GENERALIZED LEAST SQUARES ESTIMATION OF PANEL WITH COMMON SHOCKS [1]

Paolo Zaffaroni
Imperial College London
This draft: 13 February 2009

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 temporally and cross-sectionally, of the idiosyncratic innovation is permitted. Our results are presented separately for time regressions with unit-specific coefficients as well as for cross-section regressions with time-specific coefficients. As particular cases of our set up, we establish primitive conditions of our assumptions for Andrews (2005) and Pesaran (2006) regression models. A set of Monte Carlo experiments corroborate our results.

1 Introduction

Factor models represent one of the most popular and successful way to capture cross-sectional dependence, especially when facing a large 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 consistency of the OLS estimator is preserved, as N goes to infinity, when both the error and the regressors have a factor structure with uncorrelated factor loadings. The parameters estimate has a mixed normal asymptotic distribution. 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 by cross-sectional averages of the dependent and the individual-specific regressors. The conventional 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. It is here noted that, in either cases, the unfeasible generalized least squares (henceforth UGLS) estimator, based on the presumption that the covariance matrix of the factor structure is known, would be 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 GLS 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 would also be 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, 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.

The GLS estimator exhibits a number of 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 since it is not based on a preliminary estimation of the factor structure. Thus, we do not need to make use of the recent advances in estimation of (dynamic) factor models such as Forni, Hallin, Lippi, and L. (2000), Bai and Ng (2002), Stock and Watson (2002) and Bai (2003), which in turn would require preliminary testing of the number of factors (Bai and Ng (2002) and Hallin and Liska (2007) for tests designed for static and dynamic factor models, respectively).

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), and Phillips and Sul (2007). With the exception of Ahn, Hoon Lee, and Schmidt (2001), who focus on generalized method of moment estimation of cross-sectional regressions with independent and identically distributed (*i.i.d.*) regressors for fixed T the other papers are all 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 $O(T^2)$. 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 of with the idiosyncratic part. They do not treat the factor structure as a nuisance parameter but build their test on pre-estimated factors and idiosyncratic component by principal 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 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.

Factor models is not the only way to describe cross-sectional dependence. Weaker, in the sense of local, forms of dependence can be achieved by spatial econometrics approaches, in particular spatial autoregressive models (see Anselin (1988), Case (1991), Conley (1999), Chen and Conley (2001), Lee

(2004), Robinson (2006)).

This paper, which studies separately the cases of estimation of linear regressions with either individual-specific or time-specific parameters, proceeds as follows. The next section 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, 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 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 primite conditions for at least some our general assumptions. The full set of required primitive conditions is then described in a proposition. The 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, $\| \mathbf{A} \| = (\text{tr}(\mathbf{A}\mathbf{A}'))^{\frac{1}{2}}$ indicates the Euclidean norm of the matrix $\mathbf{A}$, ι_n is a $n \times 1$ vector of ones, $\mu_{i,a} = E\mathbf{a}_i$ for a random vector $\mathbf{a}_i$ and $\mathbf{\Sigma}_{i,aCb'}$ is the limit in probability of $\mathbf{A}_i'\mathbf{C}_i\mathbf{B}_i/T$ for random matrices $\mathbf{A}_i, \mathbf{B}_i$ with T rows and a finite number of columns and for the random $T \times T$ matrix $\mathbf{C}_i$ all possibly dependent on an index i. When $\mathbf{C}_i$ equals the identity matrix $\mathbf{I}_T$, we write $\mathbf{\Sigma}_{i,ab'}$. We skip dependence on the index i 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 ith unit can be expressed, in matrix form, as

$$\mathbf{y}_i = \mathbf{X}_i \beta_{i0} + \mathbf{u}_i, \tag{1}$$

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

$$\mathbf{u}_i = \mathbf{F}\mathbf{b}_i + \varepsilon_i, \tag{2}$$

for an unobserved $m \times 1$ vector of factor loadings $\mathbf{b}_i$, an unobserved $T \times m$ matrix of common factors $\mathbf{F} = (\mathbf{f}_1, ..., \mathbf{f}_T)'$ and an unobserved $T \times 1$ vector of idiosyncratic innovations $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$. 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, we assume that in fact a panel of observations $\{\mathbf{y}, \mathbf{X}\} = \{\mathbf{y}_1, ..., \mathbf{y}_i, ..., \mathbf{y}_N, \mathbf{X}_1, ..., \mathbf{X}_i, ..., \mathbf{X}_N\}$ 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 $E(\mathbf{b_i b_i'} \mid \mathbf{X}_i, \mathbf{F}) = \mathcal{B}_i > 0$ with $N^{-1} \sum_{i=1}^{N} \mathcal{B}_i \to_p \mathcal{B} > 0$ as $N \to \infty$.

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

For every i, $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{it}, ..., \varepsilon_{iT})'$ satisfies $E(\varepsilon_i \mid \mathbf{b}_i, \mathbf{X}_i, \mathbf{F}) = \mathbf{0}$ and

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

$$N^{-1} \sum_{i=1}^{N} \mathbf{H}_i \rightarrow_p \mathcal{H}_T > 0 \quad \text{as } N \rightarrow \infty. \tag{4}$$

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:

$$\frac{\mathbf{X}_i'\mathbf{X}_i}{T} \rightarrow_p \mathbf{\Sigma}_{i,xx'} > 0, \ \frac{\mathbf{X}_i'\mathbf{H}_i\mathbf{X}_i}{T} \rightarrow_p \mathbf{\Sigma}_{i,xHx'} > 0, \ \frac{\mathbf{X}_i'\mathbf{F}}{T} \rightarrow_p \mathbf{\Sigma}_{i,xf'}, \tag{5}$$

$$\frac{\mathbf{X}_i'\mathbf{H}_i^{-1}\mathbf{X}_i}{T} \rightarrow_p \mathbf{\Sigma}_{i,xH^{-1}x'} > 0, \frac{\mathbf{F}'\mathbf{H}_i^{-1}\mathbf{F}}{T} \rightarrow_p \mathbf{\Sigma}_{i,fH^{-1}f'} > 0, \ \frac{\mathbf{X}_i'\mathbf{H}_i^{-1}\mathbf{F}}{T} \rightarrow_p \mathbf{\Sigma}_{i,xH^{-1}f'}, \tag{6}$$

such that

$$\mathbf{\Sigma}_{i,xH^{-1}x'} - \mathbf{\Sigma}_{i,xH^{-1}f'}\mathbf{\Sigma}_{i,fH^{-1}f'}\mathbf{\Sigma}_{i,xH^{-1}f'}' > 0. \tag{7}$$

Assumption 5. $\mathcal{H}$ (limit conditions for GLS)
All the limit matrices below, as $N \rightarrow \infty$ and arbitrary T, are $a.s.$ finite:

$$\sum_{i=1}^{N} \frac{\mathbf{X}_i\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{\Sigma}_{i,xf'}\mathbf{b}_i\mathbf{b}_i'\mathbf{\Sigma}_{i,xf'}'\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{X}_i'}{N} = \mathcal{A}_{1T}(1+o_p(1)), \sum_{i=1}^{N} \frac{\mathbf{X}_i\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{\Sigma}_{i,xf'}\mathbf{b}_i\mathbf{b}_i'\mathbf{F}'}{N} = \mathcal{A}_{2T}(1+o_p(1)),$$

$$\sum_{i=1}^{N} \frac{\mathbf{X}_i\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{X}_i'\varepsilon_i\varepsilon_i'\mathbf{X}_i\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{X}_i'}{NT} = \mathcal{A}_{3T}(1 + o_p(1)), \sum_{i=1}^{N} \frac{\mathbf{X}_i\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{X}_i'\varepsilon_i\varepsilon_i'}{N} = \mathcal{A}_{4T}(1 + o_p(1)),$$

$$\sum_{i=1}^{N} \frac{\mathbf{b}_i\varepsilon_i'}{N^{\frac{1}{2}}} = \mathcal{C}_{1T}(1 + o_p(1)), \sum_{i=1}^{N} \frac{\mathbf{X}_i\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{\Sigma}_{i,xf'}\mathbf{b}_i\varepsilon_i'}{N^{\frac{1}{2}}} = \mathcal{C}_{2T}(1 + o_p(1)), \tag{8}$$

$$\sum_{i=1}^{N} \frac{\mathbf{b}_i\varepsilon_i'\mathbf{X}_i\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{X}_i'}{N^{\frac{1}{2}}T^{\frac{1}{2}}} = \mathcal{C}_{3T}(1 + o_p(1)), \sum_{i=1}^{N} \frac{\mathbf{X}_i\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{\Sigma}_{i,xf'}\mathbf{b}_i\varepsilon_i'\mathbf{X}_i\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{X}_i'}{N^{\frac{1}{2}}T^{\frac{1}{2}}} = \mathcal{C}_{4T}(1 + o_p(1)). \tag{9}$$

Assumption 6. $\mathcal{H}$ (distribution conditions for OLS and UGLS)

As $T \to \infty$:

$$\frac{\mathbf{X}_i'\varepsilon_{\mathbf{i}}}{T^{\frac{1}{2}}} \to_d \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{i,xHx'}), \tag{10}$$

$$\frac{\mathbf{X}_i'\mathbf{H}_i^{-1}\varepsilon_{\mathbf{i}}}{T^{\frac{1}{2}}} \to_d \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{i,xH^{-1}x'}), \frac{\mathbf{F}'\mathbf{H}_i^{-1}\varepsilon_{\mathbf{i}}}{T^{\frac{1}{2}}} \to_d \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{i,fH^{-1}f'}). \tag{11}$$

Assumption 7.$\mathcal{H}$ (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

$$\mathcal{A}_{1T} - (\mathcal{A}_{2T} + \mathcal{A}_{2T}') = \mathbf{F}\mathcal{D}_{1T}\mathbf{F}' + \mathcal{D}_{2T} \text{ and } \mathcal{A}_{3T} - (\mathcal{A}_{4T} + \mathcal{A}_{4T}') = \mathbf{F}\mathcal{E}_{1T}\mathbf{F}' + \mathcal{E}_{2T} \tag{12}$$

with an $m \times m$ non-singular symmetric $\mathcal{I}_{1T} = \mathcal{D}_{1T} + T^{-1}\mathcal{E}_{1T} + \mathcal{B}$ and a $T \times T$ non-singular $\mathcal{I}_{2T} = \mathcal{D}_{2T} + T^{-1}\mathcal{E}_{2T} + \mathcal{H}_T$ satisfying $a.s.$:

$$\frac{\mathbf{F}'\mathcal{I}_{2T}^{-1}\mathbf{F}}{T} \to_p \mathbf{\Sigma}_{f\mathcal{I}_2^{-1}f'} \text{ (non-singular)}, \frac{\mathbf{F}'\mathcal{I}_{2T}^{-1}\mathbf{H}_i\mathcal{I}_{2T}^{-1}\mathbf{F}}{T} \to_p \mathbf{\Sigma}_{i,f\mathcal{I}^{-1}\mathcal{H}\mathcal{I}_2^{-1}f'}, \tag{13}$$
$$\frac{\mathbf{X}_i'\mathcal{I}_{2T}^{-1}\mathbf{X}_i}{T} \to_p \mathbf{\Sigma}_{i,x\mathcal{I}_2^{-1}x'} \text{ (non-singular)}, \frac{\mathbf{X}_i'\mathcal{I}_{2T}^{-1}\mathbf{H}_i\mathcal{I}_{2T}^{-1}\mathbf{X}_i}{T} \to_p \mathbf{\Sigma}_{i,x\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}x'},$$
$$\frac{\mathbf{X}_i'\mathcal{I}_{2T}^{-1}\mathbf{F}}{T} \to_p \mathbf{\Sigma}_{i,x\mathcal{I}_2^{-1}f'}, \frac{\mathbf{X}_i'\mathcal{I}_{2T}^{-1}\mathbf{H}_i\mathcal{I}_{2T}^{-1}\mathbf{F}}{T} \to_p \mathbf{\Sigma}_{i,x\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}f'},$$
$$\frac{\mathbf{X}_i'\mathcal{I}_{2T}^{-1}\varepsilon_i}{T^{\frac{1}{2}}} \to_d \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{i,x\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}x'}), \frac{\mathbf{F}'\mathcal{I}_{2T}^{-1}\varepsilon_i}{T^{\frac{1}{2}}} \to_d \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{i,f\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}f'}),$$

where all the limits above hold as $T \to \infty$ with $a.s.$ finite limit matrices and, setting

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

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

$$\mathbf{X}_i'\mathbf{\Sigma}_T^{-1}(\mathbf{F}\mathcal{C}_{1T} + \mathcal{C}_{1T}'\mathbf{F}' + \mathcal{C}_{2T} + \mathcal{C}_{2T}')\mathbf{\Sigma}_T^{-1}(\mathbf{F}\mathbf{b}_i + \varepsilon_i) = O_p(T^a\iota_k), \tag{14}$$
$$\mathbf{X}_i'\mathbf{\Sigma}_T^{-1}(\mathbf{F}\mathcal{C}_{3T} + \mathcal{C}_{3T}'\mathbf{F}' + \mathcal{C}_{4T} + \mathcal{C}_{4T}')\mathbf{\Sigma}_T^{-1}(\mathbf{F}\mathbf{b}_i + \varepsilon_i) = O_p(T^b\iota_k), \tag{15}$$
$$\mathbf{X}_i'\mathbf{\Sigma}_T^{-1}(\mathbf{F}\mathcal{C}_{1T} + \mathcal{C}_{1T}'\mathbf{F}' + \mathcal{C}_{2T} + \mathcal{C}_{2T}')\mathbf{\Sigma}_T^{-1}\mathbf{X}_i = O_p(T^c\iota_k\iota_k'), \tag{16}$$
$$\mathbf{X}_i'\mathbf{\Sigma}_T^{-1}(\mathbf{F}\mathcal{C}_{3T} + \mathcal{C}_{3T}'\mathbf{F}' + \mathcal{C}_{4T} + \mathcal{C}_{4T}')\mathbf{\Sigma}_T^{-1}\mathbf{X}_i = O_p(T^d\iota_k\iota_k'). \tag{17}$$

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 for concrete models.

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 $\mathbf{f}_t$, who can be autocorrelated. One of the sufficient conditions for Assumption 7.$\mathcal{H}$ will be, however, bounded-ness of $\mathbf{\Sigma}_{ff'}$. 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 $\mathbf{X}_i$. 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 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 OLS problems.
5. When $\mathbf{\Sigma}_{i,xx'}$ and $\mathbf{\Sigma}_{ff'}$ are finite, then the other limit matrices are finite by Schwartz inequality requiring, for certain cases, that the maximum eigenvalue of $\mathbf{H}_i$ is bounded and its minimum eigenvalue is bounded away from zero, uniformly in T. Bounded-ness of the maximum eigenvalue is implied when $\mathbf{u}_i$ satisfy an approximate factor structure Chamberlain (1983).

Note that $\mathbf{\Sigma}_{i,xf'}$ 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, unless it is a matrix of zeros or, if not, 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

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

and, as seen below, 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, as stated in (7).
6. 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. Similarly, the limiting distribution results stated in Assumptions 6.$\mathcal{H}$ and 7.$\mathcal{H}$, are required for OLS, UGLS and GLS respectively. Since we aim at providing general results, we do not specify here the primitive conditions required, although these can be relatively easily established when one considers particular cases of (1) such as for Pesaran (2006)'s model, examined in section 2.3.
7. As explained below, considering the GLS will imply to consider $\mathbf{\Sigma}_T$ 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 $\mathbf{\Sigma}_T^{-1}$ will be (as $\mathbf{S}_i^{-1}$) 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.
8. Conditions (14)-(17) 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

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

setting

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

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 $\hat{u}_1'\mathbf{D} = \mathbf{0}$ for all i. As a consequence, $\hat{\mathbf{\Sigma}}_T$ 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

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

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

Theorem 1 *(unit-specific parameters)*

(i) (OLS) When Assumptions 3.$\mathcal{H}$, 4.$\mathcal{H}$.(5), 6.$\mathcal{H}$.(10) hold

$$T^{\frac{1}{2}}(\hat{\beta}_i^{OLS} - \beta_{i0} - \gamma_i^{OLS}) \rightarrow_d \mathcal{N}_k(\mathbf{0}, \mathcal{V}_i^{OLS}) \text{ as } T \rightarrow \infty,$$

setting

$$\gamma_i^{OLS} = \mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{\Sigma}_{i,xf'}\mathbf{b}_i, \;\; \mathcal{V}_i^{OLS} = \mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{\Sigma}_{i,xHx'}\mathbf{\Sigma}_{i,xx'}^{-1}.$$

(ii) (UGLS) When Assumptions 1.$\mathcal{H}$, 2.$\mathcal{H}$.(3), 3.$\mathcal{H}$, 4.$\mathcal{H}$.(6), 6.$\mathcal{H}$.(11)

$$T^{\frac{1}{2}}(\hat{\beta}_i^{UGLS} - \beta_{i0}) \rightarrow_d \mathcal{N}_k(\mathbf{0}, \mathcal{V}_i^{UGLS}) \text{ as } T \rightarrow \infty,$$

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*

$$(\mathcal{M}_i^{UGLS})^{-1} = \mathcal{N}_i^{UGLS}.$$

(iii) (GLS) When Assumptions 1.$\mathcal{H}$, 2.$\mathcal{H}$.(4), 3.$\mathcal{H}$, 4.$\mathcal{H}$.(5) and (7), 5.$\mathcal{H}$, 7.$\mathcal{H}$

$$\hat{\beta}_i^{GLS} \rightarrow_p \beta_{i0} \qquad \text{as } \frac{1}{T} + \frac{T^{max(a-1,b-\frac{3}{2})}}{N^{\frac{1}{2}}} + \frac{T^{max(c-\frac{3}{2},d-2)}}{N^{\frac{1}{2}}} \rightarrow 0,$$

$$T^{\frac{1}{2}}(\hat{\beta}_i^{GLS} - \beta_{i0}) \rightarrow_d \mathcal{N}_k(\mathbf{0}, \mathcal{V}_i^{GLS}) \qquad \text{as } \frac{1}{T} + \frac{T^{max(a-\frac{1}{2},b-1)}}{N^{\frac{1}{2}}} + \frac{T^{max(c-1,d-\frac{3}{2})}}{N^{\frac{1}{2}}} \rightarrow 0,$$

setting

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

with $\mathcal{M}_i^{GLS} = \text{plim}_{(N,T)\rightarrow\infty} T^{-1}(\mathbf{X}_i'\hat{\mathbf{\Sigma}}_T^{-1}\mathbf{X}_i)$, $\mathcal{N}_i^{GLS} = \text{plim}_{(N,T)\rightarrow\infty} T^{-1}\mathbf{X}_i'\hat{\mathbf{\Sigma}}_T^{-1}\mathbf{H}_i\hat{\mathbf{\Sigma}}_T^{-1}\mathbf{X}_i$.

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 $\gamma_i^{OLS} = \mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{\Sigma}_{i,xf'}\mathbf{b}_i$ is a random variable. Consistency is achieved if either $\mathbf{b}_i = \mathbf{0}$, meaning no factor structure, or $\mathbf{\Sigma}_{i,xf'} = \mathbf{0}$, 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 estimator improves efficiency with respect the the OLS estimator for non-spherical innovations. Here we find that UGLS exhibits a more profound property: it completely eliminates the factor structure's adverse effect on OLS of inducing an asymptotic bias. 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. There, the GLS estimator is $T^{\frac{1}{2}}$-consistent and asymptotically normal whereas the same properties are not warranted for the OLS estimator, under the same set of assumptions.
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. This result has been used, in the different context of financial portfolio optimization, by Pesaran and Zaffaroni (2008) who establish that mean-variance trading strategies do allow complete diversification of both idiosyncratic and common shocks to asset returns.
4. The feasible GLS estimator here proposed is denoted GLS since it 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 why GLS works is that , although $\hat{\mathbf{\Sigma}}_T$ is a non-consistent estimate of the true covariance matrix $\mathbf{S}_i$ (in the sense of element by element), its limit $\mathbf{\Sigma}_T = \mathbf{F}\mathcal{I}_{1T}\mathbf{F}' + \mathcal{I}_{2T},$, once taking the inverse, 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 would suggest to use OLS.
5. Since GLS delivers consistent parameter estimates, this suggests a two-step approach, achieving a more efficient estimator. The first stage consists

of getting the GLS estimator $\hat{\beta}_i$ as described above. Next, one can evaluate $\tilde{\mathbf{\Sigma}}_T = N^{-1}\sum_{i=1}^{N}\tilde{u}_i\tilde{u}_i'$ for $\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\beta}_i^{GLS}$ in order to get the *second-step* GLS estimator $\tilde{\beta}_i^{GLS} = (\mathbf{X}_i'\tilde{\mathbf{\Sigma}}_T^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i\tilde{\mathbf{\Sigma}}_T^{-1}\mathbf{y}_i$. Given a set of conditions that build on Assumptions $\mathcal{H}$, one can show that $\tilde{\beta}_i^{GLS}$ is also consistent and asymptotically normal, as N, T diverge to infinity (as some rate). Moreover, it can be shown that $\tilde{\mathbf{\Sigma}}_T \rightarrow_p \mathbf{S}_i + T^{-1}\mathbf{R}_i$, as $N \rightarrow \infty$, for a $T \times T$ matrix $\mathbf{R}_i$ satisfying $\sup_T \| T^{-1}\mathcal{R}_i \| = O(1)\, a.s.$ and where each element of $T^{-1}\mathbf{R}_i$ goes to zero as $T \rightarrow \infty$. Hence, $\tilde{\Sigma}_T$ 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 be achieved by using the two-stage GLS estimator $\tilde{\beta}_i$ and, indeed, such improvement can be substantial when N, T are both sizeable. Below we report some Monte Carlo results in order to gauge these possible improvements of efficiency in finite samples.

2.3 Particular model: Pesaran (2006)

The model is

$$y_{it} = \alpha_{0i}'\mathbf{d}_t + \beta_{0i}'\mathbf{x}_{it} + e_{it}, \tag{18}$$

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

$$\mathbf{x}_{it} = \mathbf{A}_i'\mathbf{d}_t + \mathbf{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{it} \tag{19}$$

where $\mathbf{f}_t$ is the $m \times 1$ vector of *unobserved* factors, $\mathbf{A}_i$, $\mathbf{\Gamma}_i$ 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

$$e_{it} \quad = \quad \mathbf{f}_t'\gamma_i + \varepsilon_{it}, \tag{20}$$

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, (19)-(20) imply

$$\mathbf{F} = (\mathbf{f}_1 ... \mathbf{f}_t ... \mathbf{f}_T)', \ \mathbf{B} = (\gamma_1 ... \gamma_i ... \gamma_N)', \ \mathbf{X}_i = (\mathbf{D}, \mathbf{X}_i^*),$$

where we set $\mathbf{X}_i^* = \mathbf{D}\mathbf{A}_i + \mathbf{F}\mathbf{\Gamma}_i + \mathbf{V}_i$ with $\mathbf{D} = (\mathbf{d}_1 ... \mathbf{d}_t ... \mathbf{d}_T)'$, $\mathbf{V}_i = (\mathbf{v}_{i1} ... \mathbf{v}_{it} ... \mathbf{v}_{iT})'$.

We now verify the extent to which the assumptions of Pesaran (2006) imply our Theorem 1, part (iii). It turns out that our conditions are both weaker and stronger than Pesaran (2006) depending on the circumstances. 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 $\mathcal{I}_{2T}$.

Assumption 1.$\mathcal{H}$ follows by the strong law of large numbers (LLN) and Pesaran (2006, Assumption 3) where $\mathcal{B}_i = \mathcal{B}$ equal 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 (3) is, 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.$\mathcal{H}$, (5) follows by strengthening Pesaran (2006, Assumption 1 and 2) to fourth-order covariance stationarity with absolute summable autocovariances,

$$\frac{\mathbf{X}_i'\mathbf{X}_i}{T} \rightarrow_p \mathbf{\Sigma}_{i,xx'} ==$$
$$\begin{pmatrix} \mathbf{\Sigma}_{dd'} & \mathbf{\Sigma}_{dd'}\mathbf{A}_i + \mathbf{\Sigma}_{df'}\mathbf{\Gamma}_i \\ \mathbf{A}_i'\mathbf{\Sigma}_{dd'} + \mathbf{\Gamma}_i'\mathbf{\Sigma}_{fd'} & \mathbf{\Sigma}_{vv'} + \mathbf{A}_i'\mathbf{\Sigma}_{dd'}\mathbf{A}_i + \mathbf{\Gamma}_i'\mathbf{\Sigma}_{ff'}\mathbf{\Gamma}_i + \mathbf{A}_i'\mathbf{\Sigma}_{df'}\mathbf{\Gamma}_i + \mathbf{\Gamma}_i'\mathbf{\Sigma}_{fd'}\mathbf{A}_i \end{pmatrix},$$
$$\frac{\mathbf{X}_i'\mathbf{F}}{T} \rightarrow_p \mathbf{\Sigma}_{i,xf'} = \begin{pmatrix} \mathbf{\Sigma}_{df'} \\ \mathbf{\Gamma}_i'\mathbf{\Sigma}_{ff'} + \mathbf{A}_i'\mathbf{\Sigma}_{df'} \end{pmatrix},$$

since $\mathbf{\Sigma}_{fv'}$ and $\mathbf{\Sigma}_{dv'}$ are both matrices of zeros by Pesaran (2006, Assumption 1 and 2). By the same assumptions, $\mathbf{\Sigma}_{i,xx'}$ is bounded and, using the block matrix decomposition (Magnus and Neudecker 1988), is non-singular whenever both matrices

$$\mathbf{\Sigma}_{dd'},\ \mathbf{\Sigma}_{vv'} - \mathbf{\Gamma}_i'\mathbf{\Sigma}_{(fd')}\mathbf{\Gamma}_i,$$

are non-singular, where we set $\mathbf{\Sigma}_{(fd')} = \mathbf{\Sigma}_{fd'}\mathbf{\Sigma}_{dd'}^{-1}\mathbf{\Sigma}_{df'} - \mathbf{\Sigma}_{ff'}$. The latter requires $\mathbf{\Sigma}_{vv'} > 0$, implied by Pesaran (2006, Assumption 2) who defines it as $\mathbf{\Sigma}_i$, since $-\mathbf{\Sigma}_{(fd')}$ is positive semi definite, in fact at most $\mathbf{0}$ for perfectly correlated $\mathbf{f}_t$, $\mathbf{d}_t$. However, we require in addition $\mathbf{\Sigma}_{dd'} > 0$. Expression for $\mathbf{\Sigma}_{i,xHx'}$ 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 $\mathbf{\Sigma}_{i,xx'}$ implies, by the spectral decomposition of positive definite matrices, boundedness of $\mathbf{\Sigma}_{i,xHx'}$. Note, however, that the UGLS estimator does need $\mathbf{H}_i > 0$, as in Assumption 2.$\mathcal{H}$.(3), which in turn requires the spectral density of the ε_{it} to be bounded away from zero, ensuring boundedness of $\mathbf{\Sigma}_{i,xH^{-1}x'}$, $\mathbf{\Sigma}_{i,fH^{-1}f'}$, $\mathbf{\Sigma}_{i,xH^{-1}f'}$. Concerning Assumption 5.$\mathcal{H}$, setting $\mathbf{C}_i = (\mathbf{A}_i + \mathbf{\Sigma}_{dd'}^{-1}\mathbf{\Sigma}_{df'}\mathbf{\Gamma}_i)$, one obtains

$$\mathbf{\Sigma}_{i,xx'}^{-1} = \begin{pmatrix} \mathbf{\Sigma}_{dd'}^{-1} + \mathbf{C}_i(\mathbf{\Sigma}_{vv'} - \mathbf{\Gamma}_i'\mathbf{\Sigma}_{(fd')}\mathbf{\Gamma}_i)^{-1}\mathbf{C}_i' & -\mathbf{C}_i(\mathbf{\Sigma}_{vv'} - \mathbf{\Gamma}_i'\mathbf{\Sigma}_{(fd')}\mathbf{\Gamma}_i)^{-1} \\ -(\mathbf{\Sigma}_{vv'} - \mathbf{\Gamma}_i'\mathbf{\Sigma}_{(fd')}\mathbf{\Gamma}_i)^{-1}\mathbf{C}_i' & (\mathbf{\Sigma}_{vv'} - \mathbf{\Gamma}_i'\mathbf{\Sigma}_{(fd')}\mathbf{\Gamma}_i)^{-1} \end{pmatrix}$$

and

$$\mathbf{X}_i\boldsymbol{\Sigma}_{i,xx'}^{-1}\boldsymbol{\Sigma}_{i,xf'} = \mathbf{D}\boldsymbol{\Sigma}_{dd}^{-1}\boldsymbol{\Sigma}_{df'}+(\mathbf{D}\boldsymbol{\Sigma}_{dd}^{-1}\boldsymbol{\Sigma}_{df'}\boldsymbol{\Gamma}_i-\mathbf{F}\boldsymbol{\Gamma}_i-\mathbf{V}_i)(\boldsymbol{\Sigma}_{vv'}-\boldsymbol{\Gamma}_i'\boldsymbol{\Sigma}_{(fd')}\boldsymbol{\Gamma}_i)^{-1}\boldsymbol{\Gamma}_i'\boldsymbol{\Sigma}_{(fd')}.$$

. Further manipulations yield

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

setting

$$\begin{aligned}
&N^{-1}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i(\boldsymbol{\Sigma}_{vv'}-\boldsymbol{\Gamma}_i'\boldsymbol{\Sigma}_{(fd')}\boldsymbol{\Gamma}_i)^{-1}\boldsymbol{\Gamma}_i' \to_p \mathbf{P}_{1T},\\
&N^{-1}\sum_{i=1}^{N}\left(\boldsymbol{\Gamma}_i(\boldsymbol{\Sigma}_{vv'}-\boldsymbol{\Gamma}_i'\boldsymbol{\Sigma}_{(fd')}\boldsymbol{\Gamma}_i)^{-1}\boldsymbol{\Gamma}_i'\boldsymbol{\Sigma}_{(fd')}\mathcal{B}\boldsymbol{\Sigma}_{(fd')}\boldsymbol{\Gamma}_i(\boldsymbol{\Sigma}_{vv'}-\boldsymbol{\Gamma}_i'\boldsymbol{\Sigma}_{(fd')}\boldsymbol{\Gamma}_i)^{-1}\boldsymbol{\Gamma}_i'\right) \to_p \mathbf{P}_{2T},\\
&N^{-1}\sum_{i=1}^{N}\left(\mathbf{V}_i(\boldsymbol{\Sigma}_{vv'}-\boldsymbol{\Gamma}_i'\boldsymbol{\Sigma}_{(fd')}\boldsymbol{\Gamma}_i)^{-1}\boldsymbol{\Gamma}_i'\boldsymbol{\Sigma}_{(fd')}\mathcal{B}\boldsymbol{\Sigma}_{(fd')}\boldsymbol{\Gamma}_i(\boldsymbol{\Sigma}_{vv'}-\boldsymbol{\Gamma}_i'\boldsymbol{\Sigma}_{(fd')}\boldsymbol{\Gamma}_i)^{-1}\mathbf{V}_i'\right) \to_p \mathbf{P}_{3T},\\
&N^{-1}\sum_{i=1}^{N}\mathbf{V}_i(\boldsymbol{\Sigma}_{vv'}-\boldsymbol{\Gamma}_i'\boldsymbol{\Sigma}_{(fd')}\boldsymbol{\Gamma}_i)^{-1}\mathbf{V}_i' \to_p \mathbf{P}_{4T}.
\end{aligned}$$

