \mathbf{C}_{2t}')^{-1} \mathbf{C}_{2t} \mu_x & (\mathbf{C}_{2t} \Sigma_{xx'} \mathbf{C}_{2t}')^{-1} (\mathbf{I}_k + q_t^{-1} \mathbf{C}_{2t} \mu_x \mu_x' \mathbf{C}_{2t}' (\mathbf{C}_{2t} \Sigma_{xx'} \mathbf{C}_{2t}')^{-1}) \end{pmatrix},$$

where $q_t = 1 - \mu_x' \mathbf{C}_{2t}' (\mathbf{C}_{2t} \Sigma_{xx'} \mathbf{C}_{2t}')^{-1} \mathbf{C}_{2t} \mu_x > 0$, *a.s.* since the $\mathbf{x}_{it} = \mathbf{C}_{2t} \mathbf{x}_i^*$ have an non-singular distribution. In fact, for any non-degenerate random vector \mathbf{x} one gets $E \mathbf{x} \mathbf{x}' > E \mathbf{x} E \mathbf{x}'$, equivalent to $E \mathbf{x}' (E \mathbf{x} \mathbf{x}')^{-1} E \mathbf{x} < 1$. Formulae for $\Sigma_{tX'HX}$, $\Sigma_{tX'H^{-1}X}$, $\Sigma_{tB'H^{-1}B}$, $\Sigma_{tX'H^{-1}B}$ easily follows. Concerning

Assumption 5. \mathcal{T} , the expression for \mathcal{A}_{1N} , \mathcal{A}_{2N} follows by using

$$\begin{aligned} \mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \boldsymbol{\Sigma}_{tX'B} &= [\mathbf{X}^* \mathbf{C}_{2t}' (\mathbf{C}_{2t} \boldsymbol{\Sigma}_{xx'} \mathbf{C}_{2t}')^{-1} \mathbf{C}_{2t} \boldsymbol{\Sigma}_{xu'} \\ &+ q_t^{-1} (\iota_N - \mathbf{X}^* \mathbf{C}_{2t}' (\mathbf{C}_{2t} \boldsymbol{\Sigma}_{xx'} \mathbf{C}_{2t}')^{-1} \mathbf{C}_{2t} \mu_x) (\mu'_u - \mu'_x \mathbf{C}_{2t}' (\mathbf{C}_{2t} \boldsymbol{\Sigma}_{xx'} \mathbf{C}_{2t}')^{-1} \mathbf{C}_{2t} \boldsymbol{\Sigma}_{xu'})]. \end{aligned}$$

By Greville (1965), setting \mathbf{A}^+ to be the Moore-Penrose of a matrix \mathbf{A} ,

$$(\mathbf{C}_{2t} \boldsymbol{\Sigma}_{xx'} \mathbf{C}_{2t}')^{-1} = (\mathbf{C}_{2t}')^+ \boldsymbol{\Sigma}_{xx'}^{-1} (\mathbf{C}_{2t})^+ = (\mathbf{C}_{2t} \mathbf{C}_{2t}')^{-1} \mathbf{C}_{2t} \boldsymbol{\Sigma}_{xx'}^{-1} \mathbf{C}_{2t}' (\mathbf{C}_{2t} \mathbf{C}_{2t}')^{-1},$$

where the last equality follows since \mathbf{C}_{2t} is assumed full row rank. Substituting the last expression into $\mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \boldsymbol{\Sigma}_{tX'B} \mathcal{F}_t \boldsymbol{\Sigma}'_{tX'B} \boldsymbol{\Sigma}_{tX'X}^{-1} \mathbf{X}'_t$ yields many terms like $\mathbf{C}_{2t}' (\mathbf{C}_{2t} \mathbf{C}_{2t}')^{-1} \mathbf{C}_{2t}$. Now the matrix $\mathbf{M}_{2t} = \mathbf{I}_k - \mathbf{C}_{2t}' (\mathbf{C}_{2t} \mathbf{C}_{2t}')^{-1} \mathbf{C}_{2t}$ is idempotent positive semi-definite, implying

$$\mathbf{C}_{2t}' (\mathbf{C}_{2t} \mathbf{C}_{2t}')^{-1} \mathbf{C}_{2t} = \mathbf{I}_k - \mathbf{M}_{2t} \leq \mathbf{I}_k. \quad (50)$$

This implies $\| \mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \boldsymbol{\Sigma}_{tX'B} \mathcal{F}_t \boldsymbol{\Sigma}'_{tX'B} \boldsymbol{\Sigma}_{tX'X}^{-1} \mathbf{X}'_t \| = O(\| \mathcal{F}_t \|)$ *a.s.* Similarly, for \mathcal{A}_{2N} and \mathcal{A}_{3N} one uses $\| \mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \boldsymbol{\Sigma}_{tX'B} \mathcal{F}_t \| = O(\| \mathcal{F}_t \|)$ *a.s.* and $\| \mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \boldsymbol{\Sigma}_{tX'HX} \boldsymbol{\Sigma}_{tX'X}^{-1} \mathbf{X}'_t \| = O(\sigma_t^2)$ *a.s.*, respectively. Notice that $\mathcal{A}_{4N} = \mathcal{A}_{3N}$ since \mathbf{H}_t is diagonal, where

$$\begin{aligned} \mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \mathbf{X}'_t &= \\ q_t^{-1} \iota_N \iota'_N - q_t^{-1} \mathbf{X}^* \mathbf{C}'_{2t} (\mathbf{C}_{2t} \boldsymbol{\Sigma}_{xx'} \mathbf{C}'_{2t})^{-1} \mathbf{C}_{2t} \mu_x \iota'_N - q_t^{-1} \iota_N \mu'_x \mathbf{C}'_{2t} (\mathbf{C}_{2t} \boldsymbol{\Sigma}_{xx'} \mathbf{C}'_{2t})^{-1} \mathbf{C}_{2t} \mathbf{X}^* \\ &+ \mathbf{X}^* \mathbf{C}'_{2t} (\mathbf{C}_{2t} \boldsymbol{\Sigma}_{xx'} \mathbf{C}'_{2t})^{-1} [\mathbf{I}_k + q_t^{-1} \mathbf{C}_{2t} \mu_x \mu'_x \mathbf{C}'_{2t} (\mathbf{C}_{2t} \boldsymbol{\Sigma}_{xx'} \mathbf{C}'_{2t})^{-1}] \mathbf{C}_{2t} \mathbf{X}^*. \end{aligned}$$

Hence boundedness does not require any moment conditions in \mathbf{C}_{2t} .

For (35) and (36), they follow by using the martingale CLT (see Brown (1971)) if we assume that, for each i , the ε_{it} can be written as linear processes of a martingale difference sequence with absolute summable coefficients. In turn, this is implied when $\mathbf{H}_i = [h_{ts,i}]_{t,s=1}^T$ (defined in (4)) has bounded maximum eigenvalue. Then for the i th column of \mathcal{C}_{1N} one gets $T^{-\frac{1}{2}} \sum_{t=1}^T \mathbf{f}_t \varepsilon_{it} = \mathbf{C}_{1iN}^{\frac{1}{2}} \zeta_{mi} (1 + o_p(1))$ when $T^{-1} \sum_{t,s=1}^T \mathbf{f}_t h_{ts,i} \mathbf{f}'_t \rightarrow_p \mathbf{C}_{1iN}$ and $E \|\mathbf{f}_t\|^{2+\delta} < \infty$, $E|\varepsilon_{it}|^{2+\delta} < \infty$. Likewise, the i th column of \mathcal{C}_{2N} can be written as $\mathbf{C}_{2iN}^{\frac{1}{2}} \zeta_{Ni}$ where no additional moment conditions are required because of (50). The same results apply for (36). Unless $\boldsymbol{\Sigma}_{tV'V} = \mathbf{0}$, one also needs $E \|\mathbf{v}_{i,t}\|^{6+\delta} < \infty$.

Notice that when $d_2 = k$, \mathbf{C}_{2t} is a square full rank matrix and both $\mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \boldsymbol{\Sigma}_{tX'B}$ and $\mathbf{X}_t \boldsymbol{\Sigma}_{tX'X}^{-1} \mathbf{X}'_t$ are not time-varying, simplifying the above results.

We do not verify Assumption 6. \mathcal{T} since not required by the GLS estimator. For Assumption 7. \mathcal{T} , in particular (40), set

$$\begin{aligned}\mathbf{D} &= \mathbf{X}^* \boldsymbol{\Sigma}_{xx'}^{-1} \boldsymbol{\Sigma}_{xu'} + q^{-1} (\iota_N - \mathbf{X}^* \boldsymbol{\Sigma}_{xx'}^{-1} \boldsymbol{\mu}_x) (\boldsymbol{\mu}'_u - \boldsymbol{\mu}'_x \boldsymbol{\Sigma}_{xx'}^{-1} \boldsymbol{\Sigma}_{xu'}), \\ \mathbf{E} &= \sigma^2 q^{-1} (\iota_N \iota'_N - \mathbf{X}^* \boldsymbol{\Sigma}_{xx'}^{-1} \boldsymbol{\mu}_x \iota'_N - \iota_N \boldsymbol{\mu}'_x \boldsymbol{\Sigma}_{xx'}^{-1} \mathbf{X}^{*'} + \mathbf{X}^* \boldsymbol{\Sigma}_{xx'}^{-1} [q \mathbf{I}_k + \boldsymbol{\mu}_x \boldsymbol{\mu}'_x \boldsymbol{\Sigma}_{xx'}^{-1}] \mathbf{X}^{*'}), \\ q &= 1 - \boldsymbol{\mu}'_x \boldsymbol{\Sigma}_{xx'}^{-1} \boldsymbol{\mu}_x.\end{aligned}$$

It follows that

$$\begin{aligned}\mathcal{A}_{1N} - \mathcal{A}_{2N} - \mathcal{A}'_{2N} &= (\mathbf{D} - \mathbf{B}) \mathcal{F} (\mathbf{D} - \mathbf{B})' - \mathbf{B} \mathcal{F} \mathbf{B}', \\ \mathcal{A}_{3N} - \mathcal{A}_{4N} - \mathcal{A}'_{4N} &= -\mathbf{E}, \\ \boldsymbol{\Sigma}_{xu'} &= \mathbf{0}, \quad \boldsymbol{\mu}_u = \mathbf{0}\end{aligned}\tag{51}$$

which in turn is implied when \mathbf{x}_i^* and \mathbf{u}_i^* are uncorrelated with $\boldsymbol{\mu}_u = \mathbf{0}$.

When $k < d_2$ which implies full row rank \mathbf{C}_{2t} , then obviously (40) is satisfied when (51) hold. However, now it is also possible that \mathbf{x}_i^* and \mathbf{u}_i^* are correlated, for instance even perfectly correlated such as

$$\mathbf{x}_i^* = \mathbf{A} \mathbf{u}_i^*, \quad \boldsymbol{\mu}_u = \mathbf{0}, \quad \boldsymbol{\Sigma}_{tV'V} > 0,\tag{52}$$

for a non-random full row rank matrix \mathbf{A} . As an example, set for simplicity $d_2 = d_1 > k$, $\mathbf{A} = \mathbf{I}_{d_1}$ yielding $\boldsymbol{\Sigma}_{xx'} = \boldsymbol{\Sigma}_{xu'} = \mathbf{I}_{d_1}$. Notice that now we require $\boldsymbol{\Sigma}_{tV'V} = E \mathbf{v}_{it} \mathbf{v}'_{it} > 0$, to ensure (34) holds, and also \mathbf{v}_{it} and u_i^* to be mutually independent. The previous derivations still apply by replacing $(\mathbf{C}_{2t} \boldsymbol{\Sigma}_{xx'} \mathbf{C}'_{2t})^{-1}$ by $(\mathbf{C}_{2t} \boldsymbol{\Sigma}_{xx'} \mathbf{C}'_{2t} + \boldsymbol{\Sigma}_{tV'V})^{-1}$ and $\mathbf{X}_t = (\iota_N, \mathbf{X}^* \mathbf{C}'_{2t} + \mathbf{V}_t)$. Then

$$\begin{aligned}T^{-1} \sum_{t=1}^T (\mathbf{I}_{d_1} - \mathbf{C}'_{2t} (\mathbf{C}_{2t} \mathbf{C}'_{2t} + \boldsymbol{\Sigma}_{tV'V})^{-1} \mathbf{C}_{2t}) \mathcal{F} (\mathbf{I}_{d_1} - \mathbf{C}'_{2t} (\mathbf{C}_{2t} \mathbf{C}'_{2t} + \boldsymbol{\Sigma}_{tV'V})^{-1} \mathbf{C}_{2t}) \\ - N^{-1} T^{-1} \sum_{t=1}^T \sigma_t^2 \mathbf{C}'_{2t} (\mathbf{C}_{2t} \mathbf{C}'_{2t} + \boldsymbol{\Sigma}_{tV'V})^{-1} \mathbf{C}_{2t} \rightarrow_p \mathcal{D}_{1N} + \mathcal{F} + N^{-1} \mathcal{E}_{1N} = \mathcal{I}_{1N},\end{aligned}$$

and non-singularity of \mathcal{I}_{1N} follows under mild conditions on \mathbf{C}_{2t} and $\boldsymbol{\Sigma}_{tV'V}$. Moreover, given $\mathcal{D}_{2N} = \mathbf{0}$, $\mathcal{E}_{2N} = -\sigma^2 \iota_N \iota'_N$ one obtains

$$\mathcal{I}_{2N} = \sigma^2 \mathbf{I}_N - N^{-1} \sigma^2 \iota_N \iota'_N + N^{-1} \iota_N \iota'_N$$

where \mathcal{I}_{2N} is non-singular for all values of $\sigma^2 < \infty$. Then, by the Sherman-Morrison-Woodbury formula (reported in the Appendix),

$$\boldsymbol{\Sigma}_{tX' \mathcal{I}_2^{-1} X} = \begin{pmatrix} 1 & \boldsymbol{\mu}'_x \mathbf{C}'_{2t} \\ \mathbf{C}_{2t} \boldsymbol{\mu}_x & \sigma^{-2} \mathbf{C}_{2t} (\boldsymbol{\Sigma}_{xx'} + (\sigma^2 - 1) \boldsymbol{\mu}_x \boldsymbol{\mu}'_x) \mathbf{C}'_{2t} \end{pmatrix}.$$

Similar calculations lead to $\Sigma_{tX'T_2^{-1}HT_2^{-1}X}$, $\Sigma_{tB'T_2^{-1}HT_2^{-1}B}$, $\Sigma_{tX'T_2^{-1}X}$, $\Sigma_{tX'T_2^{-1}HT_2^{-1}B}$.

It remains to verify (43)-(46). Under the assumptions made \mathcal{C}_{1N} is a random, mean zero, matrix of dimension $m \times N$, whose are uncorrelated with each ε_t , \mathbf{X}_t and with each row of Σ_N^{-1} . In addition, denoting by \mathcal{C}_{1Nj} the j th row of \mathcal{C}_{1N} , we will require $\sup_N E \|\mathcal{C}'_{1Nj}\mathcal{C}_{1Nj}\| = O(1)$ for all $1 \leq j \leq m$. The same assumptions are required for all the zero mean random matrices introduced below. Thus, for all t , $\mathbf{X}'_t \Sigma_N^{-1} \mathcal{C}'_{1N} = O_p(N^{\frac{1}{2}l_{1+k}l'_m})$, $\mathcal{C}_{1N} \Sigma_N^{-1} \varepsilon_t = O_p(N^{\frac{1}{2}l_m})$ and, by Lemma 2, $\mathbf{B}' \Sigma_N^{-1} \varepsilon_t = O_p(N^{-\frac{1}{2}l_m})$, $\mathbf{B}' \Sigma_N^{-1} \mathcal{C}'_{1N} = O_p(N^{-\frac{1}{2}l_m l'_m})$, $\mathbf{X}'_t \Sigma_N^{-1} \mathbf{B} = O_p(l_{1+k}l'_m)$, $\mathbf{B}' \Sigma_N^{-1} \mathbf{B} = O_p(l_m l'_m)$ yielding

$$\begin{aligned} \mathbf{X}'_t \Sigma_N^{-1} (\mathbf{B} \mathcal{C}_{1N} + \mathcal{C}'_{1N} \mathbf{B}') \Sigma_N^{-1} \mathbf{X}_t &= O_p(N^{\frac{1}{2}l_{1+k}l'_m}), \\ \mathbf{X}'_t \Sigma_N^{-1} (\mathbf{B} \mathcal{C}_{1N} + \mathcal{C}'_{1N} \mathbf{B}') \Sigma_N^{-1} (\mathbf{B} \mathbf{f}_t + \varepsilon_t) &= O_p(N^{\frac{1}{2}l_{1+k}}). \end{aligned}$$

We discuss only the case when (52) hold, since case (51) is much simpler. Setting for simplicity $d_2 = d_1 > k$, $\mathbf{A} = \mathbf{I}_{d_1}$ yields $\mathbf{X}_t \Sigma_{tX'X}^{-1} \Sigma_{tX'B} = (\mathbf{B} \mathcal{C}_{2t}' + \mathbf{V}_t) (\mathcal{C}_{2t} \mathcal{C}_{2t}' + \Sigma_{tV'V})^{-1} \mathcal{C}_{2t}$ and, in turn, one gets $\mathcal{C}_{2N} = \mathbf{B} \mathcal{C}_{21N} + \mathcal{C}_{22N}$ for *a.s.* random, mean zero, matrixes of dimension $m \times N$ and $N \times N$ respectively. The previous bounds apply substituting \mathcal{C}_{1N} with \mathcal{C}_{21N} . Moreover, $\mathbf{X}'_t \Sigma_N^{-1} \mathcal{C}_{22N} \Sigma_N^{-1} \mathbf{B} = O_p(l_{1+k}l'_m)$, $\mathbf{X}'_t \Sigma_N^{-1} \mathcal{C}_{22N} \Sigma_N^{-1} \varepsilon_t = O_p(N^{\frac{1}{2}l_{1+k}})$ then

$$\begin{aligned} \mathbf{X}'_t \Sigma_N^{-1} (\mathcal{C}_{2N} + \mathcal{C}'_{2N}) \Sigma_N^{-1} \mathbf{X}_t &= O_p(N l_{1+k} l'_m), \\ \mathbf{X}'_t \Sigma_N^{-1} (\mathcal{C}_{2N} + \mathcal{C}'_{2N}) \Sigma_N^{-1} (\mathbf{B} \mathbf{f}_t + \varepsilon_t) &= O_p(N^{\frac{1}{2}l_{1+k}}). \end{aligned}$$

Again, when (52) holds

$$\begin{aligned} \mathbf{X}_t \Sigma_{tX'X}^{-1} \mathbf{X}'_t &= q_t^{-1} l_N l'_N - q_t^{-1} (\mathbf{B} \mathcal{C}'_{2t} + \mathbf{V}_t) (\mathcal{C}_{2t} \Sigma_{xx'} \mathcal{C}'_{2t} + \Sigma_{tV'V})^{-1} \mathcal{C}_{2t} \mu_x l'_N \\ &\quad - q_t^{-1} l_N \mu'_x \mathcal{C}'_{2t} (\mathcal{C}_{2t} \Sigma_{xx'} \mathcal{C}'_{2t} + \Sigma_{tV'V})^{-1} (\mathcal{C}_{2t} \mathbf{X}^* + \mathbf{V}'_t) \\ &\quad + (\mathbf{X}^* \mathcal{C}'_{2t} + \mathbf{V}_t) (\mathcal{C}_{2t} \Sigma_{xx'} \mathcal{C}'_{2t} + \Sigma_{tV'V})^{-1} [\mathbf{I}_k + q_t^{-1} \mathcal{C}_{2t} \mu_x \mu'_x \mathcal{C}'_{2t} (\mathcal{C}_{2t} \Sigma_{xx'} \mathcal{C}'_{2t} + \Sigma_{tV'V})^{-1}] (\mathcal{C}_{2t} \mathbf{X}^* + \mathbf{V}'_t), \end{aligned}$$

yielding $\mathcal{C}_{3N} = \mathcal{C}_{31N} l'_N + \mathcal{C}_{32N} \mathbf{B}' + \mathcal{C}_{33N}$ for a $m \times 1$ matrix \mathcal{C}_{31N} , a $m \times d_1$ matrix \mathcal{C}_{32N} and a $m \times T$ matrix \mathcal{C}_{33N} , all zero mean random. Using the previous bounds, with \mathcal{C}_{33N} in place of \mathcal{C}_{1N} , as well as $\mathbf{X}'_t \Sigma_N^{-1} l_N = O_p(N l_{1+k})$, $\mathbf{B}' \Sigma_N^{-1} l_N = O_p(l_m)$, $\varepsilon'_t \Sigma_N^{-1} l_N = O_p(N^{\frac{1}{2}})$ yields

$$\begin{aligned} \mathbf{X}'_t \Sigma_N^{-1} (\mathbf{B} \mathcal{C}_{3N} + \mathcal{C}'_{3N} \mathbf{B}') \Sigma_N^{-1} \mathbf{X}_t &= O_p(N l_{1+k} l'_m), \\ \mathbf{X}'_t \Sigma_N^{-1} (\mathbf{B} \mathcal{C}_{3N} + \mathcal{C}'_{3N} \mathbf{B}') \Sigma_N^{-1} (\mathbf{B} \mathbf{f}_t + \varepsilon_t) &= O_p(N l_{1+k}). \end{aligned}$$

Finally $\mathcal{C}_{4N} = \mathbf{B} \mathcal{C}_{41N} l'_N + \mathbf{B} \mathcal{C}_{42N} \mathbf{B}' + \mathcal{C}_{43N} l'_N + \mathcal{C}_{44N} \mathbf{B}' + \mathcal{C}_{45N}$ for a $m \times 1$ matrix \mathcal{C}_{41N} , a $m \times m$ matrix \mathcal{C}_{42N} , a $N \times 1$ matrix \mathcal{C}_{43N} , a $N \times m$ matrix \mathcal{C}_{44N}

and a $N \times N$ matrix \mathcal{C}_{45N} , all zero mean random. Thus using the previous bounds, with \mathcal{C}_{43N} , \mathcal{C}_{44N} in place of \mathcal{C}_{1N} and \mathcal{C}_{45N} in place of \mathcal{C}_{22N} yields

$$\begin{aligned}\mathbf{X}'_t \boldsymbol{\Sigma}_N^{-1} (\mathcal{C}_{4N} + \mathcal{C}'_{4N}) \boldsymbol{\Sigma}_N^{-1} \mathbf{X}_t &= O_p(N^{\frac{3}{2}l_{1+k}l'_{1+k}}), \\ \mathbf{X}'_t \boldsymbol{\Sigma}_N^{-1} (\mathcal{C}_{4N} + \mathcal{C}'_{4N}) \boldsymbol{\Sigma}_N^{-1} (\mathbf{B}\mathbf{f}_t + \varepsilon_t) &= O_p(N^{\frac{3}{2}l_{1+k}}).\end{aligned}$$

Hence, (43), (44), (45), (46) hold with $a = 1/2, b = 1, c = 3/2, d = 3/2$. Therefore, primitive conditions for Assumption 7. \mathcal{T} can be found, in particular such as (52).

We summarize the result of this section as follows:

Proposition 2 *Assume that Andrews (2005, Assumptions 1, 2, 3) hold and, in addition, for any i the $\varepsilon_{i,t}$ have bounded $(2 + \delta)$ th moment and are linear processes of a martingale difference innovation with summable coefficients, the \mathbf{f}_t have bounded $(2 + \delta)$ th moment, the $\mathbf{v}_{i,t}$ have bounded $(6 + \delta)$ th moment. Finally let Assumptions 1. \mathcal{T} , 2. \mathcal{T} , and 7. \mathcal{T} hold.*

Then Theorem 2, (iii) applies to the GLS estimator for β_{0t} of model (47)-(48)-(49) when

$$\frac{1}{N} + \frac{1}{T} \rightarrow 0$$

for consistency and

$$\frac{1}{N} + \frac{N}{T} \rightarrow 0$$

for asymptotic normality.

No other conditions of Andrews (2005) is required, such as uncorrelatedness between the \mathbf{x}_i^ and the \mathbf{u}_i^* . Moreover, we do not require moment conditions for the \mathbf{C}_{2t} .*

Notice that the bias term of the OLS for β_{0t} is, from Theorem 2 (i), $\gamma_t^{OLS} = \boldsymbol{\Sigma}_{tX'X}^{-1} \boldsymbol{\Sigma}_{tX'B}^{-1} \mathbf{f}_t$ which is zero only if $\mathbf{f}_t = \mathbf{0}$ *a.s.* (no factor structure in the regression error) or, alternatively, if $\boldsymbol{\Sigma}_{tX'X}^{-1} \boldsymbol{\Sigma}_{tX'B} = \mathbf{0}$ *a.s.* The first row of the latter matrix, corresponding to the intercept parameter, is precisely equal to $q_t^{-1}(\mu'_u - \mu'_x \mathbf{C}'_{2t} (\mathbf{C}_{2t} \boldsymbol{\Sigma}_{xx'} \mathbf{C}'_{2t})^{-1} \mathbf{C}_{2t} \boldsymbol{\Sigma}_{xu'})$, simplifying to μ'_u under Andrews (2005, Condition SF2), that is when $\boldsymbol{\Sigma}_{xu'} = \mu_x \mu'_u$ holds. Therefore zero bias for the OLS estimator of the intercept also requires his condition SF3, viz. $\mu_u = \mathbf{0}$. However, as noted by Andrews (2005), consistency of the regression parameters only (the last k entries of β_{t0}) requires just zero correlation between the \mathbf{x}_i^* and the \mathbf{u}_i^* (his condition SF2). In fact the sub-matrix made from the second to the last row of $\boldsymbol{\Sigma}_{tX'X}^{-1} \boldsymbol{\Sigma}_{tX'B}$ is

$(\mathbf{C}_{2t}\boldsymbol{\Sigma}_{xx'}\mathbf{C}_{2t}')^{-1}[\mathbf{C}_{2t}\mu_x\mu_x'\mathbf{C}_{2t}'(\mathbf{C}_{2t}\boldsymbol{\Sigma}_{xx'}\mathbf{C}_{2t}')^{-1} - \mu_x'\mathbf{C}_{2t}'(\mathbf{C}_{2t}\boldsymbol{\Sigma}_{xx'}\mathbf{C}_{2t}')^{-1}\mathbf{C}_{2t}\mu_x\mathbf{I}_k] \times$
 $\mathbf{C}_{2t}\boldsymbol{\Sigma}_{xu'}(\mathbf{C}_{2t}\boldsymbol{\Sigma}_{xx'}\mathbf{C}_{2t}')^{-1}\mathbf{C}_{2t}(\boldsymbol{\Sigma}_{xu'} - \mu_x\mu_u')$ which, when $\boldsymbol{\Sigma}_{xu'} = \mu_x\mu_u'$ holds,
 equals $\mu_x'\mathbf{C}_{2t}'(\mathbf{C}_{2t}\boldsymbol{\Sigma}_{xx'}\mathbf{C}_{2t}')^{-1}\mathbf{C}_{2t}\mu_x(\mathbf{C}_{2t}\boldsymbol{\Sigma}_{xx'}\mathbf{C}_{2t}')^{-1}(\mathbf{C}_{2t}\mu_x - \mathbf{C}_{2t}\mu_x)\mu_u' +$
 $(\mathbf{C}_{2t}\boldsymbol{\Sigma}_{xx'}\mathbf{C}_{2t}')^{-1}\mathbf{C}_{2t}(\mu_x\mu_u' - \mu_x\mu_u')$, that is a matrix of zeros, independently
 from whether μ_u is zero or not.

4 Monte Carlo

We conduct a set of Monte Carlo experiments to appreciate the relevance of our asymptotic results for the GLS estimator in finite samples. We consider both the case of time regression with unit-specific coefficients as well as cross-sectional regression with time-specific coefficients.

4.1 Design

For the time regression case the data generating process is

$$y_{it} = \alpha_{i0} + \beta_{i0}x_{it} + b_{i10}f_{1t} + b_{i20}f_{2t} + \varepsilon_{it}, \quad (53)$$

where the single regressor satisfies

$$x_{it} = 0.5 + \delta_{i10}f_{1t} + \delta_{i30}f_{3t} + v_{it}. \quad (54)$$

Note that the model implies an observed common factor equal to 1 for all observations. The single regressor is allowed to be contemporaneously correlated with the innovation through one of the latent common factors (whenever $b_{i10}\delta_{i10} \neq 0$ a.s.). The factor loadings are normally distributed random variables, *i.i.d.* across unit:

$$\begin{pmatrix} b_{i10} \\ b_{i20} \end{pmatrix} \sim NID \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix} \right), \quad (55)$$

$$\begin{pmatrix} \delta_{i10} \\ \delta_{i30} \end{pmatrix} \sim NID \left(\begin{pmatrix} 0.5 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix} \right), \quad (56)$$

and the latent common factors and the idiosyncratic components are stationary stochastic processes, mutually independent to each other, satisfying

$$f_{j,t} = 0.5f_{j,t-1} + \sqrt{0.5}\eta_{jf,t}, \quad j = 1, 2, 3,$$

where each $\eta_{j,t} \sim NID(0, 1)$, mutually independent for $j = 1, 2, 3$, and

$$\begin{aligned}\varepsilon_{it} &= \rho_{i\varepsilon}\varepsilon_{it-1} + \eta_{i\varepsilon,t}, & \eta_{i\varepsilon,t} &\sim NID(0, \sigma_i^2(1 - \rho_{i\varepsilon}^2)), i = 1, \dots, N, \\ v_{it} &= \rho_{iv}v_{it-1} + \eta_{iv,t}, & \eta_{iv,t} &\sim NID(0, (1 - \rho_{iv}^2)), i = 1, \dots, N,\end{aligned}$$

with $\rho_{i\varepsilon} \sim UID(0.05, 0.95)$, $\rho_{iv} \sim UID(0.05, 0.95)$, $\sigma_{i\varepsilon}^2 \sim UID(0.5, 1.5)$ where NID, UID means *iid* normally and uniformly distributed respectively. Finally, the parameters of interest are constant across replications and equal to $\alpha_{i0} = 1, \gamma_{i0} = 0.5$ and, assuming N even,

$$\beta_{i0} = \begin{cases} 1 & \text{for } i = 1, \dots, \frac{N}{2}, \\ 3 & \text{for } i = \frac{N}{2} + 1, \dots, N. \end{cases}$$

This Monte Carlo design is a simplified version of Pesaran (2006), designed in such a way that (through (56)) the rank condition in Pesaran (2006, eq. (21)) is not satisfied. Pesaran (2006) shows that under this circumstance his individual specific estimator for β_{i0} is invalid whereas his pooled estimators for $\beta_0 = E\beta_{i0}$ remain consistent.

For the cross-sectional regression case we consider

$$y_{it} = \alpha_{t0} + \beta_{t0}x_{it} + b_{i10}f_{1t} + b_{i20}f_{2t} + \varepsilon_{it}. \quad (57)$$

The regressor x_{it} , the factors $f_{j,t}$, $j = 1, 2, 3$ and the idiosyncratic innovations ε_{it} , $i = 1, \dots, N$ are obtained as in the previous case. The parameters of interest are constant across replications and equal to $\alpha_{t0} = 1$ and, assuming T even,

$$\beta_{t0} = \begin{cases} 1 & \text{for } i = 1, \dots, \frac{T}{2}, \\ 2 & \text{for } i = \frac{T}{2} + 1, \dots, T. \end{cases}$$

Finally,

$$\begin{pmatrix} b_{i10} \\ b_{i20} \end{pmatrix} \sim NID\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}\right), \quad (58)$$

$$\begin{pmatrix} \delta_{i10} \\ \delta_{i30} \end{pmatrix} = \begin{pmatrix} b_{i10} \\ b_{i20} \end{pmatrix}. \quad (59)$$

implying that the factor loadings pertinent to the innovation of (57) are (perfectly) correlated with the factor loadings corresponding to the regressor x_{it} . Under this condition Andrews (2005) shows that the OLS estimator of the regression parameters α_{t0}, β_{t0} is non consistent.

We consider 2000 Monte Carlo replications with sample sizes $(N, T) \in \{60, 200, 600\} \times \{30, 100, 300\}$, with $N > T$, for the time regression and $(N, T) \in \{30, 100, 300\} \times \{60, 200, 600\}$, with $N < T$, for the cross-sectional regression.

The results are summarized in Tables 1 to 4, where we report the sample mean and the root mean square error for the estimates of the parameter α_{i0}, β_{i0} and α_{t0}, β_{t0} for time and cross-sectional regression respectively, averaged across the Monte Carlo iterations. We consider four estimators which corresponds to four panels of each table: the GLS, the multi-step GLS (described Remark 5 to Theorem 1) where the iteration is carried out $J = 4$ times, the OLS and the UGLS estimators. In particular, regarding the time regression results reported in Tables 1-2, for each of these four estimators, we report the average across all N units of the sample mean (mean)

$M^{-1} \sum_{m=1}^M \hat{\alpha}_i^m$ and of the root mean square error (**rmse**) $\left(M^{-1} \sum_{m=1}^M (\hat{\alpha}_i^m - 1)^2 \right)^{\frac{1}{2}}$

and the average across the units $i = N/2 + 1, \dots, N$ of $M^{-1} \sum_{m=1}^M \hat{\beta}_i^m$ and $\left(M^{-1} \sum_{m=1}^M (\hat{\beta}_i^m - 3)^2 \right)^{\frac{1}{2}}$ with $M = 2,000$. Recall that we assumed that the true intercept coefficients are constant across units whereas the regression coefficients take two different values for the first half and second half of the N units. Here $\hat{\alpha}_i^m$ and $\hat{\beta}_i^m$ denote, respectively, the estimates of the intercept and regression coefficients corresponding to the m th Monte Carlo iteration for a generic estimator. The same description applies to the cross-sectional regression results although now Tables 3 and 4 report, respectively, for each of these four estimators, the average across all T periods of $M^{-1} \sum_{m=1}^M \hat{\alpha}_t^m$ and $\left(M^{-1} \sum_{m=1}^M (\hat{\alpha}_t^m - 1)^2 \right)^{\frac{1}{2}}$ and the average across the periods $t = T/2 + 1, \dots, T$

of $M^{-1} \sum_{m=1}^M \hat{\beta}_t^m$ and $\left(M^{-1} \sum_{m=1}^M (\hat{\beta}_t^m - 2)^2 \right)^{\frac{1}{2}}$ with $M = 2,000$. (The results for the regression coefficients corresponding to the units $i = 1, \dots, N/2$, for time regression, and to periods $t = 1, \dots, T/2$, for cross-sectional regression, are not reported but are available.)

4.2 Results

We start by looking the estimation results for time regression with unit-specific intercept term (Table 1) and regression coefficient (Table 2), respectively. Since the GLS and multi-step GLS estimators requires $N \geq T$, each

panel is made by a lower triangular matrix. Obviously, the OLS and the UGLS estimator do not require this constraint since they can be also evaluated when $N < T$ but we did not report the results for this case. The upper left panel describes the GLS results. One can see how the bias diminishes as both N, T grow or when N increases for a given T . This is because the inverse of the pseudo-covariance matrix Σ_T^{-1} is better estimated in these circumstances. In contrast, although still negligible in absolute terms, the bias, if any, tends to increase when T grows for a given N . Instead, as expected, the rmse always diminishes when T increases for a given N or when they both increase. The rmse of the regression coefficient estimates (Table 2) diminishes also when N increases for given T . In general terms these results suggest that the bias of the estimates varies mainly with N and their variance varies with T . The same pattern is observed with respect to the multi-step GLS results, reported in the upper right panel. The only difference is that now the bias and the rmse are always much smaller than the GLS case. The lower right panel reports the results for the UGLS which is unfeasible in practice since it involves the true covariance matrix \mathbf{S}_i . As a consequence, the results do not depend on N but only on T . The bias is negligible even for small samples and, for larger sample sizes, it is remarkably comparable to the iterated GLS although the latter exhibit a slightly larger rmse. Finally, the lower left panel reports the OLS results which also do not depend on N , as expected. Under our design, the OLS estimator is non-consistent obtaining a bias which is much larger than for any other estimators and, more importantly, only marginally varying as N or T increases. The rmse diminishes suggesting that the variance of the OLS estimator is converging to zero with the squared bias converging to $(\gamma_i^{OLS})^2$.

The cross-section regression results are in Tables 3 and 4. Now the GLS and iterated GLS estimators requires $N \leq T$ and thus each panel is made by an upper triangular matrix. The results are specular to the ones obtained for the time regression case. Regarding the GLS results in the upper left panel, the bias diminishes as both N, T grow or when T increases for a given N whereas it does not necessarily decreases when N grows for a given T since in this latter circumstance Σ_N^{-1} is more poorly estimated. The rmse diminishes when either T increases for a given N or when they both increase and, for for the regression coefficient case (Table 3) when N increases for given T as well. The performance of the iterated GLS and of the UGLS, respectively reported in the upper and lower right panels, is remarkably similar, except perhaps when N, T are either both very small or very large. Obviously, UGLS

carries the best results especially in terms of rmse where, as expected, the figures do not depend on T but vary only with N . The OLS estimator, whose results are in the lower left panel, is non-consistent under our design, with a sizeable bias only marginally varying with either N and T . Again, the rmse diminishes as N increases indicating that the OLS will eventually converge to the sum of the true parameter value and the non-zero bias. Although the results have not been reported for easy reference, the OLS and the UGLS estimators can be evaluated also for $N > T$.

5 Concluding remarks

This paper proposes a feasible GLS estimator for linear panel with common factor structure in both the regressors and the innovation. We develop our results separately for time regressions with unit-specific coefficients as well as for cross-section regressions with time varying coefficients. The GLS estimator is consistent and asymptotically normal, when both the cross-section N and time series T dimensions diverge to infinity, under circumstances that make the OLS non-consistent, hence providing more than an efficiency gain. Noticeably, the GLS estimator does not require preliminary estimation of the latent factors nor of their dimension. It uses all the panel data information but computationally it only requires to estimate $N+1$ time or $T+1$ cross-sectional regressions, respectively. We provide a set of general regularity assumptions which allows both temporal and cross-sectional dependence of the idiosyncratic innovation. We then derive primitive conditions of our general assumptions for the specific models investigated by Andrews (2005) and Pesaran (2006). Our results are corroborated by a set of Monte Carlo experiments. Several other issues appear of interest. For one, estimation of the regression coefficients mean $E\beta_i$ or $E\beta_t$, depending on whether one considers unit- or time-specific coefficients, attracted a great deal of interest. Another interesting topic would be to develop joint testing procedures of all the β_i , $i = 1, \dots, N$ or the β_t , $t = 1, \dots, T$. Also, the effect of trending regressors and, more in general, nonstationarity would be of great interest. These issues will be the focus of future research.

6 Mathematical Appendix

For random matrices \mathbf{A} non-singular of dimension $m_1 \times m_1$, \mathbf{B} of dimension $m_1 \times m_2$, \mathbf{C} non-singular of dimension $m_2 \times m_2$, \mathbf{D} of dimension $m_1 \times m_3$, with $m_1 \geq m_2$ and any m_3 , we present the well-known Sherman-Morrison-Woodbury formula, followed by two lemmas whose proofs are reported in the Supplementary Material. Note that Lemma 1 is basically reproducing Lemma B in Pesaran and Zaffaroni (2009). Throughout the paper we will refer to the lemmas without reference to the matrixes $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ when there is no risk of ambiguity.

Sherman-Morrison-Woodbury formula.

$$(\mathbf{BCB}' + \mathbf{A})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{A}^{-1} \quad a.s.$$

Lemma 1($\mathbf{A}, \mathbf{B}, \mathbf{C}, m_1$). Set $\mathbf{E} = \mathbf{BCB}' + \mathbf{A}$ a.s. Let \mathbf{G} a random positive definitive matrix such that as $m_1 \rightarrow \infty$:

$$\frac{\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}}{m_1} \rightarrow_{a.s.} \mathbf{G} \text{ non-singular.} \quad (60)$$

Then

$$\mathbf{E}^{-1}\mathbf{B} = \mathbf{A}^{-1}\mathbf{B}\left(\frac{\mathbf{C}^{-1}}{m_1} + \frac{\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}}{m_1}\right)^{-1}\frac{\mathbf{C}^{-1}}{m_1} \quad (61)$$

and, denoting by $\mathbf{e}_{m_1}^{(i)}$ the i -th column of the identity matrix \mathbf{I}_{m_1} , then for any $1 \leq i \leq m_1$:

$$\mathbf{e}_{m_1}^{(i)'}\mathbf{E}^{-1}\mathbf{b}^{(j)} \rightarrow_p 0, \quad 1 \leq j \leq m_2, \quad \text{as } m_1 \rightarrow \infty, \quad (62)$$

where $\mathbf{b}^{(j)} = \mathbf{B}\mathbf{e}_{m_2}^{(j)}$ is the j th column of \mathbf{B} .

When (60) and

$$\frac{\mathbf{B}'\mathbf{A}^{-1}\mathbf{A}^{-1}\mathbf{B}}{m_1} \rightarrow_{a.s.} \mathbf{L} \geq 0, \quad (63)$$

for some a.s. finite random positive semi-definite matrix \mathbf{L} , then

$$\|\mathbf{E}^{-1}\mathbf{B}\|^2 = O_p(m_1^{-1}) \quad \text{as } m_1 \rightarrow \infty. \quad (64)$$

Lemma 2($\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, m_1$). Set $\mathbf{E} = \mathbf{BCB}' + \mathbf{A}$ a.s. When (60) and $\mathbf{D}'\mathbf{A}^{-1}\mathbf{B} = O_p(m_1^{-1/2}l_{m_3}l'_{m_2})$ then

$$\mathbf{D}'\mathbf{E}^{-1}\mathbf{B} = O_p(l_{m_3}l'_{m_2}) \quad \text{as } m_1 \rightarrow \infty. \quad (65)$$

When (60) and $\mathbf{D}'\mathbf{A}^{-1}\mathbf{B} = O_p(m_1^{\frac{1}{2}}\iota_{m_3}\iota'_{m_2})$ then

$$\mathbf{D}'\mathbf{E}^{-1}\mathbf{B} = O_p(m_1^{-\frac{1}{2}}\iota_{m_3}\iota'_{m_2}) \quad \text{as } m_1 \rightarrow \infty. \quad (66)$$

Proof of Theorem 1. (i) All the limits below hold as $T \rightarrow \infty$. The results follow since $\hat{\beta}_i^{OLS} - \beta_{i0} = (\mathbf{X}'_i\mathbf{X}_i)^{-1}\mathbf{X}'_i\mathbf{u}_i$ can be written as $\hat{\beta}_i^{OLS} - \beta_{i0} - \hat{\gamma}_i^{OLS} = T^{-\frac{1}{2}}(\frac{\mathbf{X}'_i\mathbf{X}_i}{T})^{-1}T^{-\frac{1}{2}}\mathbf{X}'_i\varepsilon_i$. setting $\hat{\gamma}_i^{OLS} = (\mathbf{X}'_i\mathbf{X}_i)^{-1}\mathbf{X}'_i\mathbf{F}\mathbf{b}_i$. The result then easily follows by our assumptions.

(ii) All the limits below hold as $T \rightarrow \infty$. Simple manipulations yield

$$\begin{aligned} \hat{\beta}_i^{UGLS} - \beta_{i0} &= (\mathbf{X}'_i\mathcal{S}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}'_i\mathcal{S}_i^{-1}(\mathbf{F}\mathbf{b}_i + \varepsilon_i) \\ &= (\mathbf{X}'_i\mathcal{S}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}'_i(\mathbf{H}_i^{-1} - \mathbf{H}_i^{-1}\mathbf{F}(\mathcal{B}_i^{-1} + \mathbf{F}'\mathbf{H}_i^{-1}\mathbf{F})^{-1}\mathbf{F}'\mathbf{H}_i^{-1})\mathbf{F}\mathbf{b}_i + (\mathbf{X}'_i\mathcal{S}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}'_i\mathcal{S}_i^{-1}\varepsilon_i \\ &= (\mathbf{X}'_i\mathcal{S}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}'_i\mathbf{H}_i^{-1}\mathbf{F}(\mathcal{B}_i^{-1} + \mathbf{F}'\mathbf{H}_i^{-1}\mathbf{F})^{-1}\mathcal{B}_i^{-1}\mathbf{b}_i + (\mathbf{X}'_i\mathcal{S}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}'_i\mathcal{S}_i^{-1}\varepsilon_i \\ &= \left(\frac{\mathbf{X}'_i\mathcal{S}_i^{-1}\mathbf{X}_i}{T}\right)^{-1}\frac{\mathbf{X}'_i\mathbf{H}_i^{-1}\mathbf{F}}{T}\left(\frac{\mathcal{B}_i^{-1}}{T} + \frac{\mathbf{F}'\mathbf{H}_i^{-1}\mathbf{F}}{T}\right)^{-1}\frac{\mathcal{B}_i^{-1}}{T}\mathbf{b}_i + T^{-\frac{1}{2}}\left(\frac{\mathbf{X}'_i\mathcal{S}_i^{-1}\mathbf{X}_i}{T}\right)^{-1}\frac{\mathbf{X}'_i\mathcal{S}_i^{-1}\varepsilon_i}{T^{\frac{1}{2}}} \\ &= \hat{\gamma}_i^{UGLS} + T^{-\frac{1}{2}}\left(\frac{\mathbf{X}'_i\mathcal{S}_i^{-1}\mathbf{X}_i}{T}\right)^{-1}\frac{\mathbf{X}'_i\mathcal{S}_i^{-1}\varepsilon_i}{T^{\frac{1}{2}}}, \end{aligned}$$

where $\hat{\gamma}_i^{UGLS}$ denote the first term on the right hand side of the second last equality sign. The first equality is warranted by the Sherman-Morrison-Woodbury (hereafter SMW) formula and the fourth equality makes use of the Lemma 1($\mathbf{H}_i, \mathbf{F}, \mathcal{B}_i, T$). Using the SMW formula again

$$\frac{\mathbf{X}'_i\mathcal{S}_i^{-1}\mathbf{X}_i}{T} = \frac{\mathbf{X}'_i\mathbf{H}_i^{-1}\mathbf{X}_i}{T} - \frac{\mathbf{X}'_i\mathbf{H}_i^{-1}\mathbf{F}}{T}\left(\frac{\mathcal{B}_i^{-1}}{T} + \frac{\mathbf{F}'\mathbf{H}_i^{-1}\mathbf{F}}{T}\right)^{-1}\frac{\mathbf{F}'\mathbf{H}_i^{-1}\mathbf{X}_i}{T},$$

implying $(\mathbf{X}'_i\mathcal{S}_i^{-1}\mathbf{X}_i/T)^{-1} = (\boldsymbol{\Sigma}_{iX'H^{-1}X} - \boldsymbol{\Sigma}_{iX'H^{-1}F}\boldsymbol{\Sigma}_{iF'H^{-1}F}^{-1}\boldsymbol{\Sigma}'_{iX'H^{-1}F})^{-1}(1 + o_p(1))$, yielding $T^{\frac{1}{2}}\hat{\gamma}_i^{UGLS} = O_p(T^{-\frac{1}{2}}\iota_k)$. For the second (dominant) term of the right hand side of $T^{\frac{1}{2}}(\hat{\beta}_i^{UGLS} - \beta_{i0})$ we start verifying that

$$\frac{\mathbf{X}'_i\mathcal{S}_i^{-1}\varepsilon_i}{T^{\frac{1}{2}}} \rightarrow_d (\boldsymbol{\Sigma}_{iX'H^{-1}X} - \boldsymbol{\Sigma}_{iX'H^{-1}F}\boldsymbol{\Sigma}_{iF'H^{-1}F}^{-1}\boldsymbol{\Sigma}'_{iX'H^{-1}F})^{\frac{1}{2}}\zeta_{ki}, \quad (67)$$

where ζ_{ki} the zero mean, unit covariance matrix, normally distributed vector is independent from $\boldsymbol{\Sigma}_{iX'H^{-1}X}, \boldsymbol{\Sigma}_{iX'H^{-1}F}, \boldsymbol{\Sigma}_{iF'H^{-1}F}$. In fact,

$$\frac{\mathbf{X}'_i\mathcal{S}_i^{-1}\varepsilon_i}{T^{\frac{1}{2}}} = \left(\mathbf{I}_k, -\hat{\boldsymbol{\Sigma}}_{iX'H^{-1}F}\hat{\boldsymbol{\Sigma}}_{iF'H^{-1}F}^{-1}\right)\frac{\mathbf{Z}'_i\mathcal{H}_i^{-1}\varepsilon_i}{T^{\frac{1}{2}}}(1 + o_p(1)),$$

setting $\hat{\Sigma}_{iX'H^{-1}F} = T^{-1}\mathbf{X}'_i\mathcal{H}_i^{-1}\mathbf{F}$ and $\hat{\Sigma}_{iX'H^{-1}F} = T^{-1}\mathbf{F}'\mathcal{H}_i^{-1}\mathbf{F}$, where the $o_p(1)$ arises since $T^{-1}(\mathbf{B}_i^{-1} + \mathbf{F}'\mathcal{H}_i^{-1}\mathbf{F}) = \hat{\Sigma}_{iX'H^{-1}F} + o_p(\iota_m\iota'_m)$. Under our assumptions

$$\frac{\mathbf{X}'_i\mathcal{S}_i^{-1}\varepsilon_i}{T^{\frac{1}{2}}} \rightarrow_d (\mathbf{I}_k, -\Sigma_{iX'H^{-1}F}\Sigma_{iF'H^{-1}F}^{-1})(\Sigma_{iZ'H^{-1}Z})^{\frac{1}{2}}\zeta_{(k+m)i}.$$

Given that the (conditional) normal distribution is uniquely defined by the first two (conditional) moments, the left hand side is equal, in distribution, to (67), since $(\mathbf{I}_k, -\Sigma_{iX'H^{-1}F}\Sigma_{iF'H^{-1}F}^{-1})\Sigma_{iZ'H^{-1}Z}(\mathbf{I}_k, -\Sigma_{iX'H^{-1}F}\Sigma_{iF'H^{-1}F}^{-1})' = (\Sigma_{iX'H^{-1}X} - \Sigma_{iX'H^{-1}F}\Sigma_{iF'H^{-1}F}^{-1}\Sigma_{iX'H^{-1}F}')\Sigma_{iZ'H^{-1}Z}$. Combining terms concludes.

(iii) All the limits below hold as $(N, T) \rightarrow \infty$. We must assume $N \geq T$ and, with no loss of generality, that there are no observed common factors implying that $\hat{\Sigma}_T = N^{-1}\sum_{i=1}^N \hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'$ where $\hat{\mathbf{u}}_i = (\mathbf{I}_T - \mathbf{M}_i)\mathbf{u}_i$ setting $\mathbf{M}_i = \mathbf{X}_i(\mathbf{X}'_i\mathbf{X}_i)^{-1}\mathbf{X}'_i$. Then

$$\begin{aligned} \hat{\mathbf{u}}_i\hat{\mathbf{u}}_i' &= \underbrace{\mathbf{F}\mathbf{b}_i\mathbf{b}_i'\mathbf{F}'}_I + \underbrace{(\mathbf{I}_T - \mathbf{M}_i)\varepsilon_i\varepsilon_i'(\mathbf{I}_T - \mathbf{M}_i)}_{II} \\ &+ \underbrace{\mathbf{M}_i\mathbf{F}\mathbf{b}_i\mathbf{b}_i'\mathbf{F}'\mathbf{M}_i}_{III} + \underbrace{(\mathbf{F}\mathbf{b}_i\varepsilon_i'(\mathbf{I}_T - \mathbf{M}_i) + (\mathbf{I}_T - \mathbf{M}_i)\varepsilon_i\mathbf{b}_i'\mathbf{F}')}_{IV} \\ &- \underbrace{((\mathbf{I}_T - \mathbf{M}_i)\varepsilon_i\mathbf{b}_i'\mathbf{F}'\mathbf{M}_i + \mathbf{M}_i\mathbf{F}\mathbf{b}_i\varepsilon_i'(\mathbf{I}_T - \mathbf{M}_i))}_{V} - \underbrace{(\mathbf{M}_i\mathbf{F}\mathbf{b}_i\mathbf{b}_i'\mathbf{F}' + \mathbf{F}\mathbf{b}_i\mathbf{b}_i'\mathbf{F}'\mathbf{M}_i)}_{VI}. \end{aligned}$$

For I one gets $\mathbf{F}(N^{-1}\sum_{i=1}^N \mathbf{b}_i\mathbf{b}_i')\mathbf{F}' = \mathbf{F}\mathcal{B}\mathbf{F}'(1 + o_p(1))$. For II

$$\begin{aligned} N^{-1}\sum_{i=1}^N (\mathbf{I}_T - \mathbf{M}_i)\varepsilon_i\varepsilon_i'(\mathbf{I}_T - \mathbf{M}_i) &= N^{-1}\sum_{i=1}^N \varepsilon_i\varepsilon_i' \\ &+ \left(N^{-1}T^{-2}\sum_{i=1}^N \mathbf{X}_i\Sigma_{iX'X}^{-1}\mathbf{X}'_i\varepsilon_i\varepsilon_i'\mathbf{X}_i\Sigma_{iX'X}^{-1}\mathbf{X}'_i \right) (1 + o_p(1)) \\ &- \left(N^{-1}T^{-1}\sum_{i=1}^N (\mathbf{X}_i\Sigma_{iX'X}^{-1}\mathbf{X}'_i\varepsilon_i\varepsilon_i' + \varepsilon_i\varepsilon_i'\mathbf{X}_i\Sigma_{iX'X}^{-1}\mathbf{X}'_i) \right) (1 + o_p(1)), \end{aligned}$$

yielding $N^{-1}\sum_{i=1}^N (\mathbf{I}_T - \mathbf{M}_i)\varepsilon_i\varepsilon_i'(\mathbf{I}_T - \mathbf{M}_i) = (\mathcal{H}_T + T^{-1}(\mathcal{A}_{3T} - (\mathcal{A}_{4T} + \mathcal{A}'_{4T}))) (1 + o_p(1))$. For III one gets $N^{-1}\sum_{i=1}^N \mathbf{M}_i\mathbf{F}\mathbf{b}_i\mathbf{b}_i'\mathbf{F}'\mathbf{M}_i = \mathcal{A}_{1T}(1 + o_p(1))$. For IV $N^{-\frac{1}{2}}\sum_{i=1}^N \mathbf{F}\mathbf{b}_i\varepsilon_i' = \mathbf{F}\mathcal{C}_{1T}(1 + o_p(1))$, $T^{\frac{1}{2}}N^{-\frac{1}{2}}\sum_{i=1}^N \mathbf{F}\mathbf{b}_i\varepsilon_i'\mathbf{M}_i = \mathbf{F}\mathcal{C}_{3T}(1 + o_p(1))$. Combining the above results yield $N^{-1}\sum_{i=1}^N \mathbf{F}\mathbf{b}_i\varepsilon_i'(\mathbf{I}_T - \mathbf{M}_i) =$

$N^{-\frac{1}{2}}\mathbf{F}\left(\mathcal{C}_{1T} - T^{-\frac{1}{2}}\mathcal{C}_{3T}\right)(1 + o_p(1))$. For V using the same arguments one gets $N^{-1}\sum_{i=1}^N\mathbf{M}_i\mathbf{F}\mathbf{b}_i\varepsilon'_i(\mathbf{I}_T - \mathbf{M}_i) = N^{-\frac{1}{2}}\left(\mathcal{C}_{2T} - T^{-\frac{1}{2}}\mathcal{C}_{4T}\right)(1 + o_p(1))$. For VI one gets $N^{-1}\sum_{i=1}^N\mathbf{M}_i\mathbf{F}\mathbf{b}_i\mathbf{b}'_i\mathbf{F}' = \mathcal{A}_{2T}(1 + o_p(1))$. Summarizing:

$$\begin{aligned}\hat{\Sigma}_T &= [\mathbf{F}\mathbf{B}\mathbf{F}' + \mathcal{H}_T + \mathcal{A}_{1T} - \mathcal{A}_{2T} - \mathcal{A}'_{2T} + T^{-1}(\mathcal{A}_{3T} - \mathcal{A}_{4T} - \mathcal{A}'_{4T}) + \mathcal{D}_{N,T}](1 + o_p(1)) \\ &= [\mathbf{F}\mathcal{I}_{1T}\mathbf{F}' + \mathcal{I}_{2T} + \mathcal{D}_{N,T}](1 + o_p(1)) = [\Sigma_T + \mathcal{D}_{N,T}](1 + o_p(1))\end{aligned}$$

setting $\mathcal{D}_{N,T} = N^{-\frac{1}{2}}(\mathbf{F}\mathcal{C}_{1T} + \mathcal{C}'_{1T}\mathbf{F}' + \mathcal{C}_{2T} + \mathcal{C}'_{2T}) - (NT)^{-\frac{1}{2}}(\mathbf{F}\mathcal{C}_{3T} + \mathcal{C}'_{3T}\mathbf{F}' + \mathcal{C}_{4T} + \mathcal{C}'_{4T})$. Since for generic non-singular matrices \mathbf{A} , $\hat{\mathbf{A}}$ one has $\hat{\mathbf{A}}^{-1} = \mathbf{A}^{-1}\hat{\mathbf{A}}\mathbf{A}^{-1}(1 + o_p(1))$ whenever $\hat{\mathbf{A}} \rightarrow_p \mathbf{A}$, one obtains

$$\begin{aligned}(\hat{\beta}_i^{GLS} - \beta_{i0}) &= (\mathbf{X}'_i\Sigma_T^{-1}\mathbf{X}_i)^{-1}\mathbf{X}'_i\Sigma_T^{-1}(\mathbf{F}\mathbf{b}_i + \varepsilon_i)(1 + o_p(1)) \\ &+ (\mathbf{X}'_i\Sigma_T^{-1}\mathbf{X}_i)^{-1}\mathbf{X}'_i\Sigma_T^{-1}\mathcal{D}_{N,T}\Sigma_T^{-1}(\mathbf{F}\mathbf{b}_i + \varepsilon_i)(1 + o_p(1)) \\ &+ (\mathbf{X}'_i\Sigma_T^{-1}\mathbf{X}_i)^{-1}(\mathbf{X}'_i\Sigma_T^{-1}\mathcal{D}_{N,T}\Sigma_T^{-1}\mathbf{X}_i)(\mathbf{X}'_i\Sigma_T^{-1}\mathbf{X}_i)^{-1}\mathbf{X}'_i\Sigma_T^{-1}(\mathbf{F}\mathbf{b}_i + \varepsilon_i)(1 + o_p(1)) \\ &+ (\mathbf{X}'_i\Sigma_T^{-1}\mathbf{X}_i)^{-1}(\mathbf{X}'_i\Sigma_T^{-1}\mathcal{D}_{N,T}\Sigma_T^{-1}\mathbf{X}_i)(\mathbf{X}'_i\Sigma_T^{-1}\mathbf{X}_i)^{-1}\mathbf{X}'_i\Sigma_T^{-1}\mathcal{D}_{N,T}\Sigma_T^{-1}(\mathbf{F}\mathbf{b}_i + \varepsilon_i)(1 + o_p(1)).\end{aligned}$$

Regarding the first term on the right hand side of $(\hat{\beta}_i^{GLS} - \beta_{i0})$, one follows precisely the same steps of part (ii) but replacing \mathbf{S}_i , \mathcal{B}_i , \mathbf{H}_i by Σ_T , \mathcal{I}_{1T} , \mathcal{I}_{2T} respectively. By Lemma 1(\mathcal{I}_{2T} , \mathbf{F} , \mathcal{I}_{1T} , T)

$$T^{\frac{1}{2}}(\mathbf{X}'_i\Sigma_T^{-1}\mathbf{X}_i)^{-1}\mathbf{X}'_i\Sigma_T^{-1}(\mathbf{F}\mathbf{b}_i + \varepsilon_i) \rightarrow_d (\mathcal{V}_i^{GLS})^{\frac{1}{2}}\zeta_{ki} \text{ as } T \rightarrow \infty$$

with $\mathcal{V}_i^{GLS} = (\mathcal{M}_i^{GLS})^{-1}\mathcal{N}_i^{GLS}(\mathcal{M}_i^{GLS})^{-1}$ where

$$\begin{aligned}\mathcal{N}_i^{GLS} &= \Sigma_{i,X'\mathcal{I}_2^{-1}X}^{-1} + \Sigma_{iX'\mathcal{I}_2^{-1}F}\Sigma_{iF'\mathcal{I}_2^{-1}F}^{-1}\Sigma_{iF'\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}F}\Sigma_{iF'\mathcal{I}_2^{-1}F}^{-1}\Sigma'_{iX'\mathcal{I}_2^{-1}F} \\ &- \left(\Sigma_{iX'\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}F}\Sigma_{iF'\mathcal{I}_2^{-1}F}^{-1}\Sigma'_{iX'\mathcal{I}_2^{-1}F} + \Sigma_{iX'\mathcal{I}_2^{-1}F}\Sigma_{iF'\mathcal{I}_2^{-1}F}^{-1}\Sigma'_{iX'\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}F}\right)\end{aligned}$$

and $\mathcal{M}_i^{GLS} = \Sigma_{iX'\mathcal{I}_2^{-1}X} - \Sigma_{iX'\mathcal{I}_2^{-1}F}\Sigma_{iF'\mathcal{I}_2^{-1}F}\Sigma'_{iX'\mathcal{I}_2^{-1}F}$. For the second and third term, using (17)-(18) and (19)-(20) respectively,

$$\begin{aligned}(\mathbf{X}'_i\Sigma_T^{-1}\mathbf{X}_i)^{-1}\mathbf{X}'_i\Sigma_T^{-1}\mathcal{D}_{N,T}\Sigma_T^{-1}(\mathbf{F}\mathbf{b}_i + \varepsilon_i) &= O_p(N^{-\frac{1}{2}}T^{-1}(T^a + T^{b-\frac{1}{2}})), \\ (\mathbf{X}'_i\Sigma_T^{-1}\mathbf{X}_i)^{-1}(\mathbf{X}'_i\Sigma_T^{-1}\mathcal{D}_{N,T}\Sigma_T^{-1}\mathbf{X}_i)(\mathbf{X}'_i\Sigma_T^{-1}\mathbf{X}_i)^{-1}\mathbf{X}'_i\Sigma_T^{-1}(\mathbf{F}\mathbf{b}_i + \varepsilon_i) &= O_p(N^{-\frac{1}{2}}T^{-\frac{3}{2}}(T^c + T^{d-\frac{1}{2}})),\end{aligned}$$

whereas the fourth can be easily seen to be $O_p\left(T^{-2}N^{-1}(T^c + T^{d-\frac{1}{2}})(T^a + T^{b-\frac{1}{2}})\right)$.

□

Proof of Theorem 2. This follows precisely the proof of Theorem 1 and is reported in the Supplementary Material.

References

- AHN, S. C., Y. HOON LEE, AND P. SCHMIDT (2001): “GMM estimation of linear panel data models with time-varying individual effects,” *Journal of Econometrics*, 101, 219–255.
- ANDREWS, D. (2005): “Cross-section regression with common shocks,” *Econometrica*, 73, 1551–1585.
- BAI, J. (2009): “Panel data models with interactive fixed effects,” *Econometrica*, 77, 1229–1279.
- BAI, J., AND S. NG (2002): “Determining the number of factors in approximate factor models,” *Econometrica*, 70, 191–221.
- BAI, J., AND S. NG (2004): “A PANIC attack on unit roots and cointegration,” *Econometrica*, 72, 1127–1177.
- BROWN, B. M. (1971): “Martingale central limit theorems,” *Annals of Mathematical Statistics*, 42, 59–66.
- CHAMBERLAIN, G. (1983): “Funds, factors, and diversification in arbitrage pricing models,” *Econometrica*, 51, 1305–1323.
- FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000): “The generalized dynamic-factor model: identification and estimation,” *The Review of Economics and Statistics*, 82, 540–554.
- HALLIN, M., AND R. LISKA (2007): “Determining the number of factors in the general dynamic factor model,” *Journal of the American Statistical Association*, 102, 603–617.
- HOLTZ-EAKIN, D., W. NEWEY, AND H. S. ROSEN (1988): “Estimating vector autoregressions with panel data,” *Econometrica*, 56, 1371–95.
- MAGNUS, J. R., AND H. NEUDECKER (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Probability & Mathematical Statistics. London: John Wiley & Sons.
- MOON, H., AND M. WEIDNER (2009): “Likelihood expansion for panel regression models with factors,” USC, preprint.

- MOON, H. R., AND B. PERRON (2003): “Testing for a unit root in panels with dynamic factors,” *Journal of Econometrics*, 122, 81–126.
- NEWBY, W., AND K. WEST (1987): “A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix,” *Econometrica*, 55, 703–708.
- ONATSKI, A. (2009): “Testing Hypotheses About the Number of Factors in Large Factor Models,” *Econometrica*, 77, 1447–1479.
- PESARAN, M. H. (2006): “Estimation and inference in large heterogeneous panels with a multifactor error structure,” *Econometrica*, 74, 967–1012.
- PESARAN, M. H., AND E. TOSETTI (2007): “Large panels with common factors and spatial correlations,” *IZA Discussion Paper n. 3032*.
- PESARAN, M. H., AND P. ZAFFARONI (2009): “Optimality and diversifiability of mean variance and arbitrage pricing portfolios,” Cambridge University and Imperial College London, preprint.
- PHILLIPS, P. C. B., AND D. SUL (2003): “Dynamic panel estimation and homogeneity testing under cross section dependence,” *Econometrics Journal*, 6, 217–259.
- (2007): “Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence,” *Journal of Econometrics*, 137, 162–188.
- ROBINSON, P. M., AND F. J. HIDALGO (1997): “Time series regression with long-range dependence,” *Annals of Statistics*, 25, 77–104.
- STOCK, J. H., AND M. W. WATSON (2002): “Macroeconomic forecasting using diffusion indexes,” *Journal of Business & Economic Statistics*, 20, 147–62.

Table 1: time regression with unit-specific coefficients
intercept term $\alpha_{i0} = 1$, $i = 1, \dots, N$.

	GLS						GLS (multi-step)					
	mean			rmse			mean			rmse		
(N, T)	30	100	300	30	100	300	30	100	300	30	100	300
60	0.944	–	–	0.523	–	–	0.976	–	–	0.531	–	–
200	0.967	0.951	–	0.518	0.315	–	0.986	0.987	–	0.527	0.309	–
600	0.981	0.982	0.955	0.524	0.308	0.200	0.994	0.998	0.991	0.531	0.310	0.184
	OLS						UGLS					
	mean			rmse			mean			rmse		
(N, T)	30	100	300	30	100	300	30	100	300	30	100	300
60	0.897	–	–	0.560	–	–	0.993	–	–	0.369	–	–
200	0.892	0.901	–	0.563	0.361	–	0.993	0.999	–	0.368	0.221	–
600	0.898	0.902	0.904	0.567	0.363	0.261	0.994	0.998	0.999	0.369	0.222	0.134

Note to Table 1: data are generated according to model

$$y_{it} = \alpha_{i0} + \beta_{i0}x_{it} + b_{i10}f_{1t} + b_{i20}f_{2t} + \varepsilon_{it}$$

with regressor $x_{it} = \gamma_{i0} + \delta_{i10}f_{1t} + \delta_{i30}f_{3t} + v_{it}$. Factor loadings are normally distributed random variables, *iid* across units and mutually independent, satisfying $b_{i10} \sim NID(1, 0.2)$, $b_{i20} \sim NID(0, 0.2)$, $\delta_{i10} \sim NID(0.5, 0.5)$, $\delta_{i30} \sim NID(0, 0.5)$. Latent common factors are $f_{j,t} = 0.5f_{j,t-1} + \sqrt{0.5}\eta_{jf,t}$, with $\eta_{jf,t} \sim NID(0, 1)$, mutually independent for $j = 1, 2, 3$, and idiosyncratic innovation are $\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{it-1} + \eta_{i\varepsilon,t}$ with $\eta_{i\varepsilon,t} \sim NID(0, \sigma_{i\varepsilon}^2(1 - \rho_{i\varepsilon}^2))$, $v_{it} = \rho_{iv}v_{it-1} + \eta_{iv,t}$, with $\eta_{iv,t} \sim NID(0, (1 - \rho_{iv}^2))$, with $\rho_{i\varepsilon} \sim UID(0.05, 0.95)$, $\rho_{iv} \sim UID(0.05, 0.95)$, $\sigma_{i\varepsilon}^2 \sim UID(0.5, 1.5)$, *iid* across $i = 1, \dots, N$ and mutually independent.

Parameters of interest are constant across replications and equal to $\alpha_{i0} = 1$, $\gamma_{i0} = 0.5$ and, assuming N even, $\beta_{i0} = 1$ for $i = 1, \dots, N/2$ and $\beta_{i0} = 3$ for $i = N/2 + 1, \dots, N$.

Panels headed by **mean** and **rmse** report, respectively, $N^{-1} \sum_{i=1}^N \left(M^{-1} \sum_{m=1}^M \hat{\alpha}_i^m \right)$ and $N^{-1} \sum_{i=1}^N \left(M^{-1} \sum_{m=1}^M (\hat{\alpha}_i^m - 1)^2 \right)^{\frac{1}{2}}$ with $M = 2,000$. Here $\hat{\alpha}_i^m$ denotes the estimate, based on either the GLS (top left panel), multi-step GLS with $J = 4$ steps (top right panel), OLS (bottom left panel) and UGLS (bottom right panel) of α_{i0} for the m Monte Carlo iteration.

**Table 3: cross-sectional regression with time-specific coefficients
intercept term $\alpha_{t0} = 1, t = 1, \dots, T$.**

	GLS						GLS (multi-step)					
	mean			rmse			mean			rmse		
(N, T)	60	200	600	60	200	600	60	200	600	60	200	600
30	0.953	0.968	0.973	0.242	0.238	0.237	0.971	0.984	0.985	0.244	0.238	0.237
100	–	0.959	0.977	–	0.141	0.130	–	0.983	0.995	–	0.132	0.128
300	–	–	0.967	–	–	0.095	–	–	0.987	–	–	0.076
	OLS						UGLS					
	mean			rmse			mean			rmse		
(N, T)	60	200	600	60	200	600	60	200	600	60	200	600
30	0.938	0.939	0.939	0.257	0.258	0.258	0.985	0.986	0.986	0.206	0.207	0.207
100	–	0.939	0.939	–	0.167	0.167	–	0.995	0.996	–	0.109	0.108
300	–	–	0.939	–	–	0.131	–	–	0.998	–	–	0.062

Note to Table 3: data are generated according to model

$$y_{it} = \alpha_{t0} + \beta_{t0}x_{it} + b_{i10}f_{1t} + b_{i20}f_{2t} + \varepsilon_{it},$$

with regressor $x_{it} = \gamma_{i0} + \delta_{i10}f_{1t} + \delta_{i30}f_{3t} + v_{it}$. Factor loadings are normally distributed random variables, *iid* across time and mutually independent, satisfying $b_{i10} \sim NID(0, 0.2)$, $b_{i20} \sim NID(0, 0.2)$, $\delta_{i10} = \beta_{i10}$, $\delta_{i30} = \beta_{i20}$. Latent common factors are $f_{j,t} = 0.5f_{j,t-1} + \sqrt{0.5}\eta_{jf,t}$, with $\eta_{jf,t} \sim NID(0, 1)$, mutually independent for $j = 1, 2, 3$, and idiosyncratic innovation are $\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{it-1} + \eta_{i\varepsilon,t}$ with $\eta_{i\varepsilon,t} \sim NID(0, \sigma_i^2(1 - \rho_{i\varepsilon}^2))$, $v_{it} = \rho_{iv}v_{it-1} + \eta_{iv,t}$, with $\eta_{iv,t} \sim NID(0, (1 - \rho_{iv}^2))$, with $\rho_{i\varepsilon} \sim UID(0.05, 0.95)$, $\rho_{iv} \sim UID(0.05, 0.95)$, $\sigma_{i\varepsilon}^2 \sim UID(0.5, 1.5)$, *iid* across $i = 1, \dots, N$ and mutually independent.

Parameters of interest are constant across replications and equal to $\alpha_{t0} = 1, \gamma_{t0} = 0.5$ and, assuming T even, $\beta_{t0} = 1$ for $t = 1, \dots, T/2$ and $\beta_{t0} = 2$ for $t = T/2 + 1, \dots, T$.

Panels headed by **mean** and **rmse** report, respectively, $T^{-1} \sum_{i=1}^T \left(M^{-1} \sum_{m=1}^M \hat{\alpha}_t^m \right)$ and $T^{-1} \sum_{i=1}^T \left(M^{-1} \sum_{m=1}^M (\hat{\alpha}_t^m - 1)^2 \right)^{\frac{1}{2}}$ with $M = 2,000$. Here $\hat{\alpha}_t^m$ denotes the estimate, based on either the GLS (top left panel), multi-step GLS with $J = 4$ steps (top right panel), OLS (bottom left panel) and UGLS (bottom right panel) of α_{t0} for the m Monte Carlo iteration.

Table 2: time regression with unit-specific coefficients
regression coefficient $\beta_{i0} = 3$, $i = N/2 + 1, \dots, N$.

	<i>GLS</i>						<i>GLS</i> (multi-step)					
	mean			rmse			mean			rmse		
(N, T)	30	100	300	30	100	300	30	100	300	30	100	300
60	3.105	–	–	0.314	–	–	3.041	–	–	0.277	–	–
200	3.053	3.095	–	0.228	0.227	–	3.013	3.024	–	0.215	0.146	–
600	3.034	3.037	3.091	0.209	0.133	0.198	3.012	3.004	3.019	0.200	0.111	0.090
	<i>OLS</i>						<i>UGLS</i>					
	mean			rmse			mean			rmse		
(N, T)	30	100	300	30	100	300	30	100	300	30	100	300
60	3.201	–	–	0.488	–	–	3.013	–	–	0.181	–	–
200	3.204	3.196	–	0.490	0.415	–	3.012	3.003	–	0.180	0.090	–
600	3.204	3.197	3.193	0.492	0.417	0.389	3.012	3.003	3.001	0.180	0.090	0.051

Note to Table 2: data are generated according to the same model described in Table 1.

Panels headed by **mean** and **rmse** report, respectively, $(N/2)^{-1} \sum_{i=N/2+1}^N \left(M^{-1} \sum_{m=1}^M \hat{\beta}_i^m \right)$ and $(N/2)^{-1} \sum_{i=N/2+1}^N \left(M^{-1} \sum_{m=1}^M (\hat{\beta}_i^m - 3)^2 \right)^{\frac{1}{2}}$ with $M = 2,000$. Here $\hat{\beta}_i^m$ denotes the estimate, based on either the GLS (top left panel), multi-step GLS with $J = 4$ steps (top right panel), OLS (bottom left panel) and UGLS (bottom right panel) of β_{i0} for the m th Monte Carlo iteration.

Table 4: cross-sectional regression with time-specific coefficients
regression coefficient $\beta_{t0} = 2$, $t = T/2 + 1, \dots, T$.

	<i>GLS</i>						<i>GLS</i> (multi-step)					
	mean			rmse			mean			rmse		
(N, T)	60	200	600	60	200	600	60	200	600	60	200	600
30	2.093	2.065	2.054	0.235	0.208	0.201	2.058	2.032	2.031	0.229	0.198	0.192
100	–	2.082	2.045	–	0.167	0.116	–	2.033	2.010	–	0.118	0.101
300	–	–	2.078	–	–	0.135	–	–	2.025	–	–	0.070
	<i>OLS</i>						<i>UGLS</i>					
	mean			rmse			mean			rmse		
(N, T)	60	200	600	60	200	600	60	200	600	60	200	600
30	2.121	2.121	2.122	0.292	0.291	0.291	2.028	2.027	2.028	0.190	0.189	0.190
100	–	2.120	2.120	–	0.241	0.240	–	2.001	2.009	–	0.099	0.098
300	–	–	2.121	–	–	0.225	–	–	2.003	–	–	0.056

Note to Table 4: data are generated according to the same model described in Table 3.

Panels headed by **mean** and **rmse** report, respectively, $(T/2)^{-1} \sum_{t=T/2+1}^T \left(M^{-1} \sum_{m=1}^M \hat{\beta}_t^m \right)$ and $(T/2)^{-1} \sum_{t=T/2+1}^T \left(M^{-1} \sum_{m=1}^M (\hat{\beta}_t^m - 2)^2 \right)^{\frac{1}{2}}$ with $M = 2,000$. Here $\hat{\beta}_t^m$ denotes the estimate, based on either the GLS (top left panel), multi-step GLS with $J = 4$ steps (top right panel), OLS (bottom left panel) and UGLS (bottom right panel) of β_{t0} for the m th Monte Carlo iteration.