Generalised Canonical Correlation Estimation of the Multilevel Factor Model*

Rui Lin† Yongcheol Shin‡

November 2022, preliminary and incomplete

Abstract

We develop a novel approach based on the generalised canonical correlation (GCC) analysis to consistently estimating the multilevel factor model and providing the proper inference theory. Importantly, our approach is shown to be robust to a non-zero correlation between the local factors across the different blocks and valid even if some blocks share the same local factors. We also propose a novel selection criterion for identifying the number of the global factors. Relevant asymptotic theories are derived under fairly standard conditions. Via Monte Carlo simulations, we show the satisfactory and dominant performance of the GCC estimator relative to existing approaches. Finally, we demonstrate its usefulness with an application to the housing market in England and Wales using a large disaggregated panel data of the real house price growth rates for the 331 local authorities over the period 1996Q1 to 2021Q2.

JEL: C55, R31.

Keywords: Multilevel Factor Models, Principal Components, Generalised Canonical Correlation, Housing Market Cycles.

*We are mostly grateful for the insightful comments by Luis Serven, Michael Thornton, Takashi Yamagata, and Chaowen Zheng. We are also grateful to the seminar participants at University of York. Shin acknowledges partial financial support from the Economic Social Research Council in the UK (Grant number ES/T01573X/1). Lin is grateful to the financial support from China Scholarship Council. The usual disclaimer applies.

†Department of Economics, University of York. Email: rui.lin@york.ac.uk

‡Department of Economics, University of York. Email: yongcheol.shin@york.ac.uk

1 Introduction

In a data-rich environment with large cross-section units and time periods, the factor model is a useful technique for dimension reduction, e.g. Chamberlain and Rothschild (1982), Stock and Watson (2002) and Bai and Ng (2002). Recently, the multilevel factor models have gained increasing attention, in which the factors are not only pervasive (i.e. common to all groups) but also semi-pervasive (i.e. common to a subset of groups only). They are referred to as the global and local factors, respectively. Kose et al. (2003) advance the multilevel factor model for characterising the global business cycle, documenting evidence that the global factors play an important role in explaining macroeconomic activities. Barrot and Serven (2018) find that the common factors are the main driving force behind advanced-country capital flows whilst idiosyncratic components dominate the emerging/developing country capital flows. Andreou et al. (2019) show that the industrial production is still the most important workhorse in the US economy, using the two-block factor model with a mixed-frequency data.

Although the principal component (PC) estimation is a popular method in the single-level factor model, it is not directly applicable to the multilevel setting, because it can only estimate the whole factor space consistently but fails to separately identify the global and local factors. This renders the estimation of the multilevel factor model a challenging issue. Wang (2008) proposes a sequential PC approach which updates the global and local factors iteratively, though this approach does not guarantee convergence to the global minimum unless the initial estimate is consistent. Breitung and Eickmeier (2016) and Choi et al. (2018) propose the use of the canonical correlation analysis (CCA) for obtaining an initial consistent estimate of the global factors by employing CCA using any two blocks. Once the (estimated) global factors are projected out, the local factors can be consistently estimated for each block. The global and local factors are iteratively updated until convergence.

Consider, however, the more general multilevel factor models in which some blocks share the common regional factors, see for example, Moench et al. (2013) and Beck et al. (2016). Another case is provided by Hallin and Liška (2011) and Rodríguez-Caballero and Caporin (2019), where the blocks share the pairwise common local factors. In such cases, CCA does not always produce consistent estimate of the global factors because the common local factors can be misidentified as the global factors.

As the main contribution, we propose the generalised canonical correlation analysis (GCC), which extends the standard CCA using any two blocks through constructing the system-wide matrix, denoted $\mathbf{\Phi}$, that contains all the factor spaces from all blocks. As the pairwise canonical correlation between any two blocks is now satisfied simultaneously for all pairs of the blocks, this approach is shown to overcome the aforementioned issue associated with the common local factors. Moreover, unlike most existing studies, GCC is computationally convenient as it does not involve any iteration.

We provide an asymptotic theory that establishes the consistency of the estimated factors and loadings based on the matrix perturbation theory, and derives the asymptotic normal distributions of the factors and loadings estimates. Andreou et al. (2019) develop an asymptotic theory for the factors and loadings estimators under rather stringent conditions, though their theory can be applied to the case with the two blocks only. In this regard, we highlight that our theories are derived under fairly standard assumptions, and the GCC approach can be applied to the more general cases.

Furthermore, we develop a GCC-based consistent selection criteria for identifying the number of the global factors by evaluating the ratios of adjacent singular values of the matrix $\mathbf{\Phi}$. As shown by Han (2021), the standard approaches for selecting the number of factors (r_0) in the single-level factor literature (e.g. Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013)), fail to generate reliable model selection in the multilevel case. Recently, a few approaches have been proposed to deal with an issue of consistently estimating r_0 under the multilevel setting. Andreou et al. (2019) propose a testing procedure by deriving the asymptotic distribution of the canonical correlation between the factor spaces in a two block model. Choi et al. (2021) develop consistent selection criteria for determining the number of the global factors based on the average pairwise canonical correlation among all blocks. Chen (2022) proposes a selection criteron based on the average residual sum of square ($ARSS$) from a regression of (estimated) global factors on the factor spaces in each block. It is important to notice that our approach does not require either the orthogonality between the global and local factors or the selection of any tuning parameters. This makes the GCC criterion more general than existing studies.

Via Monte Carlo simulations, we first focus on the consistent estimation of the global factors and the number of the global factors, finding that GCC outperforms the CCA approach by Andreou et al. (2019) and Choi et al. (2021), and the circular projection matrix estimation (CPE) approach by Chen (2022) under all experiments we consider. Next, we document evidence that the GCC estimator of the global factors and loadings is well-centered and tends to the standard normal density, confirming the validity of our asymptotic theory.

We apply the GCC approach to estimating the multilevel factor model and characterising the national and regional housing market cycles in England and Wales using a large disaggregated panel data of the real house price growth rates for the 331 local authorities over the period 1996Q1 to 2021Q2. The main empirical findings are summarised as follows:

We first detect one global (national) factor, one local factor in the seven regions (NE, NW, YH, EE, LD, SE and WA) but no local factor in the three regions (EM, WM and SW) (see Table 1). Second, the national factor explains a considerable portion of the hosue price inflation variation with a mean of 46.6% while the regional factor contribution is much weaker with its average at 8.3% only. This suggests that the house market in England and Wales appears to be more integrated than the U.S. market (e.g. Del Negro and Otrok (2007)). Third, we can identify that the regional factor components of EE, LD and SE (Area 1) co-move closely while those of NE, NW, YH and WA (Area 2) tend to cluster, confirming that the regional factors are common across some regions. Fourth, the national housing market cycle captured by the global factor components displays a typical boom-bust-recover behaviour, which is in line with the conventional view that the national housing market cycle is pro-cyclical and closely related to economic fundamentals (e.g. Chodorow-Reich et al. (2021)). By contrast, the regional housing market cycles captured by the regional/areal factor components display a heterogeneous and opposition pattern unrelated to fundamentals, demonstrating a housing market segmentation in the North and the South. Finally, we document evidence that the growth rate of the (lagged) population gap between areas strongly comoves with the areal components gap, suggesting that the population gap growth may be an important driver behind the regional house price gap.

The rest of the paper is structured as follows. Section 2 introduces the multilevel factor model and provides a review of the related literature. Section 3 proposes the novel GCC approach and presents the main estimation algorithms. Section 4 develops the asymptotic of the GCC estimator. We also advance a new selection criterion for identifying the number of the global factors. Section 5 reports Monte Carlo simulation results. Section 6 presents an empirical application to the house price inflation data in England and Wales. Section 7 offers concluding remarks. The mathematical proofs, the additional simulation results and theoretical derivations are relegated to the Online Appendix.

2 The Multilevel Factor Model

Consider the multilevel factor model:

$$y_{ijt} = \boldsymbol{\gamma}_{ij}' \boldsymbol{G}_t + \boldsymbol{\lambda}_{ij}' \boldsymbol{F}_{it} + e_{ijt}, i = 1, ..., R, j = 1, ..., N_i, t = 1, ..., T \tag{1}$$

where $\boldsymbol{G}_t = \left[G_t^1, ..., G_t^{r_0}\right]'$ is the $r_0 \times 1$ vector of the global factors, $\boldsymbol{F}_{it} = \left[F_{it}^1, ..., F_{it}^{r_i}\right]'$ is the $r_i \times 1$ vector of the local factors in the block i, $\boldsymbol{\gamma}_{ij}$ and $\boldsymbol{\lambda}_{ij}$ are the corresponding factor loadings, and e_{ijt} is the idiosyncratic error. Stacking (1) across the N_i individuals in block i, we have:

$$\boldsymbol{Y}_{it} = \boldsymbol{\Gamma}_i \boldsymbol{G}_t + \boldsymbol{\Lambda}_i \boldsymbol{F}_{it} + \boldsymbol{e}_{it}, \tag{2}$$

where $\underset{N_i \times 1}{\boldsymbol{Y}_{it}} = [y_{i1t}, ..., y_{iN_it}]'$, $\underset{N_i \times 1}{\boldsymbol{e}_{it}} = [e_{i1t}, ..., e_{iN_it}]'$, $\underset{N_i \times r_0}{\boldsymbol{\Gamma}_i} = (\boldsymbol{\gamma}_{i1}, ..., \boldsymbol{\gamma}_{iN_i})'$ and $\underset{N_i \times r_i}{\boldsymbol{\Lambda}_i} = [\boldsymbol{\lambda}_{i1}, ..., \boldsymbol{\lambda}_{iN_i}]'$. The model can also be written as

$$\boldsymbol{Y}_t = \boldsymbol{\Theta}^+ \boldsymbol{K}_t^+ + \boldsymbol{e}_t, \tag{3}$$

where

$$\underset{N \times 1}{\boldsymbol{Y}_t} = \begin{bmatrix} \boldsymbol{Y}_{1t} \\ \vdots \\ \boldsymbol{Y}_{Rt} \end{bmatrix}, \underset{N \times 1}{\boldsymbol{e}_t} = \begin{bmatrix} \boldsymbol{e}_{1t} \\ \vdots \\ \boldsymbol{e}_{Rt} \end{bmatrix}, \underset{r^+ \times 1}{\boldsymbol{K}_t^+} = \begin{bmatrix} \boldsymbol{G}_t \\ \boldsymbol{F}_{1t} \\ \vdots \\ \boldsymbol{F}_{Rt} \end{bmatrix}, \underset{N \times r^+}{\boldsymbol{\Theta}^+} = \begin{bmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Lambda}_1 & 0 & \cdots & 0 \\ \boldsymbol{\Gamma}_2 & 0 & \boldsymbol{\Lambda}_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_R & 0 & 0 & \cdots & \boldsymbol{\Lambda}_R \end{bmatrix}$$

with $N = \sum_{i=1}^R N_i$ and $r^+ = r_0 + \sum_{i=1}^R r_i$. Further, the model is written in a matrix form:

$$\boldsymbol{Y} = \boldsymbol{K}^+ \boldsymbol{\Theta}^{+\prime} + \boldsymbol{e}, \tag{4}$$

where $\underset{T \times N}{\boldsymbol{Y}} = [\boldsymbol{Y}_1, ..., \boldsymbol{Y}_T]'$, $\underset{T \times r^+}{\boldsymbol{K}^+} = [\boldsymbol{K}_1, ..., \boldsymbol{K}_T]'$, and $\underset{T \times N}{\boldsymbol{e}} = [\boldsymbol{e}_1, ..., \boldsymbol{e}_T]'$.

Alternatively, stacking (1) over time period t, we can rewrite the model as

$$\boldsymbol{Y}_{ij} = \boldsymbol{G}\boldsymbol{\gamma}_{ij} + \boldsymbol{F}_i \boldsymbol{\lambda}_{ij} + \boldsymbol{e}_{ij} = \boldsymbol{K}_i \boldsymbol{\theta}_{ij} + \boldsymbol{e}_{ij} \tag{5}$$

where $\underset{T \times 1}{\boldsymbol{Y}_{ij}} = [y_{ij1}, ..., y_{ijT}]'$, $\underset{T \times 1}{\boldsymbol{e}_{ij}} = [e_{ij1}, ..., e_{ijT}]'$, $\underset{T \times r_0}{\boldsymbol{G}} = [\boldsymbol{G}_1, ..., \boldsymbol{G}_T]'$, $\underset{T \times r_i}{\boldsymbol{F}_i} = [\boldsymbol{F}_{i1}, ..., \boldsymbol{F}_{iT}]'$, $\boldsymbol{\theta}_{ij} = \left[\boldsymbol{\gamma}_{ij}', \boldsymbol{\lambda}_{ij}'\right]'$ and $\boldsymbol{K}_i = [\boldsymbol{G}, \boldsymbol{F}_i]$. For each block i, we then have:

$$\boldsymbol{Y}_i = \boldsymbol{G}\boldsymbol{\Gamma}_i' + \boldsymbol{F}_i \boldsymbol{\Lambda}_i' + \boldsymbol{e}_i = \boldsymbol{K}_i \boldsymbol{\Theta}_i' + \boldsymbol{e}_i \tag{6}$$

where $\boldsymbol{Y}_i = [\boldsymbol{Y}_{i1}, \boldsymbol{Y}_{i2}, \ldots, \boldsymbol{Y}_{iN_i}]$, $\boldsymbol{e}_i = [\boldsymbol{e}_{i1}, \boldsymbol{e}_{i2}, \ldots, \boldsymbol{e}_{iN_i}]$ and $\boldsymbol{\Theta}_i = [\boldsymbol{\Gamma}_i, \boldsymbol{\Lambda}_i]$.

The primary issue in the multilevel factor model is to identify the global and local factors, separately. Suppose that we express the model (2) as

$$\boldsymbol{Y}_{it} = \boldsymbol{\Gamma}_i \boldsymbol{G}_t + \boldsymbol{u}_{it},\ \boldsymbol{u}_{it} = \boldsymbol{\Lambda}_i \boldsymbol{F}_{it} + \boldsymbol{e}_{it}, \tag{7}$$

where the local factors are treated as the part of the error components. The first r_0 factors extracted from the PC estimation applied to the whole data $\boldsymbol{Y}_t = [\boldsymbol{Y}_{1t}', \ldots, \boldsymbol{Y}_{Rt}']'$, will be inconsistent estimates of $\boldsymbol{G}_t$ because the weak correlation condition among the error components in $\boldsymbol{u}_t = [\boldsymbol{u}_{1t}', \ldots, \boldsymbol{u}_{Rt}']'$ is violated due to the presence of the local factors (see Breitung and Eickmeier (2016)). Alternatively, if we apply the PC estimation to each block $\boldsymbol{Y}_i$ in (6), the factor space spanned by $\boldsymbol{K}_i = [\boldsymbol{G}, \boldsymbol{F}]$ can be consistently estimated up to rotation, though the global and local factors cannot be separately identified.[1]

[1]Moreover, the r^+ factors extracted from $\boldsymbol{Y}_t$ in (3) are not necessarily consistent estimates of $\boldsymbol{K}^+$. Lemma 2 in Freyaldenhoven (2021) establishes that the local factors can be consistently estimated only if the number of individuals within that group is larger than $\sqrt{N}$.

A number of alternative methods have been developed to separately identify the global and local factors. Wang (2008) proposed an iterative sequential approach. Given the estimated global factors and loadings, denoted $\widehat{\boldsymbol{G}}$ and $\widehat{\boldsymbol{\Gamma}}_i$, then the local factors and loadings for each block i can be estimated from the following PC estimation:

$$\boldsymbol{Y}_i - \widehat{\boldsymbol{G}}\widehat{\boldsymbol{\Gamma}}_i' = \boldsymbol{F}_i \boldsymbol{\Lambda}_i' + \boldsymbol{e}_i \tag{8}$$

Given the estimated local factors and loadings, denoted $\widehat{\boldsymbol{F}}_i$ and $\widehat{\boldsymbol{\Lambda}}_i$, then the global factors and loadings can be updated by the following PC estimation:

$$\left[\boldsymbol{Y}_1 - \widehat{\boldsymbol{F}}_1 \widehat{\boldsymbol{\Lambda}}_1', \ldots, \boldsymbol{Y}_R - \widehat{\boldsymbol{F}}_R \widehat{\boldsymbol{\Lambda}}_R'\right] = \boldsymbol{G}\left[\boldsymbol{\Gamma}_1', \ldots, \boldsymbol{\Gamma}_R'\right] + \boldsymbol{e}$$

This procedure will be repeated until convergence. However, this approach does not guarantee consistency unless the initial estimates of the global factors and loadings are consistent, because the least square objective function is not globally convex.

To get consistent initial estimates of the global factors, Breitung and Eickmeier (2016) and Choi et al. (2018) propose the use of the canonical correlation analysis (CCA), where the canonical correlation between $\widehat{\boldsymbol{K}}_m$ and $\widehat{\boldsymbol{K}}_h$ is estimated using the PC from any two blocks m and h. For simplicity assume that r_0, r_m and r_h are known and set $r_0 + r_m = r_0 + r_h$. Then, we consider the following characteristic equation:

$$\left(\widehat{\boldsymbol{S}}_{mh}\widehat{\boldsymbol{S}}_{hh}^{-1}\widehat{\boldsymbol{S}}_{hm} - \ell \widehat{\boldsymbol{S}}_{mm}\right)\boldsymbol{v} = \boldsymbol{0} \tag{9}$$

where $\widehat{\boldsymbol{S}}_{ab}$ $(a, b = m, h)$ denotes the variance matrix between $\widehat{\boldsymbol{K}}_m$ and $\widehat{\boldsymbol{K}}_h$. We then obtain the solution ℓ by the (squared) canonical correlations between $\widehat{\boldsymbol{K}}_m$ and $\widehat{\boldsymbol{K}}_h$. Since $\widehat{\boldsymbol{K}}_m$ and $\widehat{\boldsymbol{K}}_h$ share the factor space spanned by the global factors, the r_0 largest canonical correlations will be equal to one asymptotically. Therefore, we can consistently estimate the global factors by $\widehat{\boldsymbol{G}} = \widehat{\boldsymbol{K}}_m \boldsymbol{V}_m^{r_0}$, where $\boldsymbol{V}_m^{r_0}$ is an $(r_0 + r_m) \times r_0$ matrix consisting of the characteristic vectors corresponding to the r_0 largest characteristic roots. Next, after projecting $\widehat{\boldsymbol{G}}$ out, we can consistently estimate the local factors and loadings. In practice, this estimation proceeds iteratively until convergence. Breitung and Eickmeier (2016) and Choi et al. (2018) suggest choosing the block pair (m, h) that yields the largest canonical correlation. Andreou et al. (2019) develop an asymptotic theory for the estimated factors and loadings under rather stringent conditions, though their theory can be applied to the case with the two blocks only.

However, the pairwise identification strategy, based on CCA, does not always produce the consistent estimation of the global factors. For instance, consider a two-level factor model with three blocks ($R = 3$) and $r_0 = r_i = 1$ for $i = 1, 2, 3$. Suppose that the first and second blocks share the same local factor, and we obtain the largest canonical correlation between $\widehat{\boldsymbol{K}}_1$ and $\widehat{\boldsymbol{K}}_2$. Now, we are no longer sure whether $\widehat{\boldsymbol{K}}_1 \boldsymbol{V}_1^{r_0}$ produces the consistent estimate of the global factor or the (common) local factor. Furthermore, the number of global factors tends to be overestimated. A few empirical studies show that some blocks, that share the same geographic region, are subject to (common) regional factors. Hallin and Liška (2011) find one common local factor between France and Germany in a three-country model using industrial production indices for France, Germany and Italy. Alternatively, Rodríguez-Caballero and Caporin (2019) consider the pairwise-common local factors by employing two parallel country classifications using the Debt/GDP ratio and credit ratings, in which case CCA cannot consistently estimate the global factors. See also Moench et al. (2013) and Beck et al. (2016).

Hence, to overcome this important issue, we propose the GCC by incorporating the information from all blocks simultaneously. Recently, Chen (2022) proposed a circular projection estimation (CPE) approach. The circular projection matrix is a successive product of the factor spaces of $\boldsymbol{K}_i$, given by the product inside the bracket in $\left[\left(\prod_{i=1}^{R} P\left(\boldsymbol{K}_i\right)\right)' \left(\prod_{i=1}^{R} P\left(\boldsymbol{K}_i\right)\right)\right] \boldsymbol{\zeta} = \pi\boldsymbol{\zeta}$, where $P(.)$ is the projection matrix, and π and $\boldsymbol{\zeta}$ are the eigenvalue and eigenvector. Only if $\pi = 1$, then $\boldsymbol{\zeta}$ is a global factor. Hence, the global factors can be estimated as $\sqrt{T}$ times the r_0 eigenvectors corresponding to the unit eigenvalues of the circular projection matrix by replacing $\boldsymbol{K}_i$ by $\widehat{\boldsymbol{K}}_i$. The CPE does not suffer from the issue related to the common local factors since it encompasses all blocks. By contrast, the GCC estimates the global factors by a linear combination of the factor spaces (see (19) below). This yields a simpler asymptotic expansion of the global factors, which enables us to directly derive the asymptotic normal distribution of the estimator of the global factors. Moreover, via the simulation studies, we show that GCC outperforms CPE in all cases considered (see Section 4).

3 The Generalised Canonical Correlation Analysis

We begin with the standard canonical correlation analysis (CCA) by selecting any two blocks, h and m, and letting $\boldsymbol{K}_m$ and $\boldsymbol{K}_h$ be $T \times (r_0 + r_m)$ and $T \times (r_0 + r_h)$ matrices consisting of the global and local factors. The CCA aims to find the linear combinations $\boldsymbol{v}_{mj}$ and $\boldsymbol{v}_{hj}$ such that

$$(\boldsymbol{v}_{mj}, \boldsymbol{v}_{hj}) = \underset{\boldsymbol{v}_m, \boldsymbol{v}_h}{\operatorname{argmax}}\, Corr\left(\boldsymbol{K}_m\boldsymbol{v}_m, \boldsymbol{K}_h\boldsymbol{v}_h\right). \tag{10}$$

subject to the restrictions

$$\boldsymbol{V}_m'\boldsymbol{K}_m'\boldsymbol{K}_m\boldsymbol{V}_m = \boldsymbol{I}_{r_{\min}} \text{ and } \boldsymbol{V}_h'\boldsymbol{K}_h'\boldsymbol{K}_h\boldsymbol{V}_h = \boldsymbol{I}_{r_{\min}} \tag{11}$$

where $r_{\min} = \min\{r_0 + r_m, r_0 + r_h\}$, $\boldsymbol{V}_m = [\boldsymbol{v}_{m1}, \ldots, \boldsymbol{v}_{mr_{\min}}]$ and $\boldsymbol{V}_h = [\boldsymbol{v}_{h1}, \ldots, \boldsymbol{v}_{hr_{\min}}]$. If $\boldsymbol{K}_m$ and $\boldsymbol{K}_h$ share the r_0 global factors, then there exists r_0 linear combinations such that their correlations are equal to one or equivalently

$$\boldsymbol{K}_m\boldsymbol{V}_m^{r_0} = \boldsymbol{K}_h\boldsymbol{V}_h^{r_0} \tag{12}$$

where $\boldsymbol{V}_m^{r_0} = [\boldsymbol{v}_{m1}, \ldots, \boldsymbol{v}_{mr_0}]$ and $\boldsymbol{V}_h^{r_0} = [\boldsymbol{v}_{h1}, \ldots, \boldsymbol{v}_{hr_0}]$ are the matrices collecting such linear combinations. We then solve the following characteristic equation:

$$\left(\boldsymbol{S}_{mh}\boldsymbol{S}_{hh}^{-1}\boldsymbol{S}_{hm} - \ell\boldsymbol{S}_{mm}\right)\boldsymbol{v} = \mathbf{0}$$

to obtain $\boldsymbol{V}_m^{r_0}$ that is the collection of characteristic vectors $\boldsymbol{v}$ corresponding to the r_0 largest characteristic roots.

Notice, however, that CCA cannot always identify the global factors in the presence of common local factors. To address this important issue, we propose the generalised canonical correlation (GCC) analysis by constructing the following $T(R-1)R/2 \times \sum_{l=1}^{R}(r_0 + r_l)$ system-wide matrix:

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{K}_1 & -\boldsymbol{K}_2 & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \boldsymbol{K}_1 & \mathbf{0} & -\boldsymbol{K}_3 & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ & & & & \vdots & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{K}_{R-1} & -\boldsymbol{K}_R \end{bmatrix} \tag{13}$$

where $\boldsymbol{K}_i = [\boldsymbol{G}, \boldsymbol{F}_i]$ for $i = 1, ..., R$. We then find the kernel of $\boldsymbol{\Phi}$, i.e. a set of vectors collected by the matrix $\boldsymbol{Q} = [\boldsymbol{Q}_1', \ldots, \boldsymbol{Q}_R']'$ that satisfies:

$$\boldsymbol{\Phi}\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{K}_1\boldsymbol{Q}_1 - \boldsymbol{K}_2\boldsymbol{Q}_2 \\ \boldsymbol{K}_1\boldsymbol{Q}_1 - \boldsymbol{K}_3\boldsymbol{Q}_3 \\ \vdots \\ \boldsymbol{K}_{R-1}\boldsymbol{Q}_{R-1} - \boldsymbol{K}_R\boldsymbol{Q}_R \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}$$

To this end we consider the following singular value decomposition (SVD) of $\boldsymbol{\Phi}$:

$$\boldsymbol{\Phi} = \boldsymbol{P}\boldsymbol{\Delta}\boldsymbol{Q}' \tag{14}$$

such that $\boldsymbol{\Phi}\boldsymbol{Q} = \boldsymbol{P}\boldsymbol{\Delta}$, where $\boldsymbol{P}$ and $\boldsymbol{Q}$ are the $TR(R-1)/2 \times \sum_{l=1}^{R}(r_0+r_l)$ and $\sum_{l=1}^{R}(r_0+r_l) \times \sum_{l=1}^{R}(r_0+r_l)$ orthonormal matrices, and $\boldsymbol{\Delta} = diag\{\delta_1, \delta_2, \ldots, \delta_{\sum_{l=1}^{R}(r_0+r_l)}\}$ is a $\sum_{l=1}^{R}(r_0+r_l) \times \sum_{l=1}^{R}(r_0+r_l)$ diagonal matrix consisting of the singular values in *ascending order*. If we can find a set of vectors $\boldsymbol{q}$ and the singular values $\delta = 0$ such that $\boldsymbol{\Phi}\boldsymbol{q} = \delta\boldsymbol{p} = \boldsymbol{0}$, then we obtain $\boldsymbol{Q}$ by the set of vectors, $\boldsymbol{q}$.

We establish the existence of the r_0 zero singular values and the corresponding eigenvectors, denoted $\boldsymbol{Q}^{r_0}$ in the following proposition. [2] A direct example of $\boldsymbol{Q}^{r_0}$ is such that each $\boldsymbol{Q}_i^{r_0} = [\boldsymbol{I}_{r_0}, \boldsymbol{0}]'$ is a selection matrix. To rule out an infeasible case where the global factors can be expressed as a linear combination of the local factors, we assume that $\boldsymbol{G}\boldsymbol{\alpha}_0 = \boldsymbol{F}_1\boldsymbol{\alpha}_1 + \cdots + \boldsymbol{F}_R\boldsymbol{\alpha}_R$ if and only if $\boldsymbol{\alpha}_0 = \boldsymbol{0}$, $\boldsymbol{\alpha}_1 = \boldsymbol{0}$, $\ldots, \boldsymbol{\alpha}_R = \boldsymbol{0}$, which resembles the rank condition in Assumption A of Wang (2008).

Proposition 1. *There exists a $\sum_{l=1}^{R}(r_0+r_l) \times r_0$ matrix, $\boldsymbol{Q}^{r_0} = [\boldsymbol{Q}_1^{r_0\prime}, \boldsymbol{Q}_2^{r_0\prime}, \ldots, \boldsymbol{Q}_R^{r_0\prime}]'$ containing the right eigenvectors of $\boldsymbol{\Phi}$, such that $\boldsymbol{\Phi}\boldsymbol{Q}^{r_0} = \boldsymbol{0}$ with the r_0 zero singular values. Moreover, the remaining singular values of $\boldsymbol{\Phi}$ are larger than zero and of stochastic order $O_p(\sqrt{T})$.*

From Proposition 1 we have:

$$\boldsymbol{K}_1\boldsymbol{Q}_1^{r_0} = \boldsymbol{K}_2\boldsymbol{Q}_2^{r_0} = \cdots = \boldsymbol{K}_R\boldsymbol{Q}_R^{r_0} \tag{15}$$

which shows that the pairwise canonical correlation in (12) is simultaneously satisfied for all pairs of the blocks. This important result demonstrates that all $\boldsymbol{K}_i\boldsymbol{Q}_i^{r_0}$ for $i = 1, ..., R$, obtained by the system approach, can consistently estimate the factor space spanned by $\boldsymbol{G}$.

Let $\boldsymbol{\Psi} = [\boldsymbol{K}_1\boldsymbol{Q}_1^{r_0}, \ldots, \boldsymbol{K}_R\boldsymbol{Q}_R^{r_0}]$ and consider the eigen-decomposition,

$$T^{-1}\boldsymbol{\Psi}\boldsymbol{\Psi}' = \boldsymbol{L}\boldsymbol{\Xi}\boldsymbol{L}', \tag{16}$$

where $\boldsymbol{\Xi}$ is a diagonal matrix containing the eigenvalues of $T^{-1}\boldsymbol{\Psi}\boldsymbol{\Psi}'$ in descending order.

Proposition 2. *The first r_0 columns of $\boldsymbol{L}$, denoted $\boldsymbol{L}^{r_0}$, consists of the factor space spanned by $\boldsymbol{G}$.*

Proposition 2 shows that the global factors can be identified by a linear combination of appropriately rotated block factor spaces. Importantly, the factor space spanned by the r_0 global factors can be consistently estimated so long as the factor spaces of $\boldsymbol{K}_i$ are consistently estimated for $i = 1, ..., R$.

The estimation algorithm proceeds as follows.

[2] We note that the solution $\boldsymbol{Q}_i$'s are equivalent to

$$(\boldsymbol{Q}_1^{r_0}, \boldsymbol{Q}_2^{r_0}, \ldots, \boldsymbol{Q}_R^{r_0}) = \underset{\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots, \boldsymbol{W}_R}{\operatorname{argmin}} \sum_{i=1}^{R} \|\boldsymbol{G} - \boldsymbol{K}_i\boldsymbol{W}_i\|^2 \,,$$

which is more common in the *GCC* literature (see Yang et al. (2019)). Therefore, we name our approach after *GCC* dispite the slight difference in the problem formulation.

Estimation of global factors and loadings We first obtain the PC estimate of $\boldsymbol{K}_i$ for each block i, denoted $\widehat{\boldsymbol{K}}_i$, by $\sqrt{T}$ times the $r_{\max}$ eigenvectors of $\boldsymbol{Y}_i\boldsymbol{Y}_i'$ corresponding to the $r_{\max}$ largest eigenvalues, where $r_{\max} \geq \max_{i=1,\ldots,R}\{r_0 + r_i\}$ is a common positive integer. We then construct the $TR(R-1)/2 \times Rr_{\max}$ matrix, $\widehat{\boldsymbol{\Phi}}$ by replacing $\boldsymbol{K}_i$ with $\widehat{\boldsymbol{K}}_i$ in (13), and evaluate the SDV of $\widehat{\boldsymbol{\Phi}}$ as

$$\widehat{\boldsymbol{\Phi}} = \widehat{\boldsymbol{P}}\widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{Q}}', \tag{17}$$

where $\widehat{\boldsymbol{P}}$ and $\widehat{\boldsymbol{Q}}$ are the $TR(R-1)/2 \times Rr_{\max}$ and $Rr_{\max} \times Rr_{\max}$ orthonormal matrices, and $\widehat{\boldsymbol{\Delta}}$ is the $Rr_{\max} \times Rr_{\max}$ diagonal matrix consisting of the singular values in *ascending order.*

Next, denote $\widehat{\boldsymbol{Q}}^{r_0} = \left[\widehat{\boldsymbol{Q}}_1^{r_0\prime}, \ldots, \widehat{\boldsymbol{Q}}_R^{r_0\prime}\right]'$ as the first r_0 columns of $\widehat{\boldsymbol{Q}}$, and construct the $T \times Rr_0$ matrix, $\widehat{\boldsymbol{\Psi}} = \left[\widehat{\boldsymbol{K}}_1\widehat{\boldsymbol{Q}}_1^{r_0}, \ldots, \widehat{\boldsymbol{K}}_R\widehat{\boldsymbol{Q}}_R^{r_0}\right]$. We consider the eigen decomposition,

$$T^{-1}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Psi}}' = \widehat{\boldsymbol{L}}\widehat{\boldsymbol{\Xi}}\widehat{\boldsymbol{L}}' \tag{18}$$

where $\widehat{\boldsymbol{L}}$ is a $T \times Rr_0$ orthonormal matrix and $\widehat{\boldsymbol{\Xi}}$ is a $T \times T$ diagonal matrix consisting of the eigenvalues in *descending order.* Then, from (18), we obtain the consistent estimator of the global factors, denoted $\widehat{\boldsymbol{G}}$, by the r_0 vectors of $\widehat{\boldsymbol{L}}$ corresponding to the r_0 largest eigenvalues multiplied by $\sqrt{T}$; namely,

$$\widehat{\boldsymbol{G}} = \frac{1}{\sqrt{T}}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Psi}}'\widehat{\boldsymbol{J}}^{r_0} = \frac{1}{\sqrt{T}}\left(\sum_{i=1}^{R}\widehat{\boldsymbol{K}}_i\widehat{\boldsymbol{Q}}_i^{r_0}\widehat{\boldsymbol{Q}}_i^{r_0\prime}\widehat{\boldsymbol{K}}_i'\right)\widehat{\boldsymbol{J}}^{r_0} \tag{19}$$

where $\widehat{\boldsymbol{J}}^{r_0} = \widehat{\boldsymbol{L}}^{r_0}\left(\widehat{\boldsymbol{\Xi}}^{r_0}\right)^{-1}$, $\widehat{\boldsymbol{L}}^{r_0}$ collects the first r_0 columns of $\widehat{\boldsymbol{L}}$ and $\widehat{\boldsymbol{\Xi}}^{r_0}$ is an $r_0 \times r_0$ diagonal matrix consisting of the r_0 largest eigenvalues of $T^{-1}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Psi}}'$ in *descending order.*

Finally, the global factor loadings can be estimated by $\widehat{\boldsymbol{\Gamma}}_i = T^{-1}\boldsymbol{Y}_i'\widehat{\boldsymbol{G}}$.

Estimation of local factors and loadings For each block $i = 1, ..., R$, the local factors, denoted $\widehat{\boldsymbol{F}}_i$, can be consistently estimated by $\sqrt{T}$ times the r_i eigenvectors of $\widehat{\boldsymbol{Y}}_i\widehat{\boldsymbol{Y}}_i'$ corresponding to the r_i largest eigenvalues, where $\widehat{\boldsymbol{Y}}_i = \boldsymbol{Y}_i - \widehat{\boldsymbol{G}}\widehat{\boldsymbol{\Gamma}}_i'$.

The local factor loadings can be estimated by $\widehat{\boldsymbol{\Lambda}}_i = T^{-1}\widehat{\boldsymbol{Y}}_i'\widehat{\boldsymbol{F}}_i$ for each block $i = 1, ..., R$.

4 Asymptotic Theory for the GCC Estimator

Section 4.1 establishes the consistency of estimates of factors and loadings based on the matrix perturbation theory, assuming that the number of global and local factors, r_0 and r_i are known for all i. Section 4.2 develops a consistent selection criteria for determining the number of the global factors. In Section 4.3, we derive asymptotic normal distributions for the factors and loadings estimates.

4.1 Consistent estimation of factors and loadings

Let $\mathcal{M}$ be a finite constant. Following Bai and Ng (2002) and Choi et al. (2021), we assume:

Assumption A.

1. $E(e_{ijt}) = 0$ *and* $E\left(|e_{ijt}|^8\right) \leq \mathcal{M}$ *for all* i, j *and* t.

2. Let $E\left(N_i^{-1}\sum_{j=1}^{N_i} e_{ijs}e_{ijt}\right) = \omega_i(s,t)$ for all i. Then, $|\omega_{i,N_i}(s,s)| \leq \mathcal{M}$ and $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}|\omega_i(s,t)| \leq \mathcal{M}$ for all t.

3. Let $E(e_{ijt}e_{ikt}) = \tau_{i,(jk),t}$, with $|\tau_{i,(jk),t}| \leq |\tau_{i,(jk)}| < \mathcal{M}$ for all i and t. In addition, for each i, we have $N_i^{-1}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\left|\tau_{i,(jk)}\right| \leq \mathcal{M}$.

4. Let $E(e_{ijt}e_{iks}) = \tau_{i,(jk),(ts)}$. For each i, we have

$$\frac{1}{N_i T}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\sum_{t=1}^{T}\sum_{s=1}^{T}|\tau_{i,(jk),(ts)}| \leq \mathcal{M}$$

5. For every i, t and s

$$E\left(\left|\frac{1}{\sqrt{N_i}}\sum_{j=1}^{N_i}\left[e_{ijs}e_{ijt} - E(e_{ijs}e_{ijt})\right]\right|^4\right) \leq \mathcal{M}$$

Assumption B.

1. $T^{-1}\boldsymbol{G}'\boldsymbol{G}$ has distinct eigenvalues. Let $\boldsymbol{K}_{it} = (\boldsymbol{G}_t', \boldsymbol{F}_{it}')'$. For every i and t, we have $E\left(\boldsymbol{K}_{it}\right) = 0$, $E\left(\|\boldsymbol{K}_{it}\|^4\right) < \infty$ and $T^{-1}\boldsymbol{K}_i'\boldsymbol{K}_i \xrightarrow{p} \boldsymbol{\Sigma}_{K_i}$ where $\boldsymbol{\Sigma}_{K_i}$ is positive definite.

2. For each m, h and t,

$$E\left(\frac{1}{N_m}\sum_{j=1}^{N_m}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{K}_{ht}e_{mjt}\right\|^2\right) \leq \mathcal{M}$$

Assumption C.

1. $\|\boldsymbol{\gamma}_{ij}\| \leq \bar{\gamma} < \infty$ and $\|\boldsymbol{\lambda}_{ij}\| \leq \bar{\lambda} < \infty$ for all i and j, where $\bar{\gamma}$ and $\bar{\lambda}$ are constants.

2. For every $i = 1, \cdots, R$,

   (a) rank $(\boldsymbol{\Theta}_i) = r_0 + r_i$ where $\boldsymbol{\Theta}_i = [\boldsymbol{\Gamma}_i, \boldsymbol{\Lambda}_i]$;

   (b) $N_i^{-1}\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i = N_i^{-1}\begin{bmatrix}\boldsymbol{\Gamma}_i'\boldsymbol{\Gamma}_i & \boldsymbol{\Gamma}_i'\boldsymbol{\Lambda}_i \\ \boldsymbol{\Lambda}_i'\boldsymbol{\Gamma}_i & \boldsymbol{\Lambda}_i'\boldsymbol{\Lambda}_i\end{bmatrix} \longrightarrow \boldsymbol{\Sigma}_{\Theta_i} = \begin{bmatrix}\boldsymbol{\Sigma}_{\Gamma_i} & \boldsymbol{\Sigma}_{\Gamma_i\Lambda_i} \\ \boldsymbol{\Sigma}_{\Gamma_i\Lambda_i}' & \boldsymbol{\Sigma}_{\Lambda_i}\end{bmatrix}$ which is a positive-definite matrix;

   (c) $\boldsymbol{\Sigma}_{\Theta_i}\boldsymbol{\Sigma}_{K_i}$ has distinct eigenvalues;

   (d) $\boldsymbol{\Sigma}_{\Lambda_i}\boldsymbol{\Sigma}_{F_i}$ has distinct eigenvalues.

Assumption D. *The global factors are uncorrelated to the local factors; for every i,* $T^{-1}\boldsymbol{K}_i'\boldsymbol{K}_i = \begin{bmatrix}\boldsymbol{\Sigma}_G & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{F_i}\end{bmatrix} + O_p\left(T^{-1/2}\right)$ *where $\boldsymbol{\Sigma}_G$ and $\boldsymbol{\Sigma}_{F_i}$ are $r_0 \times r_0$ and $r_i \times r_i$ full rank matrices.*

Assumption A is an extended version of Assumption C in Bai and Ng (2002), which allows the idiosyncratic errors to be serially and (weakly) cross-sectionally correlated within blocks. This is less restrictive than the assumption made in Choi et al. (2018). Assumptions B and C are standard in the literature. Assumption B.2 allows weak correlation between global/local factors and idiosyncratic errors. Assumption C requires the global (local) factors to have non-trivial contributions to the variance of all individuals within the corresponding block. Assumption D ensures that the global and local factors can be separately identified. Notice that we do not require the orthogonality between global and local factors for consistently estimating the global factors and their dimension, though we need Assumption D for consistent estimation of $\boldsymbol{\Gamma}_i$, $\boldsymbol{\Lambda}_i$, $\boldsymbol{F}_i$ and r_i. More importantly, we allow the local factors to be correlated or even identical across some blocks although some existing studies require the orthogonality among local factors, e.g. Choi et al. (2018) and Han (2021). Nevertheless, the *GCC* estimator is shown to be valid in the presence of the common local factors. We focus on the practical case with a fixed number of blocks R, but the *GCC* can be valid even as $R \to \infty$.[3]

Lemma 1. *Under Assumptions A–C, as $N_i, T \to \infty$, we have:*

$$\frac{1}{\sqrt{T}} \left\| \widehat{\boldsymbol{K}}_i - \boldsymbol{K}_i \widehat{\boldsymbol{H}}_i \right\| = O_p\left(\frac{1}{C_{N_i T}}\right), \ i = 1, ..., R,$$

where $\widehat{\boldsymbol{K}}_i$ is the $T \times r_{\max}$ matrix of the PC estimates given by $\sqrt{T}$ times the $r_{\max}$ eigenvectors of $\mathbf{Y}_i \mathbf{Y}_i'$ corresponding to the $r_{\max}$ largest eigenvalues, $\boldsymbol{K}_i = [\boldsymbol{G}, \boldsymbol{F}_i]$ is the $T \times (r_0 + r_i)$ factors, $\widehat{\boldsymbol{H}}_i$ is the $(r_0 + r_i) \times r_{\max}$ rotation matrix, $C_{N_i T} = \min\left\{\sqrt{N_i}, \sqrt{T}\right\}$, and

$$\frac{1}{\sqrt{T}} \left\| \widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi} \widehat{\boldsymbol{H}} \right\| = O_p\left(\frac{1}{C_{\underline{N}, T}}\right)$$

where $\boldsymbol{\Phi}$ is the $T(R-1)R/2 \times \sum_{l=1}^R (r_0 + r_l)$ matrix defined in (13), $\widehat{\boldsymbol{\Phi}}$ is the $T(R-1)R/2 \times R r_{\max}$ matrix by replacing $\boldsymbol{K}_i$ with $\widehat{\boldsymbol{K}}_i$, $\widehat{\boldsymbol{H}} = diag\left\{\widehat{\boldsymbol{H}}_1, \widehat{\boldsymbol{H}}_2, \ldots, \widehat{\boldsymbol{H}}_R\right\}$ is a $\sum_{l=1}^R (r_0 + r_l) \times R r_{\max}$ block-diagonal rotation matrix and $C_{\underline{N}, T} = \min\{\sqrt{\underline{N}}, \sqrt{T}\}$ with $\underline{N} = \min\{N_1, N_2, \ldots, N_R\}$.

Lemma 1 establishes that as $N_i, T \to \infty$, $\widehat{\boldsymbol{K}}_i$ converges to their population counterpart up to a rotation. The rotation matrix, $\widehat{\boldsymbol{H}}_i$ is shown to exist in Bai and Ng (2002), but we do not need a specific form since any full rank rotation matrix yields the observationally equivalent model.

Lemma 2. *There exists an $R r_{\max} \times r_0$ matrix $\bar{\boldsymbol{Q}}^{r_0}$ such that $\boldsymbol{\Phi} \widehat{\boldsymbol{H}} \bar{\boldsymbol{Q}}^{r_0} = \mathbf{0}$, where the r_0 singular values are zero. The remaining singular values of $\boldsymbol{\Phi} \widehat{\boldsymbol{H}}$ are larger than zero and of stochastic order $O_p(\sqrt{T})$.*

Lemma 2 extends Proposition 1 to the case under the rotation incurred by the *PC* estimation, and enables us to apply Lemma 3 below to $\widehat{\boldsymbol{\Phi}}$ for deriving the convergence rate of the estimated eigenvectors under rotation. It also helps to estimate the number of global factors r_0 by counting the number of zero singular values of $\widehat{\boldsymbol{\Phi}}$ (see Section 4.2).

While the consistency of the estimated eigenvalues are well-established, there are the two main issues in establishing the consistency of the estimated eigenvectors. First, it is widely acknowledged that the convergence of the eigenvectors may not be well-behaved under eigenvalue-multiplicity. Second,

[3] When $R \to \infty$, the identification of global factors is simpler because each block is asymptotically negligible and the *PC* estimation can be applied to the whole data matrix.

convergence rates of the eigenvectors associated with zero eigenvalues are unclear according to Davis-Kahan theorem (see Theorem 3.4 of Stewart and Sun (1990)).

In Lemma 3 we state the perturbation theory developed by Yu et al. (2015), that is a variant of the Davis-Kahan Theorem, and necessary for deriving our consistency results.

Lemma 3. *Let $\boldsymbol{S}$ and $\widehat{\boldsymbol{S}}$ be the $p \times p$ symmetric matrices with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p$ and $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p$, respectively. Fix $1 \leq r \leq s \leq p$ and set $d = s - r + 1$. Assume that $\min\{\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1}\} > 0$, where $\lambda_0 = \infty$ and $\lambda_{p+1} = -\infty$. Let the $p \times d$ matrices $\boldsymbol{V} = [v_r, v_{r+1}, \ldots, v_s]$ and $\hat{\boldsymbol{V}} = [\hat{v}_r, \hat{v}_{r+1}, \ldots, \hat{v}_s]$ have orthogonal columns, satisfying $\boldsymbol{\Sigma}\boldsymbol{v}_j = \lambda_j \boldsymbol{v}_j$ and $\widehat{\boldsymbol{\Sigma}}\hat{\boldsymbol{v}}_j = \lambda_j \hat{\boldsymbol{v}}_j$ for $j = r, r+1, \ldots, s$. Then, there exists a $d \times d$ orthogonal matrix $\widehat{\boldsymbol{O}}$ such that*

$$\left\|\widehat{\boldsymbol{V}}\widehat{\boldsymbol{O}} - \boldsymbol{V}\right\| \leq \frac{2^{3/2}\left\|\widehat{\boldsymbol{S}} - \boldsymbol{S}\right\|}{\min\{\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1}\}}.$$

The Davis-Kahan Theorem states that the eigenvectors converge to their population counterparts corresponding to non-zero eigenvalues up to rotation under eigenvalue-multiplicity for any real symmetric matrices. However, the stochastic bound provided by the Davis-Kahan Theorem cannot be applicable to our case where the eigenvalues of interest are zero. Lemma 3 establishes that the convergence of the eigenvectors still holds up to an orthogonal rotation even if the population eigenvalues are zero.

With Lemmas 1–3, we establish the consistency of the estimated global factors and loadings (up to rotation) in Theorem 1.

Theorem 1. *1. Under Assumptions A–C, as $N_1, N_2, \ldots, N_R, T \to \infty$, we have:*

$$\frac{1}{\sqrt{T}}\left\|\widehat{\boldsymbol{G}} - \boldsymbol{G}\mathbb{H}\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right)$$

2. Under Assumptions A–D, as $N_1, N_2, \ldots, N_R, T \to \infty$, we have:

$$\frac{1}{\sqrt{N_i}}\left\|\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right)$$

where $\mathbb{H} = T^{-1/2}\boldsymbol{G}'\boldsymbol{J}^{r_0}\boldsymbol{U}$ is an $r_0 \times r_0$ rotation matrix, $\boldsymbol{J}^{r_0} = \boldsymbol{L}^{r_0}(\boldsymbol{\Xi}^{r_0})^{-1}$, $\boldsymbol{\Xi}^{r_0}$ is an $r_0 \times r_0$ diagonal matrix consisting of the r_0 non-zero eigenvalues of $T^{-1}\boldsymbol{G}\boldsymbol{G}'$ in descending order, $\boldsymbol{L}^{r_0}$ is a $T \times r_0$ matrix of the corresponding eigenvectors, $\boldsymbol{U}$ is an $r_0 \times r_0$ orthogonal matrix defined in (24), and $C_{\underline{N}T} = \min\{\sqrt{\underline{N}}, \sqrt{T}\}$ with $\underline{N} = \min\{N_1, N_2, \ldots, N_R\}$.

If the main focus is on the consistent estimation of the global factors (e.g. Del Negro and Otrok (2007)), then an orthogonality between global and local factors is not required. This feature is more general than existing studies that assume an orthogonality, see Wang (2008), Choi et al. (2018), Andreou et al. (2019) and Han (2021). But, we still need to impose such an orthogonality for consistent estimation of the global factor loadings.

Given consistent estimates of the global factors and loadings, we next establish the consistency of the estimated local factors and loadings in Theorem 2.

Theorem 2. *Under Assumptions A–D, as $N_i, T \to \infty$, for each $i = 1, \ldots, R$, we have:*

$$\frac{1}{\sqrt{T}}\left\|\widehat{\boldsymbol{F}}_i - \boldsymbol{F}_i\widehat{\mathscr{H}_i}\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right)$$

$$\frac{1}{\sqrt{N_i}}\left\|\widehat{\mathbf{\Lambda}}_i' - \widehat{\mathscr{H}_i}^{-1}\mathbf{\Lambda}_i'\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right)$$

where $\widehat{\mathscr{H}_i} = (\mathbf{\Lambda}_i'\mathbf{\Lambda}_i/N_i)\left(\widehat{\boldsymbol{F}}_i'\boldsymbol{F}/\boldsymbol{T}\right)\widehat{\mathbf{\Upsilon}}_i^{-1}$ *is an* $r_i \times r_i$ *rotation matrix,* $\widehat{\mathbf{\Upsilon}}_i$ *is an* $r_i \times r_i$ *diagonal matrix consisting of the* r_i *largest eigenvalues of* $\frac{1}{N_iT}\widehat{\mathbf{Y}}_i\widehat{\mathbf{Y}}_i'$ *in descending order,* $\widehat{\mathbf{Y}}_i = \mathbf{Y}_i - \widehat{\boldsymbol{G}}\widehat{\mathbf{\Gamma}}_i'$*, and* $C_{\underline{N},T} = \min\{\sqrt{\underline{N}}, \sqrt{T}\}$ *with* $\underline{N} = \min\{N_1, N_2, \ldots, N_R\}$*.*

We allow the local factors to be correlated or identical across some blocks, unlike many existing studies that require orthogonality among the local factors, e.g. Choi et al. (2018) and Han (2021). Theorem 2 establishes that the GCC estimator is still consistent even in the presence of the pairwise common local factors and the local factors common across some blocks.

4.2 Determining the number of global factors

We now develop the GCC criterion for identifying the number of global factors. Consider the diagonal matrix, $\widehat{\mathbf{\Delta}}$ from the SDV of $\widehat{\mathbf{\Phi}}$ defined in (17). Then, we evaluate the ratio of adjacent (squared) singular values in a similar fashion as in Ahn and Horenstein (2013).

Let $\hat{\delta}_1, \ldots, \hat{\delta}_{Rr_{\max}}$ be the diagonal elements of $\widehat{\mathbf{\Delta}}$ in *ascending order*. Then, we propose estimating the number of global factors by

$$\hat{r}_{0,GCC} = \underset{k=0,\ldots,r_{\max}}{\operatorname{argmax}} \frac{\hat{\delta}_{k+1}^2}{\hat{\delta}_k^2} \tag{20}$$

The main idea is that the ratio sharply separates the zero singular value with the positive one. Using Lemma 2, we can show that $\hat{\delta}_k = O_p\left(\sqrt{T}/C_{\underline{N}T}\right)$ for $k = 1, \ldots, r_0$ while $\hat{\delta}_k = O_p\left(\sqrt{T}\right)$ for $k = r_0 + 1, \ldots, Rr_{\max}$, where $C_{\underline{N}T} = \min\{\underline{N}, T\}$ and $\underline{N} = \min\{N_1, N_2, \ldots, N_R\}$. Hence, the ratio is bounded for $k = 0, \ldots, r_0 - 1, r_0 + 1, \ldots, r_{\max}$, but it tends to infinity for $k = r_0$.

To deal with the case with $r_0 = 0$, we set the mock singular value as

$$\hat{\delta}_0^2 = \frac{1}{C_{\underline{N}T}Rr_{\max}}\sum_{k=1}^{Rr_{\max}}\hat{\delta}_k^2$$

Since the average of squared singular values is of stochastic order $O_p\left(\sqrt{T}\right)$, we have: $\hat{\delta}_0 = O_p\left(\sqrt{T}/C_{\underline{N}T}\right)$, that has the same stochastic order as $\hat{\delta}_k$ for $k = 1, \ldots, r_0$. Hence, $\hat{\delta}_1^2/\hat{\delta}_0^2 = O_p(1)$ for $r_0 > 0$ whilst $\hat{\delta}_1^2/\hat{\delta}_0^2 \xrightarrow{p} \infty$ for $r_0 = 0$. This ensures that we do not overestimate r_0 even for $r_0 = 0$.

Theorem 3. *Under Assumptions A–C, we have:*

$$\lim_{N_1,\ldots,N_R,T\to\infty} Pr\left(\hat{r}_{0,GCC} = r_0\right) = 1$$

where $\hat{r}_{0,GCC} = \underset{k=0,\ldots,r_{\max}}{\arg\max}\ \hat{\delta}_{k+1}^2/\hat{\delta}_k^2$*,* $\hat{\delta}_1 \leq \cdots \leq \hat{\delta}_{r_{\max}} \leq \cdots \leq \hat{\delta}_{r_{\max}}$ *are the singular values of* $\widehat{\mathbf{\Phi}}$ *and* $\hat{\delta}_0^2 = (C_{\underline{N}T}Rr_{\max})^{-1}\sum_{l=1}^{Rr_{\max}}\hat{\delta}_l^2$*.*

The justification behind Theorem 3 lies in the sense of the matrix perturbation theory that the eigenvalues converge to their population counterparts under a small perturbation term (see Stewart and Sun (1990)). Notice that if our main focus is on the consistent estimation of r_0, then an orthogonality between global and local factors is not required. This make the GCC criterion more general than existing studies that require orthogonality, e.g. Andreou et al. (2019) and Han (2021).

Given $\hat{r}_0$, we can consistently estimate global factors and loadings, denoted $\widehat{\boldsymbol{G}}$ and $\widehat{\boldsymbol{\Gamma}}_i$. Then, the number of local factors, r_i can be consistently estimated by applying the existing approximate factor model to $\widehat{\boldsymbol{Y}}_i = \boldsymbol{Y}_i - \widehat{\boldsymbol{G}}\widehat{\boldsymbol{\Gamma}}_i'$ for $i = 1, ..., R$, which has been extensively studied, e.g. Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013). See Choi and Jeong (2019) for a comprehensive review.

Related literature Chen (2012) and Dias et al. (2013) develop the following information criteria to determine the number of global and local factors:

$$(\hat{r}_0, \hat{r}_1, \ldots, \hat{r}_i) = \underset{k_0, k_1, \ldots, k_R}{\operatorname{argmin}} \sum_{i=1}^{R} \left\| \boldsymbol{Y}_i - \widehat{\boldsymbol{G}}^{k_0} \widehat{\boldsymbol{\Gamma}}_i^{k_0 \prime} - \widehat{\boldsymbol{F}}_i^{k_i} \widehat{\boldsymbol{\Lambda}}_i^{k_i \prime} \right\|^2 + \text{ penalty}$$

As described in Choi et al. (2021), however, these information criteria have two shortcomings. First, it involves too many combinations of k_0 and k_i even if R is mildly large. Second, it is nontrivial to construct a proper penalty function that can discriminate the respective roles played by the global and local factors.

Andreou et al. (2019) derive the canonical correlation based test statistic given by $\hat{\xi}(r) - r$ where $\hat{\xi}(r) = \sum_{k=1}^{r} \sqrt{\hat{\ell}_k}$ and $\hat{\ell}_k$ is the k-th largest characteristic root of (9). Let $\tilde{\xi}(r)$ be the de-biased and re-scaled version of $\hat{\xi}(r) - r$. Then, it is shown that $\tilde{\xi}(r) \xrightarrow{d} N(0,1)$ for $r = 1, \ldots, r_0$. A sequence of tests can be conducted from $r = r_{\max}$ to $r = 1$ so that r_0 can be estimated by

$$\hat{r}_{0,AGGR} = \max \left\{ r : 1 \leq r \leq r_{\max}, \tilde{\xi}(r) \geq z_{\alpha_{NT}} \right\}$$

where $z_{\alpha_{NT}}$ is a threshold value depending on $(\underline{N}, T)$ and some tuning parameters. However, the main weakness of their approach lies in that it can be applied to the data with the two blocks only.

Choi et al. (2021) develop consistent selection criteria based on the average canonical correlations among all block pairs. Let $\hat{\ell}_{mh,r}$ be the r-th largest characteristic root of (9) between a block pair m and h, and construct the average (squared) canonical correlation by $\hat{s}(r) = \frac{2}{R(R-1)} \sum_{m=1}^{R-1} \sum_{h=m+1}^{R} \hat{\ell}_{mh,r}$. The following two selection criteria, CCD and MCC, are proposed:

$$\hat{r}_{0,CCD} = \underset{r=0,\ldots,r_{\max}+1}{\operatorname{argmax}} \hat{s}(r) - \hat{s}(r+1)$$

$$\hat{r}_{0,MCC} = \max \left\{ 0 \leq r \leq r_{\max} : 1 - \hat{s}(r) - C \times \text{penalty} < 0 \right\}$$

where C is a data dependent tuning parameter. CCD is consistent while imposing a slightly strong condition that the average canonical correlation has an upper bound. MCC does not require this condition but $1 - \hat{s}(r)$ needs to be modified by the product of a data dependent tuning parameter and a penalty term. We conjecture that CCD and MCC can be consistent in the presence of multi-block common local factors while they become inconsistent in the presence of the pairwise common local factors.[4]

Chen (2022) proposes a selection criteron based on the average residual sum of square ($ARSS$) from a regression of $\widehat{\boldsymbol{\zeta}}_r$ on $\widehat{\boldsymbol{K}}_i$ given by $ARSS_r = \frac{1}{R} \sum_{i=1}^{R} \widehat{\boldsymbol{\zeta}}_r' \left(\boldsymbol{I}_T - P\left(\widehat{\boldsymbol{K}}_i\right) \right) \widehat{\boldsymbol{\zeta}}_r$, where $\widehat{\boldsymbol{\zeta}}_r$ is the eigenvector corresponding to the r-th largest eigenvalue of the circular projection matrix, $\left[\left(\prod_{i=1}^{R} P\left(\widehat{\boldsymbol{K}}_i\right) \right)' \left(\prod_{i=1}^{R} P\left(\widehat{\boldsymbol{K}}_i\right) \right) \right]$.

[4]For instance, if the two blocks share the pairwise common local factors, then the $r_0 + 1$ largest canonical correlations between such a block pair is equal to one, in which case CCD and MCC tend to select the $r_0 + 1$ global factors instead of r_0. We also observe that CCD and MCC are sensitive to the excessively large $r_{\max}$ when the errors are serially correlated. By contrast, in (unreported) simulations, we find that GCC is generally insensitive to the coice of $r_{\max}$.

Chen suggests estimating r_0 by

$$\hat{r}_{0,ARSS} = \underset{r=1,\ldots,r_{\max}}{\text{argmax}} \ \text{Logistic}(\log\log(\underline{N}) \times ARSS_{r+1}) - \text{Logistic}(\log\log(\underline{N}) \times ARSS_r)$$

where the logistic function, $\text{Logistic}(x) = P_1/[1 + A\exp(-\tau x)]$ polarises $ARSS_r$ to 0 or 1 with $A = P_1/P_0 - 1$, $P_0 = 10^{-3}$, $P_1 = 1$ and $\tau = 14$. The $ARSS$ can allow non-zero correlations between local factors, but it does not cover the case with a zero global factor, implying that the $ARSS$ estimator always overestimates r_0 when $r_0 = 0$ (see the simulation evidence in Section 5).

4.3 Asymptotic distributions of the estimated factors and loadings

To develop the asymptotic distributions of the estimated factors and loadings, we need to impose slightly stronger conditions than those required for consistency in Section 4.1. Following Bai (2003), we make the additional assumptions.

Assumption E. *For each i, we have $\lim_{N_i,N\to\infty} N/N_i = \alpha_i \leq \mathcal{M}$*

Assumption F.

1. *$\sum_{s=1}^{T} |\omega_{i,N_i}(s,t)| < \mathcal{M}$ for all i and t.*
2. *Let $\tau_{(mh),(kj),t} = E\left(e_{mkt}e_{hjt}\right)$. For every t, we have $|\tau_{(mh),(kj),t}| \leq |\tau_{(mh),(kj)}| \leq \mathcal{M}$. Moreover, for every m,h,k,j, we have $\sum_{k=1}^{N_m} |\tau_{(mh),(kj)}| \leq \mathcal{M}$.*

Assumption G.

1. *For each m, h and t,*

$$E\left(\left\| \frac{1}{\sqrt{N_h T}} \sum_{s=1}^{T} \sum_{k=1}^{N_h} \boldsymbol{K}_{ms} \left[e_{hks}e_{hkt} - E(e_{hks}e_{hkt})\right] \right\|^2\right) \leq \mathcal{M}$$

2. *For each m, h and t, the $(r_0 + r_i) \times (r_0 + r_i)$ matrix satisfies*

$$E\left(\left\| \frac{1}{\sqrt{N_h T}} \sum_{t=1}^{T} \sum_{j=1}^{N_h} \boldsymbol{K}_{mt} \boldsymbol{\theta}'_{hj} e_{hjt} \right\|^2\right) \leq \mathcal{M}$$

3. *For each t, as $N_1, \ldots, N_R \to \infty$, we have*

$$\mathbb{E}_t = \begin{bmatrix} \mathbb{E}_{1t} \\ \mathbb{E}_{2t} \\ \vdots \\ \mathbb{E}_{Rt} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{N_1}} \sum_{j=1}^{N_1} \boldsymbol{\theta}_{1j} e_{1jt} \\ \frac{1}{\sqrt{N_2}} \sum_{j=1}^{N_2} \boldsymbol{\theta}_{2j} e_{2jt} \\ \vdots \\ \frac{1}{\sqrt{N_R}} \sum_{j=1}^{N_R} \boldsymbol{\theta}_{Rj} e_{Rjt} \end{bmatrix} \xrightarrow{d} N\left(\mathbf{0}, \mathbb{D}_t^{(1)}\right)$$

where

$$\mathbb{D}_t^{(1)} = \begin{bmatrix} \mathbb{D}_{11,t}^{(1)} & \mathbb{D}_{12,t}^{(1)} & \dots & \mathbb{D}_{1R,t}^{(1)} \\ \mathbb{D}_{21,t}^{(1)} & \mathbb{D}_{22,t}^{(1)} & \dots & \mathbb{D}_{2R,t}^{(1)} \\ & & \vdots & \\ \mathbb{D}_{R1,t}^{(1)} & \mathbb{D}_{R2,t}^{(1)} & \dots & \mathbb{D}_{RR,t}^{(1)} \end{bmatrix}$$

is the covariance matrix with

$$\mathbb{D}_{mh,t}^{(1)} = plim_{N_m,N_h\to\infty}(N_m N_h)^{-1/2}\sum_{j=1}^{N_m}\sum_{k=1}^{N_h}\boldsymbol{\theta}_{mj}\boldsymbol{\theta}_{hk}' E(e_{mjt}e_{hkt}) \le \mathcal{M}.$$

4. *For each i and j, as $T\to\infty$, we have:*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{G}_t\left(\boldsymbol{\lambda}_{ij}'\boldsymbol{F}_{it}+e_{ijt}\right) \xrightarrow{d} N(\mathbf{0},\mathbb{D}_{ij}^{(2)})$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{F}_t e_{ijt} \xrightarrow{d} N(\mathbf{0},\mathbb{D}_{ij}^{(3)})$$

where $\mathbb{D}_{ij}^{(2)} = plim_{T\to\infty}T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}E\left[\boldsymbol{G}_s\left(\boldsymbol{\lambda}_{ij}'\boldsymbol{F}_{is}+e_{ijs}\right)\left(\boldsymbol{\lambda}_{ij}'\boldsymbol{F}_{it}+e_{ijt}\right)\boldsymbol{G}_t'\right]$ *and* $\mathbb{D}_{ij}^{(3)} = plim_{T\to\infty}\sum_{s=1}^{T}\sum_{t=1}^{T}E\left(\boldsymbol{F}_{it}\boldsymbol{F}_{is}'e_{ijs}e_{ijt}\right)$.

Assumption E imposes that N_i is of the same order of magnitude as N for all $i=1,...,R$, similarly to Choi et al. (2018). Assumptions F and G, corresponding to Assumptions E and F in Bai (2003), are standard in the literature. Assumption F restricts the cross-sectional and serial dependence of the errors. Notice that Assumption F.2 imposes limited cross-block dependence, which is not required in Assumption A. Assumptions G.1 and G.2 are technical conditions for controlling the stochastic order of the bias terms in the asymptotic expansions, though they are not too restrictive since they are summations of zero mean random variables. Assumptions G.3 and G.4 are the central limit theorems that can be applied to several mixing processes.

With Assumptions F and G, Lemma 6 establishes that some parts in the asymptotic expansion of $\widehat{\boldsymbol{K}}_{it}$ achieve a convergence rate faster than $O_p\left(C_{N_iT}^{-1}\right)$, as previously shown in Lemma 1. This allows us to refine the convergence rates of $\widehat{\boldsymbol{Q}}^{r_0}$ and $\widehat{\boldsymbol{L}}^{r_0}$ in Lemma 7 so that they are now $O_p\left(C_{\underline{N}T}^{-2}\right)$ instead of $O_p\left(C_{\underline{N}T}^{-1}\right)$ as in the proof of Theorem 1. By applying these results, we are able to derive the asymptotic normal distributions of the estimated factors and loadings in Theorems 4-7.

Theorem 4. *Under Assumptions A–C and E–G, as $N_1,N_2,\dots,N_R,T\to\infty$ and $\sqrt{N}/T\to 0$, we have for each t:*

$$\sqrt{N}\left[\widehat{\boldsymbol{G}}_t-\left(\mathbb{H}'+\mathbb{B}'\right)\boldsymbol{G}_t\right] = \frac{1}{R}\mathbb{H}'\mathcal{I}'\widehat{\mathbb{C}}\mathbb{E}_t + o_p(1) \xrightarrow{d} N\left(\mathbf{0},\frac{1}{R^2}\mathbb{H}'\mathcal{I}'\mathbb{C}\mathbb{D}_t^{(1)}\mathbb{C}'\mathcal{I}\mathbb{H}\right)$$

where $\mathbb{H}$ is an $r_0\times r_0$ rotation matrix defined in Theorem 1, $\mathcal{I}=[\boldsymbol{I}_{r_0},\dots,\boldsymbol{I}_{r_0}]'$ is an $Rr_0\times r_0$ matrix, $\widehat{\mathbb{C}} = diag\left(\sqrt{\frac{N}{N_1}}\mathbb{I}_1'\left(\frac{\boldsymbol{\Theta}_1'\boldsymbol{\Theta}_1}{N_1}\right)^{-1},...,\sqrt{\frac{N}{N_R}}\mathbb{I}_R'\left(\frac{\boldsymbol{\Theta}_R'\boldsymbol{\Theta}_R}{N_R}\right)^{-1}\right)$ is an $Rr_0\times Rr_0$ block diagonal matrix with $\mathbb{I}_i =$

$[\boldsymbol{I}_{r_0}, \mathbf{0}]'$ *an* $(r_0 + r_i) \times r_0$ *matrix,* $\mathbb{C} = plim_{N_1,\ldots,N_R,T\to\infty}\widehat{\mathbb{C}}$*,* $\mathbb{E}_t$ *and* $\mathbb{D}_t^{(1)}$ *are defined in Assumption G.3, and* $\mathbb{B}$ *is an* $r_0 \times r_0$ *matrix given by*

$$\mathbb{B} = \frac{1}{R}\sum_{i=1}^{R}\sqrt{\frac{1}{N_i}}\mathbb{I}_i'\left(\frac{\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i}{N_i}\right)^{-1}\frac{\boldsymbol{\Theta}_i'\boldsymbol{e}_i'}{\sqrt{N_iT}}\boldsymbol{J}^{r_0}\boldsymbol{U} = O_p\left(\frac{1}{\sqrt{\underline{N}}}\right)$$

where $\boldsymbol{J}^{r_0}$ *and* $\boldsymbol{U}$ *are defined in Theorem 1 and (24).*

Theorem 5. *Under Assumptions A–G, as* $N_1, N_2, \ldots, N_R, T \to \infty$ *and* $\sqrt{T}/\underline{N} \to 0$*, we have for each* i *and* j*:*

$$\sqrt{T}\left[\widehat{\boldsymbol{\gamma}}_{ij} - (\mathbb{H} + \mathbb{B})^{-1}\boldsymbol{\gamma}_{ij}\right] = \mathbb{H}'\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{G}_t\left(\boldsymbol{\lambda}_{ij}'\boldsymbol{F}_{it} + e_{ijt}\right) + o_p(1) \xrightarrow{d} N\left(\mathbf{0}, \mathbb{H}'\mathbb{D}_{ij}^{(2)}\mathbb{H}\right)$$

where $\mathbb{D}_{ij}^{(2)}$ *is defined in Assumption G.4.*

Theorem 6. *Under Assumptions A–G, as* $N_1, N_2, \ldots, N_R, T \to \infty$*, and if* $\sqrt{N_i}/T \to 0$ *and* $0 < N_i/T < \infty$*, then we have for each* t*:*

$$\sqrt{N_i}\left(\widehat{\boldsymbol{F}}_{it} - \widehat{\mathscr{H}}_i'\boldsymbol{F}_{it} - \boldsymbol{\mathcal{B}}_{it}\right) = \widehat{\boldsymbol{\Upsilon}}_i^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\boldsymbol{F}_{is}'\right)\frac{1}{\sqrt{N_i}}\sum_{j=1}^{N_i}\boldsymbol{\lambda}_{ij}\boldsymbol{e}_{ijt} \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Upsilon}_i^{-1}\mathbb{W}_i\mathbb{D}_{ii,t}^{(4)}\mathbb{W}_i'\boldsymbol{\Upsilon}_i^{-1}\right)$$

where $\mathbb{D}_{ii,t}^{(4)} = plim_{N_i\to\infty}N_i^{-1}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\boldsymbol{\lambda}_{ij}\boldsymbol{\lambda}_{ik}'E(e_{ijt}e_{ikt})$ *is a the lower-right* $r_i \times r_i$ *matrix of* $\mathbb{D}_{ii,t}^{(1)}$*, and* $\boldsymbol{\mathcal{B}}_{it}$ *is the bias term given by*

$$\boldsymbol{\mathcal{B}}_{it} = \widehat{\boldsymbol{\Upsilon}}_i^{-1}\frac{1}{N_iT}\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\boldsymbol{F}_{is}'\boldsymbol{\Lambda}_i'\widehat{\boldsymbol{S}}_{i.t} = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

$\boldsymbol{\mathcal{I}}$*,* $\mathbb{C}$ *and* $\mathbb{E}_t$ *are defined in Theorem 3.* $\boldsymbol{\Upsilon}_i^{-1}$ *and* $\mathbb{W}_i$ *are defined in Lemma 11 and* $\boldsymbol{\Sigma}_{\Gamma_i\Lambda_i}$ *is defined in Assumption C.2b.*

Theorem 7. *Under Assumptions A–G, as* $N_1, N_2, \ldots, N_R, T \to \infty$*, and if* $\sqrt{T}/N_i \to 0$ *and* $0 < T/N_i < \infty$*, then we have each* $j = 1, \ldots, N_i$*:*

$$\sqrt{T}\left(\widehat{\boldsymbol{\lambda}}_{ij} - \widehat{\mathscr{H}}_i^{-1}\boldsymbol{\lambda}_{ij} - \mathscr{B}_{ij}\right) = \widehat{\mathscr{H}}_i'\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{F}_{it}e_{ijt} + o_p(1) \xrightarrow{d} N\left(\mathbf{0}, \left(\mathbb{W}_i^{-1}\right)'\mathbb{D}_{ij}^{(3)}\mathbb{W}_i^{-1}\right)$$

where $\mathbb{D}_{ij}^{(3)}$ *is defined in Assumption G.4,* $\mathscr{B}_{ij}$ *is the bias term given by*

$$\mathscr{B}_{ij} = \widehat{\mathscr{H}}_i'\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{F}_{it}\widehat{S}_{ijt} = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

and $\mathbb{W}_i$ *is defined in Lemma 11.*

Theorems 4 and 5 establish that the estimates of the global factors and loadings follow the asymptotic normal distributions. Unlike in Theorem 1, the rotation matrix has an additional term, $\mathbb{B}$ of order $O_p\left(\underline{N}^{-1/2}\right)$, which does not affect the asymptotic variance matrices. To the best of our knowledge, there is no studies that establish the asymptotic distributions of the global factors and loadings. One exception is Andreou et al. (2019), but their theory only applies when $R = 2$.

Theorems 6 and 7 show that there are bias terms $\boldsymbol{\mathcal{B}}_{it}$ and $\mathscr{B}_{ij}$ of order $O_p\left(C_{\underline{N}T}^{-1}\right)$ stemming from the estimation error from the global components, $\widehat{S}_{ijt}$, that is the (t, j) element of $\widehat{\boldsymbol{S}}_i = \boldsymbol{G}\boldsymbol{\Gamma}_i' - \widehat{\boldsymbol{G}}\widehat{\boldsymbol{\Gamma}}_i'$. A similar result is documented by Andreou et al. (2019), who show that the asymptotic distribution of the local factors is not centered. In principle, it is not straightforward to perform the bias correction unless the global factors and loadings are known. Notice, however, that we derive our asymptotic theories under weaker conditions than those imposed by Andreou et al. (2019); namely, we do not assume that the global factors are orthogonal to each other, and the local factors are orthogonal within blocks.

This generality brings forth the rotation matrices in the asymptotic variances, as shown in Theorem 4 and 5. To deal with this issue, we use the wild bootstrap advanced by Gonçalves and Perron (2014) for the global factors. We also use a dependent bootstrapping method developed by Shao (2010) for the global factor loadings to account for the potential serial correlation induced by the local factors as suggested in Assumption G.4 and Theorem 5. The bootstrapped covariance matrices are not consistent estimates for those in Theorems 4 and 5, because the bootstrap version of the rotation matrix $\mathbb{H}^{*(b)}$ changes in each replication and does not necessarily match $\mathbb{H}$. Therefore, we construct confidence intervals (CI) using the percentile estimates based on the back-rotated estimates by

$$\sqrt{N}\left[\left(\mathbb{H}^{*(b)\prime} + \mathbb{B}^{*(b)\prime}\right)^{-1}\widehat{\boldsymbol{G}}_t^{*(b)} - \widehat{\boldsymbol{G}}_t\right] \text{ and } \sqrt{T}\left[\left(\mathbb{H}^{*(b)} + \mathbb{B}^{*(b)}\right)\widehat{\boldsymbol{\gamma}}_{ij}^{*(b)} - \widehat{\boldsymbol{\gamma}}_{ij}\right].$$

Since the resulting CIs are unaffected by the bootstrap rotation matrix, they should provide correct coverage rates. See Appendix B for details.

5 Monte Carlo Simulation

Following Choi et al. (2021) and Han (2021), we generate the multilevel factor data as follows:

$$y_{ijt} = \boldsymbol{\gamma}_{ij}'\boldsymbol{G}_t + \sqrt{\theta_{i1}}\boldsymbol{\lambda}_{ij}'\boldsymbol{F}_{it} + \sqrt{\kappa\theta_{i2}}e_{ijt} = \sum_{z=1}^{r_0}\gamma_{ij}^z G_t^z + \sqrt{\theta_{i1}}\sum_{z=1}^{r_i}\lambda_{ij}^z F_{it}^z + \sqrt{\kappa\theta_{i2}}e_{ijt} \tag{21}$$

for $i = 1, ..., R$, $j = 1, ..., N_i$, and $t = 1, ..., T$, where the superscript z denote the z-th factor and loading. We generate the global factors/loadings, the local factors/loadings and idiosyncratic errors by

$$\boldsymbol{G}_t = \phi_G\boldsymbol{G}_{t-1} + \boldsymbol{v}_t,\ \boldsymbol{v}_t \sim \text{ i.i.d. } N(\boldsymbol{0}, \boldsymbol{I}_{r_0})$$

$$\boldsymbol{F}_{it} = \phi_F\boldsymbol{F}_{i,t-1} + \boldsymbol{w}_{it},\ \boldsymbol{w}_{it} \sim \text{ i.i.d. } N(0, \boldsymbol{I}_{r_i}) \text{ for } i = 1, \ldots, R,$$

$$\gamma_{ij}^Z \sim \text{ i.i.d. } N(0,1) \text{ for } z = 1, \ldots, r_0;\ \lambda_{ij}^z \sim \text{ i.i.d. } N(0,1) \text{ for } z = 1, \ldots, r_i$$

$$e_{ijt} = \phi_e e_{ij,t-1} + \varepsilon_{ijt} + \beta\sum_{1\leq|h|\leq 8}\varepsilon_{i,j-h,t},\ \varepsilon_{ijt} \sim \text{ i.i.d. } N(0,1)$$

We allow global and local factors to be serially correlated, but also idiosyncratic errors to be serially and cross-sectionally correlated.

We control the noise-to-signal ratio by κ. When $\kappa = 1$, the variances associated with the global factors, local factors and idiosyncratic errors are respectively given by

$$Var(\boldsymbol{\gamma}_{ij}' \boldsymbol{G}_t) = \sum_{z=1}^{r_0} Var(\gamma_{ij}^z G_t^z) = \frac{r_0}{1-\phi_G^2},$$

$$Var(\boldsymbol{\lambda}_{ij}' \boldsymbol{F}_{it}) = \sum_{z=1}^{r_i} Var(\lambda_{ij}^z F_{it}^z) = \frac{r_i}{1-\phi_F^2} \text{ and } Var(e_{ijt}) = \frac{1+16\beta^2}{1-\phi_e^2}.$$

We then make the variance contribution of each component equalised for $\kappa = 1$ (e.g. Choi et al. (2018) and Han (2021)). For $r_0 > 0$, we set:

$$\theta_{i1} = \left(\frac{r_0}{1-\phi_G^2}\right)\left(\frac{r_i}{1-\phi_F^2}\right) \text{ and } \theta_{i2} = \left(\frac{r_0}{1-\phi_G^2}\right) \Big/ \left(\frac{1+16\beta^2}{1-\phi_e^2}\right).$$

while for $r_0 = 0$ we set:

$$\theta_{i1} = 1 \text{ and } \theta_{i2} = \left(\frac{r_i}{1-\phi_G^2}\right) \Big/ \left(\frac{1+16\beta^2}{1-\phi_e^2}\right).$$

We consider five DGPs for the following combinations of sample sizes: $R \in \{3, 10\}$, $N_i \in \{20, 50, 100, 200\}$ with $N_1 = \cdots = N_R$ and $T \in \{50, 100, 200\}$. We fix $(r_0, r_i) = (2, 2)$ for $i = 1, \ldots, R$, $\phi_G = \phi_F = 0.5$ and $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ under DGP1, which serves as the benchmark case. DGP2 is the same as DGP1 except that we allow the local factors to be identical for some blocks. To generate the pairwise common local factors for $R = 3$, we set $F_{1t}^1 = F_{2t}^1$, $F_{1t}^2 = F_{3t}^2$ and $F_{2t}^2 = F_{3t}^2$. For $R = 10$, we set $F_{1t}^1 = \cdots = F_{5t}^1$ and $F_{6t}^1 = \cdots = F_{10t}^1$ to allow the presence of multi-block common local factors. DGP3 considers the noisy data with $\kappa = 3$ while the other configurations remain the same as in DGP1. DGP4 and DGP5 replicate DGP1 but allow the local factors to be correlated. Specifically, we generate the local factors by

$$\boldsymbol{F}_t = 0.5\boldsymbol{F}_{t-1} + \boldsymbol{w}_t,\ \boldsymbol{w}_t \sim \text{ i.i.d. } N(0, \boldsymbol{\Omega}_F)$$

where $\boldsymbol{F}_t = [\boldsymbol{F}_{1t}', \ldots, \boldsymbol{F}_{Rt}']'$ and $\boldsymbol{w}_t = [\boldsymbol{w}_{1t}', \ldots, \boldsymbol{w}_{Rt}']'$. We set the diagonal elements of $\boldsymbol{\Omega}_F$ at 1, and the off-diagonal elements (denoted ω_F) at 0.4 and 0.8 in DGP4 and DGP5, respectively. The number of replications of each experiment is set at 1,000.

We focus on the estimation of the global factors $\widehat{\boldsymbol{G}}$ and the number of the global factors $\hat{r}_0$. Without loss of generality we assume that the number of the global factors and local factors are known with $r_{\max} = r_0 + r_i$ for all i. To evaluate the precision of the estimated global factors, we report the trace ratio defined as

$$TR\left(\widehat{\boldsymbol{G}}\right) = \frac{tr\left\{\boldsymbol{G}'\widehat{\boldsymbol{G}}(\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{G}})^{-1}\widehat{\boldsymbol{G}}'\boldsymbol{G}\right\}}{tr\left\{\boldsymbol{G}'\boldsymbol{G}\right\}}$$

where $tr\{.\}$ is the trace of a matrix. The more precise the estimated factors are, the higher the trace ratio is. If the global factors are perfectly estimated, then $TR\left(\widehat{\boldsymbol{G}}\right) = 1$. For comparison, we also report the results generated by the CCA by Andreou et al. (2019) and the CPE by Chen (2022). Since the precision of $\widehat{\boldsymbol{F}}_i$ and $\hat{r}_i$ depend purely on the precision of $\widehat{\boldsymbol{G}}$ and $\hat{r}_0$ due to the sequential estimation, and their properties are extensively studied by existing literature, we only focus on the performance of GCC estimates for $\widehat{\boldsymbol{G}}$ and $\hat{r}_0$.

Table 6 shows the average trace ratios over 1000 repetitions. For DGP1, all three approaches can produce precise estimates of global factors. While GCC and CPE estimates are quite close to each other, GCC substantially outperforms them, especially when N_i and T are small. Under DGP2 where we allow the common local factors across some blocks, CCA is shown to be inconsistent since the largest canonical correlation between the two blocks does not necessarily refer to the presence of the global factors. On the other hand, CPE and GCC do not suffer from this issue, and they continue to be consistent while GCC still outperforms CPE in all sample sizes. For DGP3, all three approaches are negatively affected by the noisy data, but the performance of GCC improves faster as the sample size increases than CCA and CPE. We obtain qualitatively similar results under DGP4 and DGP5. Notice also that the performance of GCC improves as the number of blocks, R increases while CPE does not display this property.[5] Overall, we find that GCC dominates CCA and CPE in all cases we consider.

Table 6 about here

Next, we turn to the estimation of r_0 by GCC together with CCD and MCC advanced by Choi et al. (2021) and $ARSS$ by Chen (2022).[6] Table 7 reports the average of $\hat{r}_0$ over 1,000 replications and the percentages of over- and under-estimation, denoted $(O|U)$. For DGP1, all the four selection criteria perform satisfactory unless the sample size is too small. Under DGP2, CCD and MCC are shown to overestimate r_0 due to the presence of the pairwise common local factors in which case the canonical correlation between the common local factors from such two blocks is expected to be equal to one. While the performance of $ARSS$ is adversely affected, it improves for large N_i and T. We still find that GCC outperforms $ARSS$. For $R = 10$, CCD becomes the most vulnerable to the common regional factors. While MCC and $ARSS$ can produce relatively precise estimates, GCC outperforms them especially in a small T. Under DGP3, we obtain mixed results. CCD and MCC perform better than $ARSS$ and GCC for a small T whilst $ARSS$ and GCC produce more precise estimates than CCD and MCC for a small N_i. All the four selection methods can correctly select r_0 when N_i and T beome large. For DGP4, CCD can produce reliable estimates under the mild correlation between local factors while MCC estimates remain precise unless N_i and T are small. $ARSS$ underperforms when N_i or T is small. GCC has a similar performance to MCC but its performance is much better in small samples. Under DGP5 where the correlation between the local factors is extremely strong, CCD fails completely since the upper bound condition is violated whilst $ARSS$ does not show any sign of improvement. MCC can select r_0 precisely in large samples, but GCC still dominates with a faster convergence. Overall, we find that MCC, $ARSS$ and GCC can be reliable selection criteria, although $ARSS$ tends to over-estimate r_0 when there is no global factor in the data. Given that GCC does not rely upon the penalty function and the tuning parameters, we conclude that GCC is the most robust and reliable criterion.

Table 7 about here

As a robust check we repeat the simulation experiments for $(r_0, r_i) = (1, 1)$ and $(r_0, r_i) = (3, 3)$, and present the outcomes in Table 8 to 11. The results are qualitative similar to those with $(r_0, r_i) = (2, 2)$. As the number of factors in the data increases, we notice that the accuracy of the estimates becomes slightly lower.

Tables 8–11 about here

[5]For example, under DGP3 with $N_i = 20$ and $T = 50$, the trace ratios for CPE and GCC are 0.59 and 0.755 for $R = 3$ while they become 0.59 and 0.919 for $R = 10$.

[6]When implementing these alternative selection criteria, we follow the practical guidelines byChoi et al. (2021) and use $\hat{r}_{\max} = \max\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_R}\}$.

Finally, we investigate whether the global factors and loadings estimated by *GCC* follow the asymptotic normal distribution. For convenience, we fix $R = 3$, $(r_0, r_i) = (2, 2)$, $N_i \in \{20, 100, 200\}$ and $T \in \{50, 200\}$, and consider the benchmark case where $(\phi_G, \phi_F) = (0, 0)$ and $(\beta, \phi_e, \kappa) = (0, 0, 1)$. Using the known quantities in the asymptotic variances in Theorems 4 and 5, we standardise the estimates by

$$\left(\frac{1}{R^2}\mathbb{H}'\mathcal{I}'\mathbb{C}\mathbb{D}_t^{(1)}\mathbb{C}'\mathcal{I}\mathbb{H}\right)^{-1/2}\sqrt{N}\left[\widehat{\boldsymbol{G}}_t - (\mathbb{H}' + \mathbb{B}')\,\boldsymbol{G}_t\right]$$

$$\left(\mathbb{H}'\mathbb{D}_{ij}^{(2)}\mathbb{H}\right)^{-1/2}\sqrt{T}\left[\widehat{\boldsymbol{\gamma}}_{ij} - (\mathbb{H} + \mathbb{B})^{-1}\boldsymbol{\gamma}_{ij}\right]$$

We then compare our estimates with the standard normal density. In Figures 5 and 6 we display the histograms for the first element of $\widehat{\boldsymbol{G}}_t$ and $\widehat{\boldsymbol{\gamma}}_{ij}$ evaluated at $i = 1$, $j = N_i/2$ and $t = T/2$. We find that the standardised estimates are well centered and scaled, and tend to the standard normal density. As N_i and T increase, the approximation becomes more accurate, confirming the validity of our asymptotic theory.

Figures 5 and 6 about here

We also propose a bootstrap approach to produce the valid confidence intervals for the estimated global factors and loadings. In Appendix B, we conduct a simulation study using the bootstrap approach, and find that the coverage rates of the bootstrap CIs are getting close to the nominal 95% as the sample size increases.

6 Empirical Application

Using the multilevel factor model we apply the *GCC* approach to studying the national and regional housing market cycles in England and Wales. Residential houses are the most valuable properties of the households while house price fluctuations can put the financial system at a greater risk of default during a recession. The housing sector is also directly related to employment, investment and consumption, playing a central role in the business cycle (e.g. Leamer (2007)). While house prices are subject to nation-wide shocks, such as the business cycle and credit liquidity, they are also determined by regional characteristics such as local amenities and the land supply. Hence, te housing market cycle is likely to exist at both national and regional levels.

From the website of Office of National Statistics HPSSA Dataset 14, we download the quarterly (mean) house prices of four different types of properties, (detached, semi-detached, terraced and flats/maisonettes) for 331 local authorities over the period 1996Q1 to 2021Q2. The local authorities belong to ten regions: North East (NE), North West (NW), Yorkshire and the Humber (YH), East Midlands (EM), West Midlands (WM), East of England (EE), London (LD), South East (SE), South West (SW) and Wales (WA). Each "block" in the multilevel factor model is referred to as a region.

We construct the real house price growth in the jth local authority of the region i through deflating the nominal house price by CPI and log-differencing it as follows:

$$\pi_{ijt} = 100 \times \log\left(\frac{PRICE_{ijt}}{CPI_t}\right) - 100 \times \log\left(\frac{PRICE_{ij,t-1}}{CPI_{t-1}}\right)$$

By removing the series with missing observations, we end up with a balanced panel with $R = 10$, $N = \sum_{i=1}^{10} N_i = 1300$ and $T = 102$.

Table 1 displays the number of local authorities for each region as well as the mean and standard deviation of π_{ijt}. We observe that the average growth rates for NE, NW, YH and WA are lower than the overall mean, those for EE, LD and SE higher than the overall mean, and those for EM, WM and SW close to the mean. Notice that LD displays the highest mean growth and standard deviation.

Table 1 about here

We apply the GCC approach to estimating the multilevel factor model for the standardised series, denoted $\tilde{\pi}_{ijt}$, with 10 regions, which is referred to as the national-regional model. By setting $r_{\max} = 5$ and applying the GCC criterion in (20), we detect one global (national) factor.[7] Next, by applying BIC_3 to each region,[8] we find that there is one local factor for NE, NW, YH, EE, LD, SE and WA whereas no local factor is detected for EM, WM and SW (see Table 1). The existence of both global and local factors clearly suggests that there are housing market cycles at both national and regional levels.

To measure the strength of the factors relative to idiosyncratic errors, we evaluate the relative importance ratios of the national and regional factors for region i by

$$RIG_i = N_i^{-1} \sum_{j=1}^{N_i} \left(\widehat{\gamma}_{ij}' \widehat{\gamma}_{ij} / \left(T^{-1} \tilde{\boldsymbol{\pi}}_{ij}' \tilde{\boldsymbol{\pi}}_{ij} \right) \right) \text{ and } RIF_i = N_i^{-1} \sum_{j=1}^{N_i} \left(\widehat{\boldsymbol{\lambda}}_{ij}' \widehat{\boldsymbol{\lambda}}_{ij} / \left(T^{-1} \tilde{\boldsymbol{\pi}}_{ij}' \tilde{\boldsymbol{\pi}}_{ij} \right) \right)$$

where $\tilde{\boldsymbol{\pi}}_{ij}$ is the $T \times 1$ vector of the (standardised) real house price growth rates in the j-th local authority of the region i. The results reported in Table 1 show that the global factor explains a considerable portion of the variation, ranging between 29.6% (London) and 55.1% (South West) with a mean of 46.6%. The large variance share explained by the national factor suggests that the house market in England and Wales appears to be more integrated than the U.S. market where the national factor is dominated by the regional factors (see Del Negro and Otrok (2007)). RIGs of YH, EM, WM, EE and SW are above average, exhibiting that these regions are more responsive to national shocks. Interestingly, London is the least sensitive region to the national factor. On the other hand, the regional contribution is much weaker as its average relative importance ratio is only 8.3%. Still, the regional factor explains substantially larger time variations of the house price inflation for London and South East respectively at 22.6% and 15.1%.

To avoid the issue that the estimated global and local factors are subject to rotation/sign indeterminacy, we report the time-varying behaviour of the average global (national) and local (regional) factor-components for each region i at time t that are constructed by $\widehat{\mathcal{G}}_{it} = \bar{\widehat{\gamma}}_i' \widehat{\boldsymbol{G}}_t$ and $\widehat{\mathcal{F}}_{it} = \bar{\widehat{\boldsymbol{\lambda}}}_i' \widehat{\boldsymbol{F}}_{it}$, where $\bar{\widehat{\gamma}}_i = N_i^{-1} \sum_{j=1}^{N_i} \widehat{\gamma}_{ij}$ and $\bar{\widehat{\boldsymbol{\lambda}}}_i = N_i^{-1} \sum_{j=1}^{N_i} \widehat{\boldsymbol{\lambda}}_{ij}$.[9] The trajectories of $\widehat{\mathcal{G}}_{it}$ plotted in Figure 2, are highly persistent but exhibit a typical "boom-bust-recover" pattern of the (recent) housing market cycle.[10] The national factor-components initially displayed an upward trend until 2003Q3, followed by a long-term downturn until 2009Q2. It then made a quick recovery and became relatively stable from 2012 till 2020 when the COVID19 pandemic erupted. We also observe a surge in the national factor-components during

[7] CCD and MCC by Choi et al. (2021) also select one global factor. This result is robust to the different values of $r_{\max}$.

[8] We have also applied alternative selection criteria, IC_{p2} by Bai and Ng (2002), ER by Ahn and Horenstein (2013) and ED by Onatski (2010). First, ER surprisingly reports zero local factors for all regions whilst IC_{p2} and ED tend to produce more factors but the additional factors explain very small portions of variance. Second, BIC_3 is shown to have good finite sample performance, see Choi and Jeong (2019) and Choi et al. (2021).

[9] As the (uniquely identified) factor-components are just scaled factors, they carry qualitatively the same information.

[10] The boom-bust pattern is consistent with the economic theory suggesting that agents are over-optimistic about the fundamentals during a boom, rendering the growth continues to accelerate, whilst as the economy deteriorates following the negative shock, their expectations of capital return are reversed, resulting in the house market collapse, which is further worsened by foreclosures, see Kaplan et al. (2020) and Chodorow-Reich et al. (2021).

the COVID19 period, which was mainly prompted by a tax relief policy introduced by the UK government to boost the economy and improve liquidity.[11]

Figure 2 about here

The first two figures in Figure 3 display the time-varying patterns of the regional factor-components $\widehat{\mathcal{F}}_{it}$, from which we can identify that the regional components of EE, LD and SE (solid lines) comove closely (the upper panel) while those of NE, NW, YH and WA (dotted lines) tend to cluster together (the lower panel). These clustering patterns are corroborated by the correlation matrix among the estimated regional components in Table 2, showing that the first and second off-diagonal elements are close to one, but the other off-diagonal ones are considerably smaller. Furthermore, we observe transparent discrepancies between these two groups (referred to as Area 1 and Area 2). The regional factor-components in Area 1 appear to have an earlier turning point around 2000 than the global components during the boom, but declined sharply during the financial crisis, Brexit and COVID19 period. On the other hand, the regional components in Area 2 tend to move in an opposite direction, but remained remarkably stable since 2008.

Table 2 and Figure 3 about here

Next, we formally investigate an issue of whether there are areal factors common to some regions. We first project the estimated global factors out from the data and obtain the residuals containing only the local factors and errors, which form the new areal data. Then, we apply the GCC and MCC criterion to these areal data consisting of the different combinations of regions. For example, if the local factors of NE, NW, YH, and WA are common, then the number of common (areal) factors should be one, and zero otherwise. Alternatively, we may consider a two-block model with Area 1 and Area 2 as blocks. If the two areal factors are identical, then there should be one common factor. Otherwise, the number of common factor is zero. The results in Table 3 confirm that the local factors are common within each area, but the two areal factors are different. Thus, we can identify three areas, Area 1 (LD, EE and SW) with one areal factor, Area 2 (NE, NW, YH and WA) with one areal factor, and Area 3 (EM, WM and SW) with zero areal factor. Interestingly, these areas are adjacent geographically (see Figure 1). Notice that the existence of an areal factor around London is not in line with the notion that the "London factor" is pervasive nationally,[12] because the main impact of London is more likely to be confined to its neighbouring regions. In this regard, this finding may provide a support to the notion of "convergence club" that the house prices in regions, that are closer and more distant to London, tend to converge separately, e.g. Holmes and Grimes (2008) and Montagnoli and Nagayasu (2015).

Table 3 about here

Next, we estimate a national-areal model with 3 areas, and compare its estimation results with those obtained from the national-regional model with 10 regions. It is remarkable that the correlation between the global factors estimated from these two models is 0.996. Further, the local (areal) factor from Area 1 has correlations of 0.924, 0.974 and 0.977 with the local (regional) factors from EE, LD and SE, whereas the areal factor from Area 2 has correlations of 0.917, 0.978, 0.941 and 0.955 with the regional factors from NE, NW, YH, and W. This confirms the presence of the common local factors among some regions

[11] The residential property buyers in the U.K. pay Stamp Duty Land Tax (SDLT). The first stage of the policy started from July 2020 and ended at June 2021. The tax reduction is effectively raising the nil rate threshold of the property value from £125,000 to £500,000. See `https://www.gov.uk/guidance/stamp-duty-land-tax-temporary-reduced-rates`. As the housing demand was stimulated by the policy, the price was pushed up with the inelastic housing supply.

[12] Holly et al. (2011) propose a spatio-temporal model with the London price set as a common factor for all regions.

in which case the standard CCA-based estimates of the global and local factors may be inconsistent. The third panel in Figure 3 displays the areal factor components constructed by $\widehat{\mathcal{F}}_{at} = (N_a^{-1} \sum_{j=1}^{N_a} \widehat{\boldsymbol{\lambda}}'_{aj})\widehat{\boldsymbol{F}}_{at}$ for $a = 1, 2$. These areal components follow the quite similar time-varying patterns to the clustered regional components as shown in the first two figures in Figure 3.

To assess the information contents of the global/local factor components, we present the correlations between the national/areal factor components and a list of macroeconomic and financial variables in Table 5. The national components are positively correlated with the GDP growth, the number of buildings started and the New York house price growth rate, demonstrating the pro-cyclicality and possibly strong connection to the international housing market. Moreover, the national component is negatively correlated with the unemployment rate (the demand side), whilst they are negatively correlated with the labour force in the construction sector (the supply side). The credit market condition also plays an important role, as the national components are negatively correlated with the mortgage rate and the 20-year government bond yields while positively correlated with residential lending approvals. These results are in line with the conventional view that the national housing market cycle is pro-cyclical and closely related to economic fundamentals (see Chodorow-Reich et al. (2021)). By contrast, the areal housing market cycles captured by the areal components display a heterogeneous and opposition pattern, as shown in the last subplot of Figure 3. Although the areal component in Area 2 is still negatively and positively correlated with the unemployment rate and the residential credit supply respectively, it is positively correlated with the construction labour. Interestingly, the areal component in Area 1 shows that even tight financial market/economy conditions do not seem to suppress the housing market cycle surrounding Area 1. The opposite sign of the correlations reflect that the two areas react differently to changes of financial market/economy conditions. We may therefore conclude that the existence of such distinctive areal factors clearly indicates a housing market segmentation subject to a geographical gradient.

Table 5 about here

Finally, we investigate another important issue called the South-North house price gap, which has been a long-standing political concern. We collect the annual regional population data from Nomis and construct the areal population by the average of the regional population.[13] We also aggregate the areal factor components into the annual ones. The first two figures in Figure 4 display the areal factor components and the (lagged) population growth rate of in Area 1 and Area 2 respectively. We observe that they move closely to each other with correlations of 0.304 and 0.44 respectively for Area 1 and Area 2. Next, we construct the population gap between the two areas, calculated as the population in Area 1 minus the population in Area 2. We then compare its growth rate with the difference (gap) between their areal components. From the third panel in Figure 4, we observe that the growth rate of the (lagged) population gap strongly comoves with the areal components gap with the remarkably high correlation (0.8). This suggests that the growth rate of the previous population gap can become a strong predictor for the areal components gap.[14]

Figure 4 about here

[13]The regional population data can be found in `https://www.nomisweb.co.uk`.

[14]Howard and Liebersohn (2020) show that the expected income inequality may drive the divergence of the house prices through the channel of rent expectation. Our results suggest that the widening population gap also contribute to the house price gap.

7 Conclusion

We have developed a novel approach based on the generalised canonical correlation (GCC) analysis for consistently estimating the global/local factors and loadings in a multilevel factor model. We also introduce a new selection criteria for the number of global factors. The Monte Carlo simulation shows dominating performance of our approach. Our methodology is applied to analysing the house market in England and Wales using a large disaggregated panel data of the real house price growth rates for the 331 local authorities over the period 1996Q1 to 2021Q. We find that the national factor explains about half of the time series variation while the regional factors are less important but non-negligible. Moreover, we show that the regional factors are common to some regions and hence suggesting a national-areal model rather than a national-regional model.

Although we focus on the global-local specification, our approach can be extended to cover the multilevel factor model that has a more complicated grouping scheme. For example, the model in which the individuals can be classified to more than two layers. See the parallel grouping in Breitung and Eickmeier (2016) and the hierarchical grouping in Moench et al. (2013). Furthermore, if the block membership is unknown, it is possible to estimate the block memberships using methods developed by Ando and Bai (2017), Coroneo et al. (2020) and Uematsu and Yamagata (2022) and apply GCC thereafter.

Figure 1: Map of regions in England and Wales

Table 1: Main Empirical Results over 1996Q1–2021Q2

Region	N_i	Mean	Std	$\hat{r}_i$	RIG	RIF
North East	48	0.692	3.238	1	0.445	0.114
North West	153	0.823	3.429	1	0.436	0.082
Yorkshire and The Humber	84	0.848	3.2	1	0.501	0.073
East Midlands	136	0.969	3.75	0	0.507	0.000
West Midlands	119	0.912	2.817	0	0.527	0.000
East of England	180	1.163	2.8	1	0.501	0.092
London	122	1.45	4.362	1	0.296	0.226
South East	256	1.138	2.518	1	0.456	0.151
South West	116	1.072	2.843	0	0.551	0.000
Wales	86	0.875	3.829	1	0.437	0.094
Summary/Average	1300	1.037	3.237		0.466	0.083

N_i is the number of local authorities in each region. Meand and Std represent the mean and standard deviation of π_{ijt} from each region j. $\hat{r}_i$ is the number of local factors estimated by BIC_3 after projecting out one global factor selected by GCC. RIG_i and RIF_i are the relative importance ratios of global and local factors for block i, which are calculated as $RIG_i = N_i^{-1} \sum_{j=1}^{N_i} \left(\widehat{\boldsymbol{\gamma}}'_{ij} \widehat{\boldsymbol{\gamma}}_{ij} / T^{-1} \tilde{\boldsymbol{\pi}}'_{ij} \tilde{\boldsymbol{\pi}}_{ij} \right)$ and $RIF_i = N_i^{-1} \sum_{j=1}^{N_i} \left(\widehat{\boldsymbol{\lambda}}'_{ij} \widehat{\boldsymbol{\lambda}}_{ij} / T^{-1} \tilde{\boldsymbol{\pi}}'_{ij} \tilde{\boldsymbol{\pi}}_{ij} \right)$.

Table 2: Correlation matrix among the regional factor components

	NE	NW	YH	W	EE	LD	SE
NE	1	0.859	0.885	0.827	-0.59	-0.383	-0.512
NW	0.859	1	0.911	0.946	-0.659	-0.471	-0.585
YH	0.885	0.911	1	0.884	-0.672	-0.531	-0.628
W	0.827	0.946	0.884	1	-0.628	-0.456	-0.559
EE	-0.59	-0.659	-0.672	-0.628	1	0.859	0.948
LD	-0.383	-0.471	-0.531	-0.456	0.859	1	0.927
SE	-0.512	-0.585	-0.628	-0.559	0.948	0.927	1

Table 3: Test of the number of common local factors from new blocks after $\widehat{\boldsymbol{G}}$ being projected out

New Blocks	$\hat{r}_{MCC}$	$\hat{r}_{GCC}$
NE, NW, YH, W	1	1
EE, LD, SE	1	1
Area 1, Area 2	0	0

Table 4: Relative importance ratios from the Nation-Area model

Area	$\hat{r}_i$	RIG	RIF
Area 1	1	0.447	0.132
Area 2	1	0.429	0.104
Area 3	0	0.525	0.000
Avg		0.467	0.079

Table 5: The correlations between factor components and macro variables

	Obs	National	Area 1	Area 2
GDP (Growth Rate)	102	0.135	0.055	0.006
IP (Growth Rate)	102	0.106	0.031	−0.047
CPI (Growth Rate)	102	−0.39**	−0.156	0.003
Employment	102	0.198	−0.34	0.146
Unemployment	102	−0.439***	0.321	−0.241
Construction Labour (Log)	98	−0.304	−0.387**	0.492***
Building Started (Log)	97	0.532***	−0.028	0.298
Residential Investment (Log)	98	−0.269	−0.428***	0.272
New York House Price (Growth Rate)	102	0.655***	−0.176	0.21
M1 (Growth Rate)	102	0.228	0.166	0.103
M3 (Growth Rate)	102	0.062	0.028	0.15
Residential Lending Approvals (Log)	102	0.238	−0.434***	0.467***
Mortgage Rate	58	−0.343	0.354	0.135
Inter Bank Lending Rate Overnight	98	0.371*	0.303	0.048
Inter Bank Lending Rate 3 Months	87	0.287	0.163	0.085
Government Zero Coupon Bond Yields 5 Years	102	0.064	0.074	0.078
Government Zero Coupon Bond Yields 10 Years	102	−0.257	0.019	0.04
Government Zero Coupon Bond Yields 20 Years	100	−0.575***	−0.083	0.008

***, ** and * indicate 1%, 5% and 10% significance level respectively. The data of macro variables from GDP to Unemployment rate are downloaded from the website of Office for National Statistics: https://www.ons.gov.uk/. The financial variables from M1 to zero coupon bond yield are downloaded from the website of Bank of Endland: https://www.bankofengland.co.uk/statistics/research-datasets.

Figure 2: Estimated national components

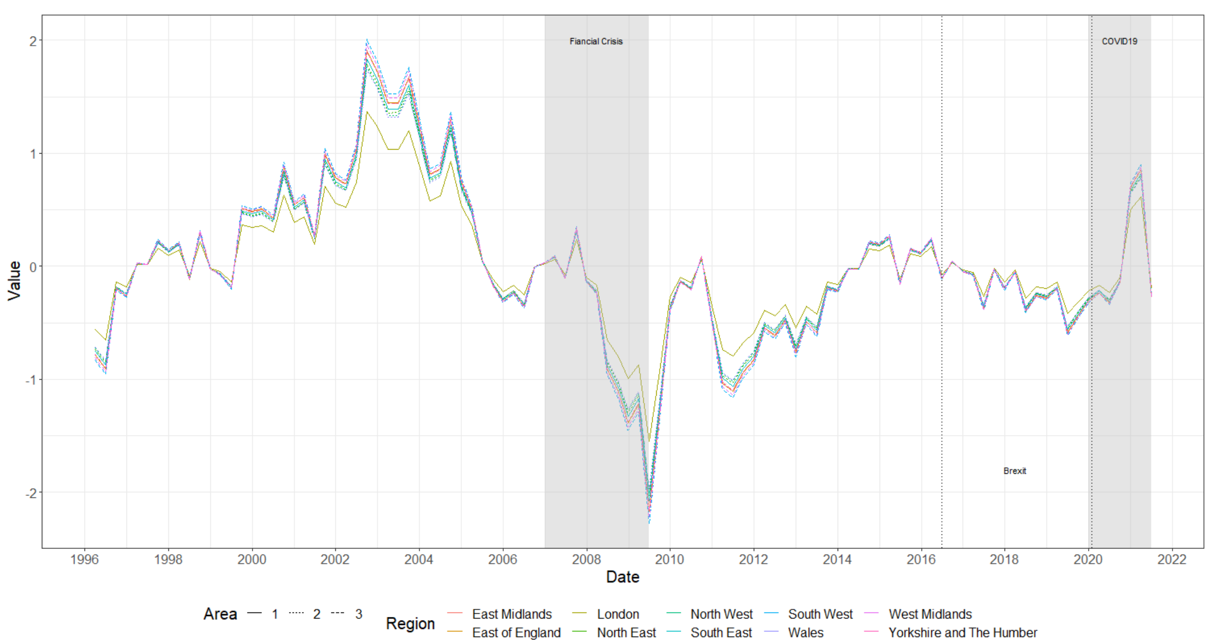

Figure 3: Estimated regional components

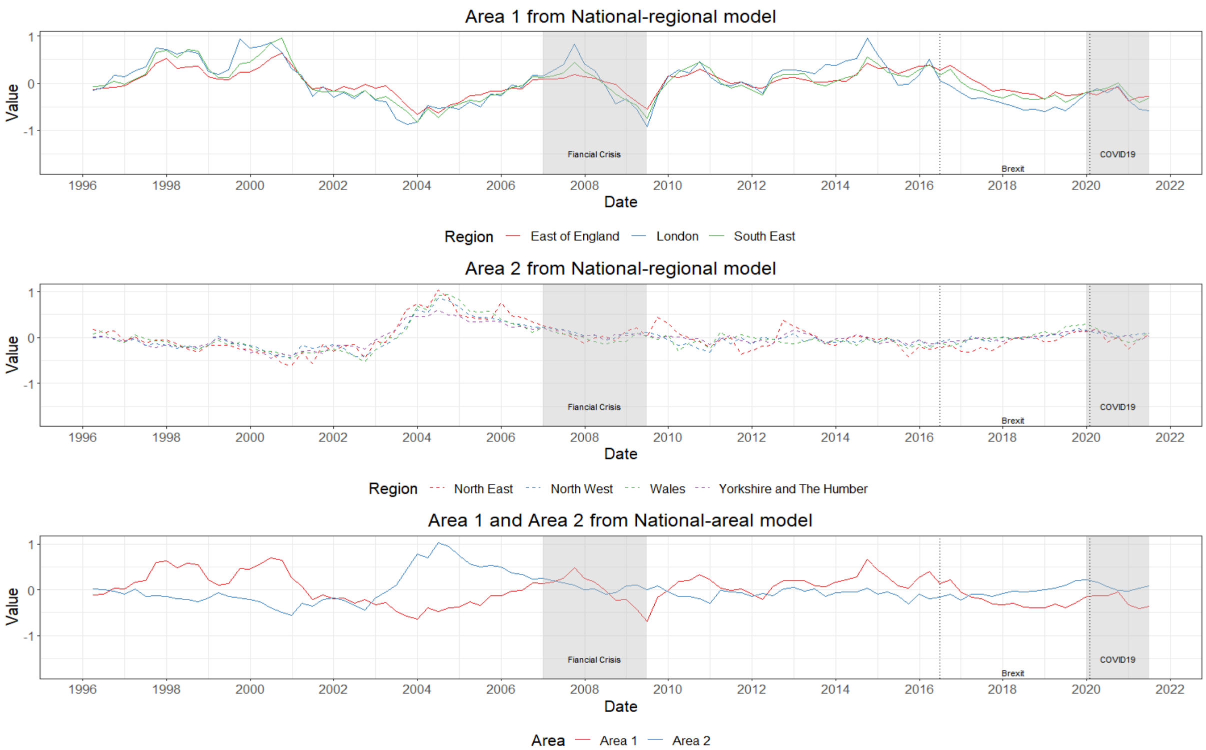

Figure 4: Areal components and population

Table 6: Average trace ratios of the global factor estimates with $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (2, 2)$

			CCA	CPE	GCC	CCA	CPE	GCC	CCA	CPE	GCC	CCA	CPE	GCC	CCA	CPE	GCC
			DGP1			DGP2			DGP3			DGP4			DGP5		
			$(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$			$(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$			$(\beta, \phi_e, \kappa) = (0.1, 0.5, 3)$			$(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$			$(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$		
R	N_i	T	benchmark			common local factors			noisy data			$\omega_F = 0.4$			$\omega_F = 0.8$		
3	20	50	0.82	0.827	0.926	0.637	0.809	0.885	0.595	0.59	0.755	0.794	0.813	0.902	0.69	0.725	0.774
3	50	50	0.93	0.942	0.977	0.661	0.941	0.971	0.727	0.744	0.861	0.911	0.94	0.974	0.784	0.894	0.926
3	100	50	0.956	0.974	0.989	0.655	0.973	0.988	0.838	0.863	0.929	0.936	0.974	0.989	0.824	0.963	0.98
3	200	50	0.969	0.987	0.994	0.658	0.987	0.993	0.904	0.931	0.962	0.955	0.987	0.994	0.844	0.984	0.991
3	20	100	0.843	0.834	0.938	0.626	0.818	0.9	0.606	0.585	0.789	0.82	0.814	0.912	0.716	0.72	0.776
3	50	100	0.949	0.95	0.982	0.654	0.949	0.98	0.772	0.761	0.898	0.944	0.949	0.98	0.87	0.925	0.957
3	100	100	0.973	0.977	0.991	0.663	0.977	0.991	0.904	0.906	0.961	0.969	0.976	0.991	0.923	0.973	0.988
3	200	100	0.985	0.989	0.996	0.666	0.988	0.995	0.953	0.957	0.982	0.982	0.989	0.996	0.939	0.987	0.995
3	20	200	0.848	0.836	0.941	0.617	0.82	0.909	0.614	0.586	0.812	0.834	0.825	0.924	0.731	0.72	0.786
3	50	200	0.954	0.952	0.983	0.649	0.951	0.982	0.8	0.785	0.916	0.952	0.952	0.982	0.921	0.939	0.971
3	100	200	0.978	0.978	0.992	0.659	0.978	0.992	0.921	0.918	0.97	0.977	0.978	0.992	0.961	0.976	0.991
3	200	200	0.989	0.989	0.996	0.664	0.989	0.996	0.963	0.963	0.986	0.988	0.989	0.996	0.976	0.989	0.996
10	20	50	0.843	0.834	0.98	0.677	0.758	0.97	0.632	0.59	0.919	0.819	0.823	0.969	0.709	0.73	0.821
10	50	50	0.933	0.944	0.992	0.709	0.932	0.991	0.751	0.744	0.948	0.914	0.945	0.991	0.793	0.917	0.963
10	100	50	0.958	0.974	0.996	0.722	0.973	0.996	0.851	0.862	0.967	0.944	0.974	0.995	0.836	0.969	0.99
10	200	50	0.971	0.987	0.997	0.721	0.986	0.997	0.911	0.932	0.979	0.956	0.987	0.997	0.845	0.986	0.996
10	20	100	0.862	0.836	0.984	0.671	0.759	0.978	0.654	0.589	0.943	0.851	0.829	0.978	0.737	0.735	0.836
10	50	100	0.954	0.949	0.994	0.715	0.947	0.994	0.798	0.765	0.969	0.949	0.949	0.994	0.875	0.94	0.983
10	100	100	0.976	0.977	0.997	0.728	0.976	0.997	0.912	0.903	0.986	0.972	0.976	0.997	0.92	0.975	0.995
10	200	100	0.986	0.989	0.998	0.731	0.989	0.998	0.956	0.957	0.992	0.983	0.989	0.998	0.939	0.988	0.998
10	20	200	0.868	0.836	0.984	0.663	0.767	0.981	0.653	0.588	0.95	0.854	0.832	0.982	0.76	0.758	0.864
10	50	200	0.958	0.951	0.995	0.716	0.95	0.995	0.823	0.784	0.976	0.956	0.951	0.995	0.924	0.947	0.99
10	100	200	0.979	0.978	0.998	0.734	0.977	0.998	0.929	0.919	0.99	0.978	0.978	0.998	0.963	0.977	0.997
10	200	200	0.989	0.989	0.999	0.736	0.989	0.999	0.966	0.963	0.995	0.989	0.989	0.999	0.977	0.989	0.999

Each entry is the average of trace ratios over 1,000 replications. r_0 and r_i are the true number of global factors and true number of local factors in group i. We set $r_1 = \cdots = r_R$, and $N_1 = \cdots = N_R$ where N_i is the number of individuals in block i. T is the number of time periods. ϕ_G and ϕ_F are AR coefficients for the global and local factors. β, ϕ_e and κ control the cross-section correlation, serial correlation and noise-to-signal ratio.

Table 7: Average estimates of the number of global factors with $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (2, 2)$

			DGP1 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$				DGP2 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ common local factors				DGP3 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 3)$			
R	N_i	T	CCD	MCC	$ARSS$	GCC	CCD	MCC	$ARSS$	GCC	CCD	MCC	$ARSS$	GCC
3	20	50	2.041(4|0.6)	2.223(22.3|0)	4.329(89.8|0)	1.872(0|12.8)	3.028(99.3|0)	3.035(100|0)	5.496(99.5|0)	1.833(1.2|17.9)	1.551(14.1|39.6)	2.183(20|1.7)	4.478(91.7|0.1)	1.597(4.6|46.6)
3	50	50	2.002(0.2|0)	2(0|0)	3.745(86|0)	1.986(0|1.4)	3.002(100|0)	3(100|0)	4.915(98.3|0)	1.978(0|2.2)	1.923(3.3|10.1)	1.95(0.4|5.4)	3.914(89.7|0)	1.825(0.9|18.5)
3	100	50	2.001(0.1|0)	2(0|0)	4.661(98.2|0)	2(0.1|0.1)	3.002(100|0)	3(100|0)	5.755(100|0)	1.994(0|0.6)	1.962(1.1|4.9)	1.921(0|7.9)	4.88(99.3|0)	1.883(0.3|12)
3	200	50	2(0|0)	2(0|0)	5.899(100|0)	1.999(0|0.1)	3(100|0)	3(100|0)	6.881(100|0)	1.999(0|0.1)	1.961(0.1|4.1)	1.947(0|5.3)	6.077(100|0)	1.944(0|5.6)
3	20	100	1.999(0|0.1)	1.994(0|0.6)	2.029(2.9|0)	1.991(0|0.9)	3(100|0)	2.984(98.4|0)	3.324(72.4|0)	1.953(0|4.7)	1.227(0|46.1)	1.281(0|69.6)	2.052(5.4|0.6)	1.796(0.1|20.5)
3	50	100	2(0|0)	2(0|0)	2.002(0.2|0)	2(0|0)	3(100|0)	3(100|0)	2.585(47.4|0)	2(0|0)	1.861(0|12.6)	1.6(0|39.7)	2.003(0.3|0)	1.991(0|0.9)
3	100	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	3(100|0)	3(100|0)	2.363(32.3|0)	2(0|0)	1.994(0|0.6)	1.943(0|5.7)	2(0|0)	2(0|0)
3	200	100	2(0|0)	2(0|0)	2.021(2|0)	2(0|0)	3(100|0)	3(100|0)	2.512(42|0)	2(0|0)	2(0|0)	2(0|0)	2.026(2.6|0)	2(0|0)
3	20	200	1.998(0|0.2)	1.914(0|8.6)	2(0|0)	1.998(0|0.2)	2.999(99.9|0)	2.827(82.7|0)	2.661(50.7|0)	1.986(0|1.4)	0.913(0|62.2)	0.663(0|98.1)	1.999(0|0.1)	1.953(0|4.7)
3	50	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	3(100|0)	3(100|0)	2.309(29.2|0)	2(0|0)	1.862(0|11.5)	1.263(0|70.1)	2(0|0)	2(0|0)
3	100	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	3(100|0)	3(100|0)	2.15(14.9|0)	2(0|0)	2(0|0)	1.97(0|3)	2(0|0)	2(0|0)
3	200	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	3(100|0)	3(100|0)	2.048(4.8|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
10	20	50	2(0|0)	2.178(17.8|0)	2.001(0.1|0)	1.992(0|0.8)	2.779(77.8|0)	2.997(99.7|0)	2.507(50.4|0)	1.978(0|2.2)	1.281(0.7|40.5)	2.28(28|0)	1.937(0.4|6.7)	1.785(0|21.5)
10	50	50	2(0|0)	2(0|0)	2.001(0.1|0)	2(0|0)	2.955(95.5|0)	2.549(54.9|0)	2.244(24.2|0)	2(0|0)	1.95(0|5)	1.988(0|1.2)	1.999(0.1|0.2)	1.944(0|5.6)
10	100	50	2(0|0)	2(0|0)	2.048(4.6|0)	2(0|0)	2.945(94.5|0)	2.021(2.1|0)	2.36(32.4|0)	2(0|0)	1.984(0|1.6)	1.983(0|1.7)	2.044(3.9|0)	1.977(0|2.3)
10	200	50	2(0|0)	2(0|0)	2.986(56.1|0)	2(0|0)	2.393(39.3|0)	2(0|0)	3.537(66.6|0)	2(0|0)	1.997(0|0.3)	1.985(0|1.5)	3.25(64.3|0)	1.987(0|1.3)
10	20	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.006(0.6|0)	2(0|0)	2.046(4.6|0)	2(0|0)	1.529(0|24.8)	1.282(0|71.8)	1.922(0|7.8)	1.98(0|2)
10	50	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.063(6.3|0)	2(0|0)	2(0|0)	2(0|0)	1.975(0|2.4)	1.695(0|30.5)	2(0|0)	1.999(0|0.1)
10	100	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.056(5.6|0)	2(0|0)	2(0|0)	2(0|0)	1.999(0|0.1)	1.976(0|2.4)	2(0|0)	2(0|0)
10	200	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.04(4|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
10	20	200	2(0|0)	1.995(0|0.5)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	0.98(0|53.4)	0.797(0|100)	1.921(0|7.9)	2(0|0)
10	50	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.983(0|1.5)	1.24(0|75.7)	2(0|0)	2(0|0)
10	100	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.001(0.1|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.99(0|1)	2(0|0)	2(0|0)
10	200	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)

			DGP4 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ $\omega_F = 0.4$				DGP5 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ $\omega_F = 0.8$			
R	N_i	T	CCD	MCC	$ARSS$	GCC	CCD	MCC	$ARSS$	GCC
3	20	50	2.297(24.4|[illegible])	2.632(62.5|0)	4.703(97.8|0)	1.829(1.3|18.5)	3.075(98.4|0)	3.039(99.8|0)	4.62(100|0)	2.576(69.9|12.8)
3	50	50	2.138(13.4|0.1)	2.08(8|0)	4.193(95.3|0)	1.978(0|2.2)	3.011(99.8|0)	2.997(99.7|0)	4.377(100|0)	2.438(48.9|5.3)
3	100	50	2.135(13.5|0)	2.009(0.9|0)	5.02(99.6|0)	1.996(0|0.4)	3.007(99.9|0)	2.973(97.3|0)	5.317(100|0)	2.15(17.8|2.8)
3	200	50	2.123(12.3|0)	2(0|0)	6.268(100|0)	2(0|0)	3.002(100|0)	2.875(87.5|0)	6.629(100|0)	2.014(2.2|0.8)
3	20	100	2.121(12.3|0.2)	1.998(0|0.2)	2.689(66.1|0)	1.948(0.5|5.7)	2.999(99.9|0)	2.95(95|0)	3.021(100|0)	2.921(94.2|2.1)
3	50	100	2.047(4.7|0)	2(0|0)	2.372(36.9|0)	1.999(0|0.1)	3(100|0)	2.976(97.6|0)	3.004(100|0)	2.772(77.3|0.1)
3	100	100	2.032(3.2|0)	2(0|0)	2.211(21.1|0)	2(0|0)	3(100|0)	2.891(89.1|0)	3(100|0)	2.305(30.5|0)
3	200	100	2.025(2.5|0)	2(0|0)	2.266(26|0)	2(0|0)	3(100|0)	2.585(58.5|0)	3.014(100|0)	2.044(4.4|0)
3	20	200	2.044(5.9|1.4)	1.928(0|7.1)	2.517(51.7|0)	1.986(0|1.4)	2.999(99.9|0)	2.598(59.9|0.1)	3(100|0)	2.994(99.5|0.1)
3	50	200	2.013(1.3|0)	2(0|0)	2.082(8.2|0)	2(0|0)	3(100|0)	2.769(76.9|0)	3(100|0)	2.937(93.7|0)
3	100	200	2.005(0.5|0)	2(0|0)	2.008(0.8|0)	2(0|0)	3(100|0)	2.412(41.2|0)	3(100|0)	2.491(49.1|0)
3	200	200	2.001(0.1|0)	2(0|0)	2.001(0.1|0)	2(0|0)	3(100|0)	2.045(4.5|0)	3(100|0)	2.041(4.1|0)
10	20	50	2.059(5.9|0)	2.882(88.2|0)	2.29(29.1|0.1)	1.962(0.5|4.3)	2.999(99.8|0)	3.018(100|0)	2.998(99.8|0)	2.904(92.2|1.8)
10	50	50	2.037(3.7|0)	2.036(3.6|0)	2.139(13.9|0)	2(0|0)	2.997(99.7|0)	2.999(99.9|0)	3(99.9|0)	2.661(66.9|0.8)
10	100	50	2.032(3.2|0)	2(0|0)	2.278(25.2|0)	2(0|0)	3(100|0)	2.993(99.3|0)	3.034(100|0)	2.265(26.8|0.3)
10	200	50	2.057(5.7|0)	2(0|0)	3.403(68.1|0)	2(0|0)	3(100|0)	2.907(90.7|0)	4.049(100|0)	2.007(0.8|0.1)
10	20	100	2.04(4|0)	2(0|0)	2.024(2.4|0)	1.998(0|0.2)	3(100|0)	2.985(98.5|0)	3(100|0)	2.999(99.9|0)
10	50	100	2.004(0.4|0)	2(0|0)	2(0|0)	2(0|0)	3(100|0)	2.992(99.2|0)	3(100|0)	2.903(90.3|0)
10	100	100	2.009(0.9|0)	2(0|0)	2(0|0)	2(0|0)	3(100|0)	2.925(92.5|0)	3(100|0)	2.492(49.2|0)
10	200	100	2.005(0.5|0)	2(0|0)	2(0|0)	2(0|0)	3(100|0)	2.566(56.6|0)	3(100|0)	2.039(3.9|0)
10	20	200	2.006(0.6|0)	1.995(0|0.5)	2(0|0)	2(0|0)	3(100|0)	2.662(66.2|0)	3(100|0)	3(100|0)
10	50	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	3(100|0)	2.788(78.8|0)	3(100|0)	2.99(99|0)
10	100	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	3(100|0)	2.369(36.9|0)	3(100|0)	2.634(63.4|0)
10	200	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	3(100|0)	2.01(1|0)	3(100|0)	2.034(3.4|0)

The average of $\hat{r}_0$ over 1,000 replications is reported together with $(O|U)$ inside the parenthesis, indicating the percentage of overestimation and underestimation. r_0 and r_i are the true numbers of global factors and local factors in group i. We set $r_1 = \cdots = r_R$ and $N_1 = \cdots = N_R$, where R is the number of groups and N_i is the number of individuals in block i. T is the number of time periods. ϕ_G and ϕ_F are AR coefficients for the global and local factors. β, ϕ_e and κ control the cross-section correlation, serial correlation and noise-to-signal ratio.

Table 8: Average trace ratios of the global factor estimates with $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (1, 1)$

			CCA	CPE	GCC	CCA	CPE	GCC	CCA	CPE	GCC	CCA	CPE	GCC	CCA	CPE	GCC
			DGP1			DGP2			DGP3			DGP4			DGP5		
			$(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$			$(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$			$(\beta, \phi_e, \kappa) = (0.1, 0.5, 3)$			$(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$			$(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$		
R	N_i	T				common local factors						$\omega_F = 0.4$			$\omega_F = 0.8$		
3	20	50	0.936	0.927	0.973	0.623	0.927	0.97	0.771	0.697	0.864	0.933	0.925	0.972	0.882	0.903	0.949
3	50	50	0.971	0.976	0.991	0.639	0.975	0.991	0.907	0.899	0.958	0.967	0.975	0.991	0.916	0.972	0.988
3	100	50	0.982	0.988	0.995	0.655	0.988	0.995	0.95	0.952	0.98	0.978	0.988	0.995	0.926	0.987	0.995
3	200	50	0.986	0.994	0.998	0.658	0.994	0.998	0.97	0.976	0.989	0.984	0.994	0.998	0.939	0.993	0.997
3	20	100	0.947	0.933	0.976	0.612	0.933	0.975	0.804	0.719	0.893	0.946	0.932	0.976	0.924	0.922	0.964
3	50	100	0.977	0.977	0.992	0.617	0.977	0.992	0.927	0.915	0.968	0.977	0.976	0.992	0.963	0.975	0.991
3	100	100	0.988	0.989	0.996	0.648	0.989	0.996	0.964	0.962	0.986	0.988	0.989	0.996	0.973	0.989	0.996
3	200	100	0.993	0.995	0.998	0.656	0.995	0.998	0.98	0.982	0.993	0.992	0.994	0.998	0.978	0.994	0.998
3	20	200	0.95	0.937	0.978	0.612	0.936	0.977	0.811	0.725	0.897	0.949	0.934	0.977	0.941	0.927	0.969
3	50	200	0.98	0.978	0.992	0.636	0.978	0.992	0.935	0.925	0.973	0.98	0.978	0.992	0.976	0.977	0.992
3	100	200	0.99	0.989	0.996	0.639	0.989	0.996	0.968	0.965	0.988	0.99	0.989	0.996	0.987	0.989	0.996
3	200	200	0.995	0.995	0.998	0.624	0.995	0.998	0.984	0.983	0.994	0.994	0.995	0.998	0.991	0.995	0.998
10	20	50	0.956	0.929	0.992	0.536	0.91	0.991	0.864	0.704	0.962	0.951	0.929	0.992	0.91	0.914	0.98
10	50	50	0.977	0.975	0.997	0.547	0.975	0.997	0.931	0.896	0.985	0.972	0.975	0.997	0.93	0.975	0.996
10	100	50	0.984	0.988	0.998	0.547	0.988	0.998	0.958	0.954	0.991	0.98	0.988	0.998	0.939	0.988	0.998
10	200	50	0.986	0.994	0.999	0.57	0.994	0.999	0.972	0.976	0.994	0.983	0.994	0.999	0.942	0.994	0.999
10	20	100	0.963	0.935	0.993	0.543	0.928	0.993	0.881	0.707	0.969	0.962	0.934	0.993	0.948	0.928	0.988
10	50	100	0.983	0.977	0.998	0.537	0.977	0.997	0.947	0.915	0.99	0.981	0.977	0.997	0.966	0.977	0.997
10	100	100	0.99	0.989	0.999	0.523	0.989	0.999	0.97	0.962	0.995	0.989	0.989	0.999	0.976	0.989	0.999
10	200	100	0.994	0.995	0.999	0.544	0.994	0.999	0.983	0.981	0.997	0.993	0.994	0.999	0.977	0.994	0.999
10	20	200	0.984	0.977	0.998	0.531	0.932	0.993	0.888	0.742	0.972	0.965	0.937	0.993	0.96	0.933	0.991
10	50	200	0.984	0.977	0.998	0.562	0.978	0.998	0.951	0.924	0.992	0.984	0.978	0.998	0.98	0.978	0.997
10	100	200	0.991	0.989	0.999	0.535	0.989	0.999	0.974	0.965	0.996	0.991	0.989	0.999	0.988	0.989	0.999
10	200	200	0.995	0.995	0.999	0.548	0.995	0.999	0.986	0.983	0.998	0.995	0.995	0.999	0.992	0.995	0.999

Each entry is the average of trace ratios over 1,000 replications. r_0 and r_i are the true numbers of the global factors and local factors in group i. We set $r_1 = \cdots = r_R$ and $N_1 = \cdots = N_R$ where N_i is the number of individuals in block i. ϕ_G and ϕ_F are AR coefficients for the global and local factors. β, ϕ_e and κ control the cross-section correlation, serial correlation and noise-to-signal ratio.

Table 9: Average estimates of the number of the global factors with $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (1, 1)$

R	N_i	T	CCD	MCC	$ARSS$	GCC	CCD	MCC	$ARSS$	GCC	CCD	MCC	$ARSS$	GCC
			DGP1 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$				DGP2 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ common local factors				DGP3 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 3)$			
3	20	50	1.004(0.4|0)	1.445(44.1|0)	1.823(51.6|0)	1(0|0)	1.35(31.9|0)	2.003(97.9|0)	1.867(53.9|0)	1(0|0)	1.023(4.1|2.8)	1.652(63.7|0)	1.819(50.2|0)	1.003(0.2|0)
3	50	50	1(0|0)	1(0|0)	1.016(1.5|0)	1(0|0)	1.261(26|0)	1.219(21.9|0)	1.029(2.9|0)	1(0|0)	1.003(0.4|0.1)	1.014(1.4|0)	1.022(2.1|0)	0.999(0|0.1)
3	100	50	1.001(0.1|0)	1(0|0)	1.011(1.1|0)	1(0|0)	1.23(22.9|0)	1.025(2.5|0)	1.025(2.5|0)	1(0|0)	1.002(0.2|0)	1(0|0)	1.01(1|0)	1(0|0)
3	200	50	1.001(0.1|0)	1(0|0)	1.002(0.2|0)	1(0|0)	1.058(5.8|0)	1(0|0)	1.001(0.1|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
3	20	100	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.007(0.7|0)	1(0|0)	1.001(0.1|0)	1(0|0)	0.998(0|0.2)	0.999(0|0.1)	1(0|0)	1(0|0)
3	50	100	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.003(0.3|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
3	100	100	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
3	200	100	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
3	20	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	0.997(0|0.3)	0.986(0|1.4)	1(0|0)	1(0|0)
3	50	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
3	100	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
3	200	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
10	20	50	1(0|0)	1.655(65.5|0)	1(0|0)	1(0|0)	1.96(96|0)	2.014(100|0)	1.633(63.3|0)	1(0|0)	0.993(0|0.7)	1.966(95.8|0)	1(0|0)	1(0|0)
10	50	50	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.988(98.8|0)	1.961(96.1|0)	1.023(2.3|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
10	100	50	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.972(97.2|0)	1.086(8.6|0)	1.012(1.2|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
10	200	50	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.601(60.1|0)	1(0|0)	1.004(0.4|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
10	20	100	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.035(3.5|0)	1(0|0)	1.01(1|0)	1(0|0)				
10	50	100	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.077(7.7|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
10	100	100	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.048(4.8|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
10	200	100	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.011(1.1|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
10	20	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	0.999(0|0.1)	1(0|0)	1(0|0)
10	50	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.002(0.2|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
10	100	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
10	200	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)

R	N_i	T	CCD	MCC	$ARSS$	GCC	CCD	MCC	$ARSS$	GCC
			DGP4 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ $\omega_F = 0.4$				DGP5 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ $\omega_F = 0.8$			
3	20	50	1.016(1.3|0)	1.618(59.9|0)	2.127(66.4|0)	1(0|0)	1.824(73.5|0)	2.118(99.7|0)	2.487(100|0)	1.101(10.1|0)
3	50	50	1.002(0.2|0)	1.006(0.6|0)	1.08(7.8|0)	1(0|0)	1.86(83.5|0)	1.895(89.5|0)	1.991(98.1|0)	1.02(2|0)
3	100	50	1.003(0.3|0)	1(0|0)	1.042(4.2|0)	1(0|0)	1.911(89.3|0)	1.609(60.9|0)	1.97(96.8|0)	1.007(0.7|0)
3	200	50	1.006(0.6|0)	1(0|0)	1.011(1.1|0)	1(0|0)	1.927(92.5|0)	1.17(17|0)	1.274(27.3|0)	1.003(0.3|0)
3	20	100	1(0|0)	1(0|0)	1.035(3.5|0)	1(0|0)	1.951(95.1|0)	1.643(64.3|0)	2(100|0)	1.149(14.9|0)
3	50	100	1(0|0)	1(0|0)	1.001(0.1|0)	1(0|0)	1.968(96.8|0)	1.297(29.7|0)	1.991(99.1|0)	1.008(0.8|0)
3	100	100	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.973(97.3|0)	1.058(5.8|0)	1.678(67.8|0)	1.008(0.8|0)
3	200	100	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.971(97.1|0)	1(0|0)	1(0|0)	1(0|0)
3	20	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.992(99.2|0)	1.056(5.6|0)	2(100|0)	1.159(15.9|0)
3	50	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.999(99.9|0)	1.002(0.2|0)	1.998(99.8|0)	1.002(0.2|0)
3	100	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.998(99.8|0)	1(0|0)	1.237(23.7|0)	1(0|0)
3	200	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	2(100|0)	1(0|0)	1(0|0)	1(0|0)
10	20	50	1(0|0)	1.938(93.4|0)	1.008(0.8|0)	1(0|0)	1.724(72.4|0)	2.118(100|0)	1.965(96.5|0)	1.09(9|0)
10	50	50	1(0|0)	1.001(0.1|0)	1(0|0)	1(0|0)	1.879(87.9|0)	1.965(96.5|0)	1.803(80.3|0)	1.007(0.7|0)
10	100	50	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.92(92|0)	1.666(66.6|0)	1.779(77.9|0)	1.001(0.1|0)
10	200	50	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.959(95.9|0)	1.112(11.2|0)	1.508(50.8|0)	1(0|0)
10	20	100	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.972(97.2|0)	1.709(70.9|0)	1.937(93.7|0)	1.153(15.3|0)
10	50	100	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.988(98.8|0)	1.251(25.1|0)	1.727(72.7|0)	1.002(0.2|0)
10	100	100	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.991(99.1|0)	1.016(1.6|0)	1.476(47.6|0)	1(0|0)
10	200	100	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.996(99.6|0)	1(0|0)	1(0|0)	1(0|0)
10	20	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.996(99.6|0)	1.025(2.5|0)	1.971(97.1|0)	1.138(13.8|0)
10	50	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	2(100|0)	1(0|0)	1.728(72.8|0)	1(0|0)
10	100	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	2(100|0)	1(0|0)	1.24(24|0)	1(0|0)
10	200	200	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1.998(99.8|0)	1(0|0)	1(0|0)	1(0|0)

The average of $\hat{r}_0$ over 1,000 replications is reported together with $(O|U)$ inside the parenthesis, indicating the percentage of overestimation and underestimation. r_0 and r_i are the true numbers of the global factors and local factors in group i. We set $r_1 = \cdots = r_R$ and $N_1 = \cdots = N_R$, where R is the number of groups and N_i is the number of individuals in block i. T is the number of time periods. ϕ_G and ϕ_F are AR coefficients for the global and local factors. β, ϕ_e and κ control the cross-section correlation, serial correlation and noise-to-signal ratio.

Table 10: Average trace ratios of the global factor estimates with $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (3, 3)$

			CCA	CPE	GCC	CCA	CPE	GCC	CCA	CPE	GCC	CCA	CPE	GCC	CCA	CPE	GCC
			DGP1			DGP2			DGP3			DGP4			DGP5		
			$(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$			$(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$			$(\beta, \phi_e, \kappa) = (0.1, 0.5, 3)$			$(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$			$(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$		
R	N_i	T				common local factors						$\omega_F = 0.4$			$\omega_F = 0.8$		
3	20	50	0.741	0.768	0.881	0.56	0.743	0.826	0.531	0.547	0.69	0.707	0.735	0.825	0.666	0.696	0.753
3	50	50	0.867	0.903	0.955	0.611	0.894	0.938	0.611	0.894	0.938	0.83	0.889	0.936	0.731	0.803	0.835
3	100	50	0.921	0.955	0.979	0.624	0.953	0.975	0.725	0.778	0.858	0.881	0.952	0.976	0.773	0.9	0.923
3	200	50	0.943	0.978	0.989	0.642	0.977	0.988	0.803	0.863	0.913	0.915	0.977	0.988	0.794	0.962	0.974
3	20	100	0.762	0.762	0.901	0.537	0.747	0.841	0.545	0.54	0.74	0.726	0.723	0.833	0.66	0.656	0.722
3	50	100	0.909	0.912	0.966	0.579	0.911	0.96	0.672	0.671	0.83	0.895	0.907	0.959	0.766	0.801	0.839
3	100	100	0.958	0.963	0.986	0.603	0.962	0.984	0.812	0.817	0.915	0.95	0.961	0.984	0.837	0.933	0.955
3	200	100	0.975	0.982	0.993	0.612	0.982	0.992	0.912	0.92	0.963	0.969	0.982	0.993	0.876	0.977	0.988
3	20	200	0.767	0.758	0.909	0.518	0.748	0.85	0.55	0.54	0.771	0.729	0.716	0.838	0.649	0.628	0.693
3	50	200	0.92	0.919	0.97	0.549	0.917	0.967	0.677	0.668	0.852	0.915	0.916	0.968	0.784	0.8	0.841
3	100	200	0.964	0.965	0.987	0.59	0.965	0.987	0.85	0.848	0.94	0.962	0.964	0.987	0.908	0.951	0.975
3	200	200	0.982	0.983	0.994	0.611	0.983	0.994	0.938	0.939	0.976	0.981	0.983	0.994	0.947	0.981	0.992
10	20	50	0.752	0.77	0.968	0.544	0.636	0.922	0.562	0.54	0.876	0.728	0.749	0.921	0.67	0.691	0.793
10	50	50	0.872	0.901	0.984	0.569	0.824	0.972	0.657	0.683	0.91	0.833	0.895	0.974	0.741	0.82	0.87
10	100	50	0.925	0.956	0.991	0.569	0.934	0.989	0.736	0.787	0.932	0.888	0.954	0.99	0.775	0.919	0.949
10	200	50	0.943	0.977	0.994	0.578	0.973	0.994	0.807	0.866	0.949	0.917	0.978	0.994	0.802	0.969	0.985
10	20	100	0.779	0.768	0.975	0.513	0.594	0.946	0.577	0.536	0.922	0.747	0.74	0.939	0.674	0.654	0.757
10	50	100	0.915	0.913	0.99	0.542	0.87	0.986	0.685	0.67	0.946	0.896	0.912	0.987	0.765	0.815	0.87
10	100	100	0.959	0.963	0.995	0.556	0.958	0.995	0.821	0.819	0.969	0.95	0.962	0.995	0.849	0.947	0.979
10	200	100	0.977	0.982	0.997	0.563	0.981	0.997	0.917	0.92	0.983	0.97	0.982	0.997	0.886	0.98	0.995
10	20	200	0.78	0.764	0.977	0.497	0.576	0.959	0.582	0.54	0.936	0.747	0.739	0.951	0.659	0.625	0.726
10	50	200	0.924	0.918	0.991	0.532	0.9	0.99	0.694	0.665	0.959	0.919	0.918	0.99	0.792	0.834	0.896
10	100	200	0.966	0.965	0.996	0.534	0.963	0.996	0.859	0.848	0.981	0.964	0.965	0.996	0.912	0.959	0.99
10	200	200	0.983	0.983	0.998	0.532	0.983	0.998	0.942	0.939	0.991	0.981	0.983	0.998	0.949	0.982	0.997

Each entry is the average of trace ratio over 1,000 replications. r_0 and r_i are the true numbers of the global factors and local factors in group i. We set $r_1 = \cdots = r_R$ and $N_1 = \cdots = N_R$ where N_i is the number of individuals in block i. T is the number of time periods. ϕ_G and ϕ_F are AR coefficients for the global and local factors. β, ϕ_e and κ control the cross-section correlation, serial correlation and noise-to-signal ratio.

Table 11: Average estimates of the number of the global factors with $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (3, 3)$

			DGP1 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$				DGP2 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ common local factors				DGP3 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 3)$			
R	N_i	T	CCD	MCC	$ARSS$	GCC	CCD	MCC	$ARSS$	GCC	CCD	MCC	$ARSS$	GCC
3	20	50	3.045(13.3|11.6)	3.027(5.4|2.7)	3.373(29.1|1.3)	2.466(0.5|37.5)	3.843(42.6|4.7)	3.534(52.7|0.3)	3.686(49.9|0.7)	2.226(2.3|53.2)	1.087(7.6|78.8)	2.444(0.6|54.9)	3.111(30.5|27.8)	1.81(3.8|78.7)
3	50	50	3(2|2.1)	2.984(0|1.6)	3.02(2.8|0.8)	2.863(0.1|11.4)	3.508(20.8|0.8)	3.13(13.3|0.3)	3.072(8.1|0.9)	2.672(0.2|24.2)	3.508(20.8|0.8)	3.13(13.3|0.3)	3.072(8.1|0.9)	2.672(0.2|24.2)
3	100	50	3(0.3|0.3)	2.994(0|0.6)	3.003(0.5|0.2)	2.967(0|3.1)	3.189(7.1|0)	3.008(0.8|0)	3.041(4.3|0.2)	2.897(0|8.5)	2.251(2.1|60.9)	1.95(0|87.7)	2.243(0.3|66.3)	1.789(0.4|80.9)
3	200	50	3(0|0)	3(0|0)	3.001(0.1|0)	2.997(0|0.3)	3.034(1.3|0.1)	3(0.1|0.1)	3.014(1.5|0.1)	2.988(0|1.1)	2.147(0.2|69.4)	1.807(0|90.7)	2.181(0|70)	1.78(0.1|83.9)
3	20	100	2.885(0|10.9)	2.418(0|57.2)	2.968(0|3.2)	2.718(0|21)	2.913(1.7|11.1)	2.736(0|26.3)	2.946(0.7|6)	2.364(0.3|45)	0.158(0|98.6)	0.902(0|100)	1.828(0.1|81.8)	1.464(0|84.7)
3	50	100	2.997(0|0.3)	2.951(0|4.9)	2.997(0|0.3)	2.978(0|2.1)	2.996(0|0.4)	2.994(0|0.6)	2.998(0|0.2)	2.924(0|6.8)	1.006(0|90.3)	1.074(0|100)	1.598(0|92.4)	1.334(0|89.4)
3	100	100	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	2.998(0|0.2)	2.133(0|67.1)	1.549(0|96.6)	1.975(0|80)	1.871(0|75.8)
3	200	100	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	2.376(0|56.6)	1.951(0|85.8)	2.301(0|63.2)	2.114(0|68.1)
3	20	200	2.788(0|18.6)	1.861(0|93.7)	2.892(0|10.4)	2.737(0|20.8)	2.787(0.2|16.4)	2.102(0|82.3)	2.903(0|9.6)	2.43(0|40.2)	0.026(0|100)	0.084(0|100)	1.449(0|95)	0.851(0|95)
3	50	200	2.997(0|0.3)	2.863(0|13.7)	2.996(0|0.4)	2.993(0|0.7)	2.998(0|0.2)	2.969(0|3.1)	2.998(0|0.2)	2.975(0|2.4)	0.51(0|97.4)	0.401(0|100)	1.354(0|97.8)	0.97(0|92.9)
3	100	200	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	2.36(0|55.2)	1.366(0|97.5)	2.173(0|67.2)	2.197(0|57.1)
3	200	200	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	2.936(0|6.3)	2.883(0|8.7)	2.924(0|6.9)	2.913(0|6.9)
10	20	50	2.987(0.2|1.5)	3.015(1.6|0.1)	2.967(0|3.2)	2.905(0|7.3)	4.825(64.7|0)	4.008(97.9|0)	3.683(47.2|7.5)	2.561(1.2|30.2)	0.286(0.5|94)	2.689(0.1|31.2)	1.911(0.3|79)	1.96(0.2|68.2)
10	50	50	2.999(0|0.1)	2.999(0|0.1)	2.985(0|1.5)	2.978(0|1.7)	5.166(72.5|0)	3.565(56.4|0)	3.133(14.1|2.2)	2.846(0.2|11.9)	2.119(0.1|59.5)	2.162(0|81.7)	1.717(0|89.1)	1.922(0|69.9)
10	100	50	3(0|0)	2.998(0|0.2)	2.997(0|0.3)	2.995(0|0.5)	4.458(48.6|0)	3.001(0.1|0)	3.162(15.4|0)	2.997(0|0.3)	2.43(0.1|50.7)	2.133(0|81.1)	1.967(0|79.9)	1.979(0|70)
10	200	50	3(0|0)	3(0|0)	3(0|0)	2.999(0|0.1)	4.458(48.6|0)	3.001(0.1|0)	3.162(15.4|0)	2.997(0|0.3)	2.362(0|56)	1.91(0|91.1)	2.065(0|75.5)	1.965(0|73.6)
10	20	100	2.987(0|1.1)	2.535(0|46.5)	2.91(0|8.2)	2.993(0|0.6)	3.015(0.9|1.2)	2.969(0|3.1)	2.901(9|20.3)	2.813(0.1|14.1)	0.014(0|99.9)	0.986(0|100)	1.443(0.3|94.2)	1.846(0|56.9)
10	50	100	2.999(0|0.1)	2.994(0|0.6)	2.997(0|0.3)	2.999(0|0.1)	3.144(4.8|0)	3(0|0)	3.003(0.9|0.6)	2.988(0|1.1)	1.044(0|88.4)	1.151(0|100)	1.346(0|97.2)	1.625(0|70.7)
10	100	100	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3.099(3.3|0)	3(0|0)	3(0|0)	3(0|0)	2.372(0|59.1)	1.676(0|96.3)	1.75(0|86.8)	2.222(0|54.9)
10	200	100	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3.117(3.9|0)	3(0|0)	3(0|0)	3(0|0)	2.538(0|45.4)	2.289(0|64.3)	2.348(0|57.2)	2.425(0|46.4)
10	20	200	2.996(0|0.4)	1.971(0|98.9)	2.896(0|9.8)	2.998(0|0.2)	2.989(0|1.1)	2.217(0|78.3)	2.786(3|23.8)	2.931(0|5.5)	0(0|100)	0.018(0|100)	1.399(0.3|94.8)	0.897(0|76.7)
10	50	200	3(0|0)	2.964(0|3.6)	2.999(0|0.1)	3(0|0)	3(0|0)	3(0|0)	2.999(0|0.1)	3(0|0)	0.32(0|97.1)	0.415(0|100)	1.282(0|98.2)	1.071(0|76.8)
10	100	200	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	2.647(0|35.1)	1.549(0|99.1)	2.144(0|60.8)	2.698(0|21.9)
10	200	200	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	2.998(0|0.2)	2.986(0|1.3)	2.997(0|0.2)	2.998(0|0.2)

			DGP4 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ $\omega_F = 0.4$				DGP5 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ $\omega_F = 0.8$			
R	N_i	T	CCD	MCC	$ARSS$	GCC	CCD	MCC	$ARSS$	GCC
3	20	50	3.65(57.9|5.9)	3.425(42.6|0.2)	4.104(86|0.1)	2.436(17.2|49.6)	4.026(91.2|0.1)	3.899(89.4|0)	4.19(98.8|0)	3.38(66|18.7)
3	50	50	3.57(54.5|0.9)	3.216(21.9|0.3)	3.657(65.2|0.2)	2.771(11.5|24.8)	4.003(99.1|0)	3.964(96.4|0)	3.999(99.6|0)	3.923(94.7|1.6)
3	100	50	3.516(51.1|0.2)	3.05(5.5|0.5)	3.674(67|0)	2.945(6.3|9.4)	4.002(99.8|0)	3.973(97.3|0)	3.993(99.5|0.2)	3.963(97.1|0.8)
3	200	50	3.454(45.5|0.1)	3.002(0.5|0.3)	3.66(66|0)	2.995(1.2|1.6)	3.998(99.8|0)	3.969(96.9|0)	3.783(80.3|1.9)	3.964(96.4|0)
3	20	100	3.462(59.3|10.4)	2.517(0.2|48)	3.735(75.3|1.8)	2.776(33.3|38.3)	3.947(94.8|0.1)	3.296(31.5|1.9)	3.972(97.2|0)	3.875(90.6|2.2)
3	50	100	3.539(54|0.1)	2.968(0.1|3.3)	3.495(49.6|0.1)	3.066(16.3|8.2)	3.998(99.8|0)	3.905(90.5|0)	3.998(99.8|0)	3.991(99.3|0.1)
3	100	100	3.427(42.7|0)	3(0|0)	3.26(26|0)	3.007(1.2|0.5)	4(100|0)	3.991(99.1|0)	3.907(92.4|1.5)	4(100|0)
3	200	100	3.344(34.4|0)	3(0|0)	3.27(27|0)	3(0|0)	4(100|0)	3.968(96.8|0)	3.826(83.4|0.7)	3.995(99.5|0)
3	20	200	3.399(62.9|14.6)	1.87(0|92.8)	3.694(72.2|2.7)	3.068(49.8|29.8)	3.922(92.2|0)	2.704(3.1|32.5)	3.948(94.9|0.1)	3.911(92.5|1.2)
3	50	200	3.545(54.6|0.1)	2.87(0|13)	3.264(26.6|0.2)	3.092(14.3|4.3)	4(100|0)	3.698(70|0.2)	3.98(98.4|0.3)	3.999(99.9|0)
3	100	200	3.338(33.8|0)	3(0|0)	3.1(10|0)	3.001(0.1|0)	4(100|0)	3.974(97.4|0)	3.881(89.6|1.5)	4(100|0)
3	200	200	3.233(23.3|0)	3(0|0)	3.027(2.7|0)	3(0|0)	4(100|0)	3.752(75.2|0)	3.985(98.5|0)	4(100|0)
10	20	50	3.634(61.8|0.5)	3.635(63.5|0)	3.367(39.8|3.1)	2.925(28.1|25.6)	3.997(99.3|0)	3.994(99.4|0)	3.957(95.8|0.1)	3.927(95.3|1.6)
10	50	50	3.531(53.2|0.1)	3.207(20.7|0)	3.113(12.7|1.4)	2.933(10.3|13.1)	4(100|0)	3.993(99.3|0)	3.983(98.3|0)	3.988(99|0.2)
10	100	50	3.465(46.5|0)	3.027(2.7|0)	3.125(12.7|0.2)	2.981(3.3|4.5)	4(100|0)	3.998(99.8|0)	3.998(99.8|0)	3.994(99.4|0)
10	200	50	3.421(42.1|0)	3(0.1|0.1)	3.11(11|0)	2.991(0|0.9)	4(100|0)	3.987(98.7|0)	3.992(99.2|0)	3.98(98|0)
10	20	100	3.685(69.4|0.9)	2.675(0|32.5)	3.027(11.2|8)	3.496(58.3|5.8)	3.998(99.8|0)	3.302(30.3|0.1)	3.915(92|0.4)	3.996(99.6|0)
10	50	100	3.523(52.3|0)	2.999(0|0.1)	3.005(0.9|0.4)	3.094(10|0.5)	4(100|0)	3.982(98.2|0)	3.993(99.3|0)	4(100|0)
10	100	100	3.384(38.4|0)	3(0|0)	3.002(0.2|0)	3.003(0.3|0)	4(100|0)	3.999(99.9|0)	4(100|0)	4(100|0)
10	200	100	3.287(28.7|0)	3(0|0)	3.003(0.3|0)	3(0|0)	4(100|0)	3.98(98|0)	3.998(99.8|0)	3.994(99.4|0)
10	20	200	3.827(84.7|1.2)	1.989(0|98.2)	2.904(3.2|11.6)	3.776(78.8|0.8)	3.998(99.8|0)	2.857(0|14.3)	3.938(94.2|0.4)	4(100|0)
10	50	200	3.603(60.3|0)	2.959(0|4.1)	2.998(0|0.2)	3.072(7.3|0.1)	4(100|0)	3.823(82.3|0)	4(100|0)	4(100|0)
10	100	200	3.278(27.8|0)	3(0|0)	3(0|0)	3(0|0)	4(100|0)	3.991(99.1|0)	3.998(99.8|0)	4(100|0)
10	200	200	3.19(19|0)	3(0|0)	3(0|0)	3(0|0)	4(100|0)	3.796(79.6|0)	4(100|0)	4(100|0)

The average of $\hat{r}_0$ over 1,000 replications is reported together with $(O|U)$ inside the parenthesis, indicating the percentage of overestimation and underestimation. r_0 and r_i are the true numbers of the lobal factors and local factors in group i. We set $r_1 = \cdots = r_R$ and $N_1 = \cdots = N_R$, where R is the number of groups and N_i is the number of individuals in block i. T is the number of time periods. $_G$ and ϕ_F are AR coefficients for the global and local factors. β, ϕ_e and κ control the cross-section correlation, serial correlation and noise-to-signal ratio.

Figure 5: Asymptotic normality of the first element of $\widehat{\boldsymbol{G}}_t$ evaluated at $T/2$

The data is simulated using $R = 3$, $(r_0, r_i) = (2, 2)$, $(\phi_G, \phi_F) = (0, 0)$ and $(\beta, \phi_e, \kappa) = (0, 0, 1)$. Standard normal density is superimposed.

Figure 6: Asymptotic normality of the first element of $\widehat{\gamma}_{ij}$ evaluated at $i = 1$ and $N_i/2$

The data is simulated using $R = 3$, $(r_0, r_i) = (2, 2)$, $(\phi_G, \phi_F) = (0, 0)$ and $(\beta, \phi_e, \kappa) = (0, 0, 1)$. Standard normal density is superimposed.

References

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors, *Econometrica* **81**(3): 1203–1227.

Ando, T. and Bai, J. (2017). Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures, *Journal of the American Statistical Association* **112**(519): 1182–1198.

Andreou, E., Gagliardini, P., Ghysels, E. and Rubin, M. (2019). Inference in group factor models with an application to mixed-frequency data, *Econometrica* **87**(4): 1267–1305.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica: Journal of the Econometric Society* pp. 817–858.

Bai, J. (2003). Inferential theory for factor models of large dimensions, *Econometrica* **71**(1): 135–171.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica* **70**(1): 191–221.

Barrot, L.-D. and Serven, L. (2018). Gross capital flows, common factors, and the global financial cycle, *World Bank Policy Research Working Paper* .

Beck, G. W., Hubrich, K. and Marcellino, M. (2016). On the importance of sectoral and regional shocks for price-setting, *Journal of Applied Econometrics* **31**(7): 1234–1253.

Breitung, J. and Eickmeier, S. (2016). Analyzing international business and financial cycles using multi-level factor models: A comparison of alternative approaches, *Dynamic Factor Models (Advances in Econometrics, Volume 35) Emerald Group Publishing Limited* **35**: 177–214.

Chamberlain, G. and Rothschild, M. (1982). Arbitrage, factor structure, and mean-variance analysis on large asset markets.

Chen, M. (2022). Circularly projected common factors for grouped data, *Journal of Business & Economic Statistics* pp. 1–14.

Chen, P. (2012). Common factors and specific factors, *MPRA Paper No. 36082* .
URL: *"https://mpra.ub.uni-muenchen.de/36114/"*

Chodorow-Reich, G., Guren, A. M. and McQuade, T. J. (2021). The 2000s housing cycle with 2020 hindsight: A neo-kindlebergerian view, *Technical report*, National Bureau of Economic Research.

Choi, I. and Jeong, H. (2019). Model selection for factor analysis: Some new criteria and performance comparisons, *Econometric Reviews* **38**(6): 577–596.

Choi, I., Kim, D., Kim, Y. J. and Kwark, N.-S. (2018). A multilevel factor model: Identification, asymptotic theory and applications, *Journal of Applied Econometrics* **33**(3): 355–377.

Choi, I., Lin, R. and Shin, Y. (2021). Canonical correlation-based model selection for the multilevel factors, *Journal of Econometrics* .

Coroneo, L., Jackson, L. E. and Owyang, M. T. (2020). International stock comovements with endogenous clusters, *Journal of Economic Dynamics and Control* **116**: 103904.

Del Negro, M. and Otrok, C. (2007). 99 luftballons: Monetary policy and the house price boom across us states, *Journal of Monetary Economics* **54**(7): 1962–1985.

Dias, F., Pinheiro, M. and Rua, A. (2013). Determining the number of global and country-specific factors in the euro area, *Studies in Nonlinear Dynamics and Econometrics* **17**(5): 573–617.

Freyaldenhoven, S. (2021). Factor models with local factors—determining the number of relevant factors, *Journal of Econometrics* .

Gonçalves, S. and Perron, B. (2014). Bootstrapping factor-augmented regression models, *Journal of Econometrics* **182**(1): 156–173.

Hallin, M. and Liška, R. (2011). Dynamic factors in the presence of blocks, *Journal of econometrics* **163**(1): 29–41.

Han, X. (2021). Shrinkage estimation of factor models with global and group-specific factors, *Journal of Business & Economic Statistics* **39**(1): 1–17.

Holly, S., Pesaran, M. H. and Yamagata, T. (2011). The spatial and temporal diffusion of house prices in the uk, *Journal of urban economics* **69**(1): 2–23.

Holmes, M. J. and Grimes, A. (2008). Is there long-run convergence among regional house prices in the uk?, *Urban studies* **45**(8): 1531–1544.

Howard, G. and Liebersohn, J. (2020). Regional divergence and house prices, *Fisher College of Business Working Paper* p. 004.

Kaplan, G., Mitman, K. and Violante, G. L. (2020). The housing boom and bust: Model meets evidence, *Journal of Political Economy* **128**(9): 3285–3345.

Kose, M. A., Otrok, C. and Whiteman, C. H. (2003). International business cycles: World, region, and country-specific factors, *The American Economic Review* **93**(4): 1216–1239.

Leamer, E. E. (2007). Housing is the business cycle.

Moench, E., Ng, S. and Potter, S. (2013). Dynamic hierarchical factor models, *Review of Economics and Statistics* **95**(5): 1811–1817.

Montagnoli, A. and Nagayasu, J. (2015). Uk house price convergence clubs and spillovers, *Journal of Housing Economics* **30**: 50–58.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues, *The Review of Economics and Statistics* **92**(4): 1004–1016.

Rodríguez-Caballero, C. V. and Caporin, M. (2019). A multilevel factor approach for the analysis of cds commonality and risk contribution, *Journal of International Financial Markets, Institutions and Money* **63**: 101144.

Shao, X. (2010). The dependent wild bootstrap, *Journal of the American Statistical Association* **105**(489): 218–235.

Stewart, G. W. and Sun, J.-g. (1990). *Matrix perturbation theory*, Academic Press, Boston.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors, *Journal of the American statistical association* **97**(460): 1167–1179.

Uematsu, Y. and Yamagata, T. (2022). Estimation of sparsity-induced weak factor models, *Journal of Business & Economic Statistics* pp. 1–15.

Wang, P. (2008). Large dimensional factor models with a multi-level factor structure: identification, estimation and inference, *Unpublished manuscript, New York University* .

Yang, X., Liu, W., Liu, W. and Tao, D. (2019). A survey on canonical correlation analysis, *IEEE Transactions on Knowledge and Data Engineering* **33**(6): 2349–2368.

Yu, Y., Wang, T. and Samworth, R. J. (2015). A useful variant of the davis–kahan theorem for statisticians, *Biometrika* **102**(2): 315–323.

Appendices

A Proofs

We use the following facts throughout the proofs. By Assumption B.1, we have: $\left\|T^{-1/2}\boldsymbol{G}\right\| = O_p(1)$ and $\left\|T^{-1/2}\boldsymbol{F}_i\right\| = O_p(1)$ for all $i = 1, ..., R$. By Assumptions C.1, we have: $\left\|N_i^{-1/2}\boldsymbol{\Gamma}_i\right\| = O_p(1)$ and $\left\|N_i^{-1/2}\boldsymbol{\Lambda}_i\right\| = O_p(1)$ for all $i = 1, ..., R$. The eigenvectors of a real $n \times n$ matrix $\boldsymbol{\Sigma}$ is scale invariant since $a\boldsymbol{\Sigma}\boldsymbol{v} = a\lambda\boldsymbol{v}$ where $\boldsymbol{v}$ is the eigenvector associated with the eigenvalue λ and a is a non-zero real number.

Proof of Proposition 1.
Using $\boldsymbol{K}_i = [\boldsymbol{G}, \boldsymbol{F}_i]$ for $i = 1, ..., R$, we can be express the matrix $\boldsymbol{\Phi}$ in (13) as

$$\boldsymbol{\Phi} = \left[\begin{array}{ccccccccccc} \boldsymbol{G} & \boldsymbol{F}_1 & -\boldsymbol{G} & -\boldsymbol{F}_2 & \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{G} & \boldsymbol{F}_1 & \boldsymbol{0} & \boldsymbol{0} & -\boldsymbol{G} & -\boldsymbol{F}_3 & \dots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ & & & & & & \vdots & & & & \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{G} & \boldsymbol{F}_{R-1} & -\boldsymbol{G} & -\boldsymbol{F}_R \end{array}\right]$$

Let

$$\underset{(r_0+r_i)\times r_0}{\boldsymbol{Q}_i^{r_0}} = \left[\begin{array}{c} \frac{1}{\sqrt{R}}\boldsymbol{A} \\ \boldsymbol{0} \end{array}\right] \text{ and } \underset{\sum_{l=1}^{R}(r_0+r_l)\times r_0}{\boldsymbol{Q}^{r_0}} = \left[\boldsymbol{Q}_1^{r_0\prime}, \boldsymbol{Q}_2^{r_0\prime}, \dots, \boldsymbol{Q}_R^{r_0\prime}\right]'$$

where $\left(1/\sqrt{R}\right)\boldsymbol{A}$ is any $r_0 \times r_0$ orthogonal matrix. For each i, it is easily see that

$$\boldsymbol{K}_i\boldsymbol{Q}_i^{r_0} = [\boldsymbol{G}, \boldsymbol{F}_i]\left[\begin{array}{c} \frac{1}{\sqrt{R}}\boldsymbol{A} \\ \boldsymbol{0} \end{array}\right] = \boldsymbol{G}\boldsymbol{B} \tag{22}$$

where $\boldsymbol{B} = \left(1/\sqrt{R}\right)\boldsymbol{A}$. This shows that $\boldsymbol{\Phi}\boldsymbol{Q}^{r_0} = \boldsymbol{0}$. Since $\boldsymbol{Q}^{r_0\prime}\boldsymbol{Q}^{r_0} = \boldsymbol{I}_{r_0}$, $\boldsymbol{Q}^{r_0}$ can serve as the right eigenvectors in the SVD of $\boldsymbol{\Phi}$. Consequently, we obtain

$$\boldsymbol{\Phi}\boldsymbol{Q}^{r_0} = \boldsymbol{P}^{r_0}\begin{bmatrix}\delta_1 & & & \\ & \delta_2 & & \\ & & \ddots & \\ & & & \delta_{r_0}\end{bmatrix} = \boldsymbol{0}$$

where $\boldsymbol{P}^{r_0}$ is the corresponding left eigenvectors. As $\boldsymbol{P}^{r_0}$ is non-zero, it follows that $\delta_1 = \cdots = \delta_R = 0$. This establishes that the first r_0 smallest singular values are zero.

We now show that the rest of the singular values are larger than zero by contradiction. Suppose that there exists an eigenvector $\boldsymbol{q}^\perp = \left[\boldsymbol{q}_1^{\perp\prime}, \ldots, \boldsymbol{q}_R^{\perp\prime}\right]'$, satisfying $\boldsymbol{\Phi}\boldsymbol{q}^\perp = \boldsymbol{0}$, $\boldsymbol{Q}^{r_0\prime}\boldsymbol{q}^\perp = \boldsymbol{0}$ and $\boldsymbol{q}^{\perp\prime}\boldsymbol{q}^\perp = 1$, where $\boldsymbol{q}_i^\perp = \left[\boldsymbol{q}_i^{G\perp\prime}, \boldsymbol{q}_i^{F\perp\prime}\right]'$. Noting $\boldsymbol{\Phi}\boldsymbol{q}^\perp = \boldsymbol{0}$, we have:

$$\boldsymbol{G}\boldsymbol{q}_m^{G\perp} + \boldsymbol{F}_m\boldsymbol{q}_m^{F\perp} = \boldsymbol{G}\boldsymbol{q}_h^{G\perp} + \boldsymbol{F}_h\boldsymbol{q}_h^{F\perp} \text{ for any } h \text{ and } m.$$

It follows that

$$R\left(\boldsymbol{G}\boldsymbol{q}_m^{G\perp} + \boldsymbol{F}_m\boldsymbol{q}_m^{F\perp}\right) = \sum_{i=1}^R\left(\boldsymbol{G}\boldsymbol{q}_i^{G\perp} + \boldsymbol{F}_i\boldsymbol{q}_i^{F\perp}\right) = \sum_{i=1}^R \boldsymbol{F}_i\boldsymbol{q}_i^{F\perp}.$$

where the second equality holds as a result of $\boldsymbol{Q}^{r_0\prime}\boldsymbol{q}^\perp = \boldsymbol{B}'\sum_{i=1}^R \boldsymbol{q}_i^{G\perp} = \boldsymbol{0}$. Consequently, we have

$$\boldsymbol{G}\left(\frac{1}{R}\boldsymbol{q}_m^{G\perp}\right) = \boldsymbol{F}_m\left(1 - \frac{1}{R}\right)\boldsymbol{q}_m^{F\perp} + \sum_{h\neq m}\boldsymbol{F}_h\boldsymbol{q}_h^{F\perp}.$$

By construction, we must have $\boldsymbol{q}_m^{G\perp} = \boldsymbol{q}_1^{F\perp} = \cdots = \boldsymbol{q}_R^{F\perp} = \boldsymbol{0}$ for all m. Hence, $\boldsymbol{q}^\perp = \boldsymbol{0}$. This contradicts the definition of an eigenvector. Since the singular values are non-negative, the remaining singular values of $\boldsymbol{\Phi}$ are larger than zero. By Assumption B.1, we have $T^{-1/2}\boldsymbol{K}_i = O_p(1)$ for all i such that $\boldsymbol{\Phi} = O_p\left(\sqrt{T}\right)$. Using $\boldsymbol{\Phi}\boldsymbol{q} = \delta\boldsymbol{p}$ and the fact that the eigenvectors $\boldsymbol{p}$ and $\boldsymbol{q}$ are bounded, we have: $\delta_{r_0+j} = O_p\left(\sqrt{T}\right)$ for $j = 1, ..., Rr_{\max} - r_0$. *Q.E.D*

Proof of Proposition 2

Using (22) we obtain:

$$\frac{1}{\sqrt{T}}\boldsymbol{\Psi} = \frac{1}{\sqrt{T}}\left[\boldsymbol{K}_1\boldsymbol{Q}_1^{r_0}, \ldots, \boldsymbol{K}_R\boldsymbol{Q}_R^{r_0}\right] = \frac{1}{\sqrt{T}}\left[\boldsymbol{G}\boldsymbol{B}, \ldots, \boldsymbol{G}\boldsymbol{B}\right] \tag{23}$$

which yields

$$\frac{\boldsymbol{\Psi}\boldsymbol{\Psi}'}{T} = \frac{\boldsymbol{G}\boldsymbol{G}'}{T} = \boldsymbol{L}\boldsymbol{\Xi}\boldsymbol{L}'$$

where $\boldsymbol{\Xi}$ is a diagonal matrix with the first r_0 elements non-zero and the remaining elements zero. Finally, it follows that

$$\boldsymbol{L}^{r_0} = \frac{1}{\sqrt{T}}\boldsymbol{G}\left(\frac{\boldsymbol{G}'\boldsymbol{L}^{r_0}(\boldsymbol{\Xi}^{r_0})^{-1}}{\sqrt{T}}\right)$$

where $\boldsymbol{\Xi}^{r_0}$ is the diagonal matrix consisting of r_0 non-zero diagonal elements of $\boldsymbol{\Xi}$. The full rank matrix inside the bracket is a rotation matrix. *Q.E.D*

Proof of Lemma 1
Since Assumptions A–D in Bai and Ng (2002) are satisfied, the stated result follows from Theorem 1 of Bai and Ng (2002).

Q.E.D

Proof of Lemma 2
Let $\underset{r_{\max}\times r_0}{\bar{\boldsymbol{Q}}_i^{r_0}} = \widehat{\boldsymbol{H}}_i^{-}\boldsymbol{Q}_i^{r_0}$ where $\widehat{\boldsymbol{H}}_i^{-}$ is the Moore-Penrose inverse of $\widehat{\boldsymbol{H}}_i$. Since $r_0+r_i \leq r_{\max}$ for all i, by the property of the Moore-Penrose inverse, it follows that $\widehat{\boldsymbol{H}}_i\widehat{\boldsymbol{H}}_i^{-} = \boldsymbol{I}_{r_0+r_i}$. Let $\underset{Rr_{\max}\times r_0}{\bar{\boldsymbol{Q}}^{r_0}} = \left[\bar{\boldsymbol{Q}}_1^{r_0\prime},\ldots,\bar{\boldsymbol{Q}}_R^{r_0\prime}\right]'$. Then, we obtain

$$\boldsymbol{\Phi}\widehat{\boldsymbol{H}}\bar{\boldsymbol{Q}}^{r_0} = \boldsymbol{\Phi}\boldsymbol{Q}^{r_0} = \boldsymbol{P}^{r_0}\boldsymbol{\Delta}^{r_0}$$

Along the same arguments in Proof of Proposition 1, we obtain the desired result.

Q.E.D

Proof of Lemma 3
See the proof of Theorem 2 in Yu et al. (2015).

Q.E.D

Lemma 4. *Under Assumption A–C, as $N_1, N_2, \ldots, N_R, T \longrightarrow \infty$, we have:*

1. *For every m and h,*
$$\frac{1}{T\sqrt{N_h}}\left\|\left(\widehat{\boldsymbol{K}}_m - \boldsymbol{K}_m\widehat{\boldsymbol{H}}_m\right)'\boldsymbol{e}_h\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right)$$
2. *For each i,*
$$\frac{1}{T\sqrt{N_i}}\left\|\widehat{\boldsymbol{G}}'\boldsymbol{e}_i\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right)$$

where $C_{\underline{N},T} = \min\{\sqrt{\underline{N}}, \sqrt{T}\}$ with $\underline{N} = \min\{N_1, N_2, \ldots, N_R\}$

Proof
1. Using the Cauchy-Schwarz inequality, we obtain:

$$\frac{1}{T\sqrt{N_h}}\left\|\left(\widehat{\boldsymbol{K}}_m - \boldsymbol{K}_m\widehat{\boldsymbol{H}}_m\right)'\boldsymbol{e}_h\right\| \leq \left\|\frac{1}{\sqrt{T}}\left(\widehat{\boldsymbol{K}}_m - \boldsymbol{K}_m\widehat{\boldsymbol{H}}_m\right)\right\|\left\|\frac{1}{\sqrt{N_hT}}\boldsymbol{e}_h\right\|$$

The first term is of stochastic order $O_p\left(C_{N_mT}^{-1}\right)$ by Lemma 1. For the second term, we have:

$$\left\|\frac{1}{\sqrt{N_hT}}\boldsymbol{e}_h\right\| = \sqrt{\frac{1}{N_hT}tr\left\{\boldsymbol{e}_h'\boldsymbol{e}_h\right\}} = \sqrt{\frac{1}{N_hT}\sum_{j=1}^{N_h}\sum_{t=1}^{T}e_{hjt}^2}$$

Since $E(e_{hjt}^2) = O(1)$, the above term is $O_p(1)$. Combining the two terms, we obtain the required result.
2. Using equation (19) and $\widehat{\boldsymbol{K}}_m = \widehat{\boldsymbol{K}}_m - \boldsymbol{K}_m\widehat{\boldsymbol{H}}_m + \boldsymbol{K}_m\widehat{\boldsymbol{H}}_m$, we have:

$$\begin{aligned}\frac{1}{T\sqrt{N_i}}\left\|\widehat{\boldsymbol{G}}'\boldsymbol{e}_i\right\| = \frac{1}{T\sqrt{N_iT}}\Bigg\|\sum_{m=1}^{R}\Big\{&\widehat{\boldsymbol{J}}^{r_0\prime}\left(\widehat{\boldsymbol{K}}_m - \boldsymbol{K}_m\widehat{\boldsymbol{H}}_m\right)\widetilde{\boldsymbol{Q}}_m^{r_0}\left(\widehat{\boldsymbol{K}}_m - \boldsymbol{K}_m\widehat{\boldsymbol{H}}_m\right)'\boldsymbol{e}_i \\ &+ \widehat{\boldsymbol{J}}^{r_0\prime}\left(\widehat{\boldsymbol{K}}_m - \boldsymbol{K}_m\widehat{\boldsymbol{H}}_m\right)\widetilde{\boldsymbol{Q}}_m^{r_0}\widehat{\boldsymbol{H}}_m'\boldsymbol{K}_m'\boldsymbol{e}_i + \widehat{\boldsymbol{J}}^{r_0\prime}\boldsymbol{K}_m\widehat{\boldsymbol{H}}_m\widetilde{\boldsymbol{Q}}_m^{r_0}\left(\widehat{\boldsymbol{K}}_m - \boldsymbol{K}_m\widehat{\boldsymbol{H}}_m\right)'\boldsymbol{e}_i \\ &+ \widehat{\boldsymbol{J}}^{r_0\prime}\boldsymbol{K}_m\widehat{\boldsymbol{H}}_m\widetilde{\boldsymbol{Q}}_m^{r_0}\widehat{\boldsymbol{H}}_m'\boldsymbol{K}_m'\boldsymbol{e}_i\Big\}\Bigg\|\end{aligned}$$

where $\widetilde{\boldsymbol{Q}}_i^{r_0} = \widehat{\boldsymbol{Q}}_i^{r_0}\widehat{\boldsymbol{Q}}_i^{r_0\prime}$. We note that $\left\|\widehat{\boldsymbol{J}}^{r_0}\right\| = O_p(1)$ since $\widehat{\boldsymbol{L}}^{r_0\prime}\widehat{\boldsymbol{L}}^{r_0} = \boldsymbol{I}_{r_0}$ and $T^{-1/2}\widehat{\boldsymbol{\Psi}} = O_p(1)$. The first term of RHS is bounded by $O_p\left(C_{\underline{N}T}^{-1}\right) \times O_p\left(C_{\underline{N}T}^{-1}\right)$ by Lemma 1.1 and Lemma 4.1. The second term is bounded by $O_p\left(T^{-1/2}C_{\underline{N}T}^{-1}\right)$ by Lemma 1.1 and the fact that $(N_mT)^{-1/2}\|\boldsymbol{K}_m'\boldsymbol{e}_i\| = O_p(1)$ under Assumption B2. The third term is bounded by $O_p\left(C_{\underline{N}T}^{-1}\right)$ by Lemma 4.1. The last term is bounded by $O_p(T^{-1/2})$ since $(N_mT)^{-1/2}\|\boldsymbol{K}_m'\boldsymbol{e}_i\| = O_p(1)$ under Assumption B.2. The proof completes by combining all these results.

Q.E.D

Proof of Theorem 1
By Lemma 1, we have:

$$\frac{1}{T}\left\|\widehat{\boldsymbol{\Phi}}'\widehat{\boldsymbol{\Phi}} - \widehat{\boldsymbol{H}}'\boldsymbol{\Phi}'\boldsymbol{\Phi}\widehat{\boldsymbol{H}}\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right)$$

Furthermore, by Lemma 2 and Lemma 3, we obtain:

$$\left\|\widehat{\boldsymbol{Q}}^{r_0} - \bar{\boldsymbol{Q}}^{r_0}\boldsymbol{D}\right\| \leq O_p(1) \times \frac{1}{T}\left\|\widehat{\boldsymbol{\Phi}}'\widehat{\boldsymbol{\Phi}} - \widehat{\boldsymbol{H}}'\boldsymbol{\Phi}'\boldsymbol{\Phi}\widehat{\boldsymbol{H}}\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right)$$

where $\boldsymbol{D}$ is an $r_0 \times r_0$ orthogonal matrix. Then, using the definition $\bar{\boldsymbol{Q}}^{r_0} = \widehat{\boldsymbol{H}}_i^{-}\boldsymbol{Q}^{r_0}$ and (22), it follows for each i that

$$\begin{aligned}
\frac{1}{\sqrt{T}}\left\|\widehat{\boldsymbol{K}}_i\widehat{\boldsymbol{Q}}_i^{r_0} - \boldsymbol{K}_i\widehat{\boldsymbol{H}}_i\bar{\boldsymbol{Q}}_i^{r_0}\boldsymbol{D}\right\| &= \frac{1}{\sqrt{T}}\left\|\widehat{\boldsymbol{K}}_i\widehat{\boldsymbol{Q}}_i^{r_0} - \boldsymbol{GBD}\right\| \\
&\leq \frac{1}{\sqrt{T}}\left\|\widehat{\boldsymbol{K}}_i\widehat{\boldsymbol{Q}}_i^{r_0} - \boldsymbol{K}_i\widehat{\boldsymbol{H}}_i\widehat{\boldsymbol{Q}}_i^{r_0} + \boldsymbol{K}_i\widehat{\boldsymbol{H}}_i\widehat{\boldsymbol{Q}}_i^{r_0} - \boldsymbol{K}_i\widehat{\boldsymbol{H}}_i\bar{\boldsymbol{Q}}_i^{r_0}\boldsymbol{D}\right\| \\
&\leq \frac{1}{\sqrt{T}}\left\|\widehat{\boldsymbol{K}}_i - \boldsymbol{K}_i\widehat{\boldsymbol{H}}_i\right\|\left\|\widehat{\boldsymbol{Q}}_i^{r_0}\right\| + \frac{1}{\sqrt{T}}\left\|\boldsymbol{K}_i\widehat{\boldsymbol{H}}_i\right\|\left\|\widehat{\boldsymbol{Q}}_i^{r_0} - \bar{\boldsymbol{Q}}_i^{r_0}\boldsymbol{D}\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right)
\end{aligned}$$

where the inequalities hold due to the Cauchy-Schwarz inequality, and the last equality follows from Lemma 1 and the fact that $\left\|\widehat{\boldsymbol{Q}}_i^{r_0}\right\| = O_p(1)$ and $\left\|\widehat{\boldsymbol{H}}_i\right\| = O_p(1)$. Using this convergence rate, we obtain:

$$\begin{aligned}
\left\|\frac{\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Psi}}'}{T} - \frac{\boldsymbol{\Psi}\boldsymbol{\Psi}'}{T}\right\| &= \left\|\frac{1}{T}\sum_{i=1}^{R}\widehat{\boldsymbol{K}}_i\widehat{\boldsymbol{Q}}_i^{r_0}\widehat{\boldsymbol{Q}}_i^{r_0\prime}\widehat{\boldsymbol{K}}_i' - \frac{R}{T}\boldsymbol{GBDD}'\boldsymbol{B}'\boldsymbol{G}'\right\| \\
&\leq \sum_{i=1}^{R}\left\|\frac{1}{T}\widehat{\boldsymbol{K}}_i\widehat{\boldsymbol{Q}}_i^{r_0}\widehat{\boldsymbol{Q}}_i^{r_0\prime}\widehat{\boldsymbol{K}}_i' - \frac{1}{T}\boldsymbol{GG}'\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right)
\end{aligned}$$

where the inequality follows from the Cauchy-Schwarz inequality. Applying Lemma 3 to the above equation, we obtain

$$\left\|\widehat{\boldsymbol{L}}^{r_0} - \boldsymbol{L}^{r_0}\boldsymbol{U}\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right) \tag{24}$$

where $\boldsymbol{U}$ is an $r_0 \times r_0$ orthogonal matrix[15]. Finally, by definition of $\widehat{\boldsymbol{G}}$ and Proposition 2, we conclude that

$$\frac{1}{\sqrt{T}}\left\|\widehat{\boldsymbol{G}} - \boldsymbol{G}\mathbb{H}\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right) \tag{25}$$

[15] If the r_0 largest eigenvalues of $\boldsymbol{GG}'/T$ are distinct, each column of $\widehat{\boldsymbol{L}}^{r_0}$ converges to its population counterpart in $\boldsymbol{L}^{r_0}$ up to sign. In such a case, $\boldsymbol{U}$ is an $r_0 \times r_0$ diagonal matrix whose diagonal elements are either 1 or -1.

where $\mathbb{H} = T^{-1/2}\boldsymbol{G}'\boldsymbol{L}^{r_0}\boldsymbol{\Xi}^{r_0,-1}\boldsymbol{U}$ is a rotation matrix.

For the global factor loadings in block i, we have:

$$\widehat{\boldsymbol{\Gamma}}_i' = \frac{1}{T}\widehat{\boldsymbol{G}}'\boldsymbol{Y}_i = \frac{1}{T}\widehat{\boldsymbol{G}}'\left(\boldsymbol{G}\boldsymbol{\Gamma}_i' + \boldsymbol{F}_i\boldsymbol{\Lambda}_i' + \boldsymbol{e}_i\right) = \frac{1}{T}\widehat{\boldsymbol{G}}'\left[\left(\boldsymbol{G} - \widehat{\boldsymbol{G}}\mathbb{H}^{-1} + \widehat{\boldsymbol{G}}\mathbb{H}^{-1}\right)\boldsymbol{\Gamma}_i' + \boldsymbol{F}_i\boldsymbol{\Lambda}_i' + \boldsymbol{e}_i\right]$$

Multiplying both sides of the above equation by $1/\sqrt{N_i}$ and rearranging the results, we have:

$$\frac{1}{\sqrt{N_i}}\left(\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right) = \frac{1}{T\sqrt{N_i}}\widehat{\boldsymbol{G}}'\left(\boldsymbol{G} - \widehat{\boldsymbol{G}}\mathbb{H}^{-1}\right)\boldsymbol{\Gamma}_i' + \frac{1}{T\sqrt{N_i}}\widehat{\boldsymbol{G}}'\boldsymbol{F}_i\boldsymbol{\Lambda}_i' + \frac{1}{T\sqrt{N_i}}\widehat{\boldsymbol{G}}'\boldsymbol{e}_i \tag{26}$$

The first term of RHS is bounded by $O_p(C_{\underline{N}T}^{-1})$ due to (25). The second term is bounded as

$$\begin{aligned}
\left\|\frac{1}{T\sqrt{N_i}}\widehat{\boldsymbol{G}}'\boldsymbol{F}_i\boldsymbol{\Lambda}_i'\right\| &= \left\|\frac{1}{T\sqrt{N_i}}\left(\widehat{\boldsymbol{G}}' - \boldsymbol{G}\mathbb{H} + \boldsymbol{G}\mathbb{H}\right)'\boldsymbol{F}_i\boldsymbol{\Lambda}_i'\right\| \\
&\leq \left\|\frac{1}{T\sqrt{N_i}}\left(\widehat{\boldsymbol{G}}' - \boldsymbol{G}\mathbb{H}\right)'\boldsymbol{F}_i\boldsymbol{\Lambda}_i'\right\| + \left\|\frac{1}{T\sqrt{N_i}}\mathbb{H}'\boldsymbol{G}'\boldsymbol{F}_i\boldsymbol{\Lambda}_i'\right\| \\
&= O_p\left(\frac{1}{C_{\underline{N}T}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) = O_p\left(\frac{1}{C_{\underline{N}T}}\right)
\end{aligned} \tag{27}$$

where the inequality follows from the Cauchy-Schwarz inequality and the second to last equalities use Lemma 1 and Assumptioin D. The last term of (27) is bounded by $O_p\left(C_{\underline{N}T}^{-1}\right)$ due to Lemma 4.2. Then,

$$\frac{1}{\sqrt{N_i}}\left\|\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right)$$

Q.E.D

Lemma 5. *Under Assumptions A–C, as $N_1, N_2, \ldots, N_R, T \longrightarrow \infty$, we have for each $i = 1, \ldots, R$:*

1.
$$\left\|\frac{1}{\sqrt{N_i T}}\boldsymbol{\Gamma}_i'\boldsymbol{e}_i'\right\| = O_p(1)$$

2.
$$\left\|\frac{1}{\sqrt{N_i T}}\boldsymbol{\Lambda}_i'\boldsymbol{e}_i'\right\| = O_p(1)$$

3.
$$\left\|\frac{1}{N_i\sqrt{T}}\left(\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right)\boldsymbol{e}_i'\right\| = O_p\left(\frac{1}{C_{\underline{N}T}\sqrt{N_i}}\right) + O_p\left(\frac{1}{\sqrt{N_i}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

Proof

1.

$$\left\|\frac{1}{\sqrt{N_i T}}\boldsymbol{\Gamma}_i'\boldsymbol{e}_i'\right\| = \frac{1}{\sqrt{N_i T}}\left(tr\left\{\sum_{j=1}^{N_i}\boldsymbol{e}_{ij}\boldsymbol{\gamma}_{ij}'\sum_{k=1}^{N_i}\boldsymbol{\gamma}_{ik}\boldsymbol{e}_{ik}'\right\}\right)^{\frac{1}{2}} = \left(\frac{1}{N_i T}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\boldsymbol{\gamma}_{ij}'\boldsymbol{\gamma}_{ik}\sum_{t=1}^{T}e_{ikt}e_{ijt}\right)^{\frac{1}{2}}$$

Taking expectations of the term inside the bracket, by Assumption A.3 and C.1, we have:

$$E\left(\frac{1}{N_i T}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\boldsymbol{\gamma}_{ij}'\boldsymbol{\gamma}_{ik}\sum_{t=1}^{T}e_{ikt}e_{ijt}\right) \leq \frac{1}{N_i T}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\boldsymbol{\gamma}_{ij}'\boldsymbol{\gamma}_{ik}\sum_{t=1}^{T}\tau_{i,(jk)} = O(1)$$

2. The proof is similar to part 1 and therefore omitted.
3. From (26) we have:

$$\frac{1}{N_i\sqrt{T}}\left(\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right)\boldsymbol{e}_i' = \frac{1}{N_iT\sqrt{T}}\widehat{\boldsymbol{G}}'\left(\boldsymbol{G} - \widehat{\boldsymbol{G}}\mathbb{H}^{-1}\right)\boldsymbol{\Gamma}_i'\boldsymbol{e}_i' + \frac{1}{N_iT\sqrt{T}}\widehat{\boldsymbol{G}}'\boldsymbol{F}_i\boldsymbol{\Lambda}_i'\boldsymbol{e}_i' + \frac{1}{N_iT\sqrt{T}}\widehat{\boldsymbol{G}}'\boldsymbol{e}_i\boldsymbol{e}_i'$$

The first term is bounded by $O_p\left(C_{\underline{N}T}^{-1}N_i^{-1/2}\right)$ by Theorem 1 and Lemma 5.1. The second term is bounded by $O_p\left(C_{\underline{N}T}^{-1}N_i^{-1/2}\right)$ due to (27) and Lemma 5.2. Using (19), the third term can be written as

$$\frac{1}{N_iT\sqrt{T}}\widehat{\boldsymbol{G}}'\boldsymbol{e}_i\boldsymbol{e}_i' = \frac{1}{N_iT}\widehat{\boldsymbol{L}}^{r_0\prime}\boldsymbol{e}_i\boldsymbol{e}_i' = \frac{1}{N_iT}\widehat{\boldsymbol{J}}^{r_0\prime}\frac{1}{T}\left(\sum_{m=1}^{R}\widehat{\boldsymbol{K}}_m\widehat{\boldsymbol{Q}}_m^{r_0}\widehat{\boldsymbol{Q}}_m^{r_0\prime}\widehat{\boldsymbol{K}}_m'\right)\boldsymbol{e}_i\boldsymbol{e}_i'$$

Following the proof of Theorem 1 in Bai and Ng (2002), we have for each m:

$$\frac{1}{N_iT\sqrt{T}}\left\|\widehat{\boldsymbol{K}}_m'\boldsymbol{e}_i\boldsymbol{e}_i'\right\| = O_p\left(\frac{1}{\sqrt{N_i}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

Therefore, it follows that

$$\frac{1}{N_iT\sqrt{T}}\left\|\widehat{\boldsymbol{G}}'\boldsymbol{e}_i\boldsymbol{e}_i'\right\| = O_p\left(\frac{1}{\sqrt{N_i}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

The proof completes by combining the above results.

Q.E.D

Proof of Theorem 2

By construction, we have the following relation for each i:

$$\widehat{\boldsymbol{F}}_i\widehat{\boldsymbol{\Upsilon}}_i = \frac{1}{N_iT}\left(\boldsymbol{Y}_i - \widehat{\boldsymbol{G}}\widehat{\boldsymbol{\Gamma}}_i'\right)\left(\boldsymbol{Y}_i - \widehat{\boldsymbol{G}}\widehat{\boldsymbol{\Gamma}}_i'\right)'\widehat{\boldsymbol{F}}_i$$

Replacing $\boldsymbol{Y}_i$ with $\boldsymbol{Y}_i = \boldsymbol{G}\boldsymbol{\Gamma}_i' + \boldsymbol{F}_i\boldsymbol{\Lambda}_i' + \boldsymbol{e}_i$, we obtain:

$$\widehat{\boldsymbol{F}}_i\widehat{\boldsymbol{\Upsilon}}_i = \frac{1}{N_iT}\left(\widehat{\boldsymbol{S}}_i + \boldsymbol{F}_i\boldsymbol{\Lambda}_i' + \boldsymbol{e}_i\right)\left(\widehat{\boldsymbol{S}}_i + \boldsymbol{F}_i\boldsymbol{\Lambda}_i' + \boldsymbol{e}_i\right)'\widehat{\boldsymbol{F}}_i$$

where $\widehat{\boldsymbol{S}}_i = \boldsymbol{G}\boldsymbol{\Gamma}_i' - \widehat{\boldsymbol{G}}\widehat{\boldsymbol{\Gamma}}_i'$. Multiplying both sides by $\left(\boldsymbol{F}_i'\widehat{\boldsymbol{F}}_i/T\right)^{-1}(\boldsymbol{\Gamma}_i'\boldsymbol{\Gamma}_i/N_i)^{-1}$ and rearranging terms:

$$\begin{aligned}\frac{1}{\sqrt{T}}\left(\widehat{\boldsymbol{F}}_i\widehat{\mathscr{H}}_i^{-1} - \boldsymbol{F}_i\right) &= \frac{1}{\sqrt{T}}\frac{1}{N_iT}\left(\boldsymbol{F}_i\boldsymbol{\Lambda}_i'\boldsymbol{e}_i + \boldsymbol{e}_i\boldsymbol{\Lambda}_i\boldsymbol{F}_i' + \boldsymbol{e}_i\boldsymbol{e}_i'\right)\widehat{\boldsymbol{F}}_i\left(\frac{\boldsymbol{F}_i'\widehat{\boldsymbol{F}}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Lambda}_i'\boldsymbol{\Lambda}_i}{N_i}\right)^{-1}\\ &+ \frac{1}{\sqrt{T}}\frac{1}{N_iT}\left(\widehat{\boldsymbol{S}}_i\widehat{\boldsymbol{S}}_i' + \widehat{\boldsymbol{S}}_i\boldsymbol{\Lambda}_i\boldsymbol{F}_i' + \widehat{\boldsymbol{S}}_i\boldsymbol{e}_i' + \boldsymbol{F}_i'\boldsymbol{\Lambda}'\widehat{\boldsymbol{S}}_i' + \boldsymbol{e}_i\widehat{\boldsymbol{S}}_i'\right)\widehat{\boldsymbol{F}}_i\left(\frac{\boldsymbol{F}_i'\widehat{\boldsymbol{F}}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Lambda}_i'\boldsymbol{\Lambda}_i}{N_i}\right)^{-1}\end{aligned}$$

The stochastic bound of the first term is $O_p(C_{\underline{N}T}^{-1})$ by Theorem 1 of Bai and Ng (2002) and the fact that $\left(\boldsymbol{F}_i'\widehat{\boldsymbol{F}}_i/T\right)$ and $(\boldsymbol{\Gamma}_i'\boldsymbol{\Gamma}_i/N_i)$ are bounded and invertible (see Proposition 1 of Bai (2003)).

Next, we study the terms in the the second line of the above equation. Using the relation that

$$\widehat{\boldsymbol{S}}_i = \boldsymbol{G}\boldsymbol{\Gamma}_i' - \widehat{\boldsymbol{G}}\widehat{\boldsymbol{\Gamma}}_i' = \boldsymbol{G}\boldsymbol{\Gamma}_i' - \left(\widehat{\boldsymbol{G}} - \boldsymbol{G}\mathbb{H} + \boldsymbol{G}\mathbb{H}\right)\left(\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i' + \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right)$$
$$= -\left(\widehat{\boldsymbol{G}} - \boldsymbol{G}\mathbb{H}\right)\left(\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right) - \left(\widehat{\boldsymbol{G}} - \boldsymbol{G}\mathbb{H}\right)\mathbb{H}^{-1}\boldsymbol{\Gamma}_i' - \boldsymbol{G}\mathbb{H}\left(\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right), \quad (28)$$

we obtain:

$$\frac{1}{\sqrt{T}}\frac{1}{N_iT}\widehat{\boldsymbol{S}}_i\widehat{\boldsymbol{S}}_i'\widehat{\boldsymbol{F}}_i = -\frac{1}{\sqrt{T}}\frac{1}{N_iT}\left(\widehat{\boldsymbol{G}} - \boldsymbol{G}\mathbb{H}\right)\left(\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right)\widehat{\boldsymbol{F}}_i$$
$$- \frac{1}{\sqrt{T}}\frac{1}{N_iT}\left(\widehat{\boldsymbol{G}} - \boldsymbol{G}\mathbb{H}\right)\mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\widehat{\boldsymbol{F}}_i - \frac{1}{\sqrt{T}}\frac{1}{N_iT}\boldsymbol{G}\mathbb{H}\left(\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right)\widehat{\boldsymbol{F}}_i$$

By Theorem 1, it follows that

$$\left\|\frac{1}{\sqrt{T}}\frac{1}{N_iT}\widehat{\boldsymbol{S}}_i\widehat{\boldsymbol{S}}_i'\widehat{\boldsymbol{F}}_i\right\| = O_p\left(\frac{1}{C_{\underline{N}T}^2\sqrt{N_iT}}\right) + O_p\left(\frac{1}{C_{\underline{N}T}\sqrt{N_iT}}\right) + O_p\left(\frac{1}{C_{\underline{N}T}\sqrt{N_iT}}\right) = O_p\left(\frac{1}{C_{\underline{N}T}\sqrt{N_iT}}\right)$$

Using (28), it follows that

$$\frac{1}{\sqrt{T}}\frac{1}{N_iT}\widehat{\boldsymbol{S}}_i\boldsymbol{\Lambda}_i\boldsymbol{F}_i'\widehat{\boldsymbol{F}}_i = -\frac{1}{\sqrt{T}}\frac{1}{N_iT}\left(\widehat{\boldsymbol{G}} - \boldsymbol{G}\mathbb{H}\right)\left(\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right)\boldsymbol{\Lambda}_i\boldsymbol{F}_i'\widehat{\boldsymbol{F}}_i$$
$$- \frac{1}{\sqrt{T}}\frac{1}{N_iT}\left(\widehat{\boldsymbol{G}} - \boldsymbol{G}\mathbb{H}\right)\mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\boldsymbol{\Lambda}_i\boldsymbol{F}_i'\widehat{\boldsymbol{F}}_i - \frac{1}{\sqrt{T}}\frac{1}{N_iT}\boldsymbol{G}\mathbb{H}\left(\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right)\boldsymbol{\Lambda}_i\boldsymbol{F}_i'\widehat{\boldsymbol{F}}_i$$

Therefore, by Theorem 1,

$$\left\|\frac{1}{\sqrt{T}}\frac{1}{N_iT}\widehat{\boldsymbol{S}}_i\boldsymbol{\Lambda}_i\boldsymbol{F}_i'\widehat{\boldsymbol{F}}_i\right\| = O_p\left(\frac{1}{C_{\underline{N}T}^2}\right) + O_p\left(\frac{1}{C_{\underline{N}T}}\right) + O_p\left(\frac{1}{C_{\underline{N}T}}\right) = O_p\left(\frac{1}{C_{\underline{N}T}}\right)$$

From (28) we obtain:

$$\frac{1}{\sqrt{T}}\frac{1}{N_iT}\widehat{\boldsymbol{S}}_i\boldsymbol{e}_i'\widehat{\boldsymbol{F}}_i = -\frac{1}{\sqrt{T}}\frac{1}{N_iT}\left(\widehat{\boldsymbol{G}} - \boldsymbol{G}\mathbb{H}\right)\left(\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right)\boldsymbol{e}_i'\widehat{\boldsymbol{F}}_i$$
$$- \frac{1}{\sqrt{T}}\frac{1}{N_iT}\left(\widehat{\boldsymbol{G}} - \boldsymbol{G}\mathbb{H}\right)\mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\boldsymbol{e}_i'\widehat{\boldsymbol{F}}_i - \frac{1}{\sqrt{T}}\frac{1}{N_iT}\boldsymbol{G}\mathbb{H}\left(\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right)\boldsymbol{e}_i'\widehat{\boldsymbol{F}}_i$$

The first term is bounded by $O_p\left(C_{\underline{N}T}^{-1}\right)\left[O_p\left(N_i^{-1/2}\right) + O_p\left(T^{-1/2}\right)\right]$ due to Theorem 1 and Lemma 5.3. The second term is bounded by $N_i^{-1/2}O_p\left(C_{\underline{N}T}^{-1}\right)$ due to Theorem 1 and Lemma 5.1. The last term is bounded by $O_p\left(N_i^{-1/2}\right) + O_p\left(T^{-1/2}\right)$. Consequently, we have:

$$\frac{1}{\sqrt{T}}\frac{1}{N_iT}\left\|\widehat{\boldsymbol{S}}_i\boldsymbol{e}_i'\widehat{\boldsymbol{F}}_i\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right)$$

It is straightforward to show that $\frac{1}{\sqrt{T}}\frac{1}{N_iT}\left\|\boldsymbol{e}_i\widehat{\boldsymbol{S}}_i'\widehat{\boldsymbol{F}}_i\right\|$ has the same stochastic order. Using (28):

$$\frac{1}{\sqrt{T}}\frac{1}{N_iT}\boldsymbol{F}_i'\boldsymbol{\Lambda}_i'\widehat{\boldsymbol{S}}_i'\widehat{\boldsymbol{F}}_i = -\frac{1}{\sqrt{T}}\frac{1}{N_iT}\boldsymbol{F}_i'\boldsymbol{\Lambda}_i'\left(\widehat{\boldsymbol{\Gamma}}_i'-\mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right)'\left(\widehat{\boldsymbol{G}}-\boldsymbol{G}\mathbb{H}\right)'\widehat{\boldsymbol{F}}_i$$
$$-\frac{1}{\sqrt{T}}\frac{1}{N_iT}\boldsymbol{F}_i'\boldsymbol{\Lambda}_i'\boldsymbol{\Gamma}_i\left(\mathbb{H}^{-1}\right)'\left(\widehat{\boldsymbol{G}}-\boldsymbol{G}\mathbb{H}\right)'\widehat{\boldsymbol{F}}_i-\frac{1}{\sqrt{T}}\frac{1}{N_iT}\boldsymbol{F}_i'\boldsymbol{\Lambda}_i'\left(\widehat{\boldsymbol{\Gamma}}_i'-\mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right)'\mathbb{H}'\boldsymbol{G}'\widehat{\boldsymbol{F}}_i$$

Using Theorem 1, we obtain:

$$\frac{1}{\sqrt{T}}\frac{1}{N_iT}\left\|\boldsymbol{F}_i'\boldsymbol{\Lambda}_i'\widehat{\boldsymbol{S}}_i'\widehat{\boldsymbol{F}}_i\right\| = O_p\left(\frac{1}{C_{\underline{N}T}^2}\right)+O_p\left(\frac{1}{C_{\underline{N}T}}\right)+O_p\left(\frac{1}{C_{\underline{N}T}}\right)=O_p\left(\frac{1}{C_{\underline{N}T}}\right)$$

Combining all the results, we conclude that

$$\frac{1}{\sqrt{T}}\left\|\widehat{\boldsymbol{F}}_i-\boldsymbol{F}_i\widehat{\mathscr{H}_i}\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right). \tag{29}$$

Next, for each i, the estimated factor loadings are:

$$\widehat{\boldsymbol{\Lambda}}_i' = \frac{1}{T}\widehat{\boldsymbol{F}}_i'\left(\boldsymbol{Y}_i-\widehat{\boldsymbol{G}}\widehat{\boldsymbol{\Gamma}}'\right)$$

Plugging $\boldsymbol{Y}_i=\boldsymbol{G}\boldsymbol{\Gamma}_i'+\boldsymbol{F}_i\boldsymbol{\Lambda}_i'+\boldsymbol{e}_i$, $\boldsymbol{F}_i=\boldsymbol{F}_i-\widehat{\boldsymbol{F}}_i\widehat{\mathscr{H}_i}^{-1}+\widehat{\boldsymbol{F}}_i\widehat{\mathscr{H}_i}^{-1}$ and (28) into the above equation, we obtain:

$$\frac{1}{\sqrt{N_i}}\left(\widehat{\boldsymbol{\Lambda}}_i'-\widehat{\mathscr{H}_i}^{-1}\boldsymbol{\Lambda}_i'\right) = -\frac{1}{T\sqrt{N_i}}\widehat{\boldsymbol{F}}_i'\left(\widehat{\boldsymbol{G}}-\boldsymbol{G}\mathbb{H}\right)\left(\widehat{\boldsymbol{\Gamma}}_i'-\mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right)-\frac{1}{T\sqrt{N_i}}\widehat{\boldsymbol{F}}_i'\left(\widehat{\boldsymbol{G}}-\boldsymbol{G}\mathbb{H}\right)\mathbb{H}^{-1}\boldsymbol{\Gamma}_i'$$
$$-\frac{1}{T\sqrt{N_i}}\widehat{\boldsymbol{F}}_i'\boldsymbol{G}\mathbb{H}\left(\widehat{\boldsymbol{\Gamma}}_i'-\mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right)+\frac{1}{T\sqrt{N_i}}\widehat{\boldsymbol{F}}_i'\left(\boldsymbol{F}_i-\widehat{\boldsymbol{F}}_i\widehat{\mathscr{H}_i}^{-1}\right)\boldsymbol{\Lambda}_i'+\frac{1}{T\sqrt{N_i}}\widehat{\boldsymbol{F}}_i'\boldsymbol{e}_i$$

The first three terms are bounded by $O_p\left(C_{\underline{N}T}^{-2}\right)$, $O_p\left(C_{\underline{N}T}^{-1}\right)$ and $O_p\left(C_{\underline{N}T}^{-1}\right)$ by Theorem 1. The fourth term is bounded by $O_p\left(C_{\underline{N}T}^{-1}\right)$ from (29). The last term can be written as

$$\frac{1}{T\sqrt{N_i}}\widehat{\boldsymbol{F}}_i'\boldsymbol{e}_i = \frac{1}{T\sqrt{N_i}}\left(\widehat{\boldsymbol{F}}_i-\boldsymbol{F}_i\widehat{\mathscr{H}_i}\right)'\boldsymbol{e}_i+\frac{1}{T\sqrt{N_i}}\widehat{\mathscr{H}_i}'\boldsymbol{F}_i'\boldsymbol{e}_i$$

The first term is bounded by $O_p\left(C_{\underline{N}T}^{-2}\right)$ that follows from Lemma B1 of Bai (2003) with a slight modification. The second term is bounded by $O_p\left(T^{-1/2}\right)$ using the fact that $(N_iT)^{-1/2}\left\|\boldsymbol{F}_i\boldsymbol{e}_i\right\|=O_p(1)$ under Assumption B.2. Collecting all the terms, we conclude that

$$\frac{1}{\sqrt{N_i}}\left(\widehat{\boldsymbol{\Lambda}}_i'-\widehat{\mathscr{H}_i}^{-1}\boldsymbol{\Lambda}_i'\right)=O_p\left(\frac{1}{C_{\underline{N}T}}\right)$$

Q.E.D

Proof of Theorem 3

By Lemmas 1 and 2 and using the continuity of the singular values, we have:

$$\hat{\delta}_k = \begin{cases} \sqrt{T}O_p(C_{\underline{N}T}^{-1}) & \text{for } k = 1, \ldots, r_0 \\ O_p\left(\sqrt{T}\right) & \text{for } k = r_0 + 1, \ldots, Rr_{\max} \\ C_{\underline{N}T}^{-1}O_p\left(\sqrt{T}\right) & \text{for } k = 0 \end{cases}$$

If $r_0 > 0$, we have:

$$\lim_{N_1,\ldots,N_R,T\to\infty} \frac{\hat{\delta}_{k+1}}{\hat{\delta}_k} = \begin{cases} O_p(C_{\underline{N}T}) & \text{for } k = r_0 \\ O_p(1) & \text{for } k = r_0 + 1, \ldots, Rr_{\max} \\ O_p(1) & \text{for } k = 0, 1, \ldots, r_0 - 1 \end{cases}$$

On the other hand, if $r_0 = 0$, we have:

$$\lim_{N_1,\ldots,N_R,T\to\infty} \frac{\hat{\delta}_{k+1}}{\hat{\delta}_k} = \begin{cases} O_p(1) & \text{for } k = 1, \ldots, Rr_{\max} \\ O_p(C_{\underline{N}T}) & \text{for } k = 0 \end{cases}$$

As $C_{\underline{N}T} \to \infty$, the ratio $\hat{\delta}_{k+1}/\hat{\delta}_k$ attains maximum at $k = r_0$. Thus, the desired results follows.

Q.E.D

Lemma 6. *Let $C_{N_i,T} = \min\{\sqrt{N_i}, \sqrt{T}\}$. Under Assumptions A-C and F–G, we have:*

1. *For each i and t, as $N_i, T \to \infty$, we have:*

$$\widehat{\boldsymbol{K}}_{it} - \widehat{\boldsymbol{H}}_i'\boldsymbol{K}_{it} = \widehat{\boldsymbol{V}}_i^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{K}}_{is}\omega_i(s,t) + \frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{K}}_{is}\zeta_{i,st} + \frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{K}}_{is}\eta_{i,st} + \frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{K}}_{is}\mu_{i,st}\right)$$

   where $\widehat{\boldsymbol{H}}_i = (\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i)\left(\boldsymbol{K}_i'\widehat{\boldsymbol{K}}_i/T\right)\widehat{\boldsymbol{V}}_i^{-1}$ is an $(r_0+r_i)\times(r_0+r_i)$ matrix with $\widehat{\boldsymbol{V}}_i$ being the diagonal matrix consisting of the first $r_0 + r_i$ eigenvalues of $(N_iT)^{-1}\mathbf{Y}_i\mathbf{Y}_i'$ in descending order. In addition,

   (a) $T^{-1}\sum_{s=1}^{T}\widehat{\boldsymbol{K}}_{is}\omega_i(s,t) = O_p\left(T^{-1/2}C_{N_iT}^{-1}\right)$ where $\omega_i(s,t) = E\left(N_i^{-1}\sum_{j=1}^{N_i}e_{ijs}e_{ijt}\right)$;

   (b) $T^{-1}\sum_{s=1}^{T}\widehat{\boldsymbol{K}}_{is}\zeta_{i,st} = O_p\left(N_i^{-1/2}C_{N_iT}^{-1}\right)$ where $\zeta_{i,st} = N_i^{-1}\boldsymbol{e}_{i.s}'\boldsymbol{e}_{i.t} - \omega_i(s,t)$;

   (c) $T^{-1}\sum_{s=1}^{T}\widehat{\boldsymbol{K}}_{is}\eta_{i,st} = O_p\left(N_i^{-1/2}\right)$ where $\eta_{i,st} = N_i^{-1}\boldsymbol{K}_{is}'\boldsymbol{\Theta}_i'\boldsymbol{e}_{i.t}$;

   (d) $T^{-1}\sum_{s=1}^{T}\widehat{\boldsymbol{K}}_{is}\mu_{i,st} = O_p\left(N_i^{-1/2}C_{N_iT}^{-1}\right)$ where $\mu_{i,st} = N_i^{-1}\boldsymbol{K}_{it}'\boldsymbol{\Theta}_i'\boldsymbol{e}_{i.s}$

2. *Let $\widehat{\boldsymbol{\mathcal{R}}}_i = T^{-1/2}\left(\boldsymbol{K}_i - \widehat{\boldsymbol{K}}_i\widehat{\boldsymbol{H}}_i\right)$. For each i, as $N_i, T \to \infty$, we have:*

$$\left\|\widehat{\boldsymbol{\mathcal{R}}}_i\right\| = O_p\left(\frac{1}{\sqrt{T}C_{N_iT}}\right) + O_p\left(\frac{1}{\sqrt{N_i}}\right)$$

3. *As $N_m, T \to \infty$, for each m and h, we have: $T^{-1/2}\widehat{\boldsymbol{\mathcal{R}}}_m'\boldsymbol{K}_h = O_p\left(C_{N_mT}^{-2}\right)$.*

4. *As $N_m, N_h, T \to \infty$, for each m and h, we have: $T^{-1/2}\widehat{\boldsymbol{\mathcal{R}}}_m'\widehat{\boldsymbol{K}}_h = O_p\left(C_{N_mT}^{-2}\right)$.*

5. *As $N_m, T \to \infty$, for each m, h and j, we have:* $T^{-1/2}\widehat{\boldsymbol{\mathcal{R}}}'_m \boldsymbol{e}_{hj} = O_p\left(C^{-2}_{N_m T}\right)$.

Proof
1. For each i, by the definition of PC, we have $\widehat{\boldsymbol{K}}_i \widehat{\boldsymbol{V}}_i = (N_i T)^{-1} \mathbf{Y}_i' \mathbf{Y}_i \widehat{\boldsymbol{K}}_i$. By plugging (6) into this equation, we obtain:

$$\widehat{\boldsymbol{K}}_i - \boldsymbol{K}_i \widehat{\boldsymbol{H}}_i = \left(\frac{1}{N_i T} \boldsymbol{e}_i \boldsymbol{\Theta}_i \boldsymbol{K}_i' \widehat{\boldsymbol{K}}_i + \frac{1}{N_i T} \boldsymbol{K}_i \boldsymbol{\Theta}_i' \boldsymbol{e}_i' \widehat{\boldsymbol{K}}_i + \frac{1}{N_i T} \boldsymbol{e}_i \boldsymbol{e}_i' \widehat{\boldsymbol{K}}_i \right) \widehat{\boldsymbol{V}}_i^{-1} \tag{30}$$

Let $\widehat{\boldsymbol{K}}_{it} - \widehat{\boldsymbol{H}}_i \boldsymbol{K}_{it}$ be the $t-$th row vector of $\widehat{\boldsymbol{K}}_i - \boldsymbol{K}_i \widehat{\boldsymbol{H}}_i$. Then, the proof follows directly from Lemma A.2 in Bai (2003).
2. For each i, we have:

$$\left\| \frac{1}{\sqrt{T}} \left(\widehat{\boldsymbol{K}}_i - \boldsymbol{K}_i \widehat{\boldsymbol{H}}_i \right) \right\|^2 = tr \left\{ \frac{1}{T} \left(\widehat{\boldsymbol{K}}_i - \boldsymbol{K}_i \widehat{\boldsymbol{H}}_i \right)' \left(\widehat{\boldsymbol{K}}_i - \boldsymbol{K}_i \widehat{\boldsymbol{H}}_i \right) \right\} =$$
$$tr \left\{ \frac{1}{T} \sum_{t=1}^{T} \left(\widehat{\boldsymbol{K}}_{it} - \widehat{\boldsymbol{H}}_i' \boldsymbol{K}_{it} \right) \left(\widehat{\boldsymbol{K}}_{it} - \widehat{\boldsymbol{H}}_i' \boldsymbol{K}_{it} \right)' \right\} = \frac{1}{T} \sum_{t=1}^{T} \left\| \widehat{\boldsymbol{K}}_{it} - \widehat{\boldsymbol{H}}_i' \boldsymbol{K}_{it} \right\|^2$$

Combining the terms of (a)-(d) in Lemma 6.1, the results follows immediately.
3. Consider the term,

$$\frac{1}{\sqrt{T}} \widehat{\mathcal{R}}'_m \boldsymbol{K}_h = \widehat{\boldsymbol{V}}_m^{-1} \left(\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \widehat{\boldsymbol{K}}_{ms} \omega_m(s,t) \boldsymbol{K}_{ht}' + \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \widehat{\boldsymbol{K}}_{ms} \zeta_{m,st} \boldsymbol{K}_{ht}' \right.$$
$$\left. + \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \widehat{\boldsymbol{K}}_{ms} \eta_{m,st} \boldsymbol{K}_{ht}' + \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \widehat{\boldsymbol{K}}_{ms} \mu_{m,st} \boldsymbol{K}_{ht}' \right)$$

where $\left\| \widehat{\boldsymbol{V}}_m^{-1} \right\| = O_p(1)$ by Lemma 8. Let $T^{-1/2} \widehat{\mathcal{R}}'_m \boldsymbol{K}_h = \widehat{\boldsymbol{V}}_m^{-1} (\boldsymbol{X1} + \boldsymbol{X2} + \boldsymbol{X3} + \boldsymbol{X4})$. $\boldsymbol{X1}$ can be written as

$$\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\widehat{\boldsymbol{K}}_{ms} - \widehat{\boldsymbol{H}}_m' \boldsymbol{K}_{ms} \right) \omega_m(s,t) \boldsymbol{K}_{ht}' + \widehat{\boldsymbol{H}}_m' \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \boldsymbol{K}_{ms} \omega_m(s,t) \boldsymbol{K}_{ht}' = \boldsymbol{X1.1} + \boldsymbol{X1.2}$$

By the Cauchy-Schwarz inequality, we have:

$$\|\boldsymbol{X1.1}\| \leq \frac{1}{\sqrt{T}} \left(\frac{1}{T} \sum_{t=1}^{T} \frac{1}{T} \sum_{s=1}^{T} \left\| \widehat{\boldsymbol{K}}_{ms} - \widehat{\boldsymbol{H}}_m' \boldsymbol{K}_{ms} \right\|^2 \right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} |\omega_m(s,t)|^2 \|\boldsymbol{K}_{ht}\|^2 \right)^{1/2}$$
$$= \left[O_p \left(\frac{1}{\sqrt{N_m}} \right) + O_p \left(\frac{1}{\sqrt{T} C_{N_m T}} \right) \right] \frac{1}{\sqrt{T}} = O_p \left(\frac{1}{\sqrt{N_m T}} \right) + O_p \left(\frac{1}{T C_{N_m T}} \right)$$

where we used Lemma 6.1, Assumption B.1 and the fact that $T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} |\omega_m(s,t)|^2 = O(1)$ (see Bai and Ng (2002) Lemma 1.(i)). The expected value of $\boldsymbol{X1.2}$ without $\widehat{\boldsymbol{H}}_m'$, is bounded by

$$\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} |\omega_m(s,t)| E \left(\|\boldsymbol{K}_{ms}\|^2 \right)^{1/2} E \left(\|\boldsymbol{K}_{ht}\|^2 \right)^{1/2} \leq \mathcal{M} \frac{1}{T} \left(\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} |\omega_m(s,t)| \right) = O \left(\frac{1}{T} \right)$$

under Assumption B.1 and Assumption A.2. Therefore, we obtain: $\|\boldsymbol{X1}\| = O_p\left(C_{N_mT}^{-2}\right)$.

Next, by the Cauchy-Schwarz inequality, $\boldsymbol{X2}$ is bounded by

$$\|\boldsymbol{X2}\| \leq \left(\frac{1}{N_mT^2}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N_mT}}\sum_{s=1}^{T}\sum_{j=1}^{N_m}\boldsymbol{K}_{ms}\left[e_{mjs}e_{mjt} - E(e_{mjs}e_{mjt})\right]\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{K}_{ht}\|^2\right)^{1/2}$$

$$= O_p\left(\frac{1}{\sqrt{N_mT}}\right)$$

under Assumptions G.1 and B.1.

$\boldsymbol{X3}$ can be expressed as

$$\boldsymbol{X3} = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widehat{\boldsymbol{K}}_{ms} - \widehat{\boldsymbol{H}}_m'\boldsymbol{K}_{ms}\right)\eta_{m,st}\boldsymbol{K}_{ht}' + \widehat{\boldsymbol{H}}_m'\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{K}_{ms}\eta_{m,st}\boldsymbol{K}_{ht}' = \boldsymbol{X3.1} + \boldsymbol{X3.2}$$

Applying the Cauchy-Schwarz inequality to $\boldsymbol{X3.1}$, we obtain:

$$\|\boldsymbol{X3.1}\| \leq \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\widehat{\boldsymbol{K}}_{ms} - \widehat{\boldsymbol{H}}_m'\boldsymbol{K}_{ms}\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{K}_{ht}\eta_{m,st}\right\|^2\right)^{1/2}$$

The second part can be expressed as

$$\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{K}_{ht}\eta_{m,st}\right\|^2\right)^{1/2} = \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{N_mT}\sum_{t=1}^{T}\boldsymbol{K}_{ht}\boldsymbol{K}_{ms}'\boldsymbol{\theta}_{mj}e_{mjt}\right\|^2\right)^{1/2}$$

$$\leq \left(\frac{1}{T}\sum_{s=1}^{T}\|\boldsymbol{K}_{ms}\|^2\left\|\frac{1}{N_mT}\sum_{t=1}^{T}\sum_{j=1}^{N_m}\boldsymbol{K}_{ht}'\boldsymbol{\theta}_{mj}e_{mjt}\right\|^2\right)^{1/2} = O_p\left(\frac{1}{\sqrt{N_mT}}\right)$$

under Assumptions B.1 and G.2. Hence, $\|\boldsymbol{X3.1}\| = O_p\left(C_{N_mT}^{-1}\right)O_p\left(N_m^{-1/2}T^{-1/2}\right)$. For $\boldsymbol{X3.2}$, we have:

$$\boldsymbol{X3.2} = \frac{1}{T}\sum_{s=1}^{T}\boldsymbol{K}_{ms}\boldsymbol{K}_{ms}'\frac{1}{N_mT}\sum_{t=1}^{T}\sum_{j=1}^{N_m}\boldsymbol{K}_{ht}'\boldsymbol{\theta}_{mj}e_{mjt} = O_p\left(\frac{1}{\sqrt{N_mT}}\right)$$

by Assumption G.2. Therefore, $\|\boldsymbol{X3}\| = O_p\left(N_m^{-1/2}T^{-1/2}\right)$.

Following similar steps, we obtain: $\boldsymbol{X4} = O_p\left(N_m^{-1/2}T^{-1/2}\right)$. Collecting all these results, we obtain: $T^{-1/2}\widehat{\boldsymbol{\mathcal{R}}}_m'\boldsymbol{K}_h = O_p\left(C_{N_mT}^{-2}\right)$.

4.

$$\frac{1}{\sqrt{T}}\widehat{\boldsymbol{\mathcal{R}}}_m\widehat{\boldsymbol{K}}_h = \frac{1}{\sqrt{T}}\widehat{\boldsymbol{\mathcal{R}}}_m\left(\widehat{\boldsymbol{K}}_h - \boldsymbol{K}_h\widehat{\boldsymbol{H}}_h\right) + \frac{1}{\sqrt{T}}\widehat{\boldsymbol{\mathcal{R}}}_m\boldsymbol{K}_h\widehat{\boldsymbol{H}}_h$$

By Lemmas 6.2 and 6.3, it follows that $T^{-1/2}\widehat{\mathcal{R}}_m'\widehat{\boldsymbol{K}}_h = O_p\left(C_{N_mT}^{-2}\right)$.

5. Consider

$$\frac{1}{\sqrt{T}}\widehat{\boldsymbol{\mathcal{R}}}'_m \boldsymbol{e}_{hj} = \widehat{\boldsymbol{V}}_m^{-1}\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widehat{\boldsymbol{K}}_{ms}\omega_m(s,t)e_{hjt} + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widehat{\boldsymbol{K}}_{ms}\zeta_{m,st}e_{hjt}\right.$$
$$\left. + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widehat{\boldsymbol{K}}_{ms}\eta_{m,st}e_{hjt} + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widehat{\boldsymbol{K}}_{ms}\mu_{m,st}e_{hjt}\right)$$

where $\left\|\widehat{\boldsymbol{V}}_m^{-1}\right\| = O_p(1)$ by Lemma 8. Let $T^{-1/2}\widehat{\mathcal{R}}'_m \boldsymbol{e}_{hj} = \widehat{\boldsymbol{V}}_m^{-1}\left(\mathcal{X}1 + \mathcal{X}2 + \mathcal{X}3 + \mathcal{X}4\right)$. As the first term $\mathcal{X}1$ is of order $O_p\left(C_{N_mT}^{-2}\right)$, the proof is the same as that of $\boldsymbol{X1}$ in Lemma 6. The second term is equal to

$$\mathcal{X}2 = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widehat{\boldsymbol{K}}_{ms} - \widehat{\boldsymbol{H}}'_m\boldsymbol{K}_{ms}\right)\zeta_{m,st}e_{hjt} + \widehat{\boldsymbol{H}}'_m\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{K}_{ms}\zeta_{m,st}e_{hjt} = \mathcal{X}2.1 + \mathcal{X}2.2$$

Using the Cauchy-Schwarz inequality, we have:

$$\|\mathcal{X}2.1\| \leq \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\widehat{\boldsymbol{K}}_{ms} - \widehat{\boldsymbol{H}}'_m\boldsymbol{K}_{ms}\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left(\frac{1}{T}\sum_{t=1}^{T}\zeta_{m,st}e_{hjt}\right)^2\right)^{1/2}$$

Notice that by Assumption A.5,

$$\frac{1}{T}\sum_{t=1}^{T}\zeta_{m,st}e_{hjt} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\sqrt{N_m}}\left(\frac{1}{\sqrt{N_m}}\sum_{k=1}^{N_m}\left[e_{mks}e_{mkt} - E(e_{mks}e_{mkt})\right]\right)e_{hjt} = O_p\left(\frac{1}{\sqrt{N_m}}\right)$$

Using Lemma 6.2, we show that

$$\|\mathcal{X}2.1\| = O_p\left(\frac{1}{\sqrt{N_mT}C_{N_mT}}\right) + O_p\left(\frac{1}{N_m}\right)$$

In addition, by Assumption G.1,

$$\mathcal{X}2.2 = \widehat{\boldsymbol{H}}'_m\frac{1}{\sqrt{N_mT}}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N_mT}}\sum_{s=1}^{T}\sum_{k=1}^{N_m}\boldsymbol{K}_{ms}\left[e_{mks}e_{mkt} - E(e_{mks}e_{mkt})\right]\right)e_{hjt} = O_p\left(\frac{1}{\sqrt{N_mT}}\right)$$

Combining these two terms, we have $\mathcal{X}2 = O_p(C_{N_mT}^{-2})$. Next, we can rewrite $\mathcal{X}3$ as

$$\mathcal{X}3 = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widehat{\boldsymbol{K}}_{ms} - \widehat{\boldsymbol{H}}'_m\boldsymbol{K}_{ms}\right)\eta_{m,st}e_{hjt} + \widehat{\boldsymbol{H}}'_m\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{K}_{ms}\eta_{m,st}e_{hjt} = \mathcal{X}3.1 + \mathcal{X}3.2$$

By the Cauchy-Schwarz inequality, we have:

$$\|\mathcal{X}3.1\| \leq \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\widehat{\boldsymbol{K}}_{ms} - \widehat{\boldsymbol{H}}'_m\boldsymbol{K}_{ms}\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left(\frac{1}{T}\sum_{t=1}^{T}\eta_{m,st}e_{hjt}\right)^2\right)^{1/2}$$

Notice that

$$\frac{1}{T}\sum_{t=1}^{T}\eta_{m,st}e_{hjt} = \frac{1}{\sqrt{N_m}}\boldsymbol{K}_{ms}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N_m}}\sum_{k=1}^{N_m}\boldsymbol{\theta}_{mk}e_{mkt}\right)e_{hjt} = O_p\left(\frac{1}{\sqrt{N_m}}\right)$$

Using Lemma 6.2, we have:

$$\|\mathcal{X}3.1\| = O_p\left(\frac{1}{\sqrt{N_m T}C_{N_mT}}\right) + O_p\left(\frac{1}{N_m}\right).$$

For the second part, by Assumption F.2, we have:

$$\mathcal{X}3.2 = \widehat{\boldsymbol{H}}_m'\left(\frac{1}{T}\sum_{s=1}^{T}\boldsymbol{K}_{ms}\boldsymbol{K}_{ms}'\right)\frac{1}{N_mT}\sum_{t=1}^{T}\sum_{k=1}^{N_m}\boldsymbol{\theta}_{mk}e_{mkt}e_{hjt} = O_p\left(\frac{1}{N_m}\right)$$

Combining these two terms, we obtain $\mathcal{X}3 = O_p\left(C_{N_mT}^{-2}\right)$. The proof of $\mathcal{X}4$ is similar to that of $\mathcal{X}3$. Finally, we conclude that $T^{-1/2}\widehat{\mathcal{R}}_m'\boldsymbol{e}_{hj} = O_p\left(C_{N_mT}^{-2}\right)$.

Q.E.D

Lemma 7. *Under Assumptions A–C and E–G, as $N_1,\ldots,N_R,T\to\infty$, we have:*

1.
$$\frac{1}{T}\left\|\widehat{\boldsymbol{\Phi}}'\widehat{\boldsymbol{\Phi}} - \widehat{\boldsymbol{H}}'\boldsymbol{\Phi}'\boldsymbol{\Phi}\widehat{\boldsymbol{H}}\right\| = O_p\left(\frac{1}{C_{\underline{N},T}^2}\right) \text{ and } \widehat{\boldsymbol{Q}}_i^{r_0} - \bar{\boldsymbol{Q}}_i^{r_0}\boldsymbol{D} = O_p\left(\frac{1}{C_{\underline{N}T}^2}\right)$$

2.
$$\frac{1}{T}\left\|\widehat{\boldsymbol{\Psi}}'\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}'\boldsymbol{\Psi}\right\| = O_p\left(\frac{1}{C_{\underline{N},T}^2}\right) \text{ and } \left\|\widehat{\boldsymbol{L}}^{r_0} - \boldsymbol{L}^{r_0}\boldsymbol{U}\right\| = O_p\left(\frac{1}{C_{\underline{N}T}^2}\right)$$

where $C_{\underline{N},T} = \min\{\sqrt{\underline{N}},\sqrt{T}\}$ and $\underline{N} = \min\{N_1,N_2,\ldots,N_R\}$.

Proof

1. By definition of $\widehat{\boldsymbol{\Phi}}$, we have:

$$\frac{1}{T}\widehat{\boldsymbol{\Phi}}'\widehat{\boldsymbol{\Phi}} = \frac{1}{T}\begin{bmatrix}(R-1)\widehat{\boldsymbol{K}}_1'\widehat{\boldsymbol{K}}_1 & -\widehat{\boldsymbol{K}}_1'\widehat{\boldsymbol{K}}_2 & \ldots & -\widehat{\boldsymbol{K}}_1'\widehat{\boldsymbol{K}}_R\\ -\widehat{\boldsymbol{K}}_2'\widehat{\boldsymbol{K}}_1 & (R-1)\widehat{\boldsymbol{K}}_2'\widehat{\boldsymbol{K}}_2 & \ldots & -\widehat{\boldsymbol{K}}_R'\widehat{\boldsymbol{K}}_R\\ & & \vdots & \\ -\widehat{\boldsymbol{K}}_R'\widehat{\boldsymbol{K}}_1 & -\widehat{\boldsymbol{K}}_R'\widehat{\boldsymbol{K}}_1 & \ldots & (R-1)\widehat{\boldsymbol{K}}_R'\widehat{\boldsymbol{K}}_R\end{bmatrix}$$

Using (30) and the definition of $\widehat{\mathcal{R}}_i$ in Lemma 6.2, we obtain:

$$\frac{1}{T}\widehat{\boldsymbol{\Phi}}'\widehat{\boldsymbol{\Phi}} = \frac{1}{T}\widehat{\boldsymbol{H}}'\boldsymbol{\Phi}'\boldsymbol{\Phi}\widehat{\boldsymbol{H}} + \widehat{\mathbb{A}}_1 + \widehat{\mathbb{A}}_2 + \widehat{\mathbb{A}}_3$$

where

$$\widehat{\mathbb{A}}_1 = \widehat{\mathbb{A}}_2' = \frac{1}{\sqrt{T}}\begin{bmatrix}(R-1)\widehat{\boldsymbol{\mathcal{R}}}_1'\boldsymbol{K}_1\widehat{\boldsymbol{H}}_1 & -\widehat{\boldsymbol{\mathcal{R}}}_1'\boldsymbol{K}_2\widehat{\boldsymbol{H}}_2 & \ldots & -\widehat{\boldsymbol{\mathcal{R}}}_1'\boldsymbol{K}_R\widehat{\boldsymbol{H}}_R\\ -\widehat{\boldsymbol{\mathcal{R}}}_2'\boldsymbol{K}_1\widehat{\boldsymbol{H}}_1 & (R-1)\widehat{\boldsymbol{\mathcal{R}}}_2'\boldsymbol{K}_2\widehat{\boldsymbol{H}}_2 & \ldots & -\widehat{\boldsymbol{\mathcal{R}}}_2'\boldsymbol{K}_R\widehat{\boldsymbol{H}}_R\\ & & \vdots & \\ -\widehat{\boldsymbol{\mathcal{R}}}_R'\boldsymbol{K}_1\widehat{\boldsymbol{H}}_1 & -\widehat{\boldsymbol{\mathcal{R}}}_R'\boldsymbol{K}_2\widehat{\boldsymbol{H}}_2 & \ldots & (R-1)\widehat{\boldsymbol{\mathcal{R}}}_R'\boldsymbol{K}_R\widehat{\boldsymbol{H}}_R\end{bmatrix}$$

and

$$\widehat{\mathbb{A}}_3 = \begin{bmatrix} (R-1)\widehat{\boldsymbol{\mathcal{R}}}_1'\widehat{\boldsymbol{\mathcal{R}}}_1 & -\widehat{\boldsymbol{\mathcal{R}}}_1\widehat{\boldsymbol{\mathcal{R}}}_2' & \dots & -\widehat{\boldsymbol{\mathcal{R}}}_1'\widehat{\boldsymbol{\mathcal{R}}}_R \\ -\widehat{\boldsymbol{\mathcal{R}}}_2'\widehat{\boldsymbol{\mathcal{R}}}_1 & (R-1)\widehat{\boldsymbol{\mathcal{R}}}_2'\widehat{\boldsymbol{\mathcal{R}}}_2 & \dots & -\widehat{\boldsymbol{\mathcal{R}}}_2'\widehat{\boldsymbol{\mathcal{R}}}_R \\ & & \vdots & \\ -\widehat{\boldsymbol{\mathcal{R}}}_R'\widehat{\boldsymbol{\mathcal{R}}}_1 & -\widehat{\boldsymbol{\mathcal{R}}}_R\widehat{\boldsymbol{\mathcal{R}}}_2' & \dots & (R-1)\widehat{\boldsymbol{\mathcal{R}}}_R'\widehat{\boldsymbol{\mathcal{R}}}_R \end{bmatrix}$$

Using Lemma 6.3 and the fact that $\widehat{\boldsymbol{H}}_i$ is $O_p(1)$, we have $\widehat{\mathbb{A}}_1 = \widehat{\mathbb{A}}_2' = O_p\left(C_{\underline{N}T}^{-2}\right)$. Furthermore, by Lemma 6.2, we have $\widehat{\mathbb{A}}_3 = O_p\left(C_{\underline{N}T}^{-2}\right)$.

2. By definition of $\widehat{\boldsymbol{\Psi}}$ and $\boldsymbol{\Psi}$, we have:

$$\left\| \frac{1}{T}\widehat{\boldsymbol{\Psi}}'\widehat{\boldsymbol{\Psi}} - \frac{1}{T}\boldsymbol{\Psi}'\boldsymbol{\Psi} \right\| \le \sum_{i=1}^{R} \left\| \frac{1}{T}\widehat{\boldsymbol{K}}_i\widehat{\boldsymbol{Q}}_i^{r_0}\widehat{\boldsymbol{Q}}_i^{r_0\prime}\widehat{\boldsymbol{K}}_i' - \frac{1}{T}\boldsymbol{G}\boldsymbol{G}' \right\|$$

Using $\widehat{\boldsymbol{K}}_i = \widehat{\boldsymbol{K}}_i - \boldsymbol{K}_i\widehat{\boldsymbol{H}}_i + \boldsymbol{K}_i\widehat{\boldsymbol{H}}_i$, we have:

$$\begin{aligned} \frac{1}{T}\widehat{\boldsymbol{K}}_i\widehat{\boldsymbol{Q}}_i^{r_0}\widehat{\boldsymbol{Q}}_i^{r_0\prime}\widehat{\boldsymbol{K}}_i' - \frac{1}{T}\boldsymbol{G}\boldsymbol{G}' = \frac{1}{T}\widehat{\boldsymbol{Q}}_i'\left(\widehat{\boldsymbol{K}}_i - \boldsymbol{K}_i\widehat{\boldsymbol{H}}_i\right)'\widehat{\boldsymbol{K}}_i\widehat{\boldsymbol{Q}}_i + \frac{1}{T}\widehat{\boldsymbol{Q}}_i'\widehat{\boldsymbol{H}}_i'\widehat{\boldsymbol{K}}_i'\left(\widehat{\boldsymbol{K}}_i - \boldsymbol{K}_i\widehat{\boldsymbol{H}}_i\right)\widehat{\boldsymbol{Q}}_i \\ + \frac{1}{T}\widehat{\boldsymbol{Q}}_i'\widehat{\boldsymbol{H}}_i'\boldsymbol{K}_i'\boldsymbol{K}_i\widehat{\boldsymbol{H}}_i\widehat{\boldsymbol{Q}}_i - \frac{1}{T}\boldsymbol{G}\boldsymbol{G}' \end{aligned}$$

The first two terms are bounded by $O_p\left(C_{N_iT}^{-2}\right)$ by Lemmas 6.3 and 6.4. Using $\widehat{\boldsymbol{Q}}_i = \widehat{\boldsymbol{H}}_i^{-1}\boldsymbol{Q}_i\boldsymbol{D} + O_p\left(C_{N_iT}^{-2}\right)$, the remaining terms can be expressed as

$$\boldsymbol{D}'\boldsymbol{B}'\frac{\boldsymbol{G}'\boldsymbol{G}}{T}\boldsymbol{B}\boldsymbol{D} + O_p\left(\frac{1}{C_{N_iT}^2}\right) - \frac{\boldsymbol{G}'\boldsymbol{G}}{T}$$

Notice that

$$\left\| \boldsymbol{D}'\boldsymbol{B}'\frac{\boldsymbol{G}'\boldsymbol{G}}{T}\boldsymbol{B}\boldsymbol{D} - \frac{\boldsymbol{G}'\boldsymbol{G}}{T} \right\| = 0$$

since $\boldsymbol{D}$ and $\boldsymbol{B}$ are orthogonal matrices. Therefore, we conclude that

$$\left\| \frac{1}{T}\widehat{\boldsymbol{\Psi}}'\widehat{\boldsymbol{\Psi}} - \frac{1}{T}\boldsymbol{\Psi}'\boldsymbol{\Psi} \right\| = O_p\left(\frac{1}{C_{\underline{N}T}^2}\right)$$

By Lemma 3 and Assumption B.1, we have $\left\|\widehat{\boldsymbol{L}}^{r_0} - \boldsymbol{L}^{r_0}\boldsymbol{U}\right\| = O_p\left(C_{\underline{N}T}^{-2}\right)$ where $\boldsymbol{U}$ is defined in (24).

Q.E.D

Lemma 8. *Under Assumptions A–C and F–G, as $N_i, T \to \infty$, we have for each i:*

1.
$$\frac{1}{T}\widehat{\boldsymbol{K}}_i'\left(\frac{1}{N_iT}\boldsymbol{Y}_i\boldsymbol{Y}_i'\right)\widehat{\boldsymbol{K}}_i = \widehat{\boldsymbol{V}}_i \xrightarrow{p} \boldsymbol{V}_i$$

where $\boldsymbol{V}_i$ is a diagonal matrix consisting of the eigenvalues of $\boldsymbol{\Sigma}_{\Theta_i}\boldsymbol{\Sigma}_{K_i}$.

2.
$$\frac{\widehat{\boldsymbol{K}}_i'\boldsymbol{K}_i}{T}\left(\frac{\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i}{N_i}\right)\frac{\boldsymbol{K}_i'\widehat{\boldsymbol{K}}_i}{T}\xrightarrow{p}\boldsymbol{V}_i$$

3.
$$plim_{N_i,T\to\infty}\frac{\widehat{\boldsymbol{K}}_i'\boldsymbol{K}_i}{T}=\mathbb{Q}_i$$

The $(r_0+r_i)\times(r_0+r_i)$ *matrix* $\mathbb{Q}_i$ *is given by* $\mathbb{Q}_i=\boldsymbol{V}_i^{1/2}\boldsymbol{\mathcal{P}}_i'\boldsymbol{\Sigma}_{\Theta_i}^{-1/2}$ *and invertible, where* $\boldsymbol{V}_i$ *is the diagonal matrix consisting of the eigenvalues of* $\boldsymbol{\Sigma}_{\Theta_i}^{1/2}\boldsymbol{\Sigma}_{K_i}\boldsymbol{\Sigma}_{\Theta_i}^{1/2}$ *and* $\mathcal{P}_i$ *is the corresponding eigenvector matrix such that* $\boldsymbol{\mathcal{P}}_i'\boldsymbol{\mathcal{P}}_i/T=\boldsymbol{I}_{r_0+r_i}$.

4.
$$plim_{N_i,T\to\infty}\widehat{\boldsymbol{H}}_i=\boldsymbol{H}_i$$

where $\boldsymbol{H}_i=\boldsymbol{\Sigma}_{\Theta_i}\mathbb{Q}_i'\boldsymbol{V}_i^{-1}$.

Proof.
The proof follows the same lines from Proposition 1 and Lemma A.3 in Bai (2003) and is thus omitted.
Q.E.D

Proof of Theorem 4
From (19), we have for each t:

$$\widehat{\boldsymbol{G}}_t=\frac{1}{\sqrt{T}}(\widehat{\boldsymbol{\Xi}}^{r_0})^{-1}\widehat{\boldsymbol{L}}^{r_0\prime}\left(\sum_i^R\widehat{\boldsymbol{K}}_i\widehat{\boldsymbol{Q}}_i^{r_0}\widehat{\boldsymbol{Q}}_i^{r_0\prime}\widehat{\boldsymbol{K}}_{it}\right)$$

Using the asymptotic expansions in Lemma 7.1 and Lemma 7.2:

$$\widehat{\boldsymbol{L}}^{r_0}=\boldsymbol{L}^{r_0}\boldsymbol{U}+O_p\left(\frac{1}{C_{\underline{N}T}^2}\right),\ \widehat{\boldsymbol{Q}}_i^{r_0}=\widehat{\boldsymbol{H}}_i^{-1}\boldsymbol{Q}_i^{r_0}\boldsymbol{D}+O_p\left(\frac{1}{C_{\underline{N}T}^2}\right)$$

and keeping the term up to order $O_p\left(C_{\underline{N}T}^{-2}\right)$, we have:

$$\widehat{\boldsymbol{G}}_t=\frac{1}{\sqrt{T}}\boldsymbol{U}(\boldsymbol{\Xi}^{r_0})^{-1}\boldsymbol{L}^{r_0\prime}\left[\sum_{i=1}^R\widehat{\boldsymbol{K}}_i\widehat{\boldsymbol{H}}_i^{-1}\boldsymbol{Q}_i^{r_0}\boldsymbol{Q}_i^{r_0\prime}\left(\widehat{\boldsymbol{H}}_i^{-1}\right)'\widehat{\boldsymbol{K}}_{it}\right]+O_p\left(\frac{1}{C_{\underline{N}T}^2}\right)$$

where we use that $(\boldsymbol{\Xi}^{r_0})^{-1}\boldsymbol{U}'=\boldsymbol{U}(\boldsymbol{\Xi}^{r_0})^{-1}$ because both matrices are diagonal. Replacing $T^{-1/2}\widehat{\boldsymbol{K}}_i$ with $T^{-1/2}\boldsymbol{K}_i\widehat{\boldsymbol{H}}_i+\widehat{\boldsymbol{\mathcal{R}}}_i$, the above equation can be written as

$$\widehat{\boldsymbol{G}}_t=\mathbb{H}'\frac{1}{R}\sum_{i=1}^R\mathbb{I}_i'\left(\widehat{\boldsymbol{H}}_i^{-1}\right)'\widehat{\boldsymbol{K}}_{it}+\boldsymbol{U}\boldsymbol{\Xi}^{r_0,-1}\boldsymbol{L}^{r_0\prime}\left[\sum_{i=1}^R\widehat{\boldsymbol{\mathcal{R}}}_i\widehat{\boldsymbol{H}}_i^{-1}\boldsymbol{Q}_i^{r_0}\boldsymbol{Q}_i^{r_0\prime}\left(\widehat{\boldsymbol{H}}_i^{-1}\right)'\widehat{\boldsymbol{K}}_{it}\right]+O_p\left(\frac{1}{C_{\underline{N}T}^2}\right)\quad(31)$$

where we use $\boldsymbol{K}_i\boldsymbol{Q}_i^{r_0}=\boldsymbol{G}\boldsymbol{B}$, $\boldsymbol{B}=R^{-1}\boldsymbol{A}$, $\boldsymbol{Q}_i^{r_0}=\left[R^{-1}\boldsymbol{A}',\boldsymbol{0}\right]'$, $\boldsymbol{B}\boldsymbol{Q}_i^{r_0}=R^{-1}\left[\boldsymbol{I}_{r_0},\boldsymbol{0}\right]=R^{-1}\mathbb{I}_i'$ and $\boldsymbol{A}$ is an orthogonal matrix. From the asymptotic expansion in Lemma 6.1, it follows that $T^{-1}\sum_{s=1}^T\widehat{\boldsymbol{K}}_{is}\eta_{i,st}$ and $(N_iT)^{-1}\boldsymbol{e}_i\boldsymbol{\Theta}_i\boldsymbol{K}_i'\widehat{\boldsymbol{K}}_i\widehat{\boldsymbol{V}}_i^{-1}$ are dominant terms in $\widehat{\boldsymbol{K}}_{it}-\widehat{\boldsymbol{H}}_i'\boldsymbol{K}_{it}$ and $\widehat{\boldsymbol{\mathcal{R}}}_i$, respectively. So we have:

$$\widehat{\boldsymbol{K}}_{it}=\widehat{\boldsymbol{H}}_i'\boldsymbol{K}_{it}+\widehat{\boldsymbol{V}}_i^{-1}\frac{1}{N_iT}\sum_{s=1}^T\widehat{\boldsymbol{K}}_{is}\boldsymbol{K}_{is}'\boldsymbol{\Theta}_i'\boldsymbol{e}_{i.t}+O_p\left(\frac{1}{C_{N_iT}^2}\right)$$

and

$$\widehat{\boldsymbol{\mathcal{R}}}_i = \frac{1}{\sqrt{T}}\frac{1}{N_iT}\boldsymbol{e}_i\boldsymbol{\Theta}_i\boldsymbol{K}_i'\widehat{\boldsymbol{K}}_i\widehat{\boldsymbol{V}}_i^{-1} + O_p\left(\frac{1}{C_{N_iT}^2}\right)$$

Plugging these expressions into (31) and multiplying both sides by $\sqrt{N}$, we can show that

$$\begin{aligned}\sqrt{N}\left(\widehat{\boldsymbol{G}}_t - \mathbb{H}'\boldsymbol{G}_t\right) = \mathbb{H}'\frac{1}{R}\sum_{i=1}^{R}\mathbb{I}_i'\left(\widehat{\boldsymbol{H}}_i^{-1}\right)'\widehat{\boldsymbol{V}}_i^{-1}\sqrt{\frac{N}{N_i}}\left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{K}}_{is}\boldsymbol{K}_{is}'\right)\frac{1}{\sqrt{N_i}}\sum_{j=1}^{N_i}\boldsymbol{\theta}_{ij}\boldsymbol{e}_{ijt} \\ + \boldsymbol{U}\boldsymbol{\Xi}^{r_0,-1}\boldsymbol{L}^{r_0\prime}\sqrt{\frac{N}{N_i}}\frac{1}{R}\sum_{i=1}^{R}\frac{1}{\sqrt{N_iT}}\boldsymbol{e}_i\boldsymbol{\Theta}_i\frac{\boldsymbol{K}_i'\widehat{\boldsymbol{K}}_i}{T}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{H}}_i^{-1}\mathbb{I}_i\boldsymbol{G}_t + O_p\left(\frac{\sqrt{N}}{C_{\underline{N}T}^2}\right)\end{aligned}$$

Using $\widehat{\boldsymbol{H}}_i = (\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i)\left(\boldsymbol{K}_i'\widehat{\boldsymbol{K}}_i/T\right)\widehat{\boldsymbol{V}}_i^{-1}$ from Lemma 6.1 and rearranging terms, the above equation can be simplified to

$$\sqrt{N}\left(\widehat{\boldsymbol{G}}_t - (\mathbb{H}' + \mathbb{B}')\boldsymbol{G}_t\right) = \mathbb{H}'\frac{1}{R}\sum_{i=1}^{R}\mathbb{I}_i'\sqrt{\frac{N}{N_i}}\left(\frac{\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i}{N_i}\right)^{-1}\frac{1}{\sqrt{N_i}}\sum_{j=1}^{N_i}\boldsymbol{\theta}_{ij}\boldsymbol{e}_{ijt} + o_p(1).$$

where

$$\mathbb{B} = \frac{1}{R}\sum_{i=1}^{R}\sqrt{\frac{1}{N_i}}\mathbb{I}_i'\left(\frac{\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i}{N_i}\right)^{-1}\frac{\boldsymbol{\Theta}_i'\boldsymbol{e}_i'}{\sqrt{N_iT}}\boldsymbol{J}^{r_0}\boldsymbol{U}.$$

Following Lemmas 5 and 8, it is straightforward to show that $\mathbb{B} = O_p\left(\underline{N}^{-1/2}\right)$.

Finally, we achieve the desired result that

$$\sqrt{N}\left(\widehat{\boldsymbol{G}}_t - (\mathbb{H}' + \mathbb{B}')\boldsymbol{G}_t\right) = \frac{1}{R}\mathbb{H}'\boldsymbol{\mathcal{I}}'\widehat{\mathbb{C}}\mathbb{E}_t + o_p(1)$$

where $\mathcal{I} = [\boldsymbol{I}_{r_0}, \ldots, \boldsymbol{I}_{r_0}]'$ is an $Rr_0 \times r_0$ matrix, $\widehat{\mathbb{C}}$ is a $Rr_0 \times Rr_0$ block diagonal matrix given by

$$\widehat{\mathbb{C}} = \begin{bmatrix} \sqrt{\frac{N}{N_1}}\mathbb{I}_1'\left(\frac{\boldsymbol{\Theta}_1'\boldsymbol{\Theta}_1}{N_1}\right)^{-1} & & \\ & \ddots & \\ & & \sqrt{\frac{N}{N_1}}\mathbb{I}_R'\left(\frac{\boldsymbol{\Theta}_R'\boldsymbol{\Theta}_R}{N_R}\right)^{-1}\end{bmatrix},$$

and $\mathbb{E}_t$ is an $Rr_0 \times 1$ vector given by

$$\mathbb{E}_t = \begin{bmatrix}\mathbb{E}_{1t} \\ \mathbb{E}_{2t} \\ \vdots \\ \mathbb{E}_{Rt}\end{bmatrix} = \begin{bmatrix}\frac{1}{\sqrt{N_1}}\sum_{j=1}^{N_1}\boldsymbol{\theta}_{1j}e_{1jt} \\ \frac{1}{\sqrt{N_2}}\sum_{j=1}^{N_2}\boldsymbol{\theta}_{2j}e_{2jt} \\ \vdots \\ \frac{1}{\sqrt{N_R}}\sum_{j=1}^{N_R}\boldsymbol{\theta}_{Rj}e_{Rjt}\end{bmatrix} \xrightarrow{d} N\left(\boldsymbol{0}, \mathbb{D}_t^{(1)}\right)$$

Using Assumptions C.2b and E, we have:

$$\widehat{\mathbb{C}} \xrightarrow{p} \mathbb{C} = \begin{bmatrix}\alpha_1^{1/2}\mathbb{I}_1'\boldsymbol{\Sigma}_{\Theta_1} & & \\ & \ddots & \\ & & \alpha_R^{1/2}\mathbb{I}_R'\boldsymbol{\Sigma}_{\Theta_R}\end{bmatrix}$$

Therefore,

$$\sqrt{N}\left[\widehat{\boldsymbol{G}}_t-\left(\mathbb{H}'+\mathbb{B}'\right)\boldsymbol{G}_t\right]\xrightarrow{d}N\left(\mathbf{0},\frac{1}{R^2}\mathbb{H}'\mathcal{I}'\mathbb{C}\mathbb{D}_t\mathbb{C}'\mathcal{I}\mathbb{H}\right).$$

Q.E.D

Lemma 9. *Under Assumptions A–G, as $N_1,\ldots,N_R,T\to\infty$, we have:*

1. *For each i, we have $T^{-1}\left[\widehat{\boldsymbol{G}}-\boldsymbol{G}\left(\mathbb{H}+\mathbb{B}\right)\right]'\boldsymbol{K}_i=O_p\left(C_{\underline{N}T}^{-2}\right)$;*
2. *For each i and j, we have $T^{-1}\left[\widehat{\boldsymbol{G}}-\boldsymbol{G}\left(\mathbb{H}+\mathbb{B}\right)\right]'\boldsymbol{e}_{ij}=O_p\left(C_{\underline{N}T}^{-2}\right)$.*

Proof
Using $\widehat{\boldsymbol{K}}_{it}-\widehat{\boldsymbol{H}}_i'\boldsymbol{K}_{it}+\widehat{\boldsymbol{H}}_i'\boldsymbol{K}_{it}$, we can write (31) as

$$\begin{aligned}\widehat{\boldsymbol{G}}_t-\left(\mathbb{H}'+\mathbb{B}'\right)\boldsymbol{G}_t=&\,\mathbb{H}'\frac{1}{R}\sum_{i=1}^{R}\mathbb{I}_i'\left(\widehat{\boldsymbol{H}}_i^{-1}\right)'\left(\widehat{\boldsymbol{K}}_{it}-\widehat{\boldsymbol{H}}_i'\boldsymbol{K}_{it}\right)\\&+\boldsymbol{\Xi}^{r_0,-1}\boldsymbol{L}^{r_0\prime}\sum_{i=1}^{R}\widehat{\mathcal{R}}_i\widehat{\boldsymbol{H}}_i^{-1}\boldsymbol{Q}_i^{r_0}\boldsymbol{Q}_i^{r_0\prime}\left(\widehat{\boldsymbol{H}}_i^{-1}\right)'\left(\widehat{\boldsymbol{K}}_{it}-\widehat{\boldsymbol{H}}_i'\boldsymbol{K}_{it}\right)+O_p\left(\frac{1}{C_{\underline{N}T}^2}\right)\end{aligned}$$

Therefore, for Lemma 9.1, we have:

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left[\widehat{\boldsymbol{G}}_t-\left(\mathbb{H}'+\mathbb{B}'\right)\boldsymbol{G}_t\right]\boldsymbol{K}_{it}'=&\,\mathbb{H}'\frac{1}{R}\sum_{m=1}^{R}\mathbb{I}_m'\left(\widehat{\boldsymbol{H}}_m^{-1}\right)'\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\boldsymbol{K}}_{mt}-\widehat{\boldsymbol{H}}_m'\boldsymbol{K}_{mt}\right)\boldsymbol{K}_{it}'\\&+\boldsymbol{\Xi}^{r_0,-1}\boldsymbol{L}^{r_0\prime}\sum_{m=1}^{R}\widehat{\mathcal{R}}_m\widehat{\boldsymbol{H}}_m^{-1}\boldsymbol{Q}_m^{r_0}\boldsymbol{Q}_m^{r_0\prime}\left(\widehat{\boldsymbol{H}}_m^{-1}\right)'\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\boldsymbol{K}}_{mt}-\widehat{\boldsymbol{H}}_m'\boldsymbol{K}_{mt}\right)\boldsymbol{K}_{it}'+O_p\left(\frac{1}{C_{\underline{N}T}^2}\right)\end{aligned}$$

By Lemma 6.2, $T^{-1}\sum_{t=1}^{T}\left(\widehat{\boldsymbol{K}}_{mt}-\widehat{\boldsymbol{H}}_m'\boldsymbol{K}_{mt}\right)\boldsymbol{K}_{it}'=O_p\left(C_{N_mT}^{-2}\right)$. Then, the required result follows.

We can prove Lemma 9.2 along similar arguments using Lemma 6.5.

Q.E.D

Proof of Theorem 5
For each i and j, we have $\widehat{\boldsymbol{\gamma}}_{ij}=T^{-1}\widehat{\boldsymbol{G}}'\boldsymbol{Y}_{ij}$. Using (5) and $\boldsymbol{G}=\boldsymbol{G}-\widehat{\boldsymbol{G}}\left(\mathbb{H}+\mathbb{B}\right)^{-1}+\widehat{\boldsymbol{G}}\left(\mathbb{H}+\mathbb{B}\right)^{-1}$, we have

$$\widehat{\boldsymbol{\gamma}}_{ij}-\left(\mathbb{H}+\mathbb{B}\right)^{-1}\boldsymbol{\gamma}_{ij}=\frac{1}{T}\widehat{\boldsymbol{G}}'\left[\boldsymbol{G}-\widehat{\boldsymbol{G}}\left(\mathbb{H}+\mathbb{B}\right)^{-1}\right]\boldsymbol{\gamma}_{ij}+\frac{1}{T}\widehat{\boldsymbol{G}}'\boldsymbol{F}_i\boldsymbol{\lambda}_{ij}+\frac{1}{T}\widehat{\boldsymbol{G}}'\boldsymbol{e}_{ij}$$

Using $\widehat{\boldsymbol{G}}=\widehat{\boldsymbol{G}}-\boldsymbol{G}\left(\mathbb{H}+\mathbb{B}\right)+\boldsymbol{G}\left(\mathbb{H}+\mathbb{B}\right)$, the above equation can be written as

$$\begin{aligned}\widehat{\boldsymbol{\gamma}}_{ij}-\left(\mathbb{H}+\mathbb{B}\right)^{-1}\boldsymbol{\gamma}_{ij}=&\,\frac{1}{T}\left[\widehat{\boldsymbol{G}}-\boldsymbol{G}\left(\mathbb{H}+\mathbb{B}\right)\right]'\left[\boldsymbol{G}-\widehat{\boldsymbol{G}}\left(\mathbb{H}+\mathbb{B}\right)^{-1}\right]\boldsymbol{\gamma}_{ij}\\&+\frac{1}{T}\left(\mathbb{H}+\mathbb{B}\right)'\boldsymbol{G}'\left[\boldsymbol{G}-\widehat{\boldsymbol{G}}\left(\mathbb{H}+\mathbb{B}\right)^{-1}\right]\boldsymbol{\gamma}_{ij}+\frac{1}{T}\left[\widehat{\boldsymbol{G}}-\boldsymbol{G}\left(\mathbb{H}+\mathbb{B}\right)\right]'\boldsymbol{F}_i\boldsymbol{\lambda}_{ij}\\&+\frac{1}{T}\left[\widehat{\boldsymbol{G}}-\boldsymbol{G}\left(\mathbb{H}+\mathbb{B}\right)\right]'\boldsymbol{e}_{ij}+\frac{1}{T}\left(\mathbb{H}+\mathbb{B}\right)'\boldsymbol{G}'\boldsymbol{F}_i\boldsymbol{\lambda}_{ij}+\frac{1}{T}\left(\mathbb{H}+\mathbb{B}\right)'\boldsymbol{G}'\boldsymbol{e}_{ij}\end{aligned}$$

The first term is bounded by $O_p\left(\underline{N}^{-1}\right)$ by Theorem 4. The second to fourth terms are $O_p\left(C_{\underline{N}T}^{-2}\right)$ by Lemma 9. Then, we obtain:

$$\widehat{\boldsymbol{\gamma}}_{ij} - (\mathbb{H}+\mathbb{B})^{-1}\boldsymbol{\gamma}_{ij} = \frac{1}{T}(\mathbb{H}+\mathbb{B})'\boldsymbol{G}'(\boldsymbol{F}_i\boldsymbol{\lambda}_{ij}+\boldsymbol{e}_{ij}) + O_p\left(\frac{1}{C_{\underline{N}T}^2}\right)$$

Multiplying both sides by $\sqrt{T}$, we have:

$$\sqrt{T}\left[\widehat{\boldsymbol{\gamma}}_{ij} - (\mathbb{H}+\mathbb{B})^{-1}\boldsymbol{\gamma}_{ij}\right] = \mathbb{H}'\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{G}_t\left(\boldsymbol{\lambda}_{ij}'\boldsymbol{F}_{it}+e_{ijt}\right) + o_p(1) \xrightarrow{d} N\left(\boldsymbol{0}, \mathbb{H}'\mathbb{D}_{ij}^{(2)}\mathbb{H}\right)$$

using Assumption G.4 and the fact that $\mathbb{B} = O_p\left(\underline{N}^{-1/2}\right)$.

Q.E.D

Lemma 10. *Under Assumptions A-G, as $N_i, T \to \infty$, we have for each i, j and t:*

$$\widehat{S}_{ijt} = -\left[\widehat{\boldsymbol{\gamma}}_{ij} - (\mathbb{H}+\mathbb{B})^{-1}\boldsymbol{\gamma}_{ij}\right]'\left(\widehat{\boldsymbol{G}}_t - (\mathbb{H}'+\mathbb{B}')\boldsymbol{G}_t\right) - \boldsymbol{\gamma}_{ij}'\left[(\mathbb{H}+\mathbb{B})^{-1}\right]'\left(\widehat{\boldsymbol{G}}_t - (\mathbb{H}'+\mathbb{B}')\boldsymbol{G}_t\right)$$
$$- \boldsymbol{G}_t'(\mathbb{H}+\mathbb{B})\left[\widehat{\boldsymbol{\gamma}}_{ij} - (\mathbb{H}+\mathbb{B})^{-1}\boldsymbol{\gamma}_{ij}\right] = O_p\left(\frac{1}{\sqrt{\underline{N}}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

where $\widehat{S}_{ijt}$ is the (t,j) element of $\widehat{\boldsymbol{S}}_i = \boldsymbol{G}\boldsymbol{\Gamma}_i' - \widehat{\boldsymbol{G}}\widehat{\boldsymbol{\Gamma}}_i'$.

Proof.
Using the expansions $\widehat{\boldsymbol{G}} = \widehat{\boldsymbol{G}} - \boldsymbol{G}(\mathbb{H}+\mathbb{B}) + \boldsymbol{G}(\mathbb{H}+\mathbb{B})$ and $\widehat{\boldsymbol{\Gamma}}_i' = \widehat{\boldsymbol{\Gamma}}_i' - (\mathbb{H}+\mathbb{B})^{-1}\boldsymbol{\Gamma}_i' + (\mathbb{H}+\mathbb{B})^{-1}\boldsymbol{\Gamma}_i'$, the result follows from Theorems 4 and 5.

Q.E.D

Lemma 11. *Under Assumptions A–G, for each i, as $N_i, T \to \infty$, we have:*

1.
$$\frac{1}{T}\widehat{\boldsymbol{F}}_i'\left(\frac{1}{N_iT}\widehat{\boldsymbol{Y}}_i\widehat{\boldsymbol{Y}}_i'\right)\widehat{\boldsymbol{F}}_i = \widehat{\boldsymbol{\Upsilon}}_i \xrightarrow{p} \boldsymbol{\Upsilon}_i$$

where $\widehat{\boldsymbol{Y}}_i = \boldsymbol{Y}_i - \widehat{\boldsymbol{G}}\widehat{\boldsymbol{\Gamma}}_i'$ and $\boldsymbol{\Upsilon}_i$ is a diagonal matrix consisting of the eigenvalues of $\boldsymbol{\Sigma}_{\Lambda_i}\boldsymbol{\Sigma}_{F_i}$.

2.
$$\frac{\widehat{\boldsymbol{F}}_i'\boldsymbol{F}_i}{T}\left(\frac{\boldsymbol{\Lambda}_i'\boldsymbol{\Lambda}_i}{N_i}\right)\frac{\boldsymbol{F}_i'\widehat{\boldsymbol{F}}_i}{T} \xrightarrow{p} \boldsymbol{\Upsilon}_i$$

3.
$$plim_{N_i,T\to\infty}\frac{\widehat{\boldsymbol{F}}_i'\boldsymbol{F}_i}{T} = \mathbb{W}_i$$

The $r_i \times r_i$ matrix $\mathbb{W}_i$ is given by $\mathbb{W}_i = \boldsymbol{\Upsilon}_i^{1/2}\boldsymbol{\mathcal{L}}_i'\boldsymbol{\Sigma}_{\Lambda_i}^{-1/2}$ and invertible, where $\boldsymbol{\Upsilon}_i$ is also an $r_i \times r_i$ diagonal matrix consisting of the eigenvalues of $\boldsymbol{\Sigma}_{\Lambda_i}^{1/2}\boldsymbol{\Sigma}_{F_i}\boldsymbol{\Sigma}_{\Lambda_i}^{1/2}$, and $\boldsymbol{\mathcal{L}}_i$ is the corresponding eigenvector matrix such that $\boldsymbol{\mathcal{L}}_i'\boldsymbol{\mathcal{L}}_i/T = \boldsymbol{I}_{r_i}$.

4.

$$plim_{N_i,T\to\infty}\widehat{\mathscr{H}_i} = \mathscr{H}_i$$

where $\mathscr{H}_i = \mathbf{\Sigma}_{\Lambda i}\mathbb{W}_i'\mathbf{\Upsilon}_i^{-1} = \mathbb{W}_i^{-1}$.

Proof.
As $\widehat{\boldsymbol{S}}_i = o_p(1)$, the proof follows directly from Proposition 1 and Lemma A.3 in Bai (2003) with slight modification.

Q.E.D

Proof of Theorem 6.
By construction of PC, we have

$$\widehat{\boldsymbol{F}}_i = \frac{1}{N_iT}\left(\widehat{\boldsymbol{S}}_i\widehat{\boldsymbol{S}}_i' + \boldsymbol{F}_i\mathbf{\Lambda}_i'\widehat{\boldsymbol{S}}_i' + \boldsymbol{e}_i\widehat{\boldsymbol{S}}_i' + \widehat{\boldsymbol{S}}_i\mathbf{\Lambda}_i\boldsymbol{F}_i' + \boldsymbol{F}_i\mathbf{\Lambda}_i'\mathbf{\Lambda}_i\boldsymbol{F}_i' + \boldsymbol{e}_i\mathbf{\Lambda}_i\boldsymbol{F}_i' + \widehat{\boldsymbol{S}}_i\boldsymbol{e}_i' + \boldsymbol{F}_i\mathbf{\Lambda}_i'\boldsymbol{e}_i' + \boldsymbol{e}_i\boldsymbol{e}_i'\right)\widehat{\boldsymbol{F}}_i\widehat{\mathbf{\Upsilon}}^{-1}$$

where $\widehat{\boldsymbol{S}}_i = \boldsymbol{G}\mathbf{\Gamma}_i' - \widehat{\boldsymbol{G}}\widehat{\mathbf{\Gamma}}_i'$. Therefore, we have

$$\begin{aligned}
&\widehat{\boldsymbol{F}}_{it} - \widehat{\mathscr{H}_i}'\boldsymbol{F}_{it} = \\
&\widehat{\mathbf{\Upsilon}}_i^{-1}\frac{1}{N_iT}\left(\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\widehat{\boldsymbol{S}}_{i.s}'\widehat{\boldsymbol{S}}_{i.t} + \sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\widehat{\boldsymbol{S}}_{i.s}'\mathbf{\Lambda}_i\boldsymbol{F}_{it} + \sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\widehat{\boldsymbol{S}}_{i.s}'\boldsymbol{e}_{i.t} + \sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\boldsymbol{F}_{is}'\mathbf{\Lambda}_i'\widehat{\boldsymbol{S}}_{i.t} + \sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\boldsymbol{e}_{i.s}'\widehat{\boldsymbol{S}}_{i.t}\right) \\
&\qquad + \widehat{\mathbf{\Upsilon}}_i^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\omega_i(s,t) + \frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\zeta_{i,st} + \frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\eta_{i,st}^* + \frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\mu_{i,st}^*\right) \quad (32)
\end{aligned}$$

where $\widehat{\mathscr{H}_i} = (\mathbf{\Lambda}_i'\mathbf{\Lambda}_i/N_i)(\boldsymbol{F}_i'\boldsymbol{F}_i/T)\widehat{\mathbf{\Upsilon}}_i^{-1}$, $\widehat{\boldsymbol{S}}_{i.t}$ is the $N_i\times 1$ vector of $\widehat{\boldsymbol{S}}_i$ (the t-th row vector), $\eta_{i,st}^* = N_i^{-1}\boldsymbol{F}_{is}'\mathbf{\Lambda}_i'\boldsymbol{e}_{i.t}$ and $\mu_{i,st}^* = N_i^{-1}\boldsymbol{F}_{it}'\mathbf{\Lambda}_i'\boldsymbol{e}_{i.s}$. $\omega_i(s,t)$ and $\zeta_{i,st}$ are defined in Lemma 6.1.

To analyse the first part of (32), we let

$$\begin{aligned}
\frac{1}{N_iT}\left(\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\widehat{\boldsymbol{S}}_{i.s}'\widehat{\boldsymbol{S}}_{i.t} + \sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\widehat{\boldsymbol{S}}_{i.s}'\widehat{\mathbf{\Lambda}}_i\boldsymbol{F}_{it} + \sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\widehat{\boldsymbol{S}}_{i.s}'\boldsymbol{e}_{i.t} + \sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\boldsymbol{F}_{is}'\mathbf{\Lambda}_i'\widehat{\boldsymbol{S}}_{i.t} + \sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\boldsymbol{e}_{i.s}'\widehat{\boldsymbol{S}}_{i.t}\right) = \\
\mathscr{X}1 + \mathscr{X}2 + \mathscr{X}3 + \mathscr{X}4 + \mathscr{X}5.
\end{aligned}$$

Using $\widehat{\boldsymbol{F}}_{is} = \widehat{\boldsymbol{F}}_{is} - \widehat{\mathscr{H}_i}'\boldsymbol{F}_{is} + \widehat{\mathscr{H}_i}'\boldsymbol{F}_{is}$ and by Theorem 1.2, we obtain:

$$\mathscr{X}1 = \frac{1}{N_iT}\sum_{s=1}^{T}\sum_{j=1}^{N_i}\left(\widehat{\boldsymbol{F}}_{is} - \widehat{\mathscr{H}_i}'\boldsymbol{F}_{is}\right)\widehat{S}_{ijs}\widehat{S}_{ijt} + \widehat{\mathscr{H}_i}'\frac{1}{N_iT}\sum_{s=1}^{T}\sum_{j=1}^{N_i}\boldsymbol{F}_{is}\widehat{S}_{ijs}\widehat{S}_{ijt} = O_p\left(\frac{1}{C_{\underline{N}T}^2}\right)$$

Similarly,

$$\mathscr{X}2 = \frac{1}{N_iT}\sum_{s=1}^{T}\sum_{j=1}^{N_i}\left(\widehat{\boldsymbol{F}}_{is} - \widehat{\mathscr{H}_i}'\boldsymbol{F}_{is}\right)\widehat{S}_{ijs}\boldsymbol{\lambda}_{ij}'\boldsymbol{F}_{it} + \widehat{\mathscr{H}_i}'\frac{1}{N_iT}\sum_{s=1}^{T}\sum_{j=1}^{N_i}\boldsymbol{F}_{is}\widehat{S}_{ijs}\boldsymbol{\lambda}_{ij}'\boldsymbol{F}_{it}$$

The first term is $O_p\left(C_{\underline{N}T}^{-2}\right)$ by Theorem 2.1 and Lemma 10. Using Lemma 10, we can express the second term as

$$\widehat{\mathscr{H}}_i' \frac{1}{N_i T} \sum_{s=1}^{T} \sum_{j=1}^{N_i} \boldsymbol{F}_{is} \widehat{S}_{ijs} \boldsymbol{\lambda}_{ij}' \boldsymbol{F}_{it} =$$

$$- \widehat{\mathscr{H}}_i' \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{1}{T} \sum_{s=1}^{T} \boldsymbol{F}_{is} \left(\widehat{\boldsymbol{G}}_t - (\mathbb{H}' + \mathbb{B}') \boldsymbol{G}_t \right)' \left[\widehat{\boldsymbol{\gamma}}_{ij} - (\mathbb{H} + \mathbb{B})^{-1} \boldsymbol{\gamma}_{ij} \right] \boldsymbol{\lambda}_{ij}' \boldsymbol{F}_{it}$$

$$- \widehat{\mathscr{H}}_i' \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{1}{T} \sum_{s=1}^{T} \boldsymbol{F}_{is} \left(\widehat{\boldsymbol{G}}_t - (\mathbb{H}' + \mathbb{B}') \boldsymbol{G}_t \right)' (\mathbb{H} + \mathbb{B})^{-1} \boldsymbol{\gamma}_{ij} \boldsymbol{\lambda}_{ij}' \boldsymbol{F}_{it}$$

$$- \widehat{\mathscr{H}}_i' \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{1}{T} \sum_{s=1}^{T} \boldsymbol{F}_{is} \boldsymbol{G}_t' (\mathbb{H} + \mathbb{B})' \left[\widehat{\boldsymbol{\gamma}}_{ij} - (\mathbb{H} + \mathbb{B})^{-1} \boldsymbol{\gamma}_{ij} \right] \boldsymbol{\lambda}_{ij}' \boldsymbol{F}_{it}$$

The first term of the above expression is $O_p\left(C_{\underline{N}T}^{-2}\right)\left[O_p\left(T^{-1/2}\right) + O_p\left(\underline{N}^{-1}\right)\right]$ by Lemma 9.1 and Theorem 5. The second term is $O_p\left(C_{\underline{N}T}^{-2}\right)$ by Lemma 9.1 while the last term is $O_p\left(T^{-1/2}\right)\left[O_p\left(T^{-1/2}\right) + O_p\left(\underline{N}^{-1}\right)\right]$ by Assumption D. Therefore, we obtain: $\mathscr{X}2 = O_p\left(C_{\underline{N}T}^{-2}\right)$. Using $\widehat{\boldsymbol{F}}_{is} = \widehat{\boldsymbol{F}}_{is} - \widehat{\mathscr{H}}_i' \boldsymbol{F}_{is} + \widehat{\mathscr{H}}_i' \boldsymbol{F}_{is}$, we have:

$$\mathscr{X}3 = \frac{1}{N_i T} \sum_{s=1}^{T} \sum_{j=1}^{N_i} \left(\widehat{\boldsymbol{F}}_{is} - \widehat{\mathscr{H}}_i' \boldsymbol{F}_{is} \right) \widehat{S}_{ijs} e_{ijt} + \widehat{\mathscr{H}}_i' \frac{1}{N_i T} \sum_{s=1}^{T} \sum_{j=1}^{N_i} \boldsymbol{F}_{is} \widehat{S}_{ijs} e_{ijt}$$

The first term is bounded by $O_p\left(C_{\underline{N}T}^{-2}\right)$ by Theorem 2.1 and Lemma 10. The second term can be written as

$$\widehat{\mathscr{H}}_i' \frac{1}{N_i T} \sum_{s=1}^{T} \sum_{j=1}^{N_i} \boldsymbol{F}_{is} \widehat{S}_{ijs} e_{ijt} =$$

$$- \widehat{\mathscr{H}}_i' \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{1}{T} \sum_{s=1}^{T} \boldsymbol{F}_{is} \left(\widehat{\boldsymbol{G}}_t - (\mathbb{H}' + \mathbb{B}') \boldsymbol{G}_t \right)' \left[\widehat{\boldsymbol{\gamma}}_{ij} - (\mathbb{H} + \mathbb{B})^{-1} \boldsymbol{\gamma}_{ij} \right] e_{ijt}$$

$$- \widehat{\mathscr{H}}_i' \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{1}{T} \sum_{s=1}^{T} \boldsymbol{F}_{is} \left(\widehat{\boldsymbol{G}}_t - (\mathbb{H}' + \mathbb{B}') \boldsymbol{G}_t \right)' (\mathbb{H} + \mathbb{B})^{-1} \boldsymbol{\gamma}_{ij} e_{ijt}$$

$$- \widehat{\mathscr{H}}_i' \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{1}{T} \sum_{s=1}^{T} \boldsymbol{F}_{is} \boldsymbol{G}_t' (\mathbb{H} + \mathbb{B})' \left[\widehat{\boldsymbol{\gamma}}_{ij} - (\mathbb{H} + \mathbb{B})^{-1} \boldsymbol{\gamma}_{ij} \right] e_{ijt}$$

The first term of the above equation is $O_p\left(C_{\underline{N}T}^{-2}\right)\left[O_p\left(T^{-1/2}\right) + O_p\left(\underline{N}^{-1}\right)\right]$ by Lemma 9.1 and Theorem 5. The second term is $O_p\left(C_{\underline{N}T}^{-2}\right)$ by Lemma 9.1 and the last term is $O_p\left(T^{-1/2}\right)\left[O_p\left(T^{-1/2}\right) + O_p\left(\underline{N}^{-1}\right)\right]$

by Assumption D. Collecting these terms, we have $\mathscr{X}3 = O_p\left(C_{\underline{N}T}^{-2}\right)$. Next, consider

$$\mathscr{X}5 = \frac{1}{N_iT}\sum_{s=1}^{T}\sum_{j=1}^{N_i}\left(\widehat{\boldsymbol{F}}_{is} - \widehat{\mathscr{H}}_i'\boldsymbol{F}_{is}\right)e_{ijs}\widehat{S}_{ijt} + \widehat{\mathscr{H}}_i'\frac{1}{N_iT}\sum_{s=1}^{T}\sum_{j=1}^{N_i}\boldsymbol{F}_{is}e_{ijs}\widehat{S}_{ijt}$$

The first term of the above equation is of order $O_p\left(C_{\underline{N}T}^{-2}\right)$ by Theorem 2.1 and Lemma 10. For the second term, we have:

$$\|\mathscr{X}5\| \leq \left\|\widehat{\mathscr{H}}_i\right\|\frac{1}{\sqrt{T}}\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\boldsymbol{F}_{is}e_{ijs}\right\|^2\right)^{-1/2}\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\left|\widehat{S}_{ijt}\right|^2\right)^{-1/2} = O_p\left(\frac{1}{\sqrt{T}C_{\underline{N}T}}\right)$$

where the last equality follows from Assumption B2 and Lemma 10.

Collecting the results above, (32) becomes

$$\begin{aligned}\widehat{\boldsymbol{F}}_{it} - \widehat{\mathscr{H}}_i'\boldsymbol{F}_{it} &= \widehat{\boldsymbol{\Upsilon}}_i^{-1}\frac{1}{N_iT}\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\boldsymbol{F}_{is}'\boldsymbol{\Lambda}_i'\widehat{\boldsymbol{S}}_{i.t} \\ &\quad + \widehat{\boldsymbol{\Upsilon}}_i^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\omega_i(s,t) + \frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\zeta_{i,st} + \frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\eta_{i,st}^* + \frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\mu_{i,st}^*\right) + O_p\left(\frac{1}{C_{\underline{N}T}^2}\right)\end{aligned}$$

It then follows that

$$\widehat{\boldsymbol{F}}_{it} - \widehat{\mathscr{H}}_i'\boldsymbol{F}_{it} = \widehat{\boldsymbol{\Upsilon}}_i^{-1}\frac{1}{N_iT}\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\boldsymbol{F}_{is}'\boldsymbol{\Lambda}_i'\widehat{\boldsymbol{S}}_{i.t} + \widehat{\boldsymbol{\Upsilon}}_i^{-1}\frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\eta_{i,st}^* + O_p\left(\frac{1}{C_{\underline{N}T}^2}\right)$$

Then, the proof is the same as that of Lemma 6.1. Let $\boldsymbol{\mathcal{B}}_{it}$ be the bias term given by

$$\boldsymbol{\mathcal{B}}_{it} = \widehat{\boldsymbol{\Upsilon}}_i^{-1}\frac{1}{N_iT}\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\boldsymbol{F}_{is}'\boldsymbol{\Lambda}_i'\widehat{\boldsymbol{S}}_{i.t}$$

Under Assumption G3, it follows that

$$\begin{aligned}\sqrt{N_i}\left(\widehat{\boldsymbol{F}}_{it} - \widehat{\mathscr{H}}_i'\boldsymbol{F}_{it} - \boldsymbol{\mathcal{B}}_{it}\right) &= \widehat{\boldsymbol{\Upsilon}}_i^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\boldsymbol{F}}_{is}\boldsymbol{F}_{is}'\right)\frac{1}{\sqrt{N_i}}\sum_{j=1}^{N_i}\boldsymbol{\lambda}_{ij}\boldsymbol{e}_{ijt} + o_p(1) \\ &\xrightarrow{d} N\left(\boldsymbol{0}, \boldsymbol{\Upsilon}_i^{-1}\mathbb{W}_i\mathbb{D}_{ii,t}^{(3)}\mathbb{W}_i'\boldsymbol{\Upsilon}_i^{-1}\right)\end{aligned}$$

Q.E.D

Lemma 12. *Under the assumptions in Theorem 6, we have for each i and j:*

1.
$$\frac{1}{T}\left(\widehat{\boldsymbol{F}}_i - \boldsymbol{F}_i\widehat{\mathscr{H}}_i\right)'\boldsymbol{F}_i = O_p\left(\frac{1}{C_{\underline{N}T}^2}\right)$$

2.

$$\frac{1}{T}\left(\widehat{\boldsymbol{F}}_i - \boldsymbol{F}_i\widehat{\mathscr{H}_i}\right)'\boldsymbol{e}_{ij} = O_p\left(\frac{1}{C^2_{\underline{N}T}}\right)$$

3.

$$\frac{1}{T}\left(\widehat{\boldsymbol{F}}_i - \boldsymbol{F}_i\widehat{\mathscr{H}_i}\right)'\widehat{\boldsymbol{S}}_{ij} = O_p\left(\frac{1}{C^2_{\underline{N}T}}\right)$$

where $\widehat{\boldsymbol{S}}_{ij} = \boldsymbol{G}\boldsymbol{\gamma}_{ij} - \widehat{\boldsymbol{G}}\widehat{\boldsymbol{\gamma}}_{ij}$.

Proof.
1. Using (32), we have:

$$\frac{1}{T}\left(\widehat{\boldsymbol{F}}_i - \boldsymbol{F}_i\widehat{\mathscr{H}_i}\right)'\boldsymbol{F}_i = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\boldsymbol{F}}_{it} - \widehat{\mathscr{H}_i}'\boldsymbol{F}_{it}\right)\boldsymbol{F}'_{it} =$$
$$\widehat{\boldsymbol{\Upsilon}}_i^{-1}\frac{1}{N_iT^2}\left(\sum_{s=1}^{T}\sum_{t=1}^{T}\widehat{\boldsymbol{F}}_{is}\widehat{\boldsymbol{S}}'_{i.s}\widehat{\boldsymbol{S}}_{i.t}\boldsymbol{F}'_{it} + \sum_{s=1}^{T}\sum_{t=1}^{T}\widehat{\boldsymbol{F}}_{is}\widehat{\boldsymbol{S}}'_{i.s}\boldsymbol{\Lambda}_i\boldsymbol{F}_{it}\boldsymbol{F}'_{it}\right.$$
$$\left.+\sum_{s=1}^{T}\sum_{t=1}^{T}\widehat{\boldsymbol{F}}_{is}\widehat{\boldsymbol{S}}'_{i.s}\boldsymbol{e}_{i.t}\boldsymbol{F}'_{it} + \sum_{s=1}^{T}\sum_{t=1}^{T}\widehat{\boldsymbol{F}}_{is}\boldsymbol{F}'_{is}\boldsymbol{\Lambda}'_i\widehat{\boldsymbol{S}}_{i.t}\boldsymbol{F}'_{it} + \sum_{s=1}^{T}\sum_{t=1}^{T}\widehat{\boldsymbol{F}}_{is}\boldsymbol{e}'_{i.s}\widehat{\boldsymbol{S}}_{i.t}\boldsymbol{F}'_{it}\right)$$
$$+\widehat{\boldsymbol{\Upsilon}}_i^{-1}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\widehat{\boldsymbol{F}}_{is}\omega_i(s,t)\boldsymbol{F}'_{it} + \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\widehat{\boldsymbol{F}}_{is}\zeta_{i,st}\boldsymbol{F}'_{it} + \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\widehat{\boldsymbol{F}}_{is}\eta^*_{i,st}\boldsymbol{F}'_{it}\right.$$
$$\left.+\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\widehat{\boldsymbol{F}}_{is}\mu^*_{i,st}\boldsymbol{F}'_{it}\right)$$

By Lemma 11, we have $\widehat{\boldsymbol{\Upsilon}}_i = O_p(1)$. The second part of the above equation is of order $O_p\left(C^{-2}_{\underline{N}T}\right)$. The proof is the same as that of Lemma 6.3 and therefore is not repeated here.

We focus on the first part, which can be written as $\widehat{\boldsymbol{\Upsilon}}_i^{-1}\left(\mathcal{Q}1 + \mathcal{Q}2 + \mathcal{Q}3 + \mathcal{Q}4 + \mathcal{Q}5\right)$. As a result of Lemma 10, $\mathcal{Q}1 = O_p\left(C^{-2}_{\underline{N}T}\right)$. Using $\widehat{\boldsymbol{F}}_{is} = \widehat{\boldsymbol{F}}_{is} - \widehat{\mathscr{H}_i}'\boldsymbol{F}_{is} + \widehat{\mathscr{H}_i}'\boldsymbol{F}_{is}$, we have:

$$\mathcal{Q}2 = \frac{1}{N_iT}\sum_{j=1}^{N_i}\sum_{s=1}^{T}\left(\widehat{\boldsymbol{F}}_{is} - \widehat{\mathscr{H}_i}'\boldsymbol{F}_{is}\right)\widehat{S}_{ijs}\boldsymbol{\lambda}'_{ij}\left(\frac{\boldsymbol{F}'_i\boldsymbol{F}_i}{T}\right) + \widehat{\mathscr{H}_i}'\frac{1}{N_iT}\sum_{j=1}^{N_i}\sum_{s=1}^{T}\boldsymbol{F}_{is}\widehat{S}_{ijs}\boldsymbol{\lambda}'_{ij}\left(\frac{\boldsymbol{F}'_i\boldsymbol{F}_i}{T}\right)$$

Note that $\boldsymbol{F}'_i\boldsymbol{F}_i/T = O_p(1)$ by Assumption B1. The first term is $O_p\left(C^{-2}_{\underline{N}T}\right)$ by Theorem 6 and Lemma 10. Combining Lemmas 9 and 10, Theorems 4 and 5, and Assumption D, we have: $T^{-1}\sum_{s=1}^{T}\boldsymbol{F}_{is}\widehat{S}_{ijs} = O_p\left(C^{-2}_{\underline{N}T}\right)$, so the second term is also $O_p\left(C^{-2}_{\underline{N}T}\right)$. We then obtain $\mathcal{Q}2 = \left(C^{-2}_{\underline{N}T}\right)$. Along similar arguments, it is easily seen that $\mathcal{Q}3$ to $\mathcal{Q}5$ have stochastic order $O_p\left(C^{-2}_{\underline{N}T}\right)$.
2. The proof is similar to part 1 of the lemma and therefore omitted.
3. The result follows from Theorem 6 and Lemma 10.

Q.E.D

Proof of Theorem 7.
Using $\widehat{\boldsymbol{\lambda}}_i = \widehat{\boldsymbol{F}}_i'\widehat{\boldsymbol{Y}}_{ij}$, $\widehat{\boldsymbol{Y}}_{ij} = \widehat{\boldsymbol{S}}_{ij} + \boldsymbol{F}_i\lambda_{ij} + \boldsymbol{e}_{ij}$ and $\boldsymbol{F}_i = \boldsymbol{F}_i - \widehat{\boldsymbol{F}}_i\widehat{\mathscr{H}_i}^{-1} + \widehat{\boldsymbol{F}}_i\widehat{\mathscr{H}_i}^{-1}$, we obtain:

$$\widehat{\boldsymbol{\lambda}}_{ij} - \widehat{\mathscr{H}_i}^{-1}\boldsymbol{\lambda}_{ij} = \frac{1}{T}\widehat{\boldsymbol{F}}_i'\left(\boldsymbol{F}_i - \widehat{\boldsymbol{F}}_i\widehat{\mathscr{H}_i}^{-1}\right)\boldsymbol{\lambda}_{ij} + \frac{1}{T}\widehat{\boldsymbol{F}}_i'\boldsymbol{e}_{ij} + \frac{1}{T}\widehat{\boldsymbol{F}}_i'\widehat{\boldsymbol{S}}_{ij}$$

Replacing $\widehat{\boldsymbol{F}}_i$ by $\widehat{\boldsymbol{F}}_i - \boldsymbol{F}_i\widehat{\mathscr{H}_i} + \boldsymbol{F}_i\widehat{\mathscr{H}_i}$, we get:

$$\begin{aligned}\widehat{\boldsymbol{\lambda}}_{ij} - \widehat{\mathscr{H}_i}^{-1}\boldsymbol{\lambda}_{ij} = &\frac{1}{T}\left(\widehat{\boldsymbol{F}}_i - \boldsymbol{F}_i\widehat{\mathscr{H}_i}\right)'\left(\boldsymbol{F}_i - \widehat{\boldsymbol{F}}_i\widehat{\mathscr{H}_i}\right)\boldsymbol{\lambda}_{ij} + \widehat{\mathscr{H}_i}'\frac{1}{T}\boldsymbol{F}_i'\left(\boldsymbol{F}_i - \widehat{\boldsymbol{F}}_i\widehat{\mathscr{H}_i}\right)\\ &+ \frac{1}{T}\left(\widehat{\boldsymbol{F}}_i - \boldsymbol{F}_i\widehat{\mathscr{H}_i}\right)'\boldsymbol{e}_{ij} + \frac{1}{T}\left(\widehat{\boldsymbol{F}}_i - \boldsymbol{F}_i\widehat{\mathscr{H}_i}\right)'\widehat{\boldsymbol{S}}_{ij} + \widehat{\mathscr{H}_i}'\frac{1}{T}\boldsymbol{F}_i'\widehat{\boldsymbol{S}}_{ij} + \widehat{\mathscr{H}_i}'\frac{1}{T}\boldsymbol{F}_i'\boldsymbol{e}_{ij}\end{aligned}$$

Then, by Theorem 6, Lemma 12 and Assumption D, it follows that

$$\widehat{\boldsymbol{\lambda}}_{ij} - \widehat{\mathscr{H}_i}^{-1}\boldsymbol{\lambda}_{ij} = \widehat{\mathscr{H}_i}'\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{F}_{it}\widehat{S}_{ijt} + \widehat{\mathscr{H}_i}'\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{F}_{it}e_{ijt} + O_p\left(\frac{1}{C^2_{\underline{N}T}}\right)$$

Let $\mathscr{B}_{ij}$ be the bias term given by

$$\mathscr{B}_{ij} = \widehat{\mathscr{H}_i}'\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{F}_{it}\widehat{S}_{ijt}.$$

By Lemma 10.4 and Assumption G.4, we finally obtain

$$\sqrt{T}\left(\widehat{\boldsymbol{\lambda}}_{ij} - \widehat{\mathscr{H}_i}^{-1}\boldsymbol{\lambda}_{ij} - \mathscr{B}_{ij}\right) = \widehat{\mathscr{H}_i}'\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{F}_{it}e_{ijt} + o_p(1) \xrightarrow{d} N\left(\boldsymbol{0}, \left(\mathbb{W}_i^{-1}\right)'\mathbb{D}_{ij}^{(3)}\mathbb{W}_i^{-1}\right)$$

Q.E.D

B Bootstrap confidence intervals for the global factors and loadings

We outline the bootstrap procedure for constructing consistent confidence intervals for the estimates of global factors and loadings. Although their asymptotic distributions are well-established, they are not readily applicable in practice. The asymptotic covariance matrices derived in Theorems 4 and 5 are subject to the rotation matrix $\mathbb{H}$, which is unknown and cannot be estimated. Moreover, we cannot use bootstrap to consistently estimate the variances, because the bootstrap version of the rotation matrix $\mathbb{H}^{*(b)}$ varies in each replication b.

It is still possible to construct valid CIs for the global factors and loadings since $\mathbb{H}^{*(b)}$ can be replaced by known quantities in the bootstrap world. The back-rotated bootstrap factors and loadings have the same asymptotic covariance matrices over all replications $b = 1, \ldots, B$, as shown in (33) and (35). This enables us to construct CIs based on the percentile estimates. For simplicity we assume that the error terms are cross-sectionally and serially uncorrelated.[16] In Theorem 5, the asymptotic covariance matrix of $\widehat{\gamma}_{ij}$ depends on the time series variation of the local factors $\boldsymbol{F}_{it}$. Therefore, we should also bootstrap

[16] This is mainly because we make the algorithm computationally tractable.

the local factors in addition to the error term. This step will affect the bootstrap rotation matrix $\widehat{\boldsymbol{H}}_i^{*(b)}$ as well as the covariance matrix in Theorem 4, which contains the bootstrap version of $\widehat{\boldsymbol{K}}_i$, denoted $\widehat{\boldsymbol{K}}_i^*$. If the local factors are also bootstrapped, $\widehat{\boldsymbol{K}}_i^*$ is not consistent for $\widehat{\boldsymbol{K}}_i$, which results in different limiting distributions of $\widehat{\boldsymbol{G}}_t^{*(b)}$ across each repetition. Therefore, the bootstrapping for $\widehat{\boldsymbol{G}}_t$ and $\widehat{\boldsymbol{\gamma}}_{ij}$ should be done, separately.

We now outline the different bootstrap algorithms for for $\widehat{\boldsymbol{G}}_t$ and $\widehat{\boldsymbol{\gamma}}_{ij}$ for $b = 1, \ldots, B$.

Bootstrapping the global factors

1. For each i, j and t, construct $e_{ijt}^{*(b)} = \hat{e}_{ijt}\varepsilon_{ijt}^{*(b)}$ where $\hat{e}_{ijt} = y_{ijt} - \widehat{\boldsymbol{\gamma}}_{ij}'\widehat{\boldsymbol{G}}_t - \widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\boldsymbol{F}}_{it}$ and $\varepsilon_{ijt}^{*(b)} \sim$ i.i.d. $N(0,1)$.
2. Generate the re-sampled data by $y_{ijt}^{*(b)} = \widehat{\boldsymbol{\gamma}}_{ij}'\widehat{\boldsymbol{G}}_t + \widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\boldsymbol{F}}_{it} + e_{ijt}^{*(b)}$.
3. Apply the estimation procedure developed in Section 3 to the re-sampled data, and obtain the bootstrap estimates, denoted $\widehat{\boldsymbol{K}}_i^{*(b)}$ and $\widehat{\boldsymbol{G}}_t^{*(b)}$.
4. Repeat Steps 1–3 for B times.

The consistency and asymptotic normality of the *GCC* estimators for the re-sampled model are achieved since Step 1 does not change the validity of Assumptions A-G. In order to have consistent estimates of the bootstrap covariance matrices, we assume cross-section independence of the error terms e_{ijt}. For each $b = 1, \ldots, B$, we have

$$\sqrt{N}\left[\widehat{\boldsymbol{G}}_t^{*(b)} - \left(\mathbb{H}^{*(b)\prime} + \mathbb{B}^{*(b)\prime}\right)\widehat{\boldsymbol{G}}_t\right] = \frac{1}{R}\mathbb{H}^{*(b)\prime}\mathcal{I}'\widehat{\mathbb{C}}^*\mathbb{E}_t^{*(b)} + o_p(1)$$
$$\xrightarrow{d} N\left(\mathbf{0}, \frac{1}{R^2}\mathbb{H}^{*(b)\prime}\mathcal{I}'\widehat{\mathbb{C}}^*\mathbb{D}_t^{*,(1)}\widehat{\mathbb{C}}^{*\prime}\mathcal{I}\mathbb{H}^{*(b)}\right),$$

where $\mathbb{H}^{*(b)} = \boldsymbol{U}^{*(b)}$ with $\boldsymbol{U}^{*(b)} = T^{-1}\widehat{\boldsymbol{G}}^{*(b)\prime}\widehat{\boldsymbol{G}} + O_p\left(C_{\underline{N}T}^{-1}\right)$[17], and

$$\mathbb{B}^{*(b)} = \frac{1}{R}\sum_{i=1}^{R}\sqrt{\frac{1}{N_i}}\mathbb{I}_i'\left(\frac{\widehat{\boldsymbol{\Theta}}_i'\widehat{\boldsymbol{\Theta}}_i}{N_i}\right)^{-1}\frac{\widehat{\boldsymbol{\Theta}}_i'\widehat{\boldsymbol{e}}_i'}{\sqrt{N_i T}}\widehat{\boldsymbol{J}}^{r_0}\boldsymbol{U}^{*(b)}$$

with $\widehat{\boldsymbol{\Theta}}_i = T^{-1}\boldsymbol{Y}_i'\widehat{\boldsymbol{K}}_i$. Moreover, $\widehat{\mathbb{C}}^* = diag\left(\sqrt{\frac{N}{N_1}}\mathbb{I}_1'\left(\frac{\widehat{\boldsymbol{\Theta}}_1'\widehat{\boldsymbol{\Theta}}_1}{N_1}\right)^{-1}, \ldots, \sqrt{\frac{N}{N_R}}\mathbb{I}_R'\left(\frac{\widehat{\boldsymbol{\Theta}}_R'\widehat{\boldsymbol{\Theta}}_R}{N_R}\right)^{-1}\right)$ and $\mathbb{D}_t^*$ being a block diagonal matrix as

$$\mathbb{D}_t^{*,(1)} = \begin{bmatrix} \mathbb{D}_{11,t}^{*,(1)} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbb{D}_{22,t}^{*,(1)} & \ldots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbb{D}_{RR,t}^{*,(1)} \end{bmatrix}$$

[17] Using Theorem 4, we have $\mathbb{H}^{*(b)} = T^{-1/2}\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{J}}^{r_0}\boldsymbol{U}^{*(b)}$ where $\widehat{\boldsymbol{J}}^{r_0} = \widehat{\boldsymbol{L}}^{r_0}\left(\widehat{\boldsymbol{\Xi}}^{r_0}\right)^{-1}$ and $\widehat{\boldsymbol{\Xi}}^{r_0}$ is an $r_0 \times r_0$ diagonal matrix consisting of the r_0 non-zero eigenvalues of $T^{-1}\widehat{\boldsymbol{G}}\widehat{\boldsymbol{G}}'$. Because $T^{-1}\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{G}} = \boldsymbol{I}_{r_0}$, it follows that $\widehat{\boldsymbol{\Xi}}^{r_0} = \boldsymbol{I}_{r_0}$. Using $\widehat{\boldsymbol{L}}^{r_0} = T^{-1/2}\widehat{\boldsymbol{G}}$, it follows that $\mathbb{H}^{*(b)} = \boldsymbol{U}^{*(b)}$. Using Lemma 7, it is straightforward that $\boldsymbol{U}^{*(b)} = T^{-1}\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{G}}^{*(b)} + O_p\left(C_{\underline{N}T}^{-2}\right)$.

with

$$\mathbb{D}^{*,(1)}_{ii,t} = \text{plim}_{N_i \to \infty} \frac{1}{N_i} \sum_{j=1}^{N_i} \widehat{\boldsymbol{\theta}}_{ij} \widehat{\boldsymbol{\theta}}'_{ij} E(\hat{e}^2_{ijt}) \leq \mathcal{M}.$$

Notice that we cannot consistently estimate the covariance matrix in Theorem 4 in general. This is mainly because $\mathbb{H}^{*(b)}$ does not necessarily converge to $\mathbb{H}$ as the rotation matrix is subject to the data dependent matrix $\boldsymbol{U}^{*(b)}$, which does not always coincide with the population counterpart $\boldsymbol{U}$. In tis regard, we follow Gonçalves and Perron (2014) and construct the CIs using the percentile estimates based on

$$\sqrt{N}\left[\left(\mathbb{H}^{*(b)\prime} + \mathbb{B}^{*(b)\prime}\right)^{-1} \widehat{\boldsymbol{G}}^{*(b)}_t - \widehat{\boldsymbol{G}}_t\right] \xrightarrow{d} N\left(\mathbf{0}, \frac{1}{R^2}\mathcal{I}'\widehat{\mathbb{C}}^*\mathbb{D}^{*,(1)}_t\widehat{\mathbb{C}}^{*\prime}\mathcal{I}\right), \tag{33}$$

which keeps the bootstrap covariance free from the rotation matrix. Let

$$\widehat{\mathcal{D}}^*_{G_t}(\tau) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\left(\sqrt{N}\left[\left(\mathbb{H}^{*(b)\prime} + \mathbb{B}^{*(b)\prime}\right)^{-1} \widehat{\boldsymbol{G}}^{*(b)}_t - \widehat{\boldsymbol{G}}_t\right] \leq \tau\right).$$

be the empirical distribution function where $\mathbb{1}$ is the indicator function. The $1-\alpha$ CI is given by

$$\left[\widehat{\boldsymbol{G}}_t - \frac{q_{\alpha/2}}{\sqrt{N}}, \widehat{\boldsymbol{G}}_t - \frac{q_{1-\alpha/2}}{\sqrt{N}}\right] \tag{34}$$

where $q_{\alpha/2} = \widehat{\mathcal{D}}^{*,-1}_{G_t}(\alpha/2)$ and $q_{1-\alpha/2} = \widehat{\mathcal{D}}^{*,-1}_{G_t}(1-\alpha/2)$ are the inverse function of $\widehat{\mathcal{D}}^*_{G_t}$ evaluated at $\alpha/2$ and $1-\alpha/2$ respectively.

We outline the bootstrap algorithm for the global factor loadings:

Bootstrapping the global factor loadings

1. For each i, j and t, let $e^{*(b)}_{ijt} = \hat{e}_{ijt}\varepsilon^{*(b)}_{ijt}$ where $\hat{e}_{ijt} = y_{ijt} - \widehat{\boldsymbol{\gamma}}'_{ij}\widehat{\boldsymbol{G}}_t - \widehat{\boldsymbol{\lambda}}'_{ij}\widehat{\boldsymbol{F}}_{it}$ and $\varepsilon^{*(b)}_{ijt} \sim$ i.i.d. $N(0,1)$.
2. Construct the re-sampled local factors as

   $$F^{k,*(b)}_{it} = \widehat{F}^z_{it} \cdot \omega^{k,*(b)}_{it} \text{ for } i = 1,\ldots,R,\ z = 1,\ldots,r_i,\ t = 1,\ldots,T.$$

   $\omega^{k,*(b)}_{it}$ is drawn from a zero mean normal distribution independent across i and k with covariance

   $$Cov\left(\omega^{k,*(b)}_{it}, \omega^{k,*(b)}_{is}\right) = Bartlett\left(\frac{t-s}{l^k_i}\right) \text{ for } t,s = 1,\ldots,T$$

   where $Bartlett$ is the Bartlett kernel function and l^k_i is a bandwidth parameter.[18]
3. Construct the re-sampled data as $y^{*(b)}_{ijt} = \widehat{\boldsymbol{\gamma}}'_{ij}\widehat{\boldsymbol{G}}_t + \widehat{\boldsymbol{\lambda}}'_{ij}\boldsymbol{F}^{*(b)}_{it} + e^{*(b)}_{ijt}$ where $\boldsymbol{F}^{*(b)}_{it} = [F^{1,*(b)}_{it},\ldots,F^{r_i,*(b)}_{it}]'$.
4. Estimate the model from the re-sampled data using the procedure developed in Section 3 and obtain the bootstrap version estimates $\widehat{\boldsymbol{\gamma}}^{*(b)}_{ij}$.
5. Repeat Step 1–4 for B times.

[18]The bandwidth parameter can be chosen following the data dependent approach developed by Andrews (1991).

Step 2 follows the dependent wild bootstrap developed by Shao (2010), which accounts for times series dependence of the local factors. We can also consider other block bootstrapping methods to preserve the serial correlation structure of the local factors. For each $b = 1, \ldots, B$, we have:

$$\sqrt{T}\left[\widehat{\boldsymbol{\gamma}}_{ij}^{*(b)} - \left(\mathbb{H}^{*(b)} + \mathbb{B}^{*(b)}\right)^{-1}\widehat{\boldsymbol{\gamma}}_{ij}\right] = \mathbb{H}^{*(b)\prime}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{\boldsymbol{G}}_t\left(\widehat{\boldsymbol{\lambda}}_{ij}'\boldsymbol{F}_{it}^{*(b)} + e_{ijt}^{*(b)}\right) + o_p(1)$$
$$\xrightarrow{d} N\left(\mathbf{0}, \mathbb{H}^{*(b)\prime}\mathbb{D}_{ij}^{*,(2)}\mathbb{H}^{*(b)}\right)$$

where $\mathbb{D}_{ij}^{*,(2)} = \text{plim}_{T\to\infty}T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}E\left[\widehat{\boldsymbol{G}}_s\left(\widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\boldsymbol{F}}_{is} + \hat{e}_{ijs}\right)\left(\widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\boldsymbol{F}}_{it} + \hat{e}_{ijt}\right)\widehat{\boldsymbol{G}}_t'\right]$. For the same reason explained before, we construct the CI based on

$$\sqrt{T}\left[\left(\mathbb{H}^{*(b)} + \mathbb{B}^{*(b)}\right)\widehat{\boldsymbol{\gamma}}_{ij}^{*(b)} - \widehat{\boldsymbol{\gamma}}_{ij}\right] \xrightarrow{d} N\left(\mathbf{0}, \mathbb{D}_{ij}^{*,(2)}\right) \tag{35}$$

to eliminate the rotational indeterminacy. Recall that the rotation matrix $\mathbb{H}^{*(b)}$ is a diagonal matrix with elements ± 1, so $\mathbb{H}^{*(b)}\mathbb{H}^{*(b)\prime} = \boldsymbol{I}_{r_0}$. Let

$$\widehat{\mathcal{D}}_{\gamma_{ij}}^{*}(\tau) = \frac{1}{B}\sum_{b=1}^{B}\mathbb{1}\left(\sqrt{T}\left[\left(\mathbb{H}^{*(b)} + \mathbb{B}^{*(b)}\right)\widehat{\boldsymbol{\gamma}}_{ij}^{*(b)} - \widehat{\boldsymbol{\gamma}}_{ij}\right] \le \tau\right).$$

be the empirical distribution function. The $1-\alpha$ CI is given by

$$\left[\widehat{\boldsymbol{\gamma}}_{ij} - \frac{q_{\alpha/2}}{\sqrt{T}}, \widehat{\boldsymbol{\gamma}}_{ij} - \frac{q_{1-\alpha/2}}{\sqrt{T}}\right] \tag{36}$$

where $q_{\alpha/2} = \widehat{\mathcal{D}}_{\gamma_{ij}}^{*,-1}(\alpha/2)$ and $q_{1-\alpha/2} = \widehat{\mathcal{D}}_{\gamma_{ij}}^{*,-1}(1-\alpha/2)$ are the inverse functions of $\widehat{\mathcal{D}}_{\gamma_{ij}}^{*,-1}$ evaluated at $\alpha/2$ and $1-\alpha/2$ respectively.

A simulation is conducted to examine the validity of our bootstrapping procedure. We use the same DGP as in Section 5 in which we fix $R = 3$ and $(r_0, r_i) = (2, 2)$ and $(\beta, \phi_e, \kappa) = (0, 0, 1)$. The sample size varies as $N_i \in \{20, 50, 100, 200\}$ with $N_1 = \cdots = N_R$ and $T \in \{50, 100, 200\}$. Moreover, we allow $(\phi_G, \phi_F) = (0, 0)$ and $(\phi_G, \phi_F) = (0.5, 0.5)$ to address the potential serial correlation induced by the local factors. We focus on the first element of $\widehat{\boldsymbol{G}}_t$ and $\widehat{\boldsymbol{\gamma}}_{ij}$ evaluated at $t = T/2$ and $i = 1$, $j = N_i/2$, respectively. The bootstrapped CIs are generated by (34) or (36). For comparison, the CIs generated by theoretical (infeasible) variances of 4 and 5 are also reported. We choose the significance level $\alpha = 0.05$ throughout the study.

Each entry of Table 12 is the coverage rate calculated as the ratios of CIs that contains the true factors or loadings over 1000 repetitions. The top panel of Table 12 shows that the infeasible CIs for the global factors have coverage rates around 0.95 whilst the coverage rates of the bootstrapped CIs increase as the sample size increases. The bottom panel of Table 12 presents the results for the global factor loadings. On one hand, it seems that the infeasible CIs are unaffected by the serial correlation of the factors and become closer to 0.95 as the sample size grows. On the other hand, the bootstrapped CIs performs better under non-zero serial correlation of the factors, although both of them become to 0.95 eventually. The above investigation confirms that the bootstrapped CIs are reliable.

Table 12: Coverage rates for the bootstrap CIs with $R = 3$, $(r_0, r_i) = (2, 2)$ and $(\beta, \phi_e, \kappa) = (0, 0, 1)$

Global factors					
		$(\phi_G, \phi_F) = (0, 0)$		$(\phi_G, \phi_F) = (0.5, 0.5)$	
N_i	T	Infeasible	Bootstrap	Infeasible	Bootstrap
20	50	0.939	0.874	0.943	0.894
50	50	0.947	0.923	0.932	0.898
100	50	0.942	0.911	0.952	0.936
200	50	0.943	0.921	0.935	0.925
20	100	0.942	0.902	0.956	0.896
50	100	0.949	0.923	0.939	0.907
100	100	0.953	0.934	0.953	0.933
200	100	0.946	0.925	0.942	0.935
20	200	0.95	0.901	0.957	0.904
50	200	0.954	0.921	0.945	0.929
100	200	0.945	0.93	0.949	0.931
200	200	0.951	0.93	0.948	0.929
Global factor loadings					
		$(\phi_G, \phi_F) = (0, 0)$		$(\phi_G, \phi_F) = (0.5, 0.5)$	
N_i	T	Infeasible	Bootstrap	Infeasible	Bootstrap
20	50	0.972	0.925	0.963	0.897
50	50	0.961	0.915	0.966	0.862
100	50	0.974	0.922	0.979	0.909
200	50	0.965	0.916	0.978	0.887
20	100	0.955	0.931	0.960	0.910
50	100	0.963	0.929	0.958	0.915
100	100	0.966	0.934	0.956	0.909
200	100	0.9678	0.936	0.967	0.914
20	200	0.951	0.933	0.930	0.911
50	200	0.937	0.926	0.955	0.932
100	200	0.958	0.942	0.959	0.931
200	200	0.955	0.936	0.951	0.929

Each entry shows the coverage rate calculated as the ratios of CIs that contains the true factors or loadings over 1000 repetitions. The infeasible CIs are generated by the theoretical asymptotic distributions in Theorem 4 or 5, and the bootstrap CIs are generated by the by (34) or (36). We report the CIs for the first global factor and loading, evaluated at $t = T/2$ and $i = 1$, $j = N_i/2$ respectively. r_0 and r_i are the true number of global factors and true number of local factors in group i. We set $r_1 = \cdots = r_R$. We set $N_1 = \cdots = N_R$ where N_i is the number of individuals in block i. T is the number of time periods. ϕ_G and ϕ_F are the AR coefficients for the global and local factors. β, ϕ_e and κ control the cross-section correlation, serial correlation and noise-to-signal ratio.