Likewise $\mathcal{A}_{2T} = \mathbf{D}\boldsymbol{\Sigma}_{dd'}^{-1}\boldsymbol{\Sigma}_{df'}\mathcal{B}\mathbf{F}' + \mathbf{D}\boldsymbol{\Sigma}_{dd'}^{-1}\boldsymbol{\Sigma}_{df'}\mathbf{P}_{1T}\boldsymbol{\Sigma}_{(fd')}\mathcal{B}\mathbf{F}' - \mathbf{F}\mathbf{P}_{1T}\boldsymbol{\Sigma}_{(fd')}\mathcal{B}\mathbf{F}'$. 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 in implied, using a strong LLN argument, by Pesaran (2006, Assumptions 2 and 3). No additional moment conditions on the $\boldsymbol{\Gamma}_i$ are required since $\sup_{\boldsymbol{\Gamma}_i} \parallel \boldsymbol{\Gamma}_i(\boldsymbol{\Sigma}_{vv'}-\boldsymbol{\Gamma}_i'\boldsymbol{\Sigma}_{(fd')}\boldsymbol{\Gamma}_i)^{-1}\boldsymbol{\Gamma}_i' \parallel = O(1)$ a.s. and the $\mathbf{V}_i$ have bounded fourth moment by Pesaran (2006, Assumption 2). By simple manipulations, $\mathbf{X}_i\boldsymbol{\Sigma}_{i,xx'}^{-1}\mathbf{X}_i'$ equals $\mathbf{D}\boldsymbol{\Sigma}_{dd'}^{-1}\mathbf{D}'+[(\mathbf{D}\boldsymbol{\Sigma}_{dd'}^{-1}\boldsymbol{\Sigma}_{df'}-\mathbf{F})\boldsymbol{\Gamma}_i - \mathbf{V}_i](\boldsymbol{\Sigma}_{vv'}-\boldsymbol{\Gamma}_i'\boldsymbol{\Sigma}_{(fd')}\boldsymbol{\Gamma}_i')^{-1}[\boldsymbol{\Gamma}_i'(\mathbf{D}\boldsymbol{\Sigma}_{dd'}^{-1}\boldsymbol{\Sigma}_{df'}-\mathbf{F})'-\mathbf{V}_i']$, 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). Assumptions (8) and (9) follow by direct use of the CTL which holds under suitable assumptions. For instance, when $\|\mathbf{b}_i\|^{2+\delta} < \infty$ and $|\varepsilon_{it}|^{2+\delta} < \infty$, some $\delta > 0$, and Pesaran (2006, Assumption 2) hold with in addition *i.i.d.*-ness of the ε_{it} across i, then the Lyapunov condition holds and the t-th column of $\mathcal{C}_{1T}$ satisfies $\mathcal{C}_{1tT}^{\frac{1}{2}}\zeta_{1t}$ for a normally distributed $m \times 1$ vector ζ_{1t} with mean zero and unit covariance matrix and $N^{-1}\sum_{i,j=1}^{N}\mathbf{b}_i\varepsilon_{it}\varepsilon_{jt}\mathbf{b}_j' \rightarrow_p \mathcal{C}_{1tT}$ 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, $\mathbf{H}_t = [h_{ij,t}]_{i,j=1}^{N} = E(\varepsilon_t\varepsilon_t' \mid \mathbf{b}_i, \mathbf{b}_j, \mathbf{X}_i, \mathbf{X}_j, \mathbf{F})$ have 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 tth column of $\mathcal{C}_{2T}$ one gets $\mathcal{C}_{2tT}^{\frac{1}{2}}\zeta_{2t}$ for a $T \times 1$ normally distributed vector ζ_{2t} with mean zero and unit covariance matrix, where boundedness of $\mathcal{C}_{2tT}$ requires $E\ \|\mathbf{D} + \mathbf{F} + \mathbf{V}_i\|^{2+\delta} < \infty$. Similar results apply to (9) where now the Lyapunov condition require, in addition, $E\,\|\mathbf{D} + \mathbf{F} + \mathbf{V}_i\|^{6+\delta} < \infty$.

Concerning Assumption 7.$\mathcal{H}$, (12) follows for

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

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

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

Now non-singularity of $\mathcal{I}_{1T} = \mathcal{D}_{1T} + T^{-1}\mathcal{E}_{1T} + \mathcal{B}$ requires

$$\mathcal{B} + \mathbf{P}_{2T} + \mathbf{P}_{1T}\boldsymbol{\Sigma}_{(fd')}\mathcal{B} + \mathcal{B}\boldsymbol{\Sigma}_{(fd')}\mathbf{P}_{1T} - T^{-1}\mathbf{P}_{1T} \text{ non-singular.} \tag{21}$$

Moreover, for (13), given

$$\mathcal{I}_{2T} = \mathcal{H}_T + \mathbf{P}_{3T} - T^{-1}(\mathbf{D}(\boldsymbol{\Sigma}_{dd'}^{-1} - \mathbf{I}_n)\mathbf{D}' + \mathbf{P}_{4t}) + (\mathbf{F} - \mathbf{D}\boldsymbol{\Sigma}_{dd'}^{-1}\boldsymbol{\Sigma}_{df'})\mathcal{I}_{1T}(\mathbf{F}' - \boldsymbol{\Sigma}_{fd'}\boldsymbol{\Sigma}_{dd'}^{-1}\mathbf{D}') - \mathbf{F}\mathcal{I}_{1T}\mathbf{F}'.$$

one needs

$$\boldsymbol{\Sigma}_{fd'} = \mathbf{0} \tag{22}$$

for otherwise $T^{-1}\mathbf{F}'\mathcal{I}_{2T}^{-1}\mathbf{F} \to_p \mathbf{0}$ by the Central Lemma ($\mathbf{C}$, $\mathbf{F}$, $-\mathcal{I}_{1T}, T$), setting $\mathbf{C} = \mathcal{H}_T + \mathbf{P}_{3T} - T^{-1}(\mathbf{D}(\mathbf{\Sigma}_{dd'}^{-1} - \mathbf{I}_n)\mathbf{D}' + \mathbf{P}_{4t}) + (\mathbf{F} - \mathbf{D}\mathbf{\Sigma}_{dd'}^{-1}\mathbf{\Sigma}_{df'})\mathcal{I}_{1T}(\mathbf{F}' - \mathbf{\Sigma}_{fd'}\mathbf{\Sigma}_{dd'}^{-1}\mathbf{D}')$. Sufficient conditions for (22) are

$$\mu_f = \mathbf{0} \text{ and } \{\mathbf{f}_t, \mathbf{d}_t\} \text{ contemporaneously uncorrelated.}$$

Uncorrelatedness follow simply when $\mathbf{d}_t$ is deterministic, including intercept term, trends or seasonal dummies. Hence, under (22)

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

Moreover $\mathbf{\Sigma}_{(fd')} = -\mathbf{\Sigma}_{ff'}$ and, by taking into consideration the definitions of $\mathbf{P}_{1T}, \mathbf{P}_{2T}$, (21) can be expressed as the limit of

$$N^{-1}\sum_{i=1}^{N}\left(\mathbf{\Gamma}_i(\mathbf{\Sigma}_{vv'} + \mathbf{\Gamma}_i'\mathbf{\Sigma}_{ff'}\mathbf{\Gamma}_i)^{-1}\mathbf{\Gamma}_i'\mathbf{\Sigma}_{ff'} - \mathbf{I}_m\right)\mathcal{B}\left(\mathbf{\Gamma}_i(\mathbf{\Sigma}_{vv'} + \mathbf{\Gamma}_i'\mathbf{\Sigma}_{ff'}\mathbf{\Gamma}_i)^{-1}\mathbf{\Gamma}_i'\mathbf{\Sigma}_{ff'} - \mathbf{I}_m\right)'$$
$$-T^{-1}N^{-1}\sum_{i=1}^{N}\mathbf{\Gamma}_i(\mathbf{\Sigma}_{vv'} + \mathbf{\Gamma}_i'\mathbf{\Sigma}_{ff'}\mathbf{\Gamma}_i)^{-1}\mathbf{\Gamma}_i' \to_p \mathcal{C}_{1T} + \mathcal{B} + T^{-1}\mathcal{D}_{1T} = \mathcal{I}_{1T} \text{ non-singular.} \quad (23)$$

A sufficient condition for (23) is non-singularity of $(\mathbf{\Gamma}_i(\mathbf{\Sigma}_{vv'} + \mathbf{\Gamma}_i'\mathbf{\Sigma}_{ff'}\mathbf{\Gamma}_i)^{-1}\mathbf{\Gamma}_i'\mathbf{\Sigma}_{ff'} - \mathbf{I}_m)$ for any i but in fact a milder condition might suffice. Set, as an example, $\mathbf{\Sigma}_{ff'} = \mathcal{B} = \mathbf{I}_m$ and $\mathbf{\Sigma}_{vv'} = \mathbf{I}_k$. For $m > k = 1$, (23) is equivalent to obtain a non-singular limit of

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

which can be obtained under mild conditions on the $\mathbf{\Gamma}_i$ since each $\left(\mathbf{I}_m - \frac{(2+\mathbf{\Gamma}_i'\mathbf{\Gamma}_i)}{(1+\mathbf{\Gamma}_i'\mathbf{\Gamma}_i)^2}\mathbf{\Gamma}_i\mathbf{\Gamma}_i'\right)$ is non-singular for all i. Instead, when $k > m = 1$ then (23) is equivalent to obtaining a non-zero limit of

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

where it easily follows that each of the addenda is non-zero. Similar arguments follow for the case $m = k$. Finally, notice that $\mathbf{\Sigma}_{vv'} > \mathbf{0}$ is strictly required, ruling out the possibility that the regressors $\mathbf{x}_{it}$ obey a pure factor structure $\mathbf{x}_{it} = \mathbf{A}_i'\mathbf{d}_t + \mathbf{\Gamma}_i'\mathbf{f}_t$, otherwise (7) fails.

Closed-form expressions for $\mathbf{\Sigma}_{i,x\mathcal{I}_2^{-1}x'}$, $\mathbf{\Sigma}_{i,x\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}x'}$, $\mathbf{\Sigma}_{i,f\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}f'}$, $\mathbf{\Sigma}_{i,x\mathcal{I}_2^{-1}f'}$, $\mathbf{\Sigma}_{i,x\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}f'}$, however, required for the verification of the CTL conditions, 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 (14)-(17). 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 $\mathbf{\Sigma}_T^{-1}$. In addition, denoting by $\mathcal{C}_{1Tj}$ the jth row of $\mathcal{C}_{1T}$, we will require $\sup_T \parallel E\mathcal{C}'_{1Tj}\mathcal{C}_{1Tj} \parallel = O(1)$ 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, $\mathbf{X}_i'\mathbf{\Sigma}_T^{-1}\mathcal{C}'_{1T} = O_p(T^{\frac{1}{2}}\iota_{n+k}\iota'_m)$, $\mathcal{C}_{1T}\mathbf{\Sigma}_T^{-1}\varepsilon_i = O_p(T^{\frac{1}{2}}\iota_m)$ and, by repeated use of Lemma 2, $\mathbf{F}'\mathbf{\Sigma}_T^{-1}\varepsilon_i = O_p(T^{-\frac{1}{2}}\iota_m)$, $\mathbf{F}'\mathbf{\Sigma}_T^{-1}\mathcal{C}'_{1T} = O_p(T^{-\frac{1}{2}}\iota_m\iota'_m)$, $\mathbf{F}'\mathbf{\Sigma}_T^{-1}\mathbf{X}_i = O_p(\iota_{n+k}\iota'_m)$, $\mathbf{F}'\mathbf{\Sigma}_T^{-1}\mathbf{F} = O_p(\iota_m\iota'_m)$ yielding

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

Similarly, since under (22), $\mathbf{X}_i\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{\Sigma}_{i,xf'} = (\mathbf{F}\mathbf{\Gamma}_i + \mathbf{V}_i)(\mathbf{\Sigma}_{vv'} + \mathbf{\Gamma}_i'\mathbf{\Sigma}_{ff'}\mathbf{\Gamma}_i)^{-1}\mathbf{\Gamma}_i'\mathbf{\Sigma}_{ff'}$, one gets $\mathcal{C}_{2T} = \mathbf{F}\mathcal{C}_{21T} + \mathcal{C}_{22T}$ for $a.s.$ random, mean zero, matrixes 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, $\mathbf{X}_i'\mathbf{\Sigma}_T^{-1}\mathcal{C}_{22T}\mathbf{\Sigma}_T^{-1}\mathbf{F} = O_p(\iota_{n+k}\iota'_m)$, $\mathbf{X}_i'\mathbf{\Sigma}_T^{-1}\mathcal{C}_{22T}\mathbf{\Sigma}_T^{-1}\varepsilon_i = O_p(T^{\frac{1}{2}}\iota_{n+k})$ then

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

Under (22), $\mathbf{X}_i\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{X}_i' = \mathbf{D}\mathbf{\Sigma}_{dd'}^{-1}\mathbf{D}' + (\mathbf{F}\mathbf{\Gamma}_i + \mathbf{V}_i)(\mathbf{\Sigma}_{vv'} + \mathbf{\Gamma}_i'\mathbf{\Sigma}_{ff'}\mathbf{\Gamma}_i')^{-1}(\mathbf{\Gamma}_i'\mathbf{F}' + \mathbf{V}_i')$ yielding $\mathcal{C}_{3T} = T^{-\frac{1}{2}}\mathcal{C}_{1T}\mathbf{D}\mathbf{\Sigma}_{dd'}^{-1}\mathbf{D}' + \mathcal{C}_{31T}\mathbf{F}' + \mathcal{C}_{32T}$ 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 $\mathbf{X}_i'\mathbf{\Sigma}_T^{-1}\mathbf{D} = O_p(T\iota_{n+k}\iota'_d)$, $\mathcal{C}_{1T}\mathbf{D} = O_p(T^{\frac{1}{2}}\iota_m\iota'_d)$ yielding

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

Finally $\mathcal{C}_{4T} = T^{-\frac{1}{2}}\mathcal{C}_{2T}\mathbf{D}\mathbf{\Sigma}_{dd'}^{-1}\mathbf{D}' + \mathbf{F}\mathcal{C}_{41T}\mathbf{F}' + \mathbf{F}\mathcal{C}_{42T} + \mathcal{C}_{43T}\mathbf{F}' + \mathcal{C}_{44T}$ 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

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

Hence, (14),(15),(16),(17) hold with $a = 1/2, b = 1, c = 3/2, d = 3/2$. In general, primitive conditions can be derived but no assumption of Pesaran (2006) would imply (21), (22) nor any of the other conditions in 7.$\mathcal{H}$.

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* $\mathbf{\Sigma}_{dd'} > 0$*, 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, which at minimum requires* $\mathbf{\Sigma}_{fd'} = \mathbf{0}$.

Then Theorem 1,(iii) applies to the GLS estimator for $(\alpha_0', \beta_0')'$ *of model (18)-(19)-(20) 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* $E(\mathbf{b}_i\mathbf{\Gamma}_i)$ *to be full row rank* m.

Finally, notice that the bias term of the OLS for β_i is, from Theorem 1 (i), $\gamma_i^{OLS} = \mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{\Sigma}_{i,xf'}\mathbf{b}_i$ which is zero only if $\mathbf{b}_i = \mathbf{0}$ a.s. (no factor structure in the regression error) or, alternatively, if $\mathbf{\Sigma}_{i,xf'} = \mathcal{O}_i = \mathbf{0}$ a.s. This latter condition requires both $\mathbf{\Gamma}_i = \mathbf{0}$ a.s. *and* $\mathbf{\Sigma}_{fd'} = \mathbf{0}$. The GLS estimator does not require $\mathbf{\Gamma}_i = \mathbf{0}$ 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

3.1 Definitions and assumptions

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

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

$$\mathbf{y}_t = \mathbf{X}_t \beta_{t0} + \mathbf{u}_t, \tag{24}$$

for an observed $N \times 1$ vector $\mathbf{y}_t = (y_{1t}, ..., y_{it}, ..., y_{Nt})'$ and an observed $N \times k$ matrix $\mathbf{X}_t = (\mathbf{x}_{1t}, ..., \mathbf{x}_{it}, ..., \mathbf{x}_{Nt})'$ related by a $k \times 1$ vector of possibly time-specific regression coefficients β_{t0}. The unobserved $N \times 1$ vector $\mathbf{u}_t = (u_{1t}, ..., u_{it}, ..., u_{Nt})'$ 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, $\mathbf{B} = (\mathbf{b}_1, ..., \mathbf{b}_N)'$ is an unobserved $N \times m$ matrix of factor loadings and $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$ 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 (50).

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} and the subsequent section discusses a set of primitive conditions of our general assumptions for a particular case of interest of model (50) namely Andrews (2005)'s model.

Assumption 1.$\mathcal{T}$ (common factors)
For every t, the $\mathbf{f}_f$ are random vector of dimension $m \times 1$ such that $E(\mathbf{f_t f_t'} \mid \mathbf{X}_t, \mathbf{B}) = \mathcal{F}_t > 0$ with $T^{-1} \sum_{t=1}^{T} \mathcal{F}_t \to_p \mathcal{F} > 0 \quad \text{as } T \to \infty$.

Assumption 2.$\mathcal{T}$ (idiosyncratic innovation)
For every t, $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{it}, ..., \varepsilon_{Nt})'$ let $E(\varepsilon_t \mid \mathbf{f}_t, \mathbf{X}_t, \mathbf{B}) = \mathbf{0}$ and

$$\mathbf{H}_t = E(\varepsilon_t \varepsilon_t' \mid \mathbf{f}_t, \mathbf{X}_t, \mathbf{B}) > 0, \tag{25}$$

$$T^{-1} \sum_{t=1}^{T} \mathbf{H}_t \to_p \mathcal{H}_N > 0 \quad \text{as } T \to \infty. \tag{26}$$

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 $N \to \infty$, are $a.s.$ finite:

$$\frac{\mathbf{X}_t'\mathbf{X}_t}{N} \to_p \mathbf{\Sigma}_{t,xx'} > 0, \ \frac{\mathbf{X}_t'\mathbf{H}_t\mathbf{X}_t}{N} \to_p \mathbf{\Sigma}_{t,xHx'} > 0, \ \frac{\mathbf{X}_t'\mathbf{B}}{N} \to_p \mathbf{\Sigma}_{t,xb'}, \tag{27}$$

$$\frac{\mathbf{X}_t'\mathbf{H}_t^{-1}\mathbf{X}_t}{N} \to_p \mathbf{\Sigma}_{t,xH^{-1}x'} > 0, \frac{\mathbf{B}'\mathbf{H}_t^{-1}\mathbf{B}}{N} \to_p \mathbf{\Sigma}_{t,bH^{-1}b'} > 0, \ \frac{\mathbf{X}_t'\mathbf{H}_t^{-1}\mathbf{B}}{N} \to_p \mathbf{\Sigma}_{t,xH^{-1}b'} \tag{28}$$

such that

$$\mathbf{\Sigma}_{t,xH^{-1}x'} - \mathbf{\Sigma}_{t,xH^{-1}b'}\mathbf{\Sigma}_{t,bH^{-1}b'}\mathbf{\Sigma}_{t,xH^{-1}b'}' > 0. \tag{29}$$

Assumption 5.$\mathcal{T}$ (limit conditions for GLS)
All the limit matrices below, as $T \to \infty$ and arbitrary N, are $a.s.$ finite:

$$\sum_{t=1}^{T} \frac{\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{\Sigma}_{t,xb'}\mathbf{f}_t\mathbf{f}_t'\mathbf{\Sigma}_{t,xb'}'\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{X}_t'}{T} = \mathcal{A}_{1N}(1+o_p(1)), \sum_{t=1}^{T} \frac{\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{\Sigma}_{t,xb'}\mathbf{f}_t\mathbf{f}_t'\mathbf{B}'}{T} = \mathcal{A}_{2N}(1+o_p(1))$$

$$\sum_{t=1}^{T} \frac{\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{X}_t'\varepsilon_t\varepsilon_t'\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{X}_t'}{NT} = \mathcal{A}_{3N}(1+o_p(1)), \sum_{t=1}^{T} \frac{\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{X}_t'\varepsilon_t\varepsilon_t'}{T} = \mathcal{A}_{4N}(1+o_p(1)),$$

$$\sum_{t=1}^{T} \frac{\mathbf{f}_t\varepsilon_t'}{T^{\frac{1}{2}}} = \mathcal{C}_{1N}(1 + o_p(1)), \sum_{t=1}^{T} \frac{\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{\Sigma}_{t,xb'}\mathbf{f}_t\varepsilon_t'}{T^{\frac{1}{2}}} = \mathcal{C}_{2N}(1 + o_p(1)), \tag{30}$$

$$\sum_{t=1}^{T} \frac{\mathbf{f}_t\varepsilon_t'\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{X}_t'}{T^{\frac{1}{2}}N^{\frac{1}{2}}} = \mathcal{C}_{3N}(1+o_p(1)), \sum_{t=1}^{T} \frac{\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{\Sigma}_{t,xb'}\mathbf{f}_t\varepsilon_t'\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{X}_t'}{T^{\frac{1}{2}}N^{\frac{1}{2}}} = \mathcal{C}_{4N}(1+o_p(1)). \tag{31}$$

Assumption 6.$\mathcal{T}$ (distribution conditions for OLS and UGLS)
As $N \to \infty$:

$$\frac{\mathbf{X}_t'\varepsilon_{\mathbf{t}}}{N^{\frac{1}{2}}} \to_d \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{t,xHx'}), \tag{32}$$

$$\frac{\mathbf{X}_t'\mathbf{H}_t^{-1}\varepsilon_{\mathbf{t}}}{N^{\frac{1}{2}}} \to_d \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{t,xH^{-1}x'}), \frac{\mathbf{B}'\mathbf{H}_t^{-1}\varepsilon_{\mathbf{t}}}{N^{\frac{1}{2}}} \to_d \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{t,bH^{-1}b'}). \tag{33}$$

Assumption 7.$\mathcal{T}$ (distribution and identification conditions for GLS)

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

$$\mathcal{A}_{1N} - (\mathcal{A}_{2N} + \mathcal{A}'_{2N}) = \mathbf{B}\mathcal{D}_{1N}\mathbf{B}' + \mathcal{D}_{2N} \text{ and } \mathcal{A}_{3N} - (\mathcal{A}_{4N} + \mathcal{A}'_{4N}) = \mathbf{B}\mathcal{E}_{1N}\mathbf{B}' + \mathcal{E}_{2N} \quad (34)$$

with an $m \times m$ non-singular symmetric $\mathcal{I}_{1N} = \mathcal{D}_{1N} + N^{-1}\mathcal{E}_{1N} + \mathcal{F}$ and a $N \times N$ non-singular $\mathcal{I}_{2N} = \mathcal{D}_{2N} + N^{-1}\mathcal{E}_{2N} + \mathcal{H}_N$ satisfying $a.s.$:

$$\begin{aligned}
&\frac{\mathbf{B}'\mathcal{I}_{2N}^{-1}\mathbf{B}}{N} \to_p \mathbf{\Sigma}_{b\mathcal{I}_2^{-1}b'}(\text{non-singular}), \quad \frac{\mathbf{B}'\mathcal{I}_{2N}^{-1}\mathbf{H}_t\mathcal{I}_{2N}^{-1}\mathbf{B}}{N} \to_p \mathbf{\Sigma}_{t,b\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}b'}, \quad (35)\\
&\frac{\mathbf{X}_t'\mathcal{I}_{2N}^{-1}\mathbf{X}_t}{N} \to_p \mathbf{\Sigma}_{t,x\mathcal{I}_2^{-1}x'} \text{ (non-singular)}, \quad \frac{\mathbf{X}_t'\mathcal{I}_{2N}^{-1}\mathbf{H}_t\mathcal{I}_{2N}^{-1}\mathbf{X}_t}{N} \to_p \mathbf{\Sigma}_{t,x\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}x'},\\
&\frac{\mathbf{X}_t'\mathcal{I}_{2N}^{-1}\mathbf{B}}{N} \to_p \mathbf{\Sigma}_{t,x\mathcal{I}_2^{-1}b'}, \quad \frac{\mathbf{X}_t'\mathcal{I}_{2N}^{-1}\mathbf{H}_t\mathcal{I}_{2N}^{-1}\mathbf{B}}{N} \to_p \mathbf{\Sigma}_{t,x\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}b'},\\
&\frac{\mathbf{X}_t'\mathcal{I}_{2N}^{-1}\varepsilon_t}{N^{\frac{1}{2}}} \to_d \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{t,x\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}x'}), \quad \frac{\mathbf{B}'\mathcal{I}_{2N}^{-1}\varepsilon_t}{N^{\frac{1}{2}}} \to_d \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{t,b\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}b'}),
\end{aligned}$$

where all the limits above hold as $N \to \infty$ with $a.s.$ finite limit matrices and, setting

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

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

$$\begin{aligned}
\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathcal{C}_{1N} + \mathcal{C}'_{1N}\mathbf{B}' + \mathcal{C}_{2N} + \mathcal{C}'_{2N})\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathbf{f}_t + \varepsilon_t) &= O_p(N^a\iota_k), \quad (36)\\
\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathcal{C}_{3N} + \mathcal{C}'_{3N}\mathbf{B}' + \mathcal{C}_{4N} + \mathcal{C}'_{4N})\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathbf{f}_t + \varepsilon_t) &= O_p(N^b\iota_k), \quad (37)\\
\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathcal{C}_{1N} + \mathcal{C}'_{1N}\mathbf{B}' + \mathcal{C}_{2N} + \mathcal{C}'_{2N})\mathbf{\Sigma}_N^{-1}\mathbf{X}_t &= O_p(N^c\iota_k\iota_k'), \quad (38)\\
\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathcal{C}_{3N} + \mathcal{C}'_{3N}\mathbf{B}' + \mathcal{C}_{4N} + \mathcal{C}'_{4N})\mathbf{\Sigma}_N^{-1}\mathbf{X}_t &= O_p(N^d\iota_k\iota_k'). \quad (39)
\end{aligned}$$

Remark: The comments made to Assumptions 1.$\mathcal{H}$-7.$\mathcal{H}$ apply now but replacing T, $\mathbf{F}, \mathbf{b}_i, \mathbf{X}_i, \varepsilon_i, \mathbf{H}_i, \mathbf{u}_i, \mathbf{\Sigma}_{i,xx'}, \mathbf{\Sigma}_{i,xf'}, \mathbf{S}_i$ with N, $\mathbf{B}, \mathbf{f}_t, \mathbf{X}_t, \varepsilon_t, \mathbf{H}_t, \mathbf{u}_t, \mathbf{\Sigma}_{t,xx'}, \mathbf{\Sigma}_{t,xb'}, \mathbf{S}_t$ 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

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

setting

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

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

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

Theorem 2 *(time-specific parameters)*

(i) (OLS) When Assumptions 3.$\mathcal{T}$, 4.$\mathcal{T}$.(27), 6.$\mathcal{T}$.(32) hold

$$N^{\frac{1}{2}}(\hat{\beta}_t^{OLS} - \beta_{t0} - \gamma_t^{OLS}) \to_d \mathcal{N}_k(\mathbf{0}, \mathcal{V}_t^{OLS}) \;\; as \; N \to \infty,$$

setting

$$\gamma_t^{OLS} = \mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{\Sigma}_{t,xb'}\mathbf{f}_t, \;\; \mathcal{V}_t^{OLS} = \mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{\Sigma}_{t,xHx'}^{-1}\mathbf{\Sigma}_{t,xx'}^{-1}.$$

(ii) (UGLS) When Assumptions 1.$\mathcal{T}$, 2.$\mathcal{T}$.(25)*, 3.$\mathcal{T}$, 4.$\mathcal{T}$.(28), 6.$\mathcal{T}$.(33)*

$$N^{\frac{1}{2}}(\hat{\beta}_t^{UGLS} - \beta_{t0}) \to_d \mathcal{N}_k(\mathbf{0}, \mathcal{V}_t^{UGLS}) \;\; as \; N \to \infty,$$

setting

$$\mathcal{V}_t^{UGLS} = (\mathcal{M}_t^{UGLS})^{-1}\mathcal{N}_t^{UGLS}(\mathcal{M}_t^{UGLS})^{-1}$$

with $\mathcal{M}_t^{UGLS} = \mathrm{plim}_{N\to\infty} N^{-1}(\mathbf{X}_t'\mathcal{S}_t^{-1}\mathbf{X}_t)$, $\mathcal{N}_t^{UGLS} = \mathrm{plim}_{N\to\infty} N^{-1}\mathbf{X}_t'\mathcal{S}_t^{-1}\mathbf{H}_t\mathcal{S}_t^{-1}\mathbf{X}_t$. *Moreover*

$$(\mathcal{M}_t^{UGLS})^{-1} = \mathcal{N}_t^{UGLS}.$$

(iii) (GLS)

When Assumptions 1.$\mathcal{T}$, 2.$\mathcal{T}$.(26), 3.$\mathcal{T}$, 4.$\mathcal{T}$.(27) and (29), 5.$\mathcal{T}$, 7.$\mathcal{T}$

$$\hat{\beta}_t^{GLS} \to_p \beta_{t0} \qquad as \ \frac{1}{N} + \frac{N^{max(a-1,b-\frac{3}{2})}}{T^{\frac{1}{2}}} + \frac{N^{max(c-\frac{3}{2},d-2)}}{T^{\frac{1}{2}}} \to 0,$$

$$N^{\frac{1}{2}}(\hat{\beta}_t^{GLS} - \beta_{t0}) \to_d \mathcal{N}_k(\mathbf{0}, \mathcal{V}_t^{GLS}) \qquad as \ \frac{1}{N} + \frac{N^{max(a-\frac{1}{2},b-1)}}{T^{\frac{1}{2}}} + \frac{N^{max(c-1,d-\frac{3}{2})}}{T^{\frac{1}{2}}} \to 0,$$

setting

$$\mathcal{V}_t^{GLS} = (\mathcal{M}_t^{GLS})^{-1}\mathcal{N}_t^{GLS}(\mathcal{M}_t^{GLS})^{-1}$$

with $\mathcal{M}_t^{GLS} = \text{plim}_{(N,T)\to\infty} N^{-1}(\mathbf{X}_t'\hat{\mathbf{\Sigma}}_N^{-1}\mathbf{X}_t)$, $\mathcal{N}_t^{GLS} = \text{plim}_{(N,T)\to\infty} N^{-1}\mathbf{X}_t'\hat{\mathbf{\Sigma}}_N^{-1}\mathbf{H}_t\hat{\mathbf{\Sigma}}_N^{-1}\mathbf{X}_t$.

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 $\gamma_t^{OLS} = \mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{\Sigma}_{t,xb'}^{-1}\mathbf{f}_t$ which is a random variable. Consistency is achieved if either $\mathbf{f}_t = \mathbf{0}$ (no factors) or $\mathbf{\Sigma}_{t,xb'}^{-1} = \mathbf{0}$, 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. Now the GLS makes use of $\hat{\mathbf{\Sigma}}_N$ 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 $\mathbf{\Sigma}_N = \mathbf{B}\mathcal{I}_{1N}\mathbf{B}' + \mathcal{I}_{2N}$, 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

$$y_{it} = \beta_{0t}'(1\,\mathbf{x}_{it}')' + u_{it}, \tag{40}$$

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

$$u_{it} = \mathbf{c}_{1t}'\mathbf{u}_i^* + \varepsilon_{it}, \tag{41}$$

$$\mathbf{x}_{it} = \mathbf{C}_{2t}\mathbf{x}_i^* + \mathbf{v}_{it}, \tag{42}$$

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 $\mathbf{\Sigma}_{t,vv'}$. 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 as start assuming $\mathbf{\Sigma}_{t,vv'} = \mathbf{0}$ implying $\mathbf{v}_{it} = 0$ *a.s.* as in Andrews (2005), although this is not necessary for our arguments to go through. (41)-(42) imply

$$\mathbf{F} = (\mathbf{c}_{11}...\mathbf{c}_{1t}...\mathbf{c}_{1T})',\ \mathbf{B} = (\mathbf{u}_1^*...\mathbf{u}_N^*)',\ \mathbf{X}_t = (\iota_N, \mathbf{X}^*\mathbf{C}_{2t}'),$$

where we set $\mathbf{X}^* = (\mathbf{x}_1^*...\mathbf{x}_i^*...\mathbf{x}_N^*)'$. 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 $\mathbf{c}_{1t}, \mathbf{C}_{2t}, \varepsilon_t$, un-necessary 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 *iid*ness 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.$\mathcal{T}$

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

$$\mathbf{X}_t'\mathbf{H}_t^{-1}\mathbf{X}_t = \sigma_t^{-2}\mathbf{X}_t'\mathbf{X}_t,\ \mathbf{B}'\mathbf{H}_t^{-1}\mathbf{B} = \sigma_t^{-2}\sum_{i=1}^N \mathbf{u}_i^*\mathbf{u}_i^{*\prime},$$

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

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

$$\mathbf{\Sigma}_{t,xx'} = \begin{pmatrix} 1 & \mu_{\mathbf{x}}{}'\mathbf{C_{2t}}' \\ \mathbf{C_{2t}}\mu_{\mathbf{x}} & \mathbf{C_{2t}}\mathbf{\Sigma}_{xx'}\mathbf{C_{2t}}' \end{pmatrix} > 0,\ \ \mathbf{\Sigma}_{t,xb'} = \begin{pmatrix} \mu_u' \\ \mathbf{C}_{2t}\mathbf{\Sigma}_{xu'} \end{pmatrix},$$

non-singularity ensured by Andrews (2005, Assumption 2(d)), where hereafter $\mu_x = Ex_i^*$, $\mathbf{\Sigma}_{xx'} = E\mathbf{x}_i^*\mathbf{x}_i^{*\prime}$, $\mathbf{\Sigma}_{uu'} = E\mathbf{u}_i^*\mathbf{u}_i^{*\prime}$, $\mu_u = E\mathbf{u}_i^*$, $\mathbf{\Sigma}_{xu'} = E\mathbf{x}_i^*\mathbf{u}_i^{*\prime}$.

Using the block matrix decomposition (Magnus and Neudecker 1988) one gets

$$\boldsymbol{\Sigma}^{-1}_{t,xx'} =$$
$$\begin{pmatrix} q_t^{-1} & -q_t^{-1}\mu_x'\mathbf{C_{2t}}'(\mathbf{C_{2t}}\boldsymbol{\Sigma}_{xx'}\mathbf{C_{2t}}')^{-1} \\ -q_t^{-1}(\mathbf{C_{2t}}\boldsymbol{\Sigma}_{xx'}\mathbf{C_{2t}}')^{-1}\mathbf{C_{2t}}\mu_x & (\mathbf{C_{2t}}\boldsymbol{\Sigma}_{xx'}\mathbf{C_{2t}}')^{-1}(\mathbf{I}_k + q_t^{-1}\mathbf{C_{2t}}\mu_x\mu_x'\mathbf{C_{2t}}'(\mathbf{C_{2t}}\boldsymbol{\Sigma}_{xx'}\mathbf{C_{2t}}')^{-1}) \end{pmatrix},$$

where $q_t = 1 - \mu_x'\mathbf{C_{2t}}'(\mathbf{C_{2t}}\boldsymbol{\Sigma}_{xx'}\mathbf{C_{2t}}')^{-1}\mathbf{C_{2t}}\mu_x > 0$, *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 $\boldsymbol{\Sigma}_{t,xHx'}, \boldsymbol{\Sigma}_{t,xH^{-1}x'}$, $\boldsymbol{\Sigma}_{t,bH^{-1}b'}, \boldsymbol{\Sigma}_{t,xH^{-1}b'}$ easily follows. Concerning Assumption 5.$\mathcal{H}$, the expression for $\mathcal{A}_{1N}$, $\mathcal{A}_{2N}$ follows by using

$$\mathbf{X}_t\boldsymbol{\Sigma}^{-1}_{t,xx'}\boldsymbol{\Sigma}_{t,xb'} = \big[\mathbf{X}^*\mathbf{C_{2t}}'(\mathbf{C_{2t}}\boldsymbol{\Sigma}_{xx'}\mathbf{C_{2t}}')^{-1}\mathbf{C_{2t}}\boldsymbol{\Sigma}_{xu'}$$
$$+q_t^{-1}(\iota_N - \mathbf{X}^*\mathbf{C_{2t}}'(\mathbf{C_{2t}}\boldsymbol{\Sigma}_{xx'}\mathbf{C_{2t}}')^{-1}\mathbf{C_{2t}}\mu_x)(\mu_u' - \mu_x'\mathbf{C_{2t}}'(\mathbf{C_{2t}}\boldsymbol{\Sigma}_{xx'}\mathbf{C_{2t}}')^{-1}\mathbf{C_{2t}}\boldsymbol{\Sigma}_{xu'})\big].$$

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

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

where the last equality follows since $\mathbf{C_{2t}}$ is full row rank. Substituting the last expression into $\mathbf{X}_t\boldsymbol{\Sigma}^{-1}_{t,xx'}\boldsymbol{\Sigma}_{t,xb'}\mathcal{F}_t\boldsymbol{\Sigma}'_{t,xb'}\boldsymbol{\Sigma}^{-1}_{t,xx'}\mathbf{X}_t'$ 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

$$\mathbf{C_{2t}}'(\mathbf{C_{2t}}\mathbf{C_{2t}}')^{-1}\mathbf{C_{2t}} = \mathbf{I}_k - \mathbf{M}_{2t} \leq \mathbf{I}_k. \tag{43}$$

This implies $\| \mathbf{X}_t\boldsymbol{\Sigma}^{-1}_{t,xx'}\boldsymbol{\Sigma}_{t,xb'}\mathcal{F}_t\boldsymbol{\Sigma}'_{t,xb'}\boldsymbol{\Sigma}^{-1}_{t,xx'}\mathbf{X}_t' \| = O(\| \mathcal{F}_t \|)$ *a.s.* Similarly, for $\mathcal{A}_{2N}$, $\| \mathbf{X}_t\boldsymbol{\Sigma}^{-1}_{t,xx'}\boldsymbol{\Sigma}_{t,xb'}\mathcal{F}_t \| = O(\| \mathcal{F}_t \|)$ *a.s.*, for $\mathcal{A}_{3N}$, $\| \mathbf{X}_t\boldsymbol{\Sigma}^{-1}_{t,xx'}\boldsymbol{\Sigma}_{t,xHx'}\boldsymbol{\Sigma}^{-1}_{t,xx'}\mathbf{X}_t' \| = O(\sigma_t^2)$ *a.s.* Notice that $\mathcal{A}_{4N} = \mathcal{A}_{3N}$ since $\mathbf{H}_t$ is diagonal, where

$$\mathbf{X}_t\boldsymbol{\Sigma}^{-1}_{t,xx'}\mathbf{X}_t' =$$
$$q_t^{-1}\iota_N\iota_N' - q_t^{-1}\mathbf{X}^*\mathbf{C}_{2t}'(\mathbf{C}_{2t}\boldsymbol{\Sigma}_{xx'}\mathbf{C}_{2t}')^{-1}\mathbf{C}_{2t}\mu_x\iota_N' - q_t^{-1}\iota_N\mu_x'\mathbf{C}_{2t}'(\mathbf{C}_{2t}\boldsymbol{\Sigma}_{xx'}\mathbf{C}_{2t}')^{-1}\mathbf{C}_{2t}\mathbf{X}^{*'}$$
$$+\mathbf{X}^*\mathbf{C}_{2t}'(\mathbf{C}_{2t}\boldsymbol{\Sigma}_{xx'}\mathbf{C}_{2t}')^{-1}[\mathbf{I}_k + q_t^{-1}\mathbf{C}_{2t}\mu_x\mu_x'\mathbf{C}_{2t}'(\mathbf{C}_{2t}\boldsymbol{\Sigma}_{xx'}\mathbf{C}_{2t}')^{-1}]\mathbf{C}_{2t}\mathbf{X}^{*'}$$

and hence its boundedness does not require any moment conditions in $\mathbf{C}_{2t}$.

For (30) and (31), 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 (3)) has bounded maximum eigenvalue. Then for the ith column of $\mathcal{C}_{1N}$ one gets $T^{-\frac{1}{2}}\sum_{t=1}^{T}\mathbf{f}_t\varepsilon_{it}=\mathbf{C}_{1iN}^{\frac{1}{2}}\zeta_{1i}(1+o_p(1))$ for a $m\times 1$ vector ζ_{1t} normally distributed with mean zero and unit covariance matrix if $T^{-1}\sum_{t,s=1}^{T}\mathbf{f}_t h_{ts,i}\mathbf{f}_t' \rightarrow_p \mathbf{C}_{1iN}$ and $E\,\|\mathbf{f}_t\|^{2+\delta}<\infty, E|\varepsilon_{it}|^{2+\delta}<\infty$. Likewise, the ith column of $\mathcal{C}_{2N}$ can be written as $\mathbf{C}_{2iN}^{\frac{1}{2}}\zeta_{2i}$ where no additional moment conditions are required because of (43). The same results apply for (31). Unless $\mathbf{\Sigma}_{t,vv'}=\mathbf{0}$, 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 $\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{\Sigma}_{t,xb'}$ and $\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{X}_t'$ are not time-varying, simplifying the above results. For Assumption 7.$\mathcal{T}$, in particular (34), set $\mathbf{D}=\mathbf{X}^*\mathbf{\Sigma}_{xx'}^{-1}\mathbf{\Sigma}_{xu'}+q^{-1}(\iota_N-\mathbf{X}^*\mathbf{\Sigma}_{xx'}^{-1}\mu_x)(\mu_u'-\mu_x'\mathbf{\Sigma}_{xx'}^{-1}\mathbf{\Sigma}_{xu'})$, $\mathbf{E}=\sigma^2q^{-1}\left(\iota_N\iota_N'-\mathbf{X}^*\mathbf{\Sigma}_{xx'}^{-1}\mu_x\iota_N'-\iota_N\mu_x'\mathbf{\Sigma}_{xx'}^{-1}\mathbf{X}^{*\prime}\right.$ $\left.+\mathbf{X}^*\mathbf{\Sigma}_{xx'}^{-1}[q\mathbf{I}_k+\mu_x\mu_x'\mathbf{\Sigma}_{xx'}^{-1}]\mathbf{X}^{*\prime}\right)$, $q=1-\mu_x'\mathbf{\Sigma}_{xx'}^{-1}\mu_x$. It follows that

$$\mathcal{A}_{1N}-\mathcal{A}_{2N}-\mathcal{A}_{2N}'=(\mathbf{D}-\mathbf{B})\mathcal{F}(\mathbf{D}-\mathbf{B})'-\mathbf{B}\mathcal{F}\mathbf{B}',\ \ \mathcal{A}_{3N}-\mathcal{A}_{4N}-\mathcal{A}_{4N}'=-\mathbf{E}.$$

$$\mathbf{\Sigma}_{xu'}=\mathbf{0},\ \ \mu_u=\mathbf{0} \tag{44}$$

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

When $k<d_2$ which implies full row rank $\mathbf{C}_{2t}$, then obviously (34) is satisfied when (44) hold. However, now it is also possible that $\mathbf{x}_i^*$ and $\mathbf{u}_i^*$ are correlated, for instance even perfectly correlated such as

$$\mathbf{x}_i^*=\mathbf{A}\mathbf{u}_i^*,\ \ \mu_u=\mathbf{0},\ \ \mathbf{\Sigma}_{t,vv'}>0, \tag{45}$$

for a $\mathbf{A}$ non-random full row rank matrix. As an example, set for simplicity $d_2=d_1>k$, $\mathbf{A}=\mathbf{I}_{d_1}$ yielding $\mathbf{\Sigma}_{xx'}=\mathbf{\Sigma}_{xu'}=\mathbf{I}_{d_1}$. Notice that now we require $\mathbf{\Sigma}_{t,vv'}=E\mathbf{v}_{it}\mathbf{v}_{it}'>0$, to ensure (29) holds, and also $\mathbf{v}_{it}$ and u_i^* to be mutually independent. The previous derivations still apply by replacing $(\mathbf{C}_{2t}\mathbf{\Sigma}_{xx'}\mathbf{C}_{2t}')^{-1}$ by $(\mathbf{C}_{2t}\mathbf{\Sigma}_{xx'}\mathbf{C}_{2t}'+\mathbf{\Sigma}_{t,vv'})^{-1}$ and $\mathbf{X}_t=(\iota_N,\mathbf{X}^*\mathbf{C}_{2t}'+\mathbf{V}_t)$. Then

$$T^{-1}\sum_{t=1}^{T}(\mathbf{I}_{d_1}-\mathbf{C}_{2t}'(\mathbf{C}_{2t}\mathbf{C}_{2t}'+\mathbf{\Sigma}_{t,vv'})^{-1}\mathbf{C}_{2t})\mathcal{F}(\mathbf{I}_{d_1}-\mathbf{C}_{2t}'(\mathbf{C}_{2t}\mathbf{C}_{2t}'+\mathbf{\Sigma}_{t,vv'})^{-1}\mathbf{C}_{2t})$$

$$-N^{-1}T^{-1}\sum_{t=1}^{T}\sigma_t^2\mathbf{C}_{2t}'(\mathbf{C}_{2t}\mathbf{C}_{2t}'+\mathbf{\Sigma}_{t,vv'})^{-1}\mathbf{C}_{2t}\rightarrow_p \mathcal{D}_{1N}+\mathcal{F}+N^{-1}\mathcal{E}_{1N}=\mathcal{I}_{1N}$$

and non-singularity of $\mathcal{I}_{1N}$ follows under mild conditions on $\mathbf{C}_{2t}$ and $\mathbf{\Sigma}_{t,vv'}$. 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,

$$\mathbf{\Sigma}_{t,x\mathcal{I}_2^{-1}x'} = \begin{pmatrix} 1 & \mu_{\mathbf{x}}'\mathbf{C_{2t}}' \\ \mathbf{C_{2t}}\mu_{\mathbf{x}} & \sigma^{-2}\mathbf{C_{2t}}(\mathbf{\Sigma}_{xx'} + (\sigma^2-1)\mu_x\mu_x')\mathbf{C_{2t}}' \end{pmatrix}.$$

Similar calculations lead to $\mathbf{\Sigma}_{t,x\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}x'}$, $\mathbf{\Sigma}_{t,b\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}b'}$, $\mathbf{\Sigma}_{t,x\mathcal{I}_2^{-1}b'}$, $\mathbf{\Sigma}_{t,x\mathcal{I}_2^{-1}H\mathcal{I}_2^{-1}b'}$.

It remains to verify (36)-(39). 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 $\mathbf{\Sigma}_N^{-1}$. In addition, denoting by $\mathcal{C}_{1Nj}$ the jth row of $\mathcal{C}_{1N}$, we will require $\sup_N \parallel E\mathcal{C}_{1Nj}'\mathcal{C}_{1Nj} \parallel = O(1)$ 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, $\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathcal{C}_{1N}' = O_p(N^{\frac{1}{2}}\iota_{1+k}\iota_m')$, $\mathcal{C}_{1N}\mathbf{\Sigma}_N^{-1}\varepsilon_t = O_p(N^{\frac{1}{2}}\iota_m)$ and, by Lemma 2, $\mathbf{B}'\mathbf{\Sigma}_N^{-1}\varepsilon_t = O_p(N^{-\frac{1}{2}}\iota_m)$, $\mathbf{B}'\mathbf{\Sigma}_N^{-1}\mathcal{C}_{1N}' = O_p(N^{-\frac{1}{2}}\iota_m\iota_m')$, $\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathbf{B} = O_p(\iota_{1+k}\iota_m')$, $\mathbf{B}'\mathbf{\Sigma}_N^{-1}\mathbf{B} = O_p(\iota_m\iota_m')$ yielding

$$\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathcal{C}_{1N} + \mathcal{C}_{1N}'\mathbf{B}')\mathbf{\Sigma}_N^{-1}\mathbf{X}_t = O_p(N^{\frac{1}{2}}\iota_{1+k}\iota_{1+k}'),$$
$$\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathcal{C}_{1N} + \mathcal{C}_{1N}'\mathbf{B}')\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathbf{f}_t + \varepsilon_t) = O_p(N^{\frac{1}{2}}\iota_{1+k}).$$

We discuss only the case when (45) hold, since case (44) is much simpler. Setting for simplicity $d_2 = d_1 > k$, $\mathbf{A} = \mathbf{I}_{d_1}$ yields $\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{\Sigma}_{t,xb'} = (\mathbf{B}\mathbf{C_{2t}}' + \mathbf{V}_t)(\mathbf{C_{2t}}\mathbf{C_{2t}}' + \mathbf{\Sigma}_{t,vv'})^{-1}\mathbf{C_{2t}}$ 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, $\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathcal{C}_{22N}\mathbf{\Sigma}_N^{-1}\mathbf{B} = O_p(\iota_{1+k}\iota_m')$, $\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathcal{C}_{22N}\mathbf{\Sigma}_N^{-1}\varepsilon_t = O_p(N^{\frac{1}{2}}\iota_{1+k})$ then

$$\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathcal{C}_{2N} + \mathcal{C}_{2N}')\mathbf{\Sigma}_N^{-1}\mathbf{X}_t = O_p(N\iota_{1+k}\iota_{1+k}'),$$
$$\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathcal{C}_{2N} + \mathcal{C}_{2N}')\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathbf{f}_t + \varepsilon_t) = O_p(N^{\frac{1}{2}}\iota_{1+k}).$$

Again, when (45) holds

$$\begin{aligned}\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{X}_t' &= q_t^{-1}\iota_N\iota_N' - q_t^{-1}(\mathbf{B}\mathbf{C}_{2t}' + \mathbf{V}_t)(\mathbf{C}_{2t}\mathbf{\Sigma}_{xx'}\mathbf{C}_{2t}' + \mathbf{\Sigma}_{t,vv'})^{-1}\mathbf{C}_{2t}\mu_x\iota_N' \\ &-q_t^{-1}\iota_N\mu_x'\mathbf{C}_{2t}'(\mathbf{C}_{2t}\mathbf{\Sigma}_{xx'}\mathbf{C}_{2t}' + \mathbf{\Sigma}_{t,vv'})^{-1}(\mathbf{C}_{2t}\mathbf{X}^{*'} + \mathbf{V}_t') \\ &+(\mathbf{X}^*\mathbf{C}_{2t}'+\mathbf{V}_t)(\mathbf{C}_{2t}\mathbf{\Sigma}_{xx'}\mathbf{C}_{2t}'+\mathbf{\Sigma}_{t,vv'})^{-1}[\mathbf{I}_k+q_t^{-1}\mathbf{C}_{2t}\mu_x\mu_x'\mathbf{C}_{2t}'(\mathbf{C}_{2t}\mathbf{\Sigma}_{xx'}\mathbf{C}_{2t}'+\mathbf{\Sigma}_{t,vv'})^{-1}](\mathbf{C}_{2t}\mathbf{X}^{*'}+\mathbf{V}_t'),\end{aligned}$$

yielding $\mathcal{C}_{3N} = \mathcal{C}_{31N}\iota_N' + \mathcal{C}_{32N}\mathbf{B}' + \mathcal{C}_{33N}$ 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 $\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\iota_N = O_p(N\iota_{1+k})$, $\mathbf{B}'\mathbf{\Sigma}_N^{-1}\iota_N = O_p(\iota_m)$, $\varepsilon_t'\mathbf{\Sigma}_N^{-1}\iota_N = O_p(N^{\frac{1}{2}})$ yielding

$$\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathcal{C}_{3N} + \mathcal{C}_{3N}'\mathbf{B}')\mathbf{\Sigma}_N^{-1}\mathbf{X}_t = O_p(N\iota_{1+k}\iota_{1+k}'),$$
$$\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathcal{C}_{3N} + \mathcal{C}_{3N}'\mathbf{B}')\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathbf{f}_t + \varepsilon_t) = O_p(N\iota_{1+k}).$$

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

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

Hence, (36), (37), (38), (39) 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 (45).

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_0 of model (40)-(41)-(42) when

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

for consistency and

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

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), $\gamma_t^{OLS} = \mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{\Sigma}_{t,xb'}^{-1}\mathbf{f}_t$ which is zero only if $\mathbf{f}_t = \mathbf{0}$ a.s. (no factor structure in the regression error) or, alternatively, if $\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{\Sigma}_{t,xb'} = \mathbf{0}$ a.s. The first row of the latter matrix, corresponding to the intercept parameter, is precisely equal to $q_t^{-1}(\mu_u' - \mu_x' \mathbf{C}_{2t}' (\mathbf{C}_{2t} \mathbf{\Sigma}_{xx'} \mathbf{C}_{2t}')^{-1} \mathbf{C}_{2t} \mathbf{\Sigma}_{xu'})$, simplifying to μ_u' under Andrews (2005, Condition SF2), that is when $\mathbf{\Sigma}_{xu'} = \mu_x \mu_u'$ 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 considering from the second to the last row of $\mathbf{\Sigma}_{t,xx'}^{-1} \mathbf{\Sigma}_{t,xb'}$ is $(\mathbf{C_{2t}} \mathbf{\Sigma}_{xx'} \mathbf{C_{2t}}')^{-1} \mathbf{C_{2t}} (\mathbf{\Sigma}_{xu'} - \mu_x \mu_u') + (\mathbf{C_{2t}} \mathbf{\Sigma}_{xx'} \mathbf{C_{2t}}')^{-1} (\mathbf{C_{2t}} \mu_x \mu_x' \mathbf{C_{2t}}' (\mathbf{C_{2t}} \mathbf{\Sigma}_{xx'} \mathbf{C_{2t}}')^{-1} - \mu_x' \mathbf{C_{2t}}' (\mathbf{C_{2t}} \mathbf{\Sigma}_{xx'} \mathbf{C_{2t}}')^{-1} \mathbf{C_{2t}} \mu_x \mathbf{I}_k) \mathbf{C}_{2t} \mathbf{\Sigma}_{xu'}$ which, when $\mathbf{\Sigma}_{xu'} = \mu_x \mu_u'$, equals $(\mathbf{C_{2t}} \mathbf{\Sigma}_{xx'} \mathbf{C_{2t}}')^{-1} \mathbf{C_{2t}} (\mu_x \mu_u' - \mu_x \mu_u')$ $+ \mu_x' \mathbf{C_{2t}}' (\mathbf{C_{2t}} \mathbf{\Sigma}_{xx'} \mathbf{C_{2t}}')^{-1} \mathbf{C_{2t}} \mu_x (\mathbf{C_{2t}} \mathbf{\Sigma}_{xx'} \mathbf{C_{2t}}')^{-1} (\mathbf{C_{2t}} \mu_x - \mathbf{C}_{2t} \mu_x) \mu_{u'}$ and thus 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. 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

In the time regression case the data generating process is a simple regression model

$$y_{it} \quad = \alpha_{i0} + \beta_{i0} x_{it} + b_{i10} f_{1t} + b_{i20} f_{2t} + \varepsilon_{it}, \tag{46}$$

where the single regressor is given by

$$x_{it} = 0.5 + \delta_{i10} f_{1t} + \delta_{i30} f_{3t} + v_{it}. \tag{47}$$

Note that the model implies an observed common factor equal to 1 for all observations. The latent common factors, their factor loadings and the idiosyncratic errors to y_{it} and to x_{it} are assumed $i.i.d$ across time and across units as well as mutually independent. Nevertheless, note that the single regressor is allowed to be contemporaneously correlated with the innovation through one of the latent common factors (whenever $b_{i10} \delta_{i10} \neq 0 \, a.s.$). In particular the factor loadings are random variables with normal distribution, i.i.d. across unit:

$$\begin{pmatrix} b_{i10} \\ b_{i20} \end{pmatrix} \sim NID(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}), \tag{48}$$

$$\begin{pmatrix} \delta_{i10} \\ \delta_{i30} \end{pmatrix} \sim NID(\begin{pmatrix} 0.5 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}), \tag{49}$$

and the latent common factors and the idiosyncratic components are stationary stochastic processes, mutually independent to each other, such as, setting $\mathbf{f}_t = (f_{1t}, f_{2t}, f_{3t})'$,

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

where each $\eta_{jf,t} \sim NID(0, 1)$ for $j = 1, 2, 3$ mutually independent. Also, for any $i = 1, ..., N$:

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

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

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

This Monte Carlo design is a simplified version of Pesaran (2006), designed in such a way that (through (49)) 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 cross-sectional estimators, we consider the regression model with time-varying parameters

$$y_{it} \quad = \alpha_{t0} + \beta_{t0}x_{it} + b_{i10}f_{1t} + b_{i20}f_{2t} + \varepsilon_{it}. \tag{50}$$

The regressor x_{it} is defined in (47) and the factors $f_{j,t}$, $j = 1, 2, 3$ and the idiosyncratic innovations ε_{it}, $i = 1, ..., N$ 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,

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

Finally,

$$\begin{pmatrix} b_{i10} \\ b_{i20} \end{pmatrix} \sim NID(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}), \tag{51}$$

$$\begin{pmatrix} \delta_{i10} \\ \delta_{i30} \end{pmatrix} = \begin{pmatrix} b_{i10} \\ b_{i20} \end{pmatrix}. \tag{52}$$

implying that the factor loadings pertinent to the innovation of (50) 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 $(N, T) \in \{(60, 200, 600), (30, 100, 300)\}$, with $N > T$, for the time regression and $(N, T) \in \{(30, 100, 300), (60, 200, 600)\}$, with $N < T$, for the cross-sectional regression.

The results from the Monte Carlo exercise are summarized in Tables 1 to 4, where we report the sample mean a and root mean square error (rmse) 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 panel of each table: the GLS, the iterated GLS (described Remark 5 to Theorem 1) where the iteration is carried out four times, the OLS and the UGLS. 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 $M^{-1} \sum_{m=1}^{M} \hat{\alpha}_i^m$ and $\left(M^{-1} \sum_{m=1}^{M} (\hat{\alpha}_i^m - 1)^2\right)^{\frac{1}{2}}$ and the average across the units $i = N/2+1, ..., N$ 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 $M = 2,000$, since 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 mth Monte Carlo iteration for a generic estimator. The same description applies to the cross-sectional regression results although now Tables 3-4 report, 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 $t=T/2+1,...,T$ 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 $i=1,...,N/2$, for time regression, and to periods $t=1,...,T/2$, for cross-sectional regression, are not reported but are available.)

4.2 Results

We start by looking at Tables 1 and 2, which report the estimation results for time regression with unit-specific intercept term and regression coefficient respectively. Notice that since the GLS and iterated 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 $\mathbf{\Sigma}_T^{-1}$ 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. For the regression coefficient case (Table 2) the rmse diminishes also when N increases for given T. The same pattern is observed with respect to the iterated 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 clearly 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 nevertheless 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. For instance, 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 $\mathbf{\Sigma}_N^{-1}$ 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 reports, 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 potentially 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. Whereas for consistency N and T can diverge at any rate, asymptotic normality will require them to diverge at specific rates, here established. Moreover, the GLS estimator does not require preliminary estimation of the latent factors nor of their dimension. It uses all the panel data structure in an essential way, but it computationally 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, the latter being even possibly correlated with the regressors. We provide primitive conditions of our general assumptions for the specific models investigated by Pesaran (2006) and Andrews (2005), as examples of time and cross-sectional regressions respectively. Our results are corroborated by a set of Monte Carlo experiments that shows that the performance of the GLS estimator is comparable to the unfeasible UGLS estimator, that makes use of the true (yet generally unknown) innovation covariance matrix.

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 $m_1 \geq m_2$, we present the well-known Sherman-Morrison-Woodbury formula, followed by the two lemmas of this paper. In particular, the proof of Lemma 1 is basically reproducing the proof of Lemma A in Pesaran and Zaffaroni (2008) and it is here repeated for easy reference. Note that 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.

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

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* $m_1 \to \infty$*:*

$$\frac{\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}}{m_1} \to_{a.s.} \mathbf{G} \text{ non-singular }, \tag{53}$$

Then

$$\mathbf{E}^{-1}\mathbf{B} = \mathbf{A}^{-1}\mathbf{B}\left(\frac{\mathbf{C}^{-1}}{m_1} + \frac{\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}}{m_1}\right)^{-1}\frac{\mathbf{C}^{-1}}{m_1} \tag{54}$$

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

$$\mathbf{e}_{m_1}^{(i)\prime}\mathbf{E}^{-1}\mathbf{b}^{(j)} \rightarrow_p 0, \;\; 1 \leq j \leq m_2, \quad \textit{as } m_1 \rightarrow \infty, \tag{55}$$

where $\mathbf{b}^{(j)} = \mathbf{B}\mathbf{e}_{m_2}^{(j)}$ *is the* j*th column of* $\mathbf{B}$*.*
When (53) and

$$\frac{\mathbf{B}'\mathbf{A}^{-1\prime}\mathbf{A}^{-1}\mathbf{B}}{m_1} \rightarrow_{a.s.} \mathbf{L} \geq 0, \tag{56}$$

where $\mathbf{L}$ *denotes an a.s. finite random positive semi-definitive matrix, then*

$$\parallel \mathbf{E}^{-1}\mathbf{B} \parallel^2 = O_p(m_1^{-1}) \quad \textit{as } m_1 \rightarrow \infty. \tag{57}$$

Proof: This follows precisely the proof of Pesaran and Zaffaroni (2008, Lemma A). We start from the Sherman-Morrison-Woodbury formula, rewritten as

$$\mathbf{E}^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\frac{\mathbf{C}^{-1}}{m_1} + \frac{\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}}{m_1})^{-1}\frac{\mathbf{B}'\mathbf{A}^{-1}}{m_1}. \tag{58}$$

Post-multiplying both sides by $\mathbf{B}$ and simple manipulations yields (54). Pre-multiplying both sides by $\mathbf{e}_{m_1}^{(i)\prime}$ and post-multiplying both sides by $\mathbf{e}_{m_2}^{(j)}$ yields (55).

We deal with (57) more explicitly. Since $\mathbf{B}\mathbf{e}_{m_2}^{(j)} = \mathbf{b}^{(j)}$

$$\begin{aligned}
&(m_1^{-1}\mathbf{C}^{-1} + m_1^{-1}\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}m_1^{-1}\mathbf{B}'\mathbf{A}^{-1}\mathbf{b}^{(j)} - \mathbf{e}_{m_2}^{(j)} \\
&= (m_1^{-1}\mathbf{C}^{-1} + m_1^{-1}\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}m_1^{-1}\mathbf{B}'\mathbf{A}^{-1}\mathbf{b}^{(j)} - (m_1^{-1}\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}m_1^{-1}\mathbf{B}'\mathbf{A}^{-1}\mathbf{b}^{(j)} \\
&= m_1^{-1}\left[-(m_1^{-1}\mathbf{C}^{-1} + m_1^{-1}\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}^{-1}(m_1^{-1}\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}m_1^{-1}\mathbf{B}'\mathbf{A}^{-1}\mathbf{b}^{(j)}\right] \\
&= m_1^{-1}\mathbf{g}^{(j)},
\end{aligned}$$

where it is easy to see that $\mathbf{g}^{(j)} \rightarrow_p -\mathbf{G}^{-1}\mathbf{C}^{-1}\mathbf{G}^{-1}\mathbf{e}_{m_2}^{(j)}$. Therefore, substituting the latter expression into (58) yields $\mathbf{E}^{-1}\mathbf{b}^{(j)} = \mathbf{A}^{-1}\mathbf{b}^{(j)} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{e}_{m_2}^{(j)} + m_1^{-1}\mathbf{g}^{(j)}) = -m_1^{-1}\mathbf{A}^{-1}\mathbf{B}\mathbf{g}^{(j)}$ and thus

$$\parallel \mathbf{E}^{-1}\mathbf{b}^{(j)} \parallel^2 = m_1^{-1}\mathbf{g}^{(j)\prime}\left(m_1^{-1}\mathbf{B}'\mathbf{A}^{-1\prime}\mathbf{A}^{-1}\mathbf{B}\right)\mathbf{g}^{(j)} = O_p(m_1^{-1}\;\; \mathbf{e}_{m_2}^{(j)\prime}\mathbf{G}^{-1}\mathbf{C}^{-1}\mathbf{G}^{-1}\mathbf{L}\mathbf{G}^{-1}\mathbf{C}^{-1}\mathbf{G}^{-1}\mathbf{e}_{m_2}^{(j)}).$$

At last (57) simply follows from

$$\parallel \mathbf{E}^{-1}\mathbf{B} \parallel^2 \leq \sum_{j=1}^{m_2} \parallel \mathbf{E}^{-1}\mathbf{b}^{(j)} \parallel^2 . \;\square$$

Lemma 2*(*$\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}, m_1$*).*
Set

$$\mathbf{E} = \mathbf{B}\mathbf{C}\mathbf{B}' + \mathbf{A} \ \ a.s.$$

When (53) and $\mathbf{D}'\mathbf{A}^{-1\prime}\mathbf{B} = O_p(m_1\, \iota_{m_3}\iota'_{m_2})$ *then*

$$\mathbf{D}'\mathbf{E}^{-1}\mathbf{B} = O_p(\iota_{m_3}\iota'_{m_2}) \quad as\ m_1 \to \infty. \tag{59}$$

When (53) and $\mathbf{D}'\mathbf{A}^{-1\prime}\mathbf{B} = O_p(m_1^{\frac{1}{2}}\, \iota_{m_3}\iota'_{m_2})$ *then*

$$\mathbf{D}'\mathbf{E}^{-1}\mathbf{B} = O_p(m_1^{-\frac{1}{2}}\, \iota_{m_3}\iota'_{m_2}) \quad as\ m_1 \to \infty. \tag{60}$$

Proof: By (58)

$$\mathbf{D}'\mathbf{E}^{-1}\mathbf{B} = \mathbf{D}'\mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}^{-1}.$$

and (59) and (60) easily follows along the lines of the proof of Lemma 1. □
Proof of Theorem 1. (i) All the limits below hold as $T \to \infty$. The results follows since

$$\hat{\beta}_i^{OLS} - \beta_{i0} = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{u}_i,$$

can be written as

$$\hat{\beta}_i^{OLS} - \beta_{i0} - (\frac{\mathbf{X}_i'\mathbf{X}_i}{T})^{-1}\frac{\mathbf{X}_i'\mathbf{F}\mathbf{b}_i}{T} = T^{-\frac{1}{2}}(\frac{\mathbf{X}_i'\mathbf{X}_i}{T})^{-1}T^{-\frac{1}{2}}\mathbf{X}_i'\varepsilon_i.$$

(ii) All the limits below hold as $T \to \infty$.
Since

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

where the first equality is warranted by the Sherman-Morrison-Woodbury (hereafter SMW) formula and the fourth equality makes use of the Central

Lemma($\mathbf{H}_i, \mathbf{F}, \mathcal{B}_i, T$). Using the SMW formula again

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

implying

$$(\frac{\mathbf{X}_i'\mathcal{S}_i^{-1}\mathbf{X}_i}{T})^{-1} = \left(\mathbf{\Sigma}_{i,xH^{-1}x'} - \mathbf{\Sigma}_{i,xH^{-1}f'}\mathbf{\Sigma}_{i,fH^{-1}f'}^{-1}\mathbf{\Sigma}_{i,xH^{-1}f'}'\right)^{-1} + o_p(1)$$

yielding $\hat{\gamma}_i^{UGLS} = O_p(T^{-1})$. Therefore concerning the first term on the right hand side of $T^{\frac{1}{2}}(\hat{\beta}_i^{UGLS} - \beta_{i0})$

$$T^{\frac{1}{2}}\hat{\gamma}_i^{UGLS} = O_p(T^{-\frac{1}{2}}).$$

For the second term of the right hand side of $T^{\frac{1}{2}}(\hat{\beta}_i^{UGLS} - \beta_{i0})$, given

$$\text{cov}\left(\frac{\mathbf{X}_i'\mathbf{H}_i^{-1}\varepsilon_i}{T^{\frac{1}{2}}}, \frac{\varepsilon_i'\mathbf{H}_i^{-1}\mathbf{F}}{T^{\frac{1}{2}}}\right) = \mathbf{\Sigma}_{i,xH^{-1}f'} + o_p(1),$$

then

$$\frac{\mathbf{X}_i'\mathcal{S}_i^{-1}\varepsilon_i}{T^{\frac{1}{2}}} \rightarrow_d \mathbf{N}\left(\mathbf{0}, (\mathbf{\Sigma}_{i,xH^{-1}x'} - \mathbf{\Sigma}_{i,xH^{-1}f'}\mathbf{\Sigma}_{t,fH^{-1}f'}^{-1}\mathbf{\Sigma}_{t,xH^{-1}f'}')\right).$$

Combining terms

$$T^{\frac{1}{2}}(\hat{\beta}_i^{UGLS} - \beta_{i0}) \rightarrow_d \mathcal{N}_k(\mathbf{0}, \mathcal{V}_i^{UGLS}).$$

(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 $\hat{\mathbf{\Sigma}}_T = N^{-1}\sum_{i=1}^{N} \hat{u}_i\hat{u}_i'$ where

$$\hat{u}_i = (\mathbf{I}_T - \mathbf{M}_i)\mathbf{u}_i,$$

setting

$$\mathbf{M}_i = \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'.$$

Then

$$\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i' = \underbrace{\mathbf{F}\mathbf{b}_i\mathbf{b}_i'\mathbf{F}'}_{I} + \underbrace{(\mathbf{I_T}-\mathbf{M_i})\varepsilon_i\varepsilon_i'(\mathbf{I_T}-\mathbf{M_i})}_{II}$$
$$+\underbrace{\mathbf{M_i}\mathbf{F}\mathbf{b}_i\mathbf{b}_i'\mathbf{F}'\mathbf{M_i}}_{III} + \underbrace{(\mathbf{F}\mathbf{b}_i\varepsilon_i'(\mathbf{I_T}-\mathbf{M_i})+(\mathbf{I_T}-\mathbf{M_i})\varepsilon_i\mathbf{b}_i'\mathbf{F}')}_{IV}$$
$$-\underbrace{((\mathbf{I_T}-\mathbf{M_i})\varepsilon_i\mathbf{b}_i'\mathbf{F}'\mathbf{M}_i+\mathbf{M}_i\mathbf{F}\mathbf{b}_i\varepsilon_i'(\mathbf{I_T}-\mathbf{M_i}))}_{V} - \underbrace{(\mathbf{M}_i\mathbf{F}\mathbf{b}_i\mathbf{b}_i'\mathbf{F}'+\mathbf{F}\mathbf{b}_i\mathbf{b}_i'\mathbf{F}'\mathbf{M}_i)}_{VI}.$$

For II

$$N^{-1}\sum_{i=1}^{N}(\mathbf{I_T}-\mathbf{M_i})\varepsilon_i\varepsilon_i'(\mathbf{I_T}-\mathbf{M_i}) = N^{-1}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'$$
$$+\left(N^{-1}T^{-2}\sum_{i=1}^{N}\mathbf{X}_i\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{X}_i'\varepsilon_i\varepsilon_i'\mathbf{X}_i\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{X}_i'\right)(1+o_p(1))$$
$$-\left(N^{-1}T^{-1}\sum_{i=1}^{N}\left(\mathbf{X}_i\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{X}_i'\varepsilon_i\varepsilon_i'+\varepsilon_i\varepsilon_i'\mathbf{X}_i\mathbf{\Sigma}_{i,xx'}^{-1}\mathbf{X}_i'\right)\right)(1+o_p(1)),$$

yielding

$$N^{-1}\sum_{i=1}^{N}(\mathbf{I_T}-\mathbf{M_i})\varepsilon_i\varepsilon_i'(\mathbf{I_T}-\mathbf{M_i}) = \left(\mathcal{H}_T+T^{-1}(\mathcal{A}_{3T}-(\mathcal{A}_{4T}+\mathcal{A}_{4T}'))\right)(1+o_p(1)).$$

For III

$$N^{-1}\sum_{i=1}^{N}\mathbf{M_i}\mathbf{F}\mathbf{b}_i\mathbf{b}_i'\mathbf{F}'\mathbf{M_i} = \mathcal{A}_{1T}(1+o_p(1)).$$

For IV

$$N^{-\frac{1}{2}}\sum_{i=1}^{N}\mathbf{F}\mathbf{b}_i\varepsilon_i' = \mathbf{F}\mathcal{C}_{1T}(1+o_p(1)),\ \ T^{\frac{1}{2}}N^{-\frac{1}{2}}\sum_{i=1}^{N}\mathbf{F}\mathbf{b}_i\varepsilon_i'\mathbf{M}_i = \mathbf{F}\mathcal{C}_{3T}(1+o_p(1)),$$

and 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

$$N^{-1}\sum_{i=1}^{N}\mathbf{M_i}\mathbf{F}\mathbf{b}_i\mathbf{b}_i'\mathbf{F}'=\mathcal{A}_{2T}(1+o_p(1)).$$

Summarizing:

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

setting

$$\mathcal{D}_{N,T}=N^{-\frac{1}{2}}\left(\mathbf{F}\mathcal{C}_{1T}+\mathcal{C}_{1T}'\mathbf{F}'+\mathcal{C}_{2T}+\mathcal{C}_{2T}'\right)-(NT)^{-\frac{1}{2}}\left(\mathbf{F}\mathcal{C}_{3T}+\mathcal{C}_{3T}'\mathbf{F}'+\mathcal{C}_{4T}+\mathcal{C}_{4T}'\right).$$

Hence, using $\hat{\mathbf{\Sigma}}_T^{-1}=\hat{\mathbf{\Sigma}}_T^{-1}\hat{\mathbf{\Sigma}}_T\hat{\mathbf{\Sigma}}_T^{-1}=\mathbf{\Sigma}_T^{-1}\hat{\mathbf{\Sigma}}_T\mathbf{\Sigma}_T^{-1}(1+o_p(1))$,

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

Following precisely the same steps of part (ii) but replacing $\mathbf{S}_i$, $\mathcal{B}_i$, $\mathbf{H}_i$ by $\mathbf{\Sigma}_T$, $\mathcal{I}_{1T}$, $\mathcal{I}_{2T}$ respectively, and using the Central Lemma$(\mathcal{I}_{2T},\mathbf{F},\mathcal{I}_{1T},T)$, then $(\mathbf{X}_i'\mathbf{\Sigma}_T^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{\Sigma}_T^{-1}(\mathbf{F}\mathbf{b}_i+\varepsilon_i)=O_p(T^{-\frac{1}{2}})$ and

$$T^{\frac{1}{2}}(\mathbf{X}_i'\mathbf{\Sigma}_T^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{\Sigma}_T^{-1}(\mathbf{F}\mathbf{b}_i+\varepsilon_i)\to_d\mathcal{N}(\mathbf{0},(\mathcal{M}_i^{GLS})^{-1}\mathcal{N}_i^{GLS}(\mathcal{M}_i^{GLS})^{-1})\text{ as }T\to\infty$$

with

$$\begin{aligned}\mathcal{N}_i^{GLS}\quad&=\mathbf{\Sigma}^{-1}_{i,x\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}x'}+\mathbf{\Sigma}_{i,x\mathcal{I}_2^{-1}f'}\mathbf{\Sigma}^{-1}_{i,f\mathcal{I}_2^{-1}f'}\mathbf{\Sigma}_{i,f\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}f'}\mathbf{\Sigma}^{-1}_{i,f\mathcal{I}_2^{-1}f'}\mathbf{\Sigma}'_{i,x\mathcal{I}_2^{-1}f'}\\&-\left(\mathbf{\Sigma}_{i,x\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}f'}\mathbf{\Sigma}^{-1}_{i,f\mathcal{I}_2^{-1}f'}\mathbf{\Sigma}'_{i,x\mathcal{I}_2^{-1}f'}+\mathbf{\Sigma}_{i,x\mathcal{I}_2^{-1}f'}\mathbf{\Sigma}^{-1}_{i,f\mathcal{I}_2^{-1}f'}\mathbf{\Sigma}'_{i,x\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}f'}\right)\end{aligned}$$

and

$$\mathcal{M}_i^{GLS} = \mathbf{\Sigma}_{i,x\mathcal{I}_2^{-1}x'} - \mathbf{\Sigma}_{i,x\mathcal{I}_2^{-1}f'}\mathbf{\Sigma}_{i,f\mathcal{I}_2^{-1}f'}\mathbf{\Sigma}'_{i,x\mathcal{I}_2^{-1}f'}.$$

For the second and third term after the second equality sign,

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

whereas the fourth term goes to zero faster than these two terms. □

Proof of Theorem 2. (i) All the limits below hold as $N \to \infty$. The results follows since

$$\hat{\beta}_t^{OLS} - \beta_{t0} = (\mathbf{X}_t'\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{u}_t,$$

can be written as

$$\hat{\beta}_t^{OLS} - \beta_{t0} - (\frac{\mathbf{X}_t'\mathbf{X}_t}{N})^{-1}\frac{\mathbf{X}_t'\mathbf{B}\mathbf{f}_t}{N} = N^{-\frac{1}{2}}(\frac{\mathbf{X}_t'\mathbf{X}_t}{N})^{-1}N^{-\frac{1}{2}}\mathbf{X}_t'\varepsilon_t.$$

(ii) All the limits below hold as $N \to \infty$.

Since

$$\begin{aligned}
&\hat{\beta}_t^{UGLS} - \beta_{t0} = (\mathbf{X}_t'\mathcal{S}_t^{-1}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathcal{S}_t^{-1}(\mathbf{B}\mathbf{f}_t + \varepsilon_t)\\
&= (\frac{\mathbf{X}_t'\mathcal{S}_t^{-1}\mathbf{X}_t}{N})^{-1}\frac{\mathbf{X}_t'\mathbf{H}_t^{-1}\mathbf{B}}{N}(\frac{\mathcal{F}_t^{-1}}{N} + \frac{\mathbf{B}'\mathbf{H}_t^{-1}\mathbf{B}}{N})^{-1}\frac{\mathcal{F}_t^{-1}}{N}\mathbf{f}_t + N^{-\frac{1}{2}}(\frac{\mathbf{X}_t'\mathcal{S}_t^{-1}\mathbf{X}_t}{N})^{-1}\frac{\mathbf{X}_t'\mathcal{S}_t^{-1}\varepsilon_t}{N^{\frac{1}{2}}}\\
&= \hat{\gamma}_t^{UGLS} + N^{-\frac{1}{2}}(\frac{\mathbf{X}_t'\mathcal{S}_t^{-1}\mathbf{X}_t}{N})^{-1}\frac{\mathbf{X}_t'\mathcal{S}_t^{-1}\varepsilon_t}{N^{\frac{1}{2}}},
\end{aligned}$$

using the Central Lemma($\mathbf{H}_t, \mathbf{B}, \mathcal{F}_t, N$). Using the SMW formula again

$$(\frac{\mathbf{X}_t'\mathcal{S}_t^{-1}\mathbf{X}_t}{N})^{-1} = \left(\mathbf{\Sigma}_{t,xH^{-1}x'} - \mathbf{\Sigma}_{t,xH^{-1}b'}\mathbf{\Sigma}_{t,bH^{-1}b'}^{-1}\mathbf{\Sigma}'_{t,xH^{-1}b'}\right)^{-1} + o_p(1)$$

yielding

$$N^{\frac{1}{2}}\hat{\gamma}_t^{UGLS} = O_p(N^{-\frac{1}{2}}).$$

Since

$$\frac{\mathbf{X}_t'\mathcal{S}_t^{-1}\varepsilon_t}{N^{\frac{1}{2}}} \to_d \mathbf{N}\left(\mathbf{0}, (\mathbf{\Sigma}_{t,xH^{-1}x'} - \mathbf{\Sigma}_{t,xH^{-1}b'}\mathbf{\Sigma}_{t,bH^{-1}b'}^{-1}\mathbf{\Sigma}'_{t,xH^{-1}b'})\right),$$

then

$$N^{\frac{1}{2}}(\hat{\beta}_t^{UGLS} - \beta_{t0}) \to_d \mathcal{N}_k(\mathbf{0}, \mathcal{V}_t^{UGLS}).$$

(iii) All the limits below hold as $(N,T)\to\infty$. We must assume $T\geq N$ and, with no loss of generality, that there are no observed common factors implying that $\hat{\mathbf{\Sigma}}_N=T^{-1}\sum_{t=1}^{T}\hat{u}_t\hat{u}_t'$ where

$$\hat{\mathbf{u}}_t=(\mathbf{I}_N-\mathbf{M}_t)\mathbf{u}_t,$$

setting

$$\mathbf{M}_t=\mathbf{X}_t(\mathbf{X}_t'\mathbf{X}_t)^{-1}\mathbf{X}_t'.$$

Then

$$\begin{aligned}\hat{\mathbf{u}}_t\hat{\mathbf{u}}_t' \quad &= \underbrace{\mathbf{Bf}_t\mathbf{f}_t'\mathbf{B}'}_{I}+\underbrace{(\mathbf{I_N}-\mathbf{M_t})\varepsilon_t\varepsilon_t'(\mathbf{I_N}-\mathbf{M_t})}_{II}\\ &+\underbrace{\mathbf{M_t}\mathbf{Bf}_t\mathbf{f}_t'\mathbf{B}'\mathbf{M_t}}_{III}+\underbrace{(\mathbf{Bf}_t\varepsilon_t'(\mathbf{I_N}-\mathbf{M_t})+(\mathbf{I_N}-\mathbf{M_t})\varepsilon_t\mathbf{b}_t'\mathbf{B}')}_{IV}\\ &-\underbrace{((\mathbf{I_N}-\mathbf{M_t})\varepsilon_t\mathbf{f}_t'\mathbf{B}'\mathbf{M}_t+\mathbf{M}_t\mathbf{Bf}_t\varepsilon_t'(\mathbf{I_N}-\mathbf{M_t}))}_{V}-\underbrace{(\mathbf{M}_t\mathbf{Bf}_t\mathbf{f}_t'\mathbf{B}'+\mathbf{Bf}_t\mathbf{f}_t'\mathbf{B}'\mathbf{M}_t)}_{VI}.\end{aligned}$$

For II

$$\begin{aligned}T^{-1}\sum_{t=1}^{T}(\mathbf{I_N}-\mathbf{M_t})\varepsilon_t\varepsilon_t'(\mathbf{I_N}-\mathbf{M_t}) &= T^{-1}\sum_{t=1}^{T}\varepsilon_t\varepsilon_t'\\ +\left(N^{-2}T^{-1}\sum_{t=1}^{T}\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{X}_t'\varepsilon_t\varepsilon_t'\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{X}_t'\right)&(1+o_p(1))\\ -\left(N^{-1}T^{-1}\sum_{t=1}^{T}\left(\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{X}_t'\varepsilon_t\varepsilon_t'+\varepsilon_t\varepsilon_t'\mathbf{X}_t\mathbf{\Sigma}_{t,xx'}^{-1}\mathbf{X}_t'\right)\right)&(1+o_p(1)),\end{aligned}$$

yielding

$$T^{-1}\sum_{t=1}^{T}(\mathbf{I_N}-\mathbf{M_t})\varepsilon_t\varepsilon_t'(\mathbf{I_N}-\mathbf{M_t})=\mathcal{H}_N+N^{-1}\left(\mathcal{A}_{3N}-\mathcal{A}_{4N}-\mathcal{A}_{4N}'\right)(1+o_p(1)).$$

For III

$$T^{-1}\sum_{t=1}^{T}\mathbf{M_t}\mathbf{Bf}_t\mathbf{f}_t'\mathbf{B}'\mathbf{M_t}=\mathcal{A}_{1N}(1+o_p(1)).$$

For IV

$$T^{-\frac{1}{2}}\sum_{t=1}^{T}\mathbf{Bf}_t\varepsilon_t'=\mathbf{B}\mathcal{C}_{1N}(1+o_p(1)),\ (NT)^{-\frac{1}{2}}\sum_{t=1}^{T}\mathbf{Bf}_t\varepsilon_t'\mathbf{M}_t=\mathbf{B}\mathcal{C}_{3N}(1+o_p(1)),$$

and combining the above results yields

$$T^{-1}\sum_{t=1}^{T}\mathbf{B}\mathbf{f}_t\varepsilon_t'(\mathbf{I}_N-\mathbf{M}_t)=T^{-\frac{1}{2}}\mathbf{B}\left(\mathcal{C}_{1N}-N^{-\frac{1}{2}}\mathcal{C}_{3N}\right)(1+o_p(1)).$$

For V using the same arguments one gets

$$T^{-1}\sum_{t=1}^{T}\mathbf{M}_t\mathbf{B}\mathbf{f}_t\varepsilon_t'(\mathbf{I}_N-\mathbf{M}_t)=T^{-\frac{1}{2}}\left(\mathcal{C}_{2N}-N^{-\frac{1}{2}}\mathcal{C}_{4N}\right)(1+o_p(1)).$$

For VI

$$T^{-1}\sum_{t=1}^{T}\mathbf{M_t}\mathbf{B}\mathbf{f}_t\mathbf{f}_t'\mathbf{B}'=\mathcal{A}_{2N}(1+o_p(1)).$$

Summarizing:

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

setting

$$\mathcal{D}_{T,N}=T^{-\frac{1}{2}}(\mathbf{B}\mathcal{C}_{1N}+\mathcal{C}_{1N}'\mathbf{B}'+\mathcal{C}_{2N}+\mathcal{C}_{2N}')-(NT)^{-\frac{1}{2}}(\mathbf{B}\mathcal{C}_{3N}+\mathcal{C}_{3N}'\mathbf{B}'+\mathcal{C}_{4N}+\mathcal{C}_{4N}').$$

Hence, using $\hat{\mathbf{\Sigma}}_N^{-1}=\hat{\mathbf{\Sigma}}_N^{-1}\hat{\mathbf{\Sigma}}_N\hat{\mathbf{\Sigma}}_N^{-1}=\mathbf{\Sigma}_N^{-1}\hat{\mathbf{\Sigma}}_N\mathbf{\Sigma}_N^{-1}(1+o_p(1))$,

$$\begin{aligned}&(\hat{\beta}_t^{GLS}-\beta_{t0})=\left((\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)^{-1}+(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)^{-1}(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathcal{D}_{T,N}\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)^{-1}\right)\\&\times\left(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathbf{f}_t+\varepsilon_t)+\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathcal{D}_{T,N}\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathbf{f}_t+\varepsilon_t)\right)(1+o_p(1))\\=\;&(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathbf{f}_t+\varepsilon_t)(1+o_p(1))\\+\;&(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathcal{D}_{T,N}\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathbf{f}_t+\varepsilon_t)(1+o_p(1))\\+\;&(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)^{-1}(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathcal{D}_{T,N}\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathbf{f}_t+\varepsilon_t)(1+o_p(1))\\+\;&(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)^{-1}(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathcal{D}_{T,N}\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathcal{D}_{T,N}\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathbf{f}_t+\varepsilon_t)(1+o_p(1)).\end{aligned}$$

Following precisely the same steps of part (ii) but replacing $\mathbf{S}_t$, $\mathcal{F}_t$, $\mathbf{H}_t$ by $\mathbf{\Sigma}_N, \mathcal{I}_{1N}, \mathcal{I}_{2N}$ respectively, and using the Central Lemma$(\mathcal{I}_{2N},\mathbf{B},\mathcal{I}_{1N},N)$, then $(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathbf{f}_t+\varepsilon_t)=O_p(N^{-\frac{1}{2}})$ and

$$N^{\frac{1}{2}}(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathbf{b}_t+\varepsilon_t)\to_d\mathcal{N}(\mathbf{0},(\mathcal{M}_t^{GLS})^{-1}\mathcal{N}_t^{GLS}(\mathcal{M}_t^{GLS})^{-1})\text{ as }N\to\infty$$

with

$$\begin{aligned}\mathcal{N}_t^{GLS} &= \mathbf{\Sigma}^{-1}_{t,x\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}x'} + \mathbf{\Sigma}_{t,x\mathcal{I}_2^{-1}b'}\mathbf{\Sigma}^{-1}_{t,b\mathcal{I}_2^{-1}b'}\mathbf{\Sigma}_{t,b\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}b'}\mathbf{\Sigma}^{-1}_{t,b\mathcal{I}_2^{-1}b'}\mathbf{\Sigma}'_{t,x\mathcal{I}_2^{-1}b'} \\ &\quad - \left(\mathbf{\Sigma}_{t,x\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}b'}\mathbf{\Sigma}^{-1}_{t,b\mathcal{I}_2^{-1}b'}\mathbf{\Sigma}'_{t,x\mathcal{I}_2^{-1}b'} + \mathbf{\Sigma}_{t,x\mathcal{I}_2^{-1}b'}\mathbf{\Sigma}^{-1}_{t,b\mathcal{I}_2^{-1}b'}\mathbf{\Sigma}'_{t,x\mathcal{I}_2^{-1}\mathcal{H}\mathcal{I}_2^{-1}b'}\right)\end{aligned}$$

and

$$\mathcal{M}_t^{GLS} = \mathbf{\Sigma}_{t,x\mathcal{I}_2^{-1}x'} - \mathbf{\Sigma}_{t,x\mathcal{I}_2^{-1}b'}\mathbf{\Sigma}_{t,b\mathcal{I}_2^{-1}b'}\mathbf{\Sigma}'_{t,x\mathcal{I}_2^{-1}b'}.$$

For the second and third term after the second equality sign,

$$\begin{aligned}&(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathcal{D}_{T,N}\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathbf{f}_t+\varepsilon_t) = O_p(T^{-\frac{1}{2}}N^{-1}(N^a + N^{b-\frac{1}{2}})),\\ &(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)^{-1}(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathcal{D}_{T,N}\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)(\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{\Sigma}_N^{-1}(\mathbf{B}\mathbf{f}_t+\varepsilon_t) = O_p(T^{-\frac{1}{2}}N^{-\frac{3}{2}}(N^c + N^{d-\frac{1}{2}})),\end{aligned}$$

whereas the fourth term goes to zero faster than these two terms. □

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

	GLS						*GLS* (iterated)					
	bias			rmse			bias			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*					
	bias			rmse			bias			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

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

	GLS						*GLS* (iterated)					
	bias			rmse			bias			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
	OLS						*UGLS*					
	bias			rmse			bias			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

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

	GLS						*GLS* (iterated)					
	bias			rmse			bias			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*					
	bias			rmse			bias			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

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

	GLS						*GLS* (iterated)					
	bias			rmse			bias			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
	OLS						*UGLS*					
	bias			rmse			bias			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

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.

ANSELIN, L. (1988): *Spatial econometrics: methods and models.* Dordrecht: Kluwer Academic Publishers.

BAI, J. (2003): "Inferential theory for factor models of large dimensions," *Econometrica*, 71, 135–171.

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.

CASE, A. C. (1991): "Spatial patterns in household demand," *Econometrica*, 59, 953–965.

CHAMBERLAIN, G. (1983): "Funds, factors, and diversification in arbitrage pricing models," *Econometrica*, 51, 1305–1323.

CHEN, X., AND T. G. CONLEY (2001): "A new semiparametric spatial model for panel time series," *Journal of Econometrics*, 105, 59–83.

CONLEY, T. G. (1999): "GMM estimation with cross sectional dependence," *Journal of Econometrics*, 92, 1–45.

FORNI, M., M. HALLIN, M. LIPPI, AND R. L. (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.

Lee, L. F. (2004): "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models," *Econometrica*, 72, 1899–1925.

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. R., and B. Perron (2003): "Testing for a unit root in panels with dynamic factors," *Journal of Econometrics*.

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 (2008): "Optimal asset allocation with factor models for large portfolios," Imperial College London, preprint.

Phillips, P. C. B., and D. Sul (2003): "Dynamic panel estimation and homogeneity testing under cross section dependence," *Econometrics Journal*.

——— (2007): "Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence," *Journal of Econometrics*, 137, 162–188.

Robinson, P. M. (2006): "Efficient estimation of the semiparametric spatial autoregressive model," *Journal of Econometrics* forthcoming.

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.