Three Essays on Modelling Large Panel Data with a Multilevel Factor Structure

Rui Lin

Doctor of Philosophy

University of York

Economics and Related Studies

June, 2023

Abstract

This thesis presents a sequential and comprehensive development of high-dimensional panel data models with a multilevel factor structure. The multilevel factors consist of unobserved global factors that affect all individuals and unobserved local factors that affect individuals within specific blocks.

In Chapter 1, we develop a novel approach based on canonical correlation analysis to identify the number of global factors in the multilevel factor model. We propose two consistent selection criteria: the canonical correlation difference (CCD) and the modified canonical correlations (MCC). Monte Carlo simulations show that CCD and MCC correctly select the number of global factors even in small samples, and they are robust to correlated local factors. In an empirical application, we investigate a multilevel asset pricing model for stock return data in 12 industries in the U.S. market.

Chapter 2 advances a unified econometric framework for the multilevel factor model based on generalized canonical correlation (GCC) analysis. Our approach is valid even if some blocks share common local factors. We establish the consistency of the estimated factors and loadings, as well as their asymptotic normality under fairly standard conditions. As a by-product of estimation, a new selection criterion is developed to estimate the number of global factors. Through Monte Carlo simulations, we confirm the validity of our asymptotic theory and demonstrate its superior performance over existing approaches. We apply the model to a large disaggregated panel data set of house prices in England and Wales.

Chapter 3 considers a panel regression model with multilevel factors. We propose a multilevel iterative principal component ($MIPC$) method that iteratively updates the slope coefficients and factors. We also propose a model selection criterion based on eigenvalue ratios to determine the number of factors. Given consistent factor estimates, we employ GCC to separately identify the global and local factors. Under a finite number of blocks, we show the consistency of our estimates and establish the asymptotic normality of the bias-corrected estimator for the slope coefficients. Monte Carlo simulations demonstrate the good finite sample performance of $MIPC$. We apply our method to an analysis of the energy consumption and economic growth nexus using a cross-country panel data categorized by regions.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this or any other University. Where individual chapters were coauthored with other researchers, this is indicated with the necessary specifications in this declaration. Chapter 1 is a coauthored paper with Prof. In Choi and my supervisor, Prof. Yongchol Shin and has been published in the *Journal of Econometrics*. I contributed to every part of the research and my co-authors contributed at various points with revisions and comments. Chapter 2 is a coauthored paper with Prof. Yongcheol Shin. I contributed to every part of the research, and Prof. Yongcheol Shin contributed at various points with revisions and comments. I am the sole author of Chapter 3. All sources are acknowledged as references.

Rui Lin

June 2023, York

Acknowledgements

Completing this thesis has been a challenging adventure, and I owe a debt of gratitude to those who have accompanied me on this journey.

I am deeply indebted to my supervisors, Prof. Yongcheol Shin and Dr. Paola Zerilli, for their invaluable guidance and continuous encouragement over these years. Their passion for academic excellence have been contagious and motivational. Without being blessed by their expertise, patience, and dedication, I would not have been able to reach this milestone. In particular, I greatly benefited from collaborating with and discussing ideas with Prof. Yongcheol Shin. I enjoyed every minute working with my supervisors. I also acknowledge the financial support from the joint funding of the Chinese Scholarship Council (CSC) and the University of York.

Special thanks extend to the the academic and professional support staff at the Department of Economics for fostering an intellectually stimulating environment. In particular, I thank Prof. Jia Chen, Prof. Laura Coroneo, Prof. Michael Thornton, Prof. Weining Wang, Prof. Takashi Yamagata, and Dr. Michael Shallcross who provided constructive comments and other support that facilitates my pursuit of Ph.D.

I would like to express my gratitude to my friends at York: Maxine Cao, Ahmed Djeddi, Jenny Hao and Neinei, Puig Gonzalez Lluis, Akseer Hussain, Yingjia Hu, Lisha Li, Yuning Li, Cheng Li, Shuyi Lou, Jaxin Peng, Vigyan Ratnoo, Carla Niu, Ao Wang, Hui Wan, Ting Xie, Qingyi Ye, Chaowen Zheng, Meichao Zhang, and many others. They enrich the palette of this lonely planet. I am incredibly fortunate to have them by my side.

Finally, I am deeply grateful to my parents, my grandmother, and my sister for their unconditional love and unwavering support. I hope that my father and grandmother, who now reside in heaven, can be proud of my accomplishment.

Contents

List of Figures

List of Tables

Introduction

In the past two decades, panel data econometrics has experienced rapid development due to its ability to account for unobserved heterogeneity among individuals. With the increasing availability of large datasets containing a large number of cross-sectional units (N) and time periods (T), factor models have gained popularity as effective tools for dimension reduction to overcome the "curse of dimensionality". In factor models, the co-movement of variables is captured by a small number of unobserved common factors, while individuals respond with heterogeneous loadings on these factors. This is known as the "blessing of dimensionality". Factor models find wide application in empirical finance and macroeconomics, see e.g. Connor & Korajczyk (1988) and Stock & Watson (2002), among others.

Recently, the multilevel factor model has gained increasing attention, where some factors are pervasive (common to all individuals) and others are semi-pervasive (common to only a subset of individuals). These factors are referred to as global and local factors, respectively, resulting in a "block structure" in the data. The multilevel factor model allows researchers to uncover and explore a richer interconnected structure among individuals. Kose et al. (2003) find that global factors and certain country-specific factors account for a larger portion of business cycle variability. Del Negro & Otrok (2007) document that local (regional) factors drive house price co-movement in the U.S. more than the global (national) factors, suggesting that expansionary monetary policy has a limited impact on housing bubbles. Andreou et al. (2019) discover sector-specific factors in the context of mixed frequency data and demonstrate that the industrial production sector remains crucial for economic growth in the U.S.

The multilevel factor model introduces challenges for estimation and inference, particularly when the number of blocks is finite. Firstly, existing model selection methods become inapplicable in the presence of local factors, as the weak cross-section correlation assumption is violated and/or the clustering of non-zero eigenvalues is obscured by local factors (see Breitung & Eickmeier (2016)). Consequently, some studies assume a known number of global factors as in Kose et al. (2003), Breitung & Eickmeier (2016), and Choi et al. (2018). Secondly, the inferential theory for the multilevel factor model is not well established when the number of blocks is finite, unlike cases where the number of blocks tends to infinity, e.g. Wang (2008) and Jin et al. (2023). Andreou et al. (2019) develop asymptotic distributions for factor and loading estimates, but their theory only applies to the special case with two blocks. Furthermore, it remains unclear how to conduct inference. In the panel regression, the unobserved factor structure in the error components consists of interactive fixed effects (see Pesaran (2006) and Bai (2009)). If they are not appropriately controlled for, endogeneity issue arises, leading to inconsistent estimation and misleading inference. The multilevel factor structure further requires new estimation techniques and asymptotic theory. In this thesis, we contribute by addressing the aforementioned important issues in each of the three chapters. As a result, this thesis offers a unified framework for the high-dimensional panel data model with the multilevel factor structure.

Chapter 1

We propose two new selection criteria for estimating the number of global factors in the multilevel factor model based on canonical correlation analysis (CCA), called the canonical correlations difference (CCD) and the modified canonical correlations (MCC). We employ principal component (PC) estimation to extract factors from each block and calculate the average pairwise canonical correlations. As all blocks share the same global factors, the corresponding canonical correlations are equal to one, while the rest are strictly less than one. The CCD criterion selects the number of global factors that maximizes the differences between consecutive canonical correlations while the MCC criterion distinguishes between unit and non-unit canonical correlations using a data-dependent threshold. We establish the consistency of CCD and MCC and demonstrate their superior finite sample performance via Monte Carlo simulations. We apply them to the multilevel asset pricing model using weekly stock returns from firms in 12 U.S. industries during the January 2015 - December 2016. Both CCD and MCC identify the presence of only one global factor, which closely co-moves with the market factor, exhibiting a correlation of 0.95. On average, the market factor explains 22.6% of the time series variation. We also identify one local (industry) factor in Non-Durable, Energy, Health, and Money, and two industry (local) factors in Utilities. Notably, for Energy and Utilities, the local factors are more important than the global factor. This suggests that industry-specific risks are essential to avoid misallocation of assets in portfolio management.

Chapter 2

We advance the generalised canonical correlation (GCC) estimation for the multilevel factor model. To this end we collect the PC estimates from each block and construct a system-wide matrix. We then obtain the linear combinations of the factor spaces from a singular value decomposition of this matrix, that can simultaneously satisfy the pairwise unit canonical correlations between any two blocks. Thus, the global factors can be consistently estimated by summarising these linearly transformed factors.

The GCC approach has several advantages. It accommodates common local factors that impact multiple blocks, in contrast to the orthogonality assumption imposed on local factors by Choi et al. (2018) and Han (2021). Once the global factors are estimated consistently, the local factors can be obtained by applying PC to each block separately after removing the global components. This sequential approach is computationally efficient, as each step yields consistent estimates and eliminates the need for iteration. Moreover, the number of global factors can be easily estimated by evaluating the ratios of adjacent singular values of the system-wide matrix. With the finite number of blocks, we establish the consistency of our estimators using matrix perturbation theory and derive their asymptotic distributions under standard regularity conditions. However, to address the issue that the covariance matrices are subject to non-estimable identities, we propose a hybrid bootstrap procedure to construct valid confidence intervals, the validity of which is confirmed via Monte Carlo simulations. We apply GCC to a large disaggregated panel data consisting of house prices for 331 local authorities in England and Wales over the period 1996Q1 to 2021Q2. We find one global (national) factor that displays a typical (global) boom-bust-recovery cycle. We also identify one local factor in seven regions (NE, NW, YH, EE, LD, SE, and WA), but no local factor in three regions (EM, WM, and SW). Notably, the local factors of EE, LD and SE (Area 1) comove closely while those of NE, NW, YH and WA (Area 2) tend to cluster. Additional GCC estimation confirms the existence of common local factors across these regions. Finally, we document a strong co-movement between the growth rate of the (lagged) population gap and the gap in areal factor components. This finding suggests that the growth of the population gap could play an important role in driving the gap of regional house prices.

Chapter 3

This chapter focuses on the estimation of the panel data regression model with unobserved multilevel factors. We propose a multilevel iterative principal component ($MIPC$) estimation, which extends the iterative principal component (IPC) method introduced by Bai (2009). The factors are updated using PC estimation applied to each block separately, given the slope coefficients. Then, the slope coefficients are updated by the pooled OLS estimator after a linear projection of the factors from the dependent variable. The estimation proceeds iteratively until convergence. We establish the oracle $\sqrt{NT}$-consistency of the slope coefficients and derive their asymptotic normal distribution. Unlike the case where the number of blocks tends to infinity as considered in Feng et al. (2023), we establish that the (asymptotic) bias terms of the slope coefficients do not vanish. Thus, we provide a bias-corrected estimator and employ a wild dependent bootstrap procedure for inference. Our approach only requires knowledge of the total numbers of factors in each block, without separately identifying the numbers of global and local factors. Furthermore, we propose a consistent model selection criterion based on eigenvalue ratios. Via Monte Carlo simulations, we confirm the consistency of the slope coefficients, factors, and the number of factors. We also document that $MIPC$ exhibits substantially smaller size distortion than an infeasible version of IPC that ignores the block structure. The utility of our approach is demonstrated through an empirical analysis of the relationship between energy consumption and economic growth using a cross-country panel dataset consisting of 80 countries from 1972 to 2014. $MIPC$ finds no global factors and one local factor for Asia Pacific and Europe, two local factors for America, and four local factors for Africa. Our finding suggests that the coefficient for energy consumption obtained from a two-way fixed effects model is underestimated by half compared to the one estimated using $MIPC$.

Chapter 1

Canonical Correlation-based Model Selection for the Multilevel Factors

Abstract We develop a novel approach based on the canonical correlation analysis to identify the number of the global factors in the multilevel factor model. We propose the two consistent selection criteria, the canonical correlations difference (CCD) and the modified canonical correlations (MCC). Via Monte Carlo simulations, we show that CCD and MCC select the number of global factors correctly even in small samples, and they are robust to the presence of serially correlated and weakly cross-sectionally correlated idiosyncratic errors as well as the correlated local factors. Finally, we demonstrate the utility of our approach with an application to the multilevel asset pricing model for the stock return data in 12 industries in the U.S. market.

Keywords: Multilevel Factor Models, Principal Components, Canonical Correlation Difference, Modified Canonical Correlations, Multilevel Asset Pricing Models.

JEL Classification: C52, G12.

1.1 Introduction

The factor models have been popular as an effective tool for the dimension reduction for the big dataset with the large number of cross-section units (N) and time periods (T) through extracting the co-movement of the variables by a small number of common factors, e.g. Stock & Watson (2002) and Bai (2003). Recently, the literature on the multilevel factor model, also referred to as the panel data model with the block structure, has been growing rapidly. Here we have the global factors that influence all the individuals as well as the local factors that only affect those within the specific block. If the structure of the multilevel factors is ignored, the conventional (approximate) factor approach would produce inconsistent and misleading results.

Different estimation methods have been developed: the Bayesian approach by Kose et al. (2003) and Moench et al. (2013), the classical approach by Breitung & Eickmeier (2016) and Choi et al. (2018), and the LASSO approach by Han (2021). Kose et al. (2003) analyse the relative contribution of the global and regional factors to explain the business cycle whilst Moench et al. (2013) demonstrate an important role played by the level factors in explaining the U.S. real activities. Breitung & Eickmeier (2016) and Choi et al. (2018) propose a canonical correlation estimator for the identification of global and local factors in the multilevel factor model. Futhermore, Bekaert et al. (2009) examine the international stock co-movements, Ando & Bai (2014) find different factors in A share and B share in the Chinese stock market, and Beck et al. (2016) investigate the source of price changes in Europe.

A remaining yet challenging issue is how to identify the number of the global factors and the number of local factors, simultaneously. It is well-established that the existing information criteria mainly developed for the single level panel data, fail to consistently estimate the number of global factors because the weak (error) cross-section correlation condition is violated in the presence of the multilevel factors. In this regard, some studies assume that the number of global factors is known *a priori*, and develop a sequential estimation approach. For example, assuming that the number of global factors is 1, Choi et al. (2018) apply the information criteria to each block and estimate the number of local factors.

Let r_0 (r_i) be the number of global (local) factors and R the number of blocks. A few studies have attempted to deal with an important issue of consistently estimating r_0 under the multilevel setting. Wang (2008) proposes a sequential procedure by applying the existing information criteria to the whole data and to the data in each block, consequently, and estimating r_0 by the cardinal difference. Chen (2012) and Dias et al. (2013) propose the modified information criteria by penalising r_i more heavily than r_0. Andreou et al. (2019) apply the canonical correlation analysis to estimate global and local factors in a two-group model and develop a novel inference on r_0. Han (2021) proposes a shrinkage estimator that can estimate the global and local factors/loadings, and determine the number of factors, jointly. As R rises, however, an implementation of these approaches would be almost impractical or infeasible due to the heavy computational burdens as well as uncertainty of the final outcomes.

In this paper, as the main contribution, we propose a novel approach based on the canonical correlation analysis to identify the number of global factors which can be easily applied to the models with a fixed number of blocks and with $R \to \infty$. To this end, we first apply the principal component (PC) estimation to the data in each block and obtain the $r_{\max}$ factors, which are consistent for the factor space spanned by the global and local factors jointly, where $r_{\max}$ is the (common) maximum number of factors allowed in each block ($i = 1, \cdots, R$). Next, we evaluate the $r_{\max}$ canonical correlations between estimated factors from any two blocks. Then, using $R(R-1)$ pairwise canonical correlations, we construct the cross-block average of the canonical correlations, denoted $\xi(r)$.

We first develop the canonical correlation difference criterion, denoted $CCD(r)$, which is constructed by the difference between the consecutive cross-block averages. Then, r_0 can be estimated consistently by maximising $CCD(r)$ over $r = 0, 1, ..., r_{\max}$. But, in the presence of correlated local factors, we need to impose the upper bound condition on the largest average canonical correlation between the local factors across R blocks, in order to ensure that CCD is maximised at $r = r_0$. In this regard we develop the alternative estimator, called the modified canonical correlation ($MCC(r)$) using the nondegenerate distribution of $1 - \xi(r)$ for $r \leq r_0$, that can remain consistent without imposing the upper bound condition. Then, r_0 can be estimated consistently by maximising r such that $1 - \xi(r)$ is below a certain threshold.

We derive asymptotic properties of pairwise canonical correlations and the cross-block average, and show that CCD and MCC are consistent selection criteria for identifying r_0. Next, via Monte Carlo simulations, we investigate their finite sample properties together with two existing approaches advanced by Chen (2012) and Andreou et al. (2019). Overall, we find that both CCD and MCC select r_0 even in small samples, outperforming the other approaches in the presence of serially correlated and weakly cross-sectionally correlated idiosyncratic errors. Only if the correlations among the local factors are deemed to be relatively weak on average (say, less than 1/2), we recommend the use of CCD because it is very simple to implement without requiring any tuning parameter. Given that the overall performances of CCD and MCC are qualitatively similar whilst MCC does not need to meet the upper bound condition, in general, we prefer the use of MCC.

Once r_0 is consistently estimated by CCD and MCC, we remove the global factors from the data in each block, and apply the existing criteria, such as BIC_3 by Bai & Ng (2002) and ER by Ahn & Horenstein (2013), to consistently estimating the number of local factors.

Our proposed approach possesses a number of advantages. First, it is simple to apply as it involves the standard PC and CCA methods, unlike other approaches that require to assess many tuning and control parameters, e.g. Han (2021). Second, even if the number of blocks is substantially large, our approach is computationally feasible as it only evaluates the cross-block average of $R(R-1)/2$ pairwise canonical correlations, unlike other approaches that will be computationally infeasible, e.g. Chen (2012) and Andreou et al. (2019). More importantly, our approach is shown to be robust to the presence of serially correlated and weakly cross-sectionally correlated idiosyncratic errors as well as the correlated local factors.

We demonstrate the utility of our framework with an application to the multilevel asset pricing model for the weekly stock return data for the twelve industries in the U.S. over the period, Jan. 2015 to Dec. 2016. First, both CCD and MCC find that there is only one global factor, which co-moves closely with the market factor, with correlation of 0.95. Then, we apply BIC_3 to the defactored data in each group and find one local factor in NoDur, Enrgy, Hlth and Money, and two local factors in Utils. On average, the global factor, local factors and idiosyncratic components can explain 22.6%, 5.8% and 70.8% of the total variation, respectively. The global factor tends to display a higher relative importance ratio for the cyclical industries, suggesting that the higher within-correlations observed for these industries are likely to reflect the higher loadings to the global factor. On the other hand, the influence of the local factors are more important than the global factor for some industries such as Enrgy, Utils and Hlth. For these industries, the high within-industry correlations are likely to reflect co-movements with local/industry factors, suggesting that the local factors should be taken into account to avoid any misleading asset allocation in portfolio management, e.g. Bekaert et al. (2009).

The rest of the paper is structured as follows. Section 1.2 provides an overview of the related literature. Section 1.3 presents the multilevel factor model with the underlying assumptions. Section 1.4 develops CCD and MCC criteria for selecting the number of global factors and derives the asymptotic theory. Section 1.5 presents Monte Carlo simulation evidence. Section 1.6 provides an empirical application. Section 1.7 offers concluding remarks. The mathematical proofs, additional simulation results, and theoretical derivations are relegated to Appendix A.

1.2 Related Literature

For the single-level panel data model with the approximate factor structure, there have been two main approaches for identifying the number of unobserved common factors. The first is the information criteria proposed by Bai & Ng (2002), which take a form: $PC(r) = V\left(r, \widehat{\mathbf{F}}\right) + rg(N,T)$, where $V\left(r, \widehat{\mathbf{F}}\right)$ is the sum of squared residuals, $\widehat{\mathbf{F}}$ is a $T \times r$ matrix of factors estimated by the principal components and $g(N,T)$ is a penalty function of the number of cross-section units, N and the number of time periods, T.

Another popular approach attempts to make use of the fact that for the data with r_0 latent factors, the first r_0 eigenvalues of the covariance matrix of the data diverge while the rest of the eigenvalues are bounded and clustered. Onatski (2010) develops the edge distribution (ED) estimator based on the difference between the adjacent eigenvalues arranged in descending order such that $\hat{r}_0 = \max_{1\leq r\leq r_{\max}}\{r|\mu_r - \mu_{r+1} \geq \delta\}$, where μ_r is the r-th largest eigenvalue and δ is a threshold value, which is calibrated from the empirical distribution of the eigenvalues and $r_{\max}$ is the maximum value of r. Ahn & Horenstein (2013) propose the eigenvalue ratio (ER) given by $\hat{r}_0 = \arg\max_{1\leq r\leq r_{\max}}\{\mu_r/\mu_{r+1}\}$.

Choi & Jeong (2019) have conducted a comprehensive simulation study on approximate factor models, and documented evidence that BIC_3 by Bai & Ng (2002) and ER by Ahn & Horenstein (2013) outperform other competing estimators. Interestingly, Breitung & Pigorsch (2013) propose a canonical correlation-based selection procedure that consistently estimate the number of dynamic factors using the static factor representation of the dynamic factor model. See also Hallin & Liška (2007), Alessi et al. (2010) and Bailey et al. (2021).

In the presence of the multilevel factors, the existing selection criteria may fail to identify the number of the global factors. If we apply existing approaches to the $T \times M_i$ data matrix, $\mathbf{Y}_i$ in each block $i = 1, \ldots, R$, respectively, we can consistently estimate only the sum, $r_0 + r_i$, but not r_0 or r_i, separately. Suppose that we apply the existing criteria to the whole data matrix, $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_R]$ by ignoring the multilevel structure. If R is fixed (and small), then the existing selection criteria mainly developed for the single level panel data, fail to consistently estimate r_0 because the weak (error) cross-section correlation condition is violated in the presence of the local factors. As $R \to \infty$, however, the impacts of the local factors would be asymptotically negligible. In this case Han (2021) conjectures that the number of global factors can be consistently estimated asymptotically by the existing selection criteria (see Remark 4).

In Section A.2 of Appendix A we examine the finite sample performance of the four criteria, IC_{p2} and BIC_3 by Bai & Ng (2002), ED by Onatski (2010) and ER by Ahn & Horenstein (2013), through applying them directly to the whole data matrix. We find that these approaches tend to produce unreliable inference. If $R = 2$, all of the four criteria select the total number of factors, $r_0 + \sum_{i=1}^{R} r_i$, not r_0. For sufficiently large R, they tend to select r_0. However, for the moderate value of R, e.g. $R = 5$ or 10, they select the intermediate value between r_0 and $r_0 + \sum_{i=1}^{R} r_i$. Next, their performances are all adversely affected in small samples by the presence of cross-sectionally and serially correlated errors. Finally and importantly, even for large R, they overestimate r_0 significantly in the presence of even moderate correlations among the local factors.

A few studies have attempted to develop a consistent estimator of the number of global factors under the multilevel setting. Wang (2008) proposes to determine the model specification based on the principle of inclusion–exclusion for set cardinality.[1] The above simulation evidence shows that Wang's sequential procedure is unreliable. For large R, it would significantly overestimate by selecting $r_0 + \sum_{i=1}^{R} r_i/R$ instead of r_0. Further, Han (2021) provides the simulation evidence that this can lead to even negative estimates of both r_0 and r_i in small samples for $R = 3$.

Chen (2012) and Dias et al. (2013) modify the information criteria advanced by Bai & Ng (2002), and include the number of local factors as arguments in the $PC(r)$ objective function. The main modification is to penalise the global factors less than the local factors for their parsimonious structure. As R rises, however, the computation will be almost infeasible since the number of candidate models increases drastically.

Andreou et al. (2019) (AGGR) apply the canonical correlation analysis to estimate global and local factors in a two-group factor model with mixed frequency data, and develop a novel inference on r_0 via canonical correlations. AGGR first apply the existing information criteria to each of two groups and obtain the estimates, $\widehat{r_0 + r_1}$ and $\widehat{r_0 + r_2}$. They extract the $T \times r_{\min}$ matrix of factors, $\widehat{\mathbf{K}}_i$, from the data, $\mathbf{Y}_i$ for $i = 1, 2$, where $r_{\min} = \min\left\{\widehat{r_0 + r_1}, \widehat{r_0 + r_2}\right\}$. They compute the sum of the r largest canonical correlations between $\widehat{\mathbf{K}}_1$ and $\widehat{\mathbf{K}}_2$, and derive the scaled and centered test statistic. Next, by imposing the strong assumption that idiosyncratic errors are neither serially nor cross-sectionally correlated, AGGR can derive that the test follows the standard normal distribution asymptotically under the null hypothesis, $r = r_0$.[2] This procedure can be used for model selection only if the critical value diverges at a certain rate, γ with $0 < \gamma < 1$. A sequential test can be performed for $r = r_{\max}, r_{\max} - 1, \ldots, 1$ backwards, and $\hat{r}_0$ is the largest r when the null is not rejected. Finally, they propose to estimate the number of local factors by $\widehat{r_0 + r_i} - \hat{r}_0$ for $i = 1, 2$. However, it would be complicated to analytically extend their approach to cover the case with $R > 2$.

Han (2021) proposes an adopted LASSO estimator that can consistently estimate the factors/loadings, and determine the number of factors, simultaneously. The number of global (local) factors can be estimated by the number of non-zero columns in their respective factor loading ma-

[1] Using the two blocks, for example, one can apply the information criteria to the whole data and obtain $\widehat{r_0 + r_1 + r_2}$. Next, using the data for each block, one can estimate $\widehat{r_0 + r_i}$, $i = 1, 2$. Then, the number of global factors can be estimated by the difference, $\widehat{r_0 + r_1} + \widehat{r_0 + r_2} - \widehat{r_0 + r_1 + r_2}$.

[2] Andreou et al. (2019) argue that their test would work in the presence of limited correlation among errors, but also discuss how to relax this assumption.

trices. But, this approach requires the selection of tuning parameters by imposing different penalty terms for different blocks. Consequently, for large R, a large number of candidate tuning parameters need to be selected coherently. Further, as the shrinkage estimation is not invariant to the order of the blocks, we need to apply the additional information criteria to determine which block is ordered first. Hence, an extension to the model with large R would be almost infeasible due to the heavy computational burden as well as uncertainty of the final outcomes. More importantly, the shrinkage estimator is shown to be consistent only if the local factors are mutually uncorrelated, though it is challenging to develop a shrinkage estimator fully robust to the local factors correlations.[3]

In the next Section we propose a novel approach based on the canonical correlation analysis. Our method differentiates from the existing approaches in two main aspects. First, our approach can be easily applied to the models with a fixed number of blocks and with $R \to \infty$. Next, our approach will be shown to be valid in the presence of serially correlated and weakly cross-sectionally correlated idiosyncratic errors as well as the correlated local factors.

1.3 The Model and Assumptions

Consider the multilevel factor model:

$$y_{ijt} = \boldsymbol{\gamma}_{ij}'\mathbf{G}_t + \boldsymbol{\lambda}_{ij}'\mathbf{F}_{it} + e_{ijt}, i = 1, ..., R, j = 1, ..., M_i, t = 1, ..., T \tag{1.3.1}$$

where $\mathbf{G}_t = [G_{t1}, ..., G_{tr_0}]'$ comprises the $r_0 \times 1$ global factors, $\mathbf{F}_{it} = [F_{it1}, ..., F_{itr_i}]'$ is the $r_i \times 1$ vector of local factors in the block $i = 1, \cdots, R$, $\boldsymbol{\gamma}_{ij}$ and $\boldsymbol{\gamma}_{ij}$ are factor loadings and e_{ijt} is the idiosyncratic error. Stacking (1.3.1) across individuals in block i, we have:

$$\mathbf{y}_{it} = \boldsymbol{\Gamma}_i\mathbf{G}_t + \boldsymbol{\Lambda}_i\mathbf{F}_{it} + \mathbf{e}_{it}, \tag{1.3.2}$$

where M_i is the number of individuals in the block i,

$$\underset{M_i\times 1}{\mathbf{y}_{it}} = \begin{bmatrix} y_{i1t} \\ \vdots \\ y_{iM_it} \end{bmatrix}, \quad \underset{M_i\times 1}{\mathbf{e}_{it}} = \begin{bmatrix} e_{i1t} \\ \vdots \\ e_{iM_it} \end{bmatrix}, \quad \underset{M_i\times r_0}{\boldsymbol{\Gamma}_i} = \begin{bmatrix} \boldsymbol{\gamma}_{i1}' \\ \vdots \\ \boldsymbol{\gamma}_{iM_i}' \end{bmatrix}, \quad \underset{M_i\times r_i}{\boldsymbol{\Lambda}_i} = \begin{bmatrix} \boldsymbol{\lambda}_{i1}' \\ \vdots \\ \boldsymbol{\lambda}_{iM_i}' \end{bmatrix}.$$

[3] From Table 5 in Han (2021), we find that the shrinkage estimator severely overestimates (underestimates) the number of global (local) factors, even if the correlation between the local factors is as small as 0.1.

The model can also be written as

$$\mathbf{Y}_t = \boldsymbol{\gamma}^+ \mathbf{F}_t^+ + \mathbf{e}_t,$$

where

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

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

$$\mathbf{Y} = \mathbf{F}^+ \boldsymbol{\Lambda}^{+\prime} + \mathbf{e}, \tag{1.3.3}$$

where

$$\underset{T\times N}{\mathbf{Y}} = \begin{bmatrix} \mathbf{Y}_1' \\ \vdots \\ \mathbf{Y}_T' \end{bmatrix}, \underset{T\times r^+}{\mathbf{F}^+} = \begin{bmatrix} \mathbf{F}_1^{+\prime} \\ \vdots \\ \mathbf{F}_T^{+\prime} \end{bmatrix} \text{ and } \underset{T\times N}{\mathbf{e}} = \begin{bmatrix} \mathbf{e}_1' \\ \vdots \\ \mathbf{e}_T' \end{bmatrix}.$$

Alternatively, stacking (1.3.1) over time, we can rewrite the model as

$$\mathbf{Y}_{ij} = \mathbf{G}\boldsymbol{\gamma}_{ij} + \mathbf{F}_i \boldsymbol{\gamma}_{ij} + \mathbf{e}_{ij}, \tag{1.3.4}$$

where

$$\underset{T\times 1}{\mathbf{Y}_{ij}} = \begin{bmatrix} \mathbf{y}_{ij,1} \\ \vdots \\ \mathbf{y}_{ij,T} \end{bmatrix}, \underset{T\times 1}{\mathbf{e}_{ij}} = \begin{bmatrix} \mathbf{e}_{ij,1} \\ \vdots \\ \mathbf{e}_{ij,T} \end{bmatrix}, \underset{T\times r_0}{\mathbf{G}} = \begin{bmatrix} \mathbf{G}_1' \\ \vdots \\ \mathbf{G}_T' \end{bmatrix}, \underset{T\times r_i}{\mathbf{F}_i} = \begin{bmatrix} \mathbf{F}_{i1}' \\ \vdots \\ \mathbf{F}_{iT}' \end{bmatrix}$$

For each block i, we then have

$$\mathbf{Y}_i = \mathbf{G}\boldsymbol{\Gamma}_i' + \mathbf{F}_i \boldsymbol{\Lambda}_i' + \mathbf{e}_i \tag{1.3.5}$$

where $\mathbf{Y}_i = [\mathbf{Y}_{i1}, \mathbf{Y}_{i2}, \ldots, \mathbf{Y}_{iM_i}]$ and $\mathbf{e}_i = [\mathbf{e}_{i1}, \mathbf{e}_{i2}, \ldots, \mathbf{e}_{iM_i}]$.

Following Bai & Ng (2002) and Choi et al. (2018), we make the following assumptions. Let $\mathcal{M}$ be a finite constant.

Assumption 1.A.

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

2. *Let* $\mathrm{E}(N^{-1}\sum_{i=1}^{R}\sum_{j=1}^{M_i} e_{ijs}e_{ijt}) = \omega_N(s,t)$. *Then,* $|\omega_N(s,t)| < \mathcal{M}$ *for all* s, *and*

$$\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}|\omega_N(s,t)| \leq \mathcal{M}.$$

3. *Let* $\mathrm{E}(e_{mjt}e_{hkt}) = \tau_{(mj),(hk),t}$, *with* $|\tau_{(mj),(hk),t}| \leq |\tau_{(mj),(hk)}| < \mathcal{M}$ *for all* t. *In addition,*

$$\frac{1}{N}\sum_{m=1}^{R}\sum_{h=1}^{R}\sum_{j=1}^{M_m}\sum_{k=1}^{M_h}|\tau_{(mj),(hk)}| \leq \mathcal{M}.$$

4. *Let* $\mathrm{E}(e_{mjt}e_{hks}) = \tau_{(mj),(hk),(ts)}$ *with*

$$\frac{1}{NT}\sum_{m=1}^{R}\sum_{h=1}^{R}\sum_{j=1}^{M_m}\sum_{k=1}^{M_h}\sum_{t=1}^{T}\sum_{s=1}^{T}|\tau_{(mj),(hk),(ts)}| \leq \mathcal{M}.$$

5. *For every* t, s, i *and* j

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

Assumption 1.B.

1. $\mathbf{G}_t$, $\mathbf{F}_{1t}, \ldots, \mathbf{F}_{Rt}$ *are zero-mean, stationary processes that satisfy the conditions for the law of large numbers and the central limit theorem, which can be applied to their self- and cross-products.*

2. $\mathrm{E}\left(\|\mathbf{K}_{it}\|^4\right) < \infty$, *where* $\mathbf{K}_{it} = (\mathbf{G}_t', \mathbf{F}_{it}')'$.

3. $T^{-1}\mathbf{G}'\mathbf{G} \xrightarrow{p} \mathbf{\Sigma}_G$, *where* $\mathbf{\Sigma}_G$ *is a positive-definite matrix.*

4. *For every* i, $T^{-1}\mathbf{F}_i'\mathbf{F}_i \xrightarrow{p} \mathbf{\Sigma}_{F_i}$ *where* $\mathbf{\Sigma}_{F_i}$ *is a positive-definite matrix.*

5. *For* i, j *and* t,

$$\mathrm{E}\left(\frac{1}{M_i}\sum_{j=1}^{M_i}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{F}_{it}e_{ijt}\right\|^2\right) \leq \mathcal{M}; \mathrm{E}\left(\frac{1}{N}\sum_{i=1}^{R}\sum_{j=1}^{M_i}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{G}_t e_{ijt}\right\|^2\right) \leq \mathcal{M}.$$

Assumption 1.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. $N^{-1}\sum_{i=1}^{R}\mathbf{\Gamma}_i'\mathbf{\Gamma}_i \longrightarrow \mathbf{\Sigma}_\Gamma$, *where* $\mathbf{\Sigma}_\Gamma$ *is a positive-definite matrix.*

3. $\mathbf{\Sigma}_\Gamma\mathbf{\Sigma}_G$ *has distinct eigenvalues.*

4. *For every* $i = 1, \cdots, R$,

   (a) $\mathrm{rank}\left([\mathbf{\Gamma}_i, \mathbf{\Lambda}_i]\right) = r_0 + r_i$.

   (b) $M_i^{-1}\begin{bmatrix}\mathbf{\Gamma}_i'\mathbf{\Gamma}_i & \mathbf{\Gamma}_i'\mathbf{\Lambda}_i \\ \mathbf{\Lambda}_i'\mathbf{\Gamma}_i & \mathbf{\Lambda}_i'\mathbf{\Lambda}_i\end{bmatrix} \longrightarrow \begin{bmatrix}\mathbf{\Sigma}_{\Gamma_i} & \mathbf{\Sigma}_{\Gamma_i\Lambda_i} \\ \mathbf{\Sigma}_{\Gamma_i\Lambda_i}' & \mathbf{\Sigma}_{\Lambda_i}\end{bmatrix}$ *which is a positive-definite matrix.*

   (c) $M_i^{-1}\mathbf{\Lambda}_i'\mathbf{\Lambda}_i \longrightarrow \mathbf{\Sigma}_{\Lambda_i}$, *where* $\mathbf{\Sigma}_{\Lambda_i}$ *is a positive-definite matrix*

   (d) $\begin{bmatrix}\mathbf{\Sigma}_{\Gamma_i} & \mathbf{\Sigma}_{\Gamma_i\Lambda_i} \\ \mathbf{\Sigma}_{\Gamma_i\Lambda_i}' & \mathbf{\Sigma}_{\Lambda_i}\end{bmatrix}\begin{bmatrix}\mathbf{\Sigma}_G & 0 \\ 0 & \mathbf{\Sigma}_{F_i}\end{bmatrix}$ *has distinct eigenvalues.*

   (e) $\mathbf{\Sigma}_{\Lambda_i}\mathbf{\Sigma}_{F_i}$ *has distinct eigenvalues.*

Assumption 1.D.

1. *The global factors are orthogonal to the local factors;* $\mathrm{E}\left(\mathbf{G}_t\mathbf{F}_{it}'\right) = \mathbf{0}$ *for all* i *and* t.

2. *The local factors,* $\mathbf{F}_{1t}, \ldots, \mathbf{F}_{Rt}$ *are mutually uncorrelated; that is,* $\mathrm{E}\left(\mathbf{F}_{mt}\mathbf{F}_{ht}'\right) = \mathbf{0}$ *for all* t *and* $m \neq h$.

Assumption 1.A is an extended version of Assumption C in Bai & Ng (2002), which implies that the idiosyncratic errors are allowed to be serially and (weakly) cross-sectionally correlated. Assumptions 1.B.1–1.B.4 are standard in the literature. Assumption 1.B.5 allows weak correlation between global/local factors and idiosyncratic errors. Assumption 1.C is also standard. Assumption 1.C.2 allows global factors to have non-trivial contributions to the variance of all the individuals while Assumption 1.C.4(c) allows the local factors to have non-trivial contributions to the individual variances within the corresponding block. Assumption 1.D.1 ensures that the global factors and local factors can be separately identified. Initially, we make Assumption 1.D.2, but we will provide an extension in Subsection 1.4.2 where we allow nonzero correlation between the local factors. We focus on the practical case with a fixed number of blocks, R though our approach is still valid even as $R \to \infty$.

1.4 Canonical Correlation-based Model Selection

1.4.1 Estimation of the number of global factors

Using the model (1.3.5), we describe the estimation algorithms as follows: Let $\mathbf{K}_i = [\mathbf{G}, \mathbf{F}_i]$ for $i = 1, \ldots, R$. We first select a sufficiently large and common $r_{\max}$, satisfying $r_{\max} \geq \max\{r_0 + r_1, \ldots, r_0 + r_R\}$. As r_0 and r_i are finite for all $i = 1, \cdots, R$, $r_{\max}$ is also finite and does not necessarily grow with R. We then apply the PC estimation to (1.3.5) for any two blocks, m and h, and obtain the estimates of $\mathbf{K}_m$ and $\mathbf{K}_h$, denoted $\widehat{\mathbf{K}}_m$ and $\widehat{\mathbf{K}}_h$, where $\widehat{\mathbf{K}}_m$ is $\sqrt{T}$ times eigenvectors corresponding to the $r_{\max}$ largest eigenvalues of the $T \times T$ matrix, $\mathbf{Y}_m \mathbf{Y}_m'$, and similarly for $\widehat{\mathbf{K}}_h$. Under Assumptions A–D, $\widehat{\mathbf{K}}_m$ and $\widehat{\mathbf{K}}_h$ contain the factor spaces spanned by $[\mathbf{G}, \mathbf{F}_m]$ and $[\mathbf{G}, \mathbf{F}_h]$, respectively. See Lemma A.1.1 in Appendix A.

Next, we construct the sample variance/covariance matrices for $\widehat{\mathbf{K}}_m$ and $\widehat{\mathbf{K}}_h$ by $\widehat{\mathbf{S}}_{ab}$ $(a, b = m, h)$ and the characteristic equation by

$$\left(\widehat{\mathbf{S}}_{mh}\widehat{\mathbf{S}}_{hh}^{-1}\widehat{\mathbf{S}}_{hm} - \ell\widehat{\mathbf{S}}_{mm}\right)\mathbf{v} = \mathbf{0}. \tag{1.4.6}$$

Let $\ell_{mh,r}$ be the r-th largest characteristic root of (1.4.6), which is the r-th largest sample squared canonical correlation between $\widehat{\mathbf{K}}_m$ and $\widehat{\mathbf{K}}_h$.

Lemma 1.1. *Under Assumptions 1.A–1.D, as $M_m, M_h, T \to \infty$, the sample squared canonical* *correlation, $\ell_{mh,r}$ converges in probability to the population counterpart:*

$$\ell_{mh,r} \xrightarrow{p} \begin{cases} 1 & \text{for } r = 1, \ldots, r_0 \\ 0 & \text{for } r = r_0 + 1, \ldots, r_{\max} \end{cases} \tag{1.4.7}$$

Since the blocks, m and h, share the r_0 global factors, the r_0 characteristic roots from (1.4.6) are equal to one, and the remaining $r_{\max} - r_0$ roots are 0. Hence, $\ell_{mh,r}$ will be close to 1 if $r \leq r_0$, and close to 0 otherwise. As this holds for every block-pair, we construct the cross-block average of the sample squared canonical correlations as

$$\xi(r) = \frac{2}{R(R-1)} \sum_{m=1}^{R-1} \sum_{h=m+1}^{R} \ell_{mh,r}$$

and a canonical correlation difference (CCD) as

$$CCD(r) = \xi(r) - \xi(r+1) \text{ for } r = 0, 1, \ldots, r_{\max}.$$

We then propose to estimate the number of global factors consistently by

$$\hat{r}_{0,CCD} = \underset{0 \leq r \leq r_{\max}}{\arg\max}\ CCD(r).$$

To cover the cases with zero global factor and zero local factor for all $i = 1, \ldots, R$, we set two mock squared canonical correlations, $\ell_{mh,0} = 1$ at the beginning and $\ell_{mh,r_{\max}+1} = 0$ at the end.[4]

We present the asymptotic properties of $\xi(r)$ and CCD in Lemmas 2 and 3.

[4] Ahn & Horenstein (2013) set a mock eigenvalue at the beginning to cover the possibility of zero factor in the 2D model. Hence, we set $\ell_{mh,0} = 1$ to cover the possibility of $r_0 = 0$. Similarly, we may need to set $\ell_{mh,r_{\max}+1} = 0$ to cover the special case where $r_i = 0$ for all $i = 1, ..., R$. For instance, consider $R = 2$ with two global factors and zero local factor for $i = 1, 2$. Then, we find $\ell_{mh,0} = 1$, $\ell_{mh,1} = 1$ and $\ell_{mh,2} = 1$ for $m = 1$ and $h = 2$. Following the practical guideline of selecting the common maximum number of factors by $r^*_{\max} = \max\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_R}\}$ as described in Section 5, we select $r^*_{\max} = 2$ for $i = 1, 2$. Then, we only obtain: $CCD(0) = CCD(1) = 0$ but $CCD(2)$ is undefined such that $r_0 = 2$ cannot be identified. Setting the zero mock canonical correlation at the end ($\ell_{mh,3} = 0$), we obtain $CCD(2) = 1$ and select two global factors. This may not be a unique solution. In the special case where we select the same number of factors for all $i = 1, ..., R$, we may employ $r^*_{\max} + 1$ instead of $r^*_{\max}$. But, it is simpler to set $\ell_{mh,r^*_{\max}+1} = 0$ because the canonical correlation for any redundant factor is asymptotically zero. Of course, we don't need to set the zero mock canonical correlation at the end if we select the different number of factors for $i = 1, ..., R$.

Lemma 1.2. *Under Assumptions 1.A–1.D, as $M_1, ..., M_R, T \to \infty$, then*

$$\xi(r) \xrightarrow{p} \begin{cases} 1 & \text{for } r = 0, ..., r_0 \\ 0 & \text{for } r = r_0 + 1, ..., r_{\max} \end{cases}$$

Lemma 1.2 shows under Assumptions 1.A–1.D that $\xi(r)$ is equal to 1 for $r \leq r_0$ while $\xi(r)$ is 0 for $r > r_0$, asymptotically.

Lemma 1.3. *Suppose that Assumptions 1.A–1.D hold.*

(i) For $r_0 > 0$, as $M_1, ..., M_R, T \to \infty$, then

$$CCD(r) \xrightarrow{p} \begin{cases} 0 & \text{for } r = 0, ..., r_0 - 1 \\ 1 & \text{for } r = r_0 \\ 0 & \text{for } r = r_0 + 1, ..., r_{\max} \end{cases}$$

(ii) For $r_0 = 0$, as $M_1, ..., M_R, T \to \infty$, then

$$CCD(r) \xrightarrow{p} \begin{cases} 1 & \text{for } r = 0 \\ 0 & \text{for } r > 0 \end{cases}$$

The following theorem shows that $\hat{r}_{0,CCD}$ is a consistent model selection criterion.

Theorem 1.1. *Suppose that Assumptions 1.A–1.D hold. Then,*

$$\lim_{M_1,...,M_R,T\to\infty} \Pr(\hat{r}_{0,CCD} = r_0) = 1.$$

It is intuitive to apply a canonical correlation-based approach to identify the number of global factors. Our approach shares the similar idea with AGGR by developing the consistent selection criteria through using the fact that the r_0 canonical correlations are equal to one while the remaining $r_{\max} - r_0$ ones are strictly less than 1. AGGR attempted to derive the asymptotic distribution of the test statistic that is nonstandard due to a parameter being at the boundary and involves a nontrivial bias correction. Only by re-centering and re-scaling of the statistic and by imposing the strong assumption that idiosyncratic errors are neither serially nor cross-sectionally correlated, they can derive that the rescaled test statistic follows the standard normal distribution asymptotically under the null hypothesis, $r_0 = r$. AGGR's approach is developed only for two blocks while our approach easily extends to more than two blocks. Further, we share the similar idea with Onatski (2010) by employing the difference between adjacent canonical correlations as the selection criterion. But, CCD does not require calibrating any threshold because the r_0 largest canonical correlations are all bounded by unity.

1.4.2 Non-zero correlation between the local factors

Kose et al. (2003), Beck et al. (2016), Choi et al. (2018) and Han (2021) assume that the local factors are all mutually uncorrelated. Wang (2008), Breitung & Eickmeier (2016) and Andreou et al. (2019) do not rule out correlation between the local factors. Chen (2012) allows the local factors to be arbitrarily correlated by assuming that both global and local factors are spanned by an aggregate pervasive factor space.

We now allow the local factors to be mutually correlated. Let $\rho_{mh,r}$ be the r-th population canonical correlation between $\mathbf{K}_m$ and $\mathbf{K}_h$. By construction we have: $1 = \rho_{mh,0} = \rho_{mh,1} = \cdots = \rho_{mh,r_0} > \rho_{mh,r_0+1} \geq \cdots \geq \rho_{mh,r_0+r_m} \geq 0 = \rho_{mh,r_0+r_m+1} = \cdots = \rho_{mh,r_{\max}+1}$, where ρ_{mh,r_0+1} is the largest population canonical correlation between local factors in group m and h. Define the block average by $\bar{\rho}_r = \frac{2}{R(R-1)} \sum_{m=1}^{R-1} \sum_{h=m+1}^{R} \rho_{mh,r}$. Then, $\bar{\rho}_{r_0+1}$ represents the largest average canonical correlation between the local factors across R blocks.

We provide the following Lemmas, which are extensions of Lemmas 1.1–1.3 (see the proofs in Section A.9 in Appendix A).

Lemma 1∗. *Under Assumptions 1.A–1.D.1, as $M_m, M_h, T \to \infty$, then the sample squared canonical correlation, $\ell_{mh,r}$, converges in probability to the population counterpart:*

$$\ell_{mh,r} \xrightarrow{p} \begin{cases} 1 & \text{for } r = 0, 1, \ldots, r_0 \\ \rho_{mh,r} & \text{for } r = r_0 + 1, \ldots, r_{\max} \end{cases}$$

Lemma 2∗. *Under Assumptions 1.A–1.D.1, as $M_1, ..., M_R, T \to \infty$, then*

$$\xi(r) \xrightarrow{p} \begin{cases} 1 & \text{for } r = 0, ..., r_0 \\ \bar{\rho}_r & \text{for } r = r_0 + 1, ..., r_{\max} \end{cases}$$

Lemma 3∗. *Suppose that Assumptions 1.A–1.D.1 hold.*

(i) For $r_0 > 0$, as $M_1, ..., M_R, T \to \infty$, then

$$CCD(r) \xrightarrow{p} \begin{cases} 0 & \text{for } r = 0, ..., r_0 - 1 \\ 1 - \bar{\rho}_{r_0+1} & \text{for } r = r_0 \\ \bar{\rho}_r - \bar{\rho}_{r+1} & \text{for } r = r_0 + 1, ..., r_{\max} \end{cases}$$

(ii) For $r_0 = 0$, as $M_1, ..., M_R, T \to \infty$, then

$$CCD(r) \xrightarrow{p} \begin{cases} 1 - \bar{\rho}_1 & \text{for } r = 0 \\ \bar{\rho}_r - \bar{\rho}_{r+1} & \text{for } r > 0 \end{cases} .$$

From Lemma 1∗ we find that the largest population canonical correlation among local factors should be bounded in order to ensure that CCD is maximised at $r = r_0$. Thus, we need to impose a condition, $\bar{\rho}_{r_0+1} < \eta$ for the consistency of CCD, where $\eta = 1 - d_{\max}(r)$ is the upper bound with $d_{\max}(r) = \max_{r_0+1 \le r \le r_{\max}}(\bar{\rho}_r - \bar{\rho}_{r+1})$. It still allows some pairs to have canonical correlation larger than η, but the average across all pairs cannot exceed η.

Theorem 1.2. *Suppose that Assumptions 1.A–1.D.1 hold. Further, we allow non-zero correlations among the local factors and impose the upper bound on the largest average population correlation among the local factors by $\bar{\rho}_{r_0+1} < \eta$ where $\eta = 1 - d_{\max}(r)$ with $d_{\max}(r) = \max_{r_0+1 \le r \le r_{\max}}(\bar{\rho}_r - \bar{\rho}_{r+1})$. Then, we have:*

$$\lim_{M_1,...,M_R,T\to\infty} \Pr(\hat{r}_{0,CCD} = r_0) = 1.$$

Theorem 1.2 implies that if the largest block-average of canonical correlations among the local factors is smaller than η, then CCD is still a consistent selection criterion. We may argue that the correlations between the local factors should not be set too high, because such strong correlations imply that the local factors in block m would directly influence the individuals in block h, and *vice versa*. In such case it may be difficult to distinguish between the roles played by the global and local factors in the multilevel factor model. Notice that the upper bound condition is trivially satisfied if $\bar{\rho}_{r_0+1} < 1/2$.

CCD is very simple to implement without requiring any tuning parameters, but the cost may be the boundedness condition in the presence of nonzero local factors correlation. In this regard we develop the alternative estimator that can remain consistent without imposing the upper bound condition. Notice that $\xi(r) \leq 1$ and $1-\xi(r)$ is monotonically increasing with r. From Lemma 2*, it follows that

$$1-\xi(r) \xrightarrow{p} \begin{cases} 0 & \text{for } r = 0, \ldots, r_0 \\ 1-\bar{\rho}(r) & \text{for } r = r_0+1, \ldots, r_{\max} \end{cases}$$

Let $\delta^2_{\underline{M}T}$ denote the convergence rate of $1-\xi(r)$ such that $\delta^2_{\underline{M}T}(1-\xi(r))$ has a nondegenerate distribution for $r \leq r_0$, where $\delta_{\underline{M}T} = \min\left(\sqrt{\underline{M}}, \sqrt{T}\right)$ and $\underline{M} = \min\{M_1, M_2, \ldots, M_R\}$ (see the proof of Lemma 1*, where we show that $\delta^2_{\underline{M}T}$ is the convergence rate of the canonical correlation). Now, it is easily seen that

$$1-\xi(r) = O_p\left(\delta^{-2}_{\underline{M},T}\right) \text{ for } r \leq r_0.$$

and

$$\Pr(1-\xi(r) > \mathcal{M}) \to 1 \text{ for } r > r_0 \text{ and for some constant } \mathcal{M} > 0.$$

On the basis of this finding, we propose to estimate r_0 by the following modified canonical correlation (MCC) criterion:

$$\hat{r}_{0,MCC} = \max\{0 \leq r \leq r_{\max} : 1-\xi(r) - C \times P(\underline{M}, T) < 0\}$$

where $P(\underline{M}, T)$ is a threshold determined by a function of $\underline{M}$ and T and C is a (data-dependent) tuning constant. As long as $P(\underline{M}, T) \to 0$ and $\delta^2_{\underline{M}T} P(\underline{M}, T) \to \infty$, then $\hat{r}_0$ is consistent for r_0. The MCC estimator can be expressed equivalently as

$$\hat{r}_{0,MCC} = \max\{0 \leq r \leq r_{\max} : \delta^2_{\underline{M},T}(1-\xi(r)) - C \times \delta^2_{\underline{M},T} P(\underline{M}, T) < 0\}.$$

Then, it is easily seen that for $r \leq r_0$,

$$\delta^2_{\underline{M},T}(1-\xi(r)) - C \times \delta^2_{\underline{M},T} P(\underline{M}, T) \xrightarrow{p} O_p(1) - \infty < 0.$$

Hence, for $r \leq r_0$, we expect that $1-\xi(r)$ vanishes faster than $P(\underline{M}, T)$ with a slower rate towards zero such that $1-\xi(r) - CP(\underline{M}, T)$ remains negative. On the contrary, for $r > r_0$, as $N, T \to \infty$, we still have $\Pr\left((1-\xi(r)) - C \times P(\underline{M}, T) > \mathcal{M}\right) = 1$ because $1-\xi(r) \xrightarrow{p} 1-\bar{\rho}(r) > 0$ and $P(\underline{M}, T) \to 0$. The positive value of $1-\bar{\rho}(r)$ dominates the vanishing penalising term, and this confirms the presence of local factors if $1-\xi(r)-CP(\underline{M},T)$ becomes positive.

We now summarise these results in Theorem 1.3.

Theorem 1.3. *Suppose that Assumptions A–D1 hold. Further, we allow non-zero correlations among the local factors, and assume that the following conditions hold: (i)* $P(\underline{M},T)\to 0$ *and (ii)* $\delta^2_{\underline{M},T}P(\underline{M},T)\to\infty$, *where* $\delta_{\underline{M}T}=\min(\sqrt{\underline{M}},\sqrt{T})$ *and* $\underline{M}=\min\{M_1,M_2,\dots,M_R\}$*. Then,*

$$\lim_{\underline{M},T\to\infty}\Pr(\hat{r}_{0,MCC}=r_0)=1.$$

To implement the MCC criterion, we propose the use of the following penalty function:

$$P(\underline{M},T)=\frac{\ln\underline{M}+\ln T}{\sqrt{\underline{M}T}}\ln\ln(\underline{M}T). \tag{1.4.8}$$

that satisfies the condition that $P(\underline{M},T)\to 0$ and $\delta^2_{\underline{M}T}P(\underline{M},T)\to\infty$. In practice, the different penalty functions may lead to the different performance, e.g. Bai & Ng (2002) and Breitung & Pigorsch (2013). We may consider the popular penalty function in BIC_3 given by

$$BIC_3=\frac{\underline{M}+T}{\underline{M}T}\ln(\underline{M}T). \tag{1.4.9}$$

In Section A.4 in Appendix A, we provide the simulation results for MCC using BIC_3 in (1.4.9). Overall, its performance is relatively satisfactory for most cases, but it is outperformed by MCC using $P(\underline{M},T)$ in (1.4.8). Since BIC_3 does not always guarantee consistency,[5] we thus recommend the use of $P(\underline{M},T)$.

Another important issue is that the estimation precision of canonical correlations is adversely affected by the noise-to-signal ratio. If the data is noisier, then we need a larger threshold, especially

[5]Consider an extreme case with $\underline{M}=\exp(T)$. Then, $BIC_3\xrightarrow{p}1$, and $1-\xi(r)-C\times BIC_3<0$ with probability 1 for $r=0,1,\dots,r_{\max}$. This implies that we always overestimate $\hat{r}_0=r_{\max}$ even if the sample size is large. By contrast, $P(\underline{M},T)$ is not subject to this issue.

in small samples. Hence, we propose the following data-dependent tuning constant:[6]

$$C = \exp(\bar{\sigma}_e^2/\bar{\sigma}_y^2)$$

where $\bar{\sigma}_e^2/\bar{\sigma}_y^2$ is the average noise-to-signal ratio,

$$\bar{\sigma}_e^2 = \frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{M_i}\sum_{t=1}^{T}\left(y_{ijt} - \widehat{\boldsymbol{\theta}}_{ij}'\widehat{\mathbf{K}}_{it}\right)^2, \ \bar{\sigma}_y^2 = \frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{M_i}\sum_{t=1}^{T} y_{ijt}^2,$$

$N = \sum_{i=1}^{R} M_i$, $\widehat{\mathbf{K}}_{it}$ are the estimated $r_{\max}$ factors and $\hat{\boldsymbol{\theta}}_{ij}$ the corresponding factor loadings. As C is bounded between 1 and e, it does not affect the asymptotic property of MCC.

1.4.3 Estimation of the number of local factors

Once the number of global factors is consistently estimated by $\hat{r}_0$, the global factors can be consistently estimated by $\widehat{\mathbf{G}} = \widehat{\mathbf{K}}_m \mathbf{V}_m^{\hat{r}_0}$, where $\mathbf{V}_m^{\hat{r}_0}$ is an $r_{\max} \times \hat{r}_0$ matrix consisting of the characteristic vectors associated with the $\hat{r}_0$ largest characteristic roots of (1.4.6). $\widehat{\mathbf{G}}$ from any block-pair would provide a consistent estimator for $\mathbf{G}$, but, in practice, we suggest to use the block-pair that yields the maximum value of $\ell_{mh,1}$.

Next, we concentrate $\widehat{\mathbf{G}}$ out in each block by $\mathbf{Y}_i^{\widehat{G}} = \mathbf{M}^{\widehat{G}}\mathbf{Y}_i$ for $i = 1, \ldots, R$ where $\mathbf{M}^{\widehat{G}} = \mathbf{I}_T - \widehat{\mathbf{G}}(\widehat{\mathbf{G}}'\widehat{\mathbf{G}})\widehat{\mathbf{G}}'$. Then, we apply the existing approaches by Bai & Ng (2002) and Ahn & Horenstein (2013) to $\mathbf{Y}_i^{\widehat{G}}$, with the maximum number of factors set to $r_{i,\max} = r_{\max} - \hat{r}_0$, and estimate the number of the local factors consistently by $\hat{r}_i$.[7] We apply the PC estimation to $\mathbf{Y}_i^{\widehat{G}}$ and obtain $\widehat{\mathbf{F}}_i$ for $i = 1, ..., R$.

Finally, the factor loadings, $\widehat{\gamma}_{ij}$ and $\widehat{\boldsymbol{\lambda}}_{ij}$, can be estimated by the OLS regression of y_{ijt} on $\widehat{\mathbf{G}}_t$ and $\widehat{\mathbf{F}}_{it}$.

[6]Following Hallin & Liška (2007) and Alessi et al. (2010), we have also implemented the subsampling approach to selecting the tuning constant, C such that the selected model becomes a stable function of the second stability interval. But, we have encountered the two crucial issues. First, the subsampling procedure takes a huge amount of time because we need to run the subsampling (at least) 30 times for each candidate of C. For example, if there are 50 grids for C, then we have to evaluate MCC, 1500 times. Second and more importantly, this approach fails to provide the second stability interval for the large samples though it works fine for the small samples. For example, if $R = 10$, $M = 100$ and $T = 100$, we find that the variations of $\hat{r}_0$ from the subsamples become all flat at zeros, implying that we cannot identify r_0. We leave this issue for future research.

[7]Alternatively, we can estimate the number of local factors directly by $\hat{r}_i = \widehat{r_0 + r_i} - \hat{r}_0$. Via (unreported) simulations, we find that our proposed approach outperforms this approach, because the smaller $r_{i,\max}$ can be selected in the sequential approach.

1.4.4 Estimation of global and local factors and loadings

In Sections 1.4.1–1.4.3, we have obtained the consistent estimates, $\hat{r}_0$ and $\hat{r}_i$. Given the initial estimates, $\widehat{\mathbf{G}}$, $\widehat{\mathbf{\Gamma}}_i$, $\widehat{\mathbf{F}}_i$ and $\widehat{\mathbf{\Lambda}}_i$ for $i = 1, ..., R$, we follow a sequential approach by Choi et al. (2018) and update the factors and loadings as follows:

First, construct $\mathbf{Y}^{\widehat{\mathbf{F}}} = \left[\mathbf{Y}_1^{\widehat{\mathbf{F}}}, \ldots, \mathbf{Y}_R^{\widehat{\mathbf{F}}}\right]$ where $\mathbf{Y}_i^{\widehat{\mathbf{F}}} = \mathbf{Y}_i - \widehat{\mathbf{F}}_i\widehat{\mathbf{\Lambda}}_i'$ for $i = 1, ..., R$. We then apply the PC estimation to $\mathbf{Y}^{\widehat{\mathbf{F}}}$, and obtain $\widetilde{\mathbf{G}}$ as $\sqrt{T}$ times the eigenvectors corresponding to the $\hat{r}_0$ largest eigenvalues of the $T \times T$ matrix, $\mathbf{Y}^{\widehat{\mathbf{F}}}\mathbf{Y}^{\widehat{\mathbf{F}}\prime}$. The global factor loadings are then estimated by $\widetilde{\mathbf{\Gamma}}' = T^{-1}\widetilde{\mathbf{G}}'\mathbf{Y}^{\widehat{\mathbf{F}}}$.

Next, for each i, construct $\mathbf{Y}_i^{\widetilde{\mathbf{G}}} = \mathbf{Y} - \widetilde{\mathbf{G}}\widetilde{\mathbf{\Gamma}}_i'$ where $\widetilde{\mathbf{\Gamma}}_i$ is the $T \times M_i$ submatrix of $\widetilde{\mathbf{\Gamma}} = \left[\widetilde{\mathbf{\Gamma}}_1, \ldots, \widetilde{\mathbf{\Gamma}}_R\right]$. The local factors, $\widetilde{\mathbf{F}}_i$ are estimated by $\sqrt{T}$ times the eigenvectors corresponding to the $\hat{r}_i$ largest eigenvalues of the $T \times T$ matrix, $\mathbf{Y}_i^{\widetilde{\mathbf{G}}}\mathbf{Y}_i^{\widetilde{\mathbf{G}}\prime}$. The local factor loadings are then estimated by $\widetilde{\mathbf{\Lambda}}_i' = T^{-1}\widetilde{\mathbf{F}}_i'\mathbf{Y}_i^{\widetilde{\mathbf{G}}}$.

1.5 Monte Carlo Simulation

We construct the multilevel factor model by the following data generating process (DGP):

$$\begin{aligned} y_{ijt} &= \boldsymbol{\gamma}_{ij}'\mathbf{G}_t + \sqrt{h_{i1}}\boldsymbol{\lambda}_{ij}'\mathbf{F}_{it} + \sqrt{\kappa h_{i2}}e_{ijt} \\ &= \sum_{z=1}^{r_0}\gamma_{ijz}G_{tz} + \sqrt{h_{i1}}\sum_{z=1}^{r_i}\lambda_{ijz}F_{itz} + \sqrt{\kappa h_{i2}}e_{ijt} \end{aligned}$$

where we generate global factors/loadings, local factors/loadings and idiosyncratic errors by

$$\mathbf{G}_t = \phi_G\mathbf{G}_{t-1} + \mathbf{v}_t,\ \mathbf{v}_t \sim iidN(\mathbf{0}, \mathbf{I}_{r_0})$$

$$\mathbf{F}_{it} = \phi_F\mathbf{F}_{i,t-1} + \mathbf{w}_t,\ \mathbf{w}_t \sim iidN(0, \mathbf{I}_{r_i})$$

$$\gamma_{ijz} \sim iidN(0,1) \text{ for } z = 1, \ldots, r_0,\ \lambda_{ijz} \sim iidN(0,1) \text{ for } z = 1, \ldots, r_i$$

$$e_{ijt} = \phi_e e_{ij,t-1} + \varepsilon_{ijt} + \beta\sum_{1\le|h|\le8}\varepsilon_{i,j-h,t},\ \varepsilon_{ijt} \sim iidN(0,1)$$

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

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

$$\text{Var}\left(\boldsymbol{\gamma}_{ij}'\mathbf{G}_t\right) = \sum_{z=1}^{r_0} \text{Var}(\gamma_{ijz}G_{tz}) = \frac{r_0}{1-\phi_G^2},$$

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

Following Choi et al. (2018) and Han (2021), we make the variance contribution of each component equalised. For $r_0 > 0$, we set

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

For $r_0 = 0$, we set

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

We consider the following sample sizes: $R \in \{2, 5, 10\}$, $M \in \{20, 50, 100, 200\}$ with $M_1 = \cdots = M_R = M$ and $T \in \{50, 100, 200\}$. The number of replications for each simulation experiment is set at 1,000. We focus on the estimation of r_0, and report the results only for the cases with $\phi_G = \phi_F = 0.5$ to save space (We obtain qualitatively similar results for $\phi_G = \phi_F = 0$).

For comparison, we consider the alternative selection criteria proposed by Chen (2012) and Andreou et al. (2019), denoted by IC_{Chen} and $AGGR$, respectively.[8] When implementing IC_{chen} and $AGGR$ in the simulation, for simplicity, we assume that the true number of factors, $r_0 + r_i$ is known. This prevents us from selecting too many candidate models for IC_{chen}. For $AGGR$, the null hypothesis is sequentially tested from $k = r_0 + r_i$ to 0 until rejected.

It is well-established that if the maximum number of factors is set too high, the redundant factors are likely to be selected.[9] Hence, we propose a practical selection guideline. We first apply BIC_3 to the data $\mathbf{Y}_i$ in each block with a sufficiently large $r_{\max}$ (by fixing $r_{\max} = 10$), and obtain the consistent estimate of $r_0 + r_i$, denoted $\widehat{r_0 + r_i}$ for $i = 1, ..., R$. Then, we select the common maximum number of factors by $r^*_{\max} = \max\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_R}\}$. This procedure selects $r^*_{\max} \leq r_{\max}$, while ensuring that $\Pr(r^*_{\max} \geq r_0 + r_i) \xrightarrow{p} 1$ for all $i = 1, ..., R$.[10] In what follows, we report the simulation

[8] See Section A.8 in Appendix A for the detailed estimation algorithms. Unfortunately, we are unable to implement Han (2021)'s algorithm because his code can only be run on Matlab R2013b and R2014a, but not on the later versions.

[9] Ahn & Horenstein (2013) show via simulations that both BIC_3 and ED estimators are quite sensitive to the choice of $r_{\max}$ in the single level factor model.

[10] In Section A.3 in Appendix A we report the simulation results for CCD and MCC using $r^*_{\max}$ together with the fixed $r_{\max} = 10$. In particular, if idiosyncratic errors are serially correlated, then the impact of the large $r_{\max}$ on the performance of CCD is non-negligible (overestimating r_0 for small T. The performance of MCC is also adversely

results for CCD and MCC obtained by applying the common $r^*_{\max}$ for each block, $i = 1, ..., R$.

In the first experiment, we fix the number of factors as $(r_0, r_i) = (2, 2)$ for $i = 1, \ldots, R$. Panel A of Table 1.1 reports the simulation results for the benchmark case with $(\beta, \phi_e, \kappa) = (0, 0, 1)$. The average of $\hat{r}_0$ over 1,000 replications are reported together with the figures inside the parenthesis, $(O|U)$, indicating the percentage of overestimation and underestimation. For example, $(0|0)$ implies that r_0 is perfectly correctly estimated. Both CCD and MCC perform very well for all the sample sizes. IC_{Chen} performs reasonably well for $R = 2$, but underestimates by detecting only one global factor for $R = 5$ and $R = 10$. $AGGR$ overestimates r_0 if M is small, but its performance improves only for large M and T.

The second case is the same as the first one, except we allow serial correlation and cross-section correlation in idiosyncratic errors by setting $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$. The simulation results presented in Panel B of Table 1.1 demonstrate that the performance of IC_{Chen} and $AGGR$ deteriorates substantially as compared to the first case. In particular, $AGGR$ produces imprecise estimates because their approach is not valid in the case where idiosyncratic errors are serially and/or cross-sectionally correlated (see Assumption A9 and Theorem 2 in $AGGR$). Both CCD and MCC select r_0 correctly in almost all cases while CCD slightly outperforms MCC if M and R are small. In line with our theoretical prediction, the performances of CCD and MCC are mostly invariant to the presence of serially and cross-sectionally correlated idiosyncratic errors.

The third case is a very noisy DGP with $\kappa = 3$ in which the variance share explained by the global factors becomes only 20%, which matches closely with empirical evidence reported in Table 6. The other setups are the same as in the second case. From Panel C of Table 1.1, we find that all approaches are adversely affected, especially if M is small. The performance of $AGGR$ is unreliable in all cases. The performance of IC_{Chen} improves with M or T only for $R = 2$, but it severely underestimates r_0 for $R = 5$ and $R = 10$ even in large samples. CCD underestimates r_0 for small M. MCC tends to overestimate r_0 for small M and small T while underestimating r_0 for small M and large T. The performance of CCD and MCC improves sharply with M or T for all values of R. Overall CCD slightly outperforms.

affected by the presence of both cross-sectional and serial correlation in errors if T is small. On the other hand, both CCD and MCC with $r^*_{\max}$, select the number of global factors correctly even in small samples.

Table 1.1: Average estimates of the number of global factors for Experiment 1 with $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (2, 2)$ and $r^*_{\max} = \max\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_R}\}$

		CCD	MCC	IC_{chen}	$AGGR$	CCD	MCC	IC_{chen}	CCD	MCC	IC_{chen}
					Panel A: $(\beta, \phi_e, \kappa) = (0, 0, 1)$						
M	T		$R = 2$				$R = 5$			$R = 10$	
20	50	1.98(0.6|2.1)	1.98(0|1.8)	2.02(2.9|1.2)	2.83(65.3|2.3)	2(0|0.2)	2(0|0.1)	1(0|100)	2(0|0)	2(0|0)	1(0|100)
50	50	2(0.1|0)	2(0|0)	2(0.2|0)	1.98(0|2.2)	2(0|0)	2(0|0)	1.24(0|75.8)	2(0|0)	2(0|0)	1(0|100)
100	50	2(0|0)	2(0|0)	2(0|0)	2(0|0.4)	2(0|0)	2(0|0)	2(0|0.2)	2(0|0)	2(0|0)	1.02(0|97.6)
200	50	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)
20	100	2(0|0.3)	1.97(0|2.7)	1.95(0.1|5.1)	2.56(53.5|4.7)	2(0|0)	2(0|0)	1(0|100)	2(0|0)	2(0|0)	1(0|100)
50	100	2(0|0)	2(0|0)	2(0|0)	2.29(29.5|0.3)	2(0|0)	2(0|0)	1(0|100)	2(0|0)	2(0|0)	1(0|100)
100	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.79(0|21.5)	2(0|0)	2(0|0)	1(0|100)
200	100	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)	1.29(0|71.3)
20	200	2(0|0)	1.95(0|5.5)	1.92(0|7.9)	2.39(45.9|7.1)	2(0|0)	1.99(0|1.3)	1(0|100)	2(0|0)	2(0|0.4)	1(0|100)
50	200	2(0|0)	2(0|0)	2(0|0)	2.2(19.9|0.4)	2(0|0)	2(0|0)	1(0|100)	2(0|0)	2(0|0)	1(0|100)
100	200	2(0|0)	2(0|0)	2(0|0)	2.09(9|0)	2(0|0)	2(0|0)	1(0|100)	2(0|0)	2(0|0)	1(0|100)
200	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0.3)	2(0|0)	2(0|0)	1(0|100)
					Panel B: $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$						
M	T		$R = 2$				$R = 5$			$R = 10$	
20	50	2.16(13.6|1.7)	2.24(22.9|0)	1.76(0.2|24)	2.61(65.2|16.5)	2(0.4|0.2)	2.2(19.6|0)	1(0|100)	2(0|0)	2.21(21|0)	1(0|100)
50	50	2.03(3|0)	2.01(0.9|0)	2(0.3|0.5)	1.62(0|35)	2(0|0)	2(0|0)	1.11(0|89.3)	2(0|0)	2(0|0)	1(0|100)
100	50	2.02(1.9|0)	2(0.3|0)	2(0|0)	1.88(0|11)	2(0|0)	2(0|0)	1.94(0|6.1)	2(0|0)	2(0|0)	1.01(0|99.4)
200	50	2(0.1|0)	2(0|0)	2(0|0)	1.95(0|4.7)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.93(0|6.7)
20	100	1.99(0|0.8)	1.99(0|1.5)	1.64(0|36.3)	2.26(53.6|25.4)	2(0|0)	2(0|0)	1(0|100)	2(0|0)	2(0|0)	1(0|100)
50	100	2(0|0)	2(0|0)	1.99(0|0.7)	2.18(31.1|11.7)	2(0|0)	2(0|0)	1(0|100)	2(0|0)	2(0|0)	1(0|100)
100	100	2(0|0)	2(0|0)	2(0|0)	1.92(0|8.1)	2(0|0)	2(0|0)	1.5(0|50.1)	2(0|0)	2(0|0)	1(0|100)
200	100	2(0|0)	2(0|0)	2(0|0)	1.98(0|2.3)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.18(0|81.9)
20	200	1.99(0|0.8)	1.86(0|13.5)	1.54(0|46)	2.01(43.9|32.3)	2(0|0)	1.97(0|3)	1(0|100)	2(0|0)	2(0|0.3)	1(0|100)
50	200	2(0|0)	2(0|0)	1.99(0|1.2)	2.02(43.9|16.1)	2(0|0)	2(0|0)	1(0|100)	2(0|0)	2(0|0)	1(0|100)
100	200	2(0|0)	2(0|0)	2(0|0)	2.02(8.2|6.3)	2(0|0)	2(0|0)	1(0|99.7)	2(0|0)	2(0|0)	1(0|100)
200	200	2(0|0)	2(0|0)	2(0|0)	2.02(0|1.9)	2(0|0)	2(0|0)	1.97(0|3.4)	2(0|0)	2(0|0)	1(0|100)
					Panel C: $(\beta, \phi_e, \kappa) = (0.1, 0.5, 3)$						
M	T		$R = 2$				$R = 5$			$R = 10$	
20	50	1.77(23.7|37.8)	2.18(23.3|5.8)	1.65(2.7|37.9)	3.12(96.2|1.5)	1.34(5.5|41.7)	2.22(22.4|0)	1(0|100)	1.33(0.5|37.8)	2.28(28.4|0)	1(0|100)
50	50	1.94(9.8|16.1)	1.89(2.3|13.3)	1.94(7.3|13.4)	0.66(0|95.5)	1.92(0.3|7.4)	1.97(0|2.9)	1(0|100)	1.95(0|4.5)	1.98(0.1|1.7)	1(0|100)
100	50	1.99(6.3|8)	1.92(0.5|8.4)	2.16(16.1|0.3)	0.95(0|84)	1.97(0|3.1)	1.93(0|6.7)	1.07(0|93.2)	1.99(0|1.2)	1.98(0|2.3)	1(0|100)
200	50	1.96(0.5|4.6)	1.95(0|5.4)	2.12(12.4|0)	1.21(0|69)	1.99(0|1.3)	1.98(0|2.2)	1.85(0|14.7)	1.99(0|0.8)	1.98(0|1.8)	1.01(0|99.4)
20	100	1.2(0|50.9)	1.22(0|71.9)	1.29(0.1|71.6)	2.89(92|3.6)	1.34(0|36.8)	1.26(0|26.2)	1(0|100)	1.46(0|28.8)	1.28(0|72)	1(0|100)
50	100	1.81(0.1|16.2)	1.6(0|39)	1.69(0|31.4)	2.21(60.2|25.6)	1.93(0|6.1)	1.67(0|33.1)	1(0|100)	1.96(0|4)	1.73(0|27.1)	1(0|100)
100	100	1.99(0|1.5)	1.93(0|6.6)	2(0|0.4)	1.12(0|75.4)	2(0|0.5)	1.97(3.4|0)	1(0|99.9)	2(0|0.1)	1.97(0|3.2)	1(0|100)
200	100	2(0|0)	2(0|0.1)	2(0|0)	1.55(0|42.8)	2(0|0)	2(0|0.1)	1.74(0|26.5)	2(0|0)	2(0|0)	1(0|100)
20	200	0.99(0|63.8)	0.63(0|97.8)	1.14(0|85.6)	2.78(87.2|6.7)	0.92(0|58.8)	0.7(0|99.7)	1(0|100)	0.96(0|54)	0.78(0|99.9)	1(0|100)
50	200	1.82(0|15.3)	1.21(0|70.9)	1.57(0|43.5)	1.79(44.2|40.4)	1.96(0|3.2)	1.25(0|74.6)	1(0|100)	1.99(0|0.8)	1.21(0|78.7)	1(0|100)
100	200	2(0|0.1)	1.95(0|5)	2(0|0.3)	1.42(21.7|55.5)	2(0|0)	1.99(0|1.4)	1(0|99.6)	2(0|0)	1.99(0|0.8)	1(0|100)
200	200	2(0|0)	2(0|0)	2(0|0)	1.75(0|23.9)	2(0|0)	2(0|0)	1.02(0|98.3)	2(0|0)	2(0|0)	1(0|100)

The average of $\hat{r}_0$ over 1,000 replications is reported together with the figures inside the parenthesis, $(O|U)$, indicating the percentage of overestimation and underestimation. r_0 and r_i are the true number of global factors and true number of local factors in group i. We set $r_1 = r_2 = \cdots = r_R$, where R is the number of groups. M_i is the number of individuals in group i. In Experiments 1, 3 and 4, we set $M_i = M$ for all 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. For IC_{chen} and $AGGR$, we assume that the true number of factors, $r_0 + r_i$ is known. We still allow the estimation uncertainty in implementing CCD and MCC using the $r^*_{\max}$.

In the second experiment we consider the model with uneven block sizes. To this end, we set $(M_1 = 50, M_2 = 100)$ for $R = 2$, $(M_1 = 20, M_2 = 40, M_3 = 60, M_4 = 80, M_5 = 100)$ for $R = 5$, and $(M_1 = 20, M_2 = 30, M_3 = 40, M_4 = 50, M_5 = 60, M_6 = 70, M_7 = 80, M_8 = 90, M_9 = 100, M_{10} = 110)$ for $R = 10$, respectively. The results in Table 1.2 display that CCD performs satisfactory, selecting r_0 precisely in almost all cases. The performance of MCC is comparable to that of CCD, except when the data become noisier. Especially for small T, MCC significantly overestimates r_0 in the presence of cross-sectionally and serially correlated errors together with the higher noise-to-signal ratio. On the other hand, IC_{Chen} underestimates r_0 while $AGGR$ overestimates r_0 in almost all cases.

Table 1.2: Average estimates of the number of global factors for Experiment 2 with uneven block sizes, $(\phi_G, \phi_F) = (0.5, 0.5)$ and $r^*_{\max} = \max\{\widehat{r_0 + r_1}, \dots, \widehat{r_0 + r_R}\}$

β	ϕ_e	κ	T	CCD	MCC	IC_{Chen}	$AGGR$
				Panel A: $R = 2$			
0	0	1	50	2(0.2|0.5)	2(0|0)	1.01(0|99.1)	3.93(99.8|0)
0	0	1	100	2(0|0.2)	2(0|0)	1(0|100)	3.77(98.8|0)
0	0	1	200	2(0|0)	2(0|0)	1(0|100)	3.66(97|0)
0.1	0.5	1	50	2.15(11.8|0.2)	2.03(2.7|0)	1.02(0|98.4)	3.78(97.1|0)
0.1	0.5	1	100	2(0|0.1)	2(0|0)	1(0|100)	3.53(90.5|0.2)
0.1	0.5	1	200	2(0|0.3)	2(0|0)	1(0|100)	3.3(84.2|0.1)
0.1	0.5	3	50	2.21(25.9|15.3)	2(5.3|5.8)	1(0|99.9)	4(100|0)
0.1	0.5	3	100	1.7(0.2|25.6)	1.78(0|21.5)	1(0|100)	4(100|0)
0.1	0.5	3	200	1.6(0|33.7)	1.66(0|33.5)	1(0|100)	3.98(99.9|0)
				Panel B: $R = 5$			
0	0	1	50	2(0|0)	2(0|0)	1(0|100)	
0	0	1	100	2(0|0)	2(0|0)	1(0|100)	
0	0	1	200	2(0|0)	2(0|0)	1(0|100)	
0.1	0.5	1	50	2(0|0)	2.36(35.6|0)	1(0|100)	
0.1	0.5	1	100	2(0|0)	2(0|0)	1(0|100)	
0.1	0.5	1	200	2(0|0)	2(0|0)	1(0|100)	
0.1	0.5	3	50	2(3|3)	2.75(70.8|0)	1(0|100)	
0.1	0.5	3	100	1.96(0|3.5)	1.97(0|3.1)	1(0|100)	
0.1	0.5	3	200	1.94(0|4.8)	1.72(0|27.9)	1(0|100)	
				Panel C: $R = 10$			
0	0	1	50	2(0|0)	2(0|0)	1(0|100)	
0	0	1	100	2(0|0)	2(0|0)	1(0|100)	
0	0	1	200	2(0|0)	2(0|0)	1(0|100)	
0.1	0.5	1	50	2(0|0)	2.55(54.4|0)	1(0|100)	
0.1	0.5	1	100	2(0|0)	2(0|0)	1(0|100)	
0.1	0.5	1	200	2(0|0)	2(0|0)	1(0|100)	
0.1	0.5	3	50	1.99(0|1)	3.1(98.2|0)	1(0|100)	
0.1	0.5	3	100	2(0|0.5)	2(0|0)	1(0|100)	
0.1	0.5	3	200	2(0|0.1)	2(0|0.5)	1(0|100)	

We set $(M_1 = 50, M_2 = 100)$ for $R = 2$, $(M_1 = 20, M_2 = 40, M_3 = 60, M_4 = 80, M_5 = 100)$ for $R = 5$, and $(M_1 = 20, M_2 = 30, M_3 = 40, M_4 = 50, M_5 = 60, M_6 = 70, M_7 = 80, M_8 = 90, M_9 = 100, M_{10} = 110)$ for $R = 10$. See also footnotes to Table 1.1.

In the third experiment we allow the number of global factors to vary from 0 to 3 by setting $(r_0, r_i) \in \{(0,2), (1,1), (3,3)\}$ for $i = 1, ..., R$ and $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$. First, the results for the case with $(r_0, r_i) = (0, 2)$, are reported in Panel A of Table 1.3. CCD, MCC and $AGGR$ tend to select zero global factor correctly, but CCD outperforms if both M and T are small. On the other hand, IC_{chen} always selects one factor incorrectly. Second, turning to the case with $(r_0, r_i) = (1, 1)$ in Panel B of Table 1.3, we find that CCD and IC_{chen} estimate $r_0 = 1$ correctly. If M and T are small, MCC overestimates r_0 while $AGGR$ tends to underestimate r_0. Finally, the results for $(r_0, r_i) = (3, 3)$ presented in Panel C, display that for $R = 2$ the performance of CCD, MCC and IC_{chen} is satisfactory and improves sharply with the sample sizes, but CCD slightly outperforms for small M. By contrast, the performance of $AGGR$ is unreliable unless both M and T are large. Next, for $R = 5$ and 10, the performance of CCD and MCC remains satisfactory whereas IC_{chen} severely underestimates r_0.

,

Table 1.3: Average estimates of the number of global factors for Experiment 3 with $(\phi_G, \phi_F) = (0.5, 0.5)$, $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ and $r^*_{\max} = \max\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_R}\}$

		CCD	MCC	IC_{chen}	$AGGR$	CCD	MCC	IC_{chen}	CCD	MCC	IC_{chen}
					Panel A: $(r_0, r_i) = (0, 2)$						
M	T			$R = 2$			$R = 5$			$R = 10$	
20	50	0.05(3.6|0)	0.9(74.6|0)	1(100|0)	0.33(32.1|0)	0(0|0)	1.04(95.8|0)	1(100|0)	0(0|0)	1.11(99.6|0)	1(100|0)
50	50	0.03(2.1|0)	0.09(9.1|0)	1(100|0)	0(0|0)	0(0|0)	0.02(1.7|0)	1(100|0)	0(0|0)	0.01(0.7|0)	1(100|0)
100	50	0.02(2.1|0)	0.01(0.9|0)	1(100|0)	0(0|0)	0(0|0)	0(0|0)	1(100|0)	0(0|0)	0(0|0)	1(100|0)
200	50	0.01(0.8|0)	0(0|0)	1(100|0)	0(0|0)	0(0|0)	0(0|0)	1(100|0)	0(0|0)	0(0|0)	1(100|0)
20	100	0(0|0)	0(0|0)	1(100|0)	0.25(24.6|0)	0(0|0)	0(0|0)	1(100|0)	0(0|0)	0(0|0)	1(100|0)
50	100	0(0|0)	0(0|0)	1(100|0)	0.15(15|0)	0(0|0)	0(0|0)	1(100|0)	0(0|0)	0(0|0)	1(100|0)
100	100	0(0|0)	0(0|0)	1(100|0)	0(0|0)	0(0|0)	0(0|0)	1(100|0)	0(0|0)	0(0|0)	1(100|0)
200	100	0(0|0)	0(0|0)	1(100|0)	0(0|0)	0(0|0)	0(0|0)	1(100|0)	0(0|0)	0(0|0)	1(100|0)
20	200	0(0|0)	0(0|0)	1(100|0)	0.19(19.2|0)	0(0|0)	0(0|0)	1(100|0)	0(0|0)	0(0|0)	1(100|0)
50	200	0(0|0)	0(0|0)	1(100|0)	0.08(8|0)	0(0|0)	0(0|0)	1(100|0)	0(0|0)	0(0|0)	1(100|0)
100	200	0(0|0)	0(0|0)	1(100|0)	0.03(2.5|0)	0(0|0)	0(0|0)	1(100|0)	0(0|0)	0(0|0)	1(100|0)
200	200	0(0|0)	0(0|0)	1(100|0)	0(0|0)	0(0|0)	0(0|0)	1(100|0)	0(0|0)	0(0|0)	1(100|0)
					Panel B: $(r_0, r_i) = (1, 1)$						
M	T			$R = 2$			$R = 5$			$R = 10$	
20	50	1.09(7.2|0)	1.45(42|0)	1(0|0)	0.93(4.7|11.3)	1(0|0)	1.54(53.4|0)	1(0|0)	1(0|0)	1.66(66.2|0)	1(0|0)
50	50	1.03(2.4|0)	1.02(1.8|0)	1(0|0)	0.92(0|8.1)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
100	50	1.02(2|0)	1(0.4|0)	1(0|0)	0.96(0|3.9)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
200	50	1.01(0.6|0)	1(0|0)	1(0|0)	0.99(0|0.9)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
20	100	1(0|0)	1(0|0)	1(0|0)	0.87(0.4|13.1)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
50	100	1(0|0)	1(0|0)	1(0|0)	0.97(0.1|3.1)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
100	100	1(0|0)	1(0|0)	1(0|0)	0.98(0|1.8)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
200	100	1(0|0)	1(0|0)	1(0|0)	0.99(0|0.6)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
20	200	1(0|0)	1(0|0)	1(0|0)	0.83(0.1|17.6)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
50	200	1(0|0)	1(0|0)	1(0|0)	0.96(0|4.2)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
100	200	1(0|0)	1(0|0)	1(0|0)	0.98(0|2.4)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
200	200	1(0|0)	1(0|0)	1(0|0)	1(0|0.1)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
					Panel C: $(r_0, r_i) = (3, 3)$						
M	T			$R = 2$			$R = 5$			$R = 10$	
20	50	3.16(23.7|14.8)	3.02(9.7|8.2)	2.92(10.5|18.6)	4.9(88.1|7)	2.97(4.1|7.1)	3.03(3|0.5)	1(0|100)	2.99(0.1|1.5)	3.02(2.3|0.1)	1(0|100)
50	50	3.05(7.1|2.9)	2.98(0.7|3.2)	3.04(6.1|2.1)	2.1(0.6|74.1)	3(0.2|0.4)	2.99(0|0.7)	1.17(0|99.7)	3(0|0)	3(0|0.1)	1(0|100)
100	50	3.01(1.4|0.4)	2.99(0.2|0.9)	3.03(3.3|0)	2.58(0.7|40.7)	3(0|0.1)	3(0|0)	2.7(0|28.2)	3(0|0.1)	3(0|0.2)	1.01(0|100)
200	50	3(0.4|0.2)	3(0|0.2)	3.02(1.6|0)	2.82(0.4|18.7)	3(0|0)	3(0|0)	3(0|0.1)	3(0|0)	3(0|0)	2.64(0|32.3)
20	100	2.78(0.2|19)	2.38(0|57.7)	2.54(0.5|44.3)	4.25(76.9|13.9)	2.94(0|5.7)	2.45(0|55)	1(0|100)	2.98(0|1.5)	2.54(0|46)	1(0|100)
50	100	2.99(0|0.8)	2.93(0|7)	2.96(0|3.9)	3.72(55.6|20.7)	3(0|0)	2.98(0|1.6)	1(0|100)	3(0|0)	3(0|0.5)	1(0|100)
100	100	3(0|0)	3(0|0)	3(0|0)	2.66(0|30.9)	3(0|0)	3(0|0)	1.89(0|78)	3(0|0)	3(0|0)	1(0|100)
200	100	3(0|0)	3(0|0)	3(0|0)	2.91(0|8.8)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	3(0|0)	1.21(0|98.3)
20	200	2.71(0|23.6)	1.78(0|91)	2.29(0|64.2)	3.88(69|20.3)	2.95(0|4.7)	1.98(0|96.8)	1(0|100)	2.99(0|1.1)	1.97(0.7|0)	1(0|100)
50	200	3(0|0.2)	2.81(0|19.2)	2.94(0|6.5)	3.29(40.1|27.6)	3(0|0)	2.91(0|8.6)	1(0|100)	3(0|0)	2.98(0|2.5)	1(0|100)
100	200	3(0|0)	3(0|0)	3(0|0)	3.21(22.6|14.1)	3(0|0)	3(0|0)	1.01(0|100)	3(0|0)	3(0|0)	1(0|100)
200	200	3(0|0)	3(0|0)	3(0|0)	2.95(0|4.9)	3(0|0)	3(0|0)	2.87(0|12.6)	3(0|0)	3(0|0)	1(0|100)

See footnotes to Table 1.1.

In the fourth experiment we use the same DGP in the benchmark experiment but allow the local factors to be mutually correlated. We generate the local factors by

$$\mathbf{F}_t = \mathbf{\Phi}_F \mathbf{F}_{t-1} + \mathbf{w}_t,\ \mathbf{w}_t \sim iidN(0, \mathbf{\Omega}_F)$$

where $\mathbf{F}_t = [\mathbf{F}'_{1t}, \ldots, \mathbf{F}'_{Rt}]'$, $\mathbf{w}_t = [\mathbf{w}'_{1t}, \ldots, \mathbf{w}'_{Rt}]'$ and $\mathbf{\Phi}_F$ is a diagonal matrix with the common elements, 0.5. We set the common diagonal elements of $\mathbf{\Omega}_F$ at 1, and the common off-diagonal elements (denoted ω_F) at 0.2, 0.4, 0.6 and 0.8, respectively. We report these results in Table 1.4.[11] If the correlation among the local factors are relatively weak, i.e. $\omega_F = (0.2, 0.4)$, then the performance of CCD is satisfactory, and improves sharply with M and T. However, if the local factors correlation becomes stronger, i.e. $\omega_F = (0.6, 0.8)$, then CCD overestimates r_0 even in large samples. This is line with Theorem 1.2 that consistency of CCD requires the upper bound condition, $\bar{\rho}_{r_0+1} < \eta$ to be met. Next, we find that the performance of MCC is satisfactory for $\omega_F = (0.2, 0.4)$. Even if $\omega_F = 0.6$, its performance improves sharply with M and T. Only in the presence of the stronger correlation among the local factors ($\omega_F = 0.8$), MCC tends to overestimate r_0 in most sample sizes, but it becomes consistent for substantially large M and T. This is line with Theorem 1.3.[12]

[11] The performances of IC_{Chen} and $AGGR$ are qualitatively similar to those in the first experiment. These results are available upon request.

[12] In Section A.5 in Appendix A, we have conducted the additional simulations to examine the performance of CCD and MCC under experiments with heterogeneous correlations among local factors and uneven block sizes. We have obtained qualitatively similar results.

Table 1.4: Average estimates of the number of global factors for Experiment 4 with correlated local factors, $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (2, 2)$, $(\beta, \phi_e, \kappa) = (0, 0, 1)$ and $r^*_{\max} = \max\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_R}\}$

			$\omega_F = 0.2$		$\omega_F = 0.4$		$\omega_F = 0.6$		$\omega_F = 0.8$	
R	M	T	CCD	MCC	CCD	MCC	CCD	MCC	CCD	MCC
2	20	50	2.02(4.9|2.8)	2(0.9|1)	2.26(29.5|2.9)	2.1(10.9|0.6)	2.75(75.5|1)	2.57(56.8|0.1)	2.97(96.9|0.1)	2.93(93.4|0)
2	50	50	2.01(1.3|0.1)	2(0.1|0)	2.23(23.4|0)	2.04(4.3|0.1)	2.79(78.4|0)	2.45(44.8|0)	3(99.8|0)	2.97(97.1|0)
2	100	50	2.01(1|0)	2(0|0.1)	2.19(19.2|0)	2.01(1.3|0)	2.76(76.1|0.1)	2.25(25.2|0.1)	3(99.7|0)	2.93(92.7|0)
2	200	50	2(0.3|0)	2(0|0)	2.18(17.6|0)	2(0.3|0.1)	2.74(73.8|0)	2.11(10.5|0)	3(100|0)	2.79(78.9|0)
2	20	100	2(0|0.2)	1.98(0|1.7)	2.2(21.5|1)	1.99(0.2|1.1)	2.85(85.8|0.5)	2.17(17.8|0.4)	2.99(99.2|0)	2.88(87.9|0)
2	50	100	2(0|0)	2(0|0)	2.1(10.3|0)	2(0|0)	2.84(84.3|0)	2.1(10|0.1)	3(100|0)	2.94(94.1|0)
2	100	100	2(0|0)	2(0|0)	2.07(6.8|0)	2(0|0)	2.82(81.7|0)	2.02(1.5|0)	3(100|0)	2.85(84.5|0)
2	200	100	2(0|0)	2(0|0)	2.07(6.5|0)	2(0|0)	2.8(80.2|0)	2(0|0)	3(100|0)	2.58(58.2|0)
2	20	200	2(0|0.3)	1.96(0|4)	2.07(9.3|1.7)	1.93(0|7.3)	2.89(90.1|0.7)	1.93(0.1|6.6)	2.99(99.3|0)	2.58(58.9|0.7)
2	50	200	2(0|0)	2(0|0)	2.02(1.5|0)	2(0|0)	2.9(90.4|0)	2(0.1|0)	3(100|0)	2.79(79|0)
2	100	200	2(0|0)	2(0|0)	2.01(1.4|0)	2(0|0)	2.9(89.6|0)	2(0|0)	3(100|0)	2.44(43.5|0)
2	200	200	2(0|0)	2(0|0)	2.01(1.1|0)	2(0|0)	2.88(87.9|0)	2(0|0)	3(100|0)	2(0|0)
5	20	50	2(0.1|0.5)	2(0|0)	2.19(19.8|0.8)	2.02(2.2|0)	2.87(87|0.2)	2.58(57.8|0)	3(100|0)	2.98(98.2|0)
5	50	50	2(0|0)	2(0|0)	2.12(11.7|0)	2(0.4|0)	2.88(87.6|0)	2.39(39.3|0)	3(100|0)	2.99(99.4|0)
5	100	50	2(0|0)	2(0|0)	2.08(7.5|0)	2(0|0)	2.86(86.1|0)	2.17(16.6|0)	3(100|0)	2.97(97.4|0)
5	200	50	2(0|0)	2(0|0)	2.08(7.7|0)	2(0|0)	2.83(83.4|0)	2.02(1.9|0.1)	3(100|0)	2.9(89.6|0)
5	20	100	2(0|0)	2(0|0.2)	2.1(10.5|0.1)	2(0|0)	2.95(94.5|0)	2.08(8.1|0)	3(100|0)	2.96(96|0)
5	50	100	2(0|0)	2(0|0)	2.03(2.8|0)	2(0|0)	2.92(92.3|0)	2.02(1.7|0)	3(100|0)	2.98(98.1|0)
5	100	100	2(0|0)	2(0|0)	2.01(1|0)	2(0|0)	2.9(89.6|0)	2(0|0)	3(100|0)	2.93(93.1|0)
5	200	100	2(0|0)	2(0|0)	2.01(0.8|0)	2(0|0)	2.89(88.7|0)	2(0|0)	3(100|0)	2.57(56.6|0)
5	20	200	2(0|0)	1.99(0|0.8)	2.02(1.8|0)	1.99(0|0.8)	2.99(99|0)	1.99(0|0.8)	3(100|0)	2.62(37.8|0)
5	50	200	2(0|0)	2(0|0)	2(0|0)	0(0.2|100)	2.97(97.2|0)	2(0|0)	3(100|0)	2.83(82.6|0)
5	100	200	2(0|0)	2(0|0)	2(0.2|0)	2(0|0)	2.94(94.2|0)	2(0|0)	3(100|0)	2.44(43.6|0)
5	200	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.94(94.2|0)	2(0|0)	3(100|0)	2.03(2.6|0)
10	20	50	2(0|0.3)	2(0|0)	2.18(18.4|0.3)	2.01(1|0)	2.89(89.4|0)	2.6(60|0)	3(99.8|0)	2.99(99.1|0)
10	50	50	2(0|0)	2(0|0)	2.07(6.8|0)	2(0.2|0)	2.89(88.6|0)	2.38(38.1|0)	3(100|0)	3(99.7|0)
10	100	50	2(0|0)	2(0|0)	2.05(5.3|0)	2(0|0)	2.88(87.8|0)	2.1(10.1|0)	3(100|0)	2.99(98.5|0)
10	200	50	2(0|0)	2(0|0)	2.05(4.5|0)	2(0|0)	2.85(85.2|0)	2.01(1.3|0)	3(100|0)	2.89(89.3|0)
10	20	100	2(0|0)	2(0|0)	2.08(7.6|0)	2(0|0.1)	2.97(96.6|0)	2.05(5.3|0)	3(100|0)	2.99(98.7|0)
10	50	100	2(0|0)	2(0|0)	2.02(1.7|0)	2(0|0)	2.93(93.1|0)	2.01(0.8|0)	3(100|0)	2.99(98.6|0)
10	100	100	2(0|0)	2(0|0)	2(0.3|0)	2(0|0)	2.93(92.5|0)	2(0|0)	3(100|0)	2.95(94.8|0)
10	200	100	2(0|0)	2(0|0)	2.01(0.6|0)	2(0|0)	2.91(91.4|0)	2(0|0)	3(100|0)	2.59(59|0)
10	20	200	2(0|0)	2(0|0.2)	2.02(1.7|0)	2(0|0.5)	2.99(99.1|0)	2(0|0.2)	3(100|0)	2.67(67|0)
10	50	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.97(96.9|0)	2(0|0)	3(100|0)	2.86(85.5|0)
10	100	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.97(97.2|0)	2(0|0)	3(100|0)	2.4(40.2|0)
10	200	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.96(96.1|0)	2(0|0)	3(100|0)	2.01(1.2|0)

We generate the local factors by $\mathbf{F}_t = \mathbf{\Phi}_F \mathbf{F}_{t-1} + \mathbf{w}_t$ with $\mathbf{w}_t \sim iidN(0, \mathbf{\Omega}_F)$, where $\mathbf{F}_t = [\mathbf{F}'_{1t}, \ldots, \mathbf{F}'_{Rt}]'$, $\mathbf{w}_t = [\mathbf{w}'_{1t}, \ldots, \mathbf{w}'_{Rt}]'$ and $\mathbf{\Phi}_F$ is a diagonal matrix with the common elements, 0.5. We set the common diagonal elements of $\mathbf{\Omega}_F$ at 1, and the common off-diagonal elements (denoted ω_F) at 0.2, 0.4, 0.6 and 0.8, respectively. See also footnotes to Table 1.1.

In Section A.6 in Appendix A, we have conducted the additional simulations for estimating the number of the local factors, after r_0 is consistently estimated by CCD and MCC. Overall results suggest that BIC_3 by Bai & Ng (2002) and ER by Ahn & Horenstein (2013) outperform the other existing approaches.

Finally, in Section A.7 in Appendix A, we follow the anonymous referee's suggestion and split the whole data with $R > 2$ groups into the two wide groups. This simple modification enables us to apply the $AGGR$'s procedure for estimating the number of global factors even if $R > 2$. Furthermore, this scheme may improve the finite sample performance of CCD and MCC estimators by increasing the number of cross-section observations used in the estimation of the number of global factors, r_0 and the global factors, $\mathbf{G}$. We explore the performance of $AGGR$, CCD and MCC with the two wide-group division, denoted respectively by $AGGR^w$, CCD^w and MCC^w, via additional Monte Carlo experiments. We consider the same DGP employed under Experiments 1 and 3, and draw the three main conclusions. First, we can apply the $AGGR$ approach to the multilevel panel with $R > 2$, though its performance becomes satisfactory only if both M and T are substantially large. But, its performance is unreliable, especially if T is small. Second, CCD and MCC still outperform CCD^w, MCC^w and $AGGR^w$ in most cases. Third, there is a trade-off between the use of more cross-section observations and a selection of the larger $r^*_{\max}$. We find that CCD^w and MCC^w can significantly improve the estimation precision of r_0 for the multilevel panel with $R > 2$, especially if T is sufficiently large and M is much smaller than T. On the other hand, if T is small, then CCD^w and MCC^w overestimate r_0. Hence, we may recommend this 2-wide groups modification in practice, only if T is sufficiently large and M is much smaller than T.

Overall, the simulation results demonstrate that CCD and MCC tend to select the number of global factors correctly even in small samples while outperforming other existing methods even in the presence of serially correlated and weakly cross-sectionally correlated idiosyncratic errors. Only if the correlations among the local factors are deemed to be relatively weak on average (say, less than 1/2), we recommend the use of CCD because it is very simple to implement without requiring any tuning parameter and its performance is robust against noisier idiosyncratic errors. Given that the overall performances of CCD and MCC are qualitatively similar but MCC does not need to meet the upper bound condition, in general, we prefer the use of MCC.

1.6 Empirical Application

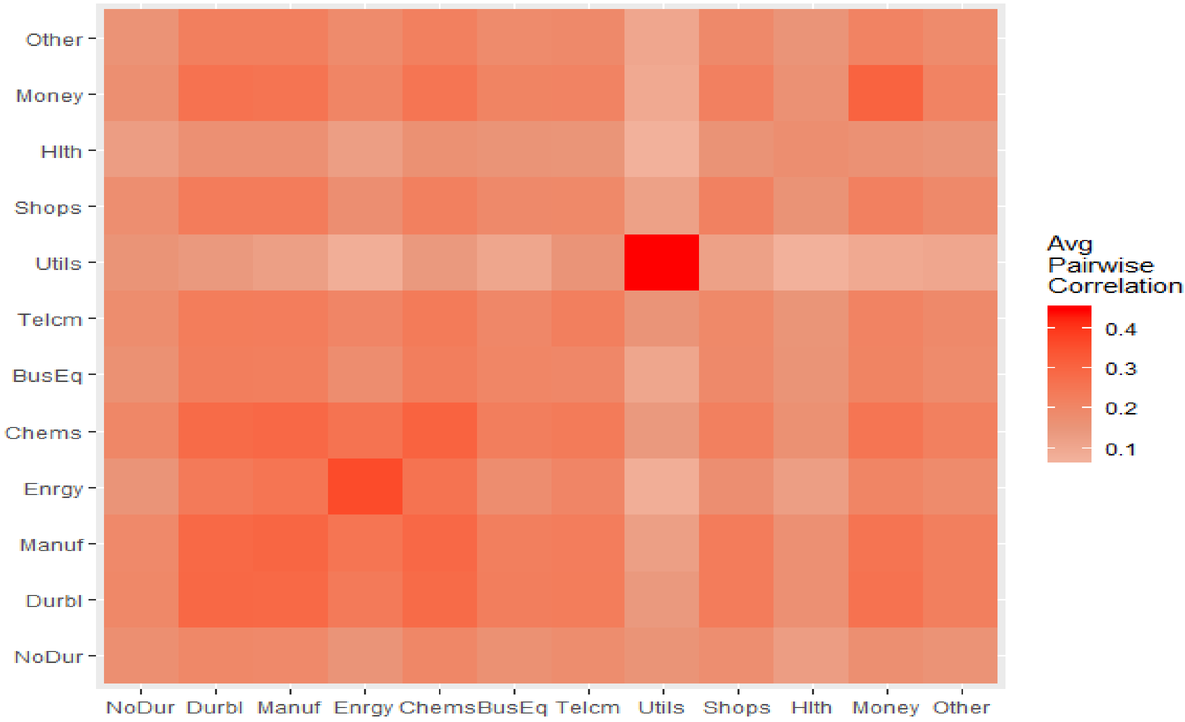
We demonstrate the utility of our approach in the context of the multilevel asset pricing model. The standard literature on asset pricing models suggests a linear relation between stock returns and common factors, e.g. Sharpe (1964), Connor & Korajczyk (1988) and Fama & French (1993). However, the studies investigating the role of industry factors explicitly in asset pricing model are relatively few. Fama & French (1997) provide evidence that both CAPM and the three factor models are unable to precisely estimate the cost of equity for industry portfolios. Lewellen et al. (2010) demonstrate that the asset pricing models are rejected for industry portfolios. Chou et al. (2012) find that the residuals of stocks from the same industry share a non-negligible correlation even after controlling for the common factors. Moskowitz & Grinblatt (1999) find that industry momentum contributes substantially to the momentum strategy such that the winners and the losers tend to belong to the same industry. These studies reveal the fact that stocks in the same industry share a strong co-movement, which cannot be explained by the common factors alone. In this regard, it would be an important issue of investigating whether there is any industry-specific factor driving the within-industry co-movement as well as how important they are relative to global factors and idiosyncratic disturbances.

We collect the weekly return data of stocks listed on NYSE and NASDAQ from Jan. 2015 to Dec. 2016 from CRSP database. We follow Fama & French (1997) and use the SIC codes to categorise the stocks into twelve industries, listed in the first column of Table 1.5.[13] We consider a balanced block panel data with unequal block sizes and include stocks that have the complete return data during the sample period. Following Fama & French (1993) we require the stocks to be listed on NYSE and NASDAQ for two years prior to Jan. 2015. We end up with twelve industries ($R = 12$), 2618 firms ($N = 2618$) and 105 weeks ($T = 105$). The number of stocks in each industry is reported in the second column of Table 1.5.

We first report the within correlations and between correlations. The former is evaluated as the average pairwise correlation of individual stock returns within the same industry while the latter is the average correlation between individual returns across two different industries. We visualise them through a heat map in Figure 1.1, where the diagonal elements represent the within correlations and the off-diagonal elements are the between correlations. Both correlations are positive and substantial across all industries. Overall, the within correlation is higher than the between correlation for all industries. For example, for Enrgy, Utils and Money, the within correlations are 0.36, 0.45 and 0.31, and the between correlations are 0.19, 0.11 and 0.21. Such differences imply that there may be some local/industry factors, rendering the assets co-move within the same industry.

[13]These are Consumer Non-Durable, Consumer Durable, Manufacturing, Energy, Chemicals, Business Equipment, Telecommunication, Utilities, Shops, Health, Money and Others. The definitions of the industries can be found on Kenneth French's website: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

Figure 1.1: Average pairwise correlations of returns



Next, we explore the correlation structure using the multilevel factor model. We standardise the data following Bai & Ng (2002) and Ahn & Horenstein (2013). First, we follow the practical guideline for $r^*_{\max}$ as described in Section 1.5. In our application we only need to run BIC_3 12 times using $r_{\max} = 10$ for $i = 1, ..., 12$, and select $r^*_{\max} = \max\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_{12}}\} = 3$. We then apply CCD and MCC with $r^*_{\max} = 3$. Both select only one global factor, which is in line with Trzcinka (1986) and Bailey et al. (2021).[14] Then, we apply BIC_3 with $r^*_{i,\max} = r^*_{\max} - \hat{r}_0 = 2$ to the defactored data in each block by concentrating out the global factor. We find that there is one local factor in NoDur, Enrgy, Hlth and Money, two local factors in Utils, and zero factor in other industries. Finally, we apply the estimation method described in Section 1.4.4, and report the full estimation results in Table 1.5.

[14] For the robustness check, we have tried the different values of $r_{\max} = 5, 10, 20$ directly applied to CCD and MCC, finding that they always select one global factor.

Table 1.5: The main empirical results

	M_i	$\hat{r}_i$	RI_G	RI_F	RI_E
NoDur	131	1	0.165	0.093	0.737
Durbl	63	0	0.328	0	0.672
Manuf	244	0	0.321	0	0.679
Enrgy	92	1	0.199	0.232	0.562
Chems	67	0	0.3	0	0.7
BusEq	368	0	0.222	0	0.778
Telcm	69	0	0.221	0	0.779
Utils	79	2	0.083	0.542	0.373
Shops	242	0	0.222	0	0.778
Hlth	240	1	0.105	0.096	0.776
Money	525	1	0.274	0.101	0.602
Other	498	0	0.214	0	0.786
Avg/Total	2618		0.226	0.058	0.708

M_i is the number of firms in each industry. $\hat{r}_i$ is the estimated number of local factors. RI_G, RI_F and RI_E stand for the relative importance ratios for the global, local factors and idiosyncratic components, respectively.

We evaluate the relative importance ratios of the global factor, the local/industry factors and idiosyncratic errors,[15] that are summarised in columns 4 - 6 in Table 1.5. On average, the global factor and local factors can explain 22.6% and 5.8% of the total variation whereas idiosyncratic disturbance components still account for 70.6% of the total variation. The global factor tends to display the higher relative importance ratios for the cyclical industries such as Durbl (32.8%), Manuf (32.1%), Chems (30%) and Money (27.4%), suggesting that the higher within correlations observed in these industries are likely to reflect the higher loadings to the global factor. On the other hand, the influence of the global factor is below average for the non-cyclical industries such as NoDur (16.5%), Utils (8.3%) and Hlth (10.5%). Interestingly, local factors are more important than the global factor for Enrgy (23.2%) and Utils (54.2%). The variance share explained by the local factors are also non-negligible for NoDur (9.3%), Hlth (9.6%) and Money (10.1%).

Next, we examine the within and between correlations after concentrating out the global and local factors, respectively. Figure 1.2 displays the results constructed using the residuals from a regression of the return data on the global factor only. In contrast to Figure 1.1, the between correlations decline drastically for all industries, indicating that the market-wide co-movement of the individual stock returns is well-captured by the global factor. Notice, however, that the within correlations for NoDur, Enrgy, Utils, Hlth and Money are still non-negligible, which implies that such co-movements may be captured by the local factors. We further project out the local factors such that the resulting residuals would be purely idiosyncratic. Figure 1.3 shows that both correlations are almost negligible, suggesting that the local/industry factors are an important driver behind the higher within correlations for NoDur, Enrgy, Utils, Hlth and Money.

[15] The time series variance decomposition for the individual stock return is given by

$$\mathrm{Var}(y_{ijt}) = \mathrm{Var}\left(\widehat{\boldsymbol{\gamma}}_{ij}'\widehat{\mathbf{G}}_t\right) + \mathrm{Var}\left(\widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\mathbf{F}}_{it}\right) + \mathrm{Var}\left(\hat{e}_{ijt}\right)$$

We construct the relative importance ratios for each industry by

$$IRG_i = M_i^{-1}\sum_{j=1}^{M_i}\frac{\mathrm{Var}\left(\widehat{\boldsymbol{\gamma}}_{ij}'\widehat{\mathbf{G}}_t\right)}{\mathrm{Var}(y_{ijt})},\quad IRF_i = M_i^{-1}\sum_{j=1}^{M_i}\frac{\mathrm{Var}\left(\widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\mathbf{F}}_{it}\right)}{\mathrm{Var}(y_{ijt})} \text{ and } IRE_i = M_i^{-1}\sum_{j=1}^{M_i}\frac{\mathrm{Var}(\hat{e}_{ijt})}{\mathrm{Var}(y_{ijt})}.$$

The average relative importance ratios across the market for these three components can be evaluated as

$$\overline{IRG} = N^{-1}\sum_{i=1}^{R}\sum_{j=1}^{M_i}\frac{\mathrm{Var}\left(\widehat{\boldsymbol{\gamma}}_{ij}'\widehat{\mathbf{G}}_t\right)}{\mathrm{Var}(y_{ijt})},\quad \overline{IRF} = N^{-1}\sum_{i=1}^{R}\sum_{j=1}^{M_i}\frac{\mathrm{Var}\left(\widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\mathbf{F}}_{it}\right)}{\mathrm{Var}(y_{ijt})} \text{ and } \overline{IRE} = N^{-1}\sum_{i=1}^{R}\sum_{j=1}^{M_i}\frac{\mathrm{Var}(\hat{e}_{ijt})}{\mathrm{Var}(y_{ijt})}.$$

Figure 1.2: Average pairwise correlations of residuals after concentrating out $\widehat{\mathbf{G}}$

Figure 1.3: Average pairwise correlations of residuals after concentrating out $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{F}}_i$'s

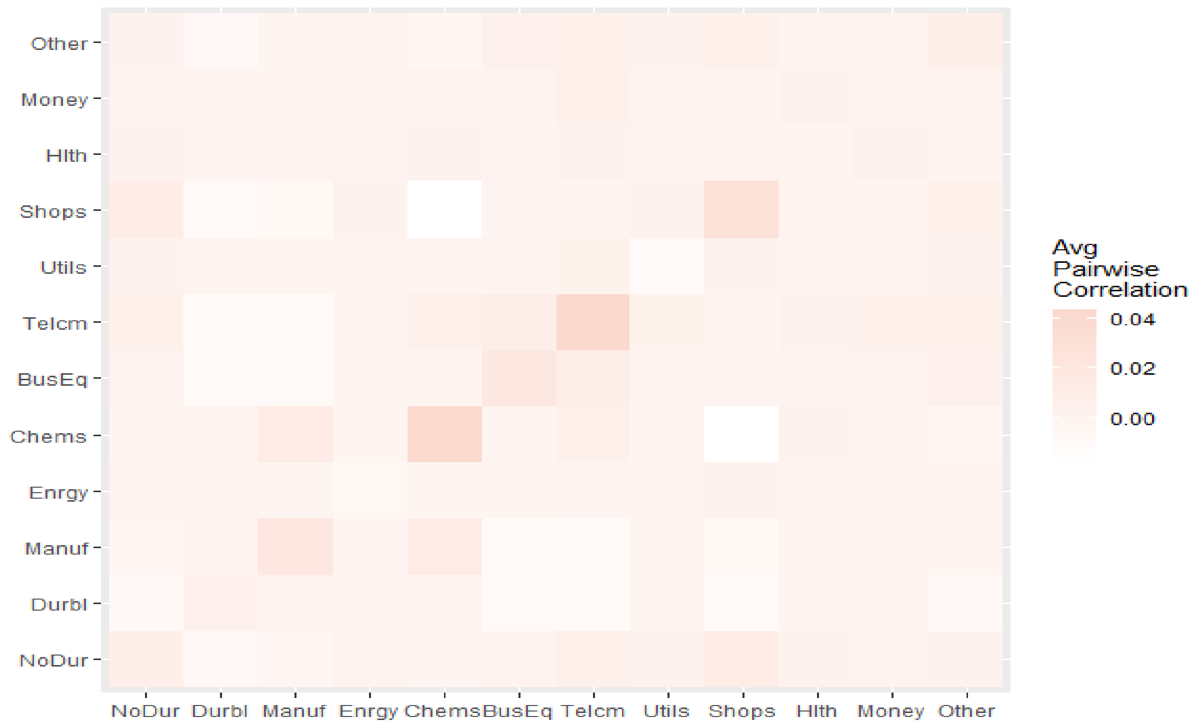
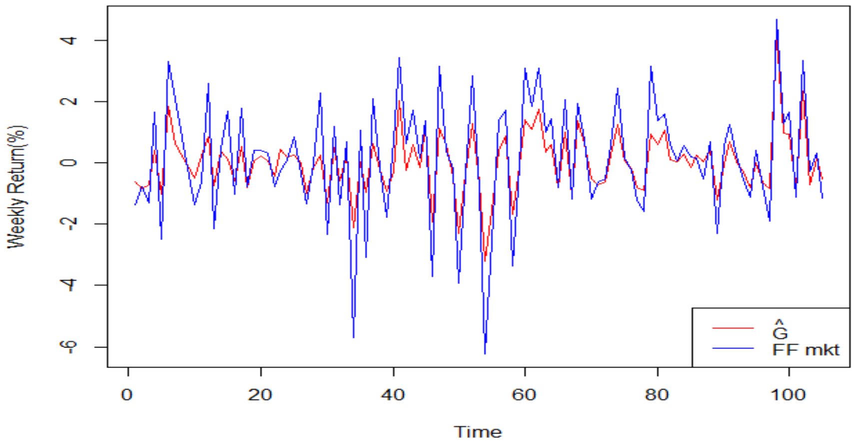


Figure 1.4 displays that the estimated global factor co-moves closely with the market factor

with correlation of 0.95,[16] though the latter is slightly more volatile. This is a well-known result since Brown (1989) that the market index plays a predominant role in the asset pricing model. However, it is more challenging to find out which financial indicators measuring local economic and financial conditions, can be connected closely to the local/industry factors. For example, we find that the local factor in Enrgy is highly correlated with the changes in WTI (an oil price index) with the correlation of 0.7. Further, we observe that the average (absolute) pairwise correlation among the local/industry factors is 0.21, which may provide an empirical support for the upper bound condition imposed in Theorem 1.2.

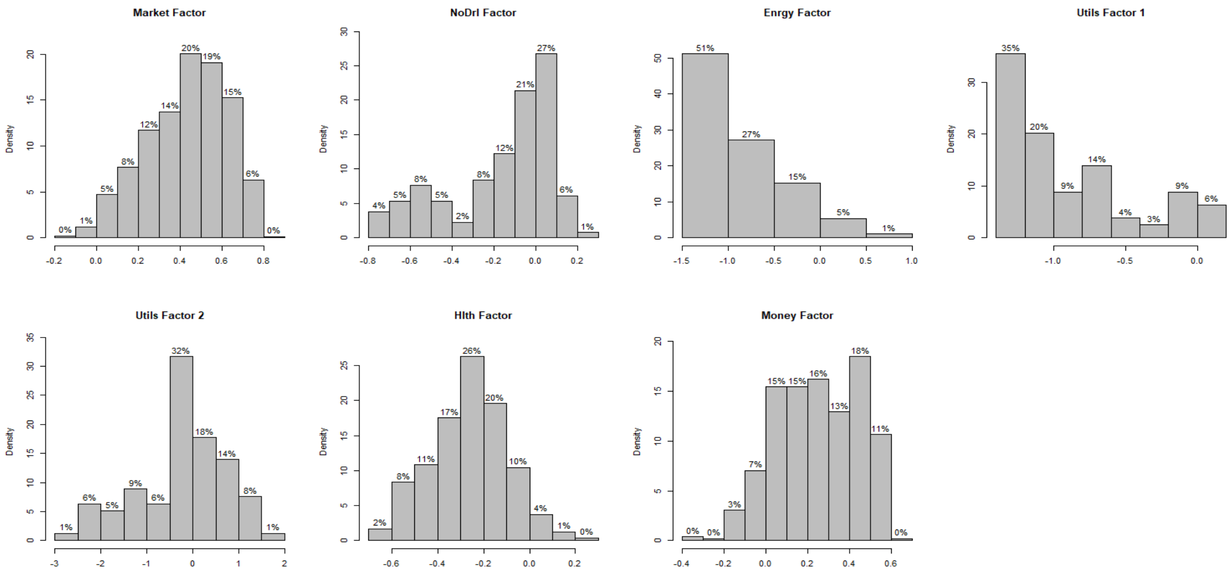
Figure 1.4: The global factor and market factor



Finally, in Figure 1.5, we plot the density of the factor loadings associated with one global factor and with six local factors. As the estimated factors/loadings are subject to a rotation and sign indeterminacy, we focus on whether the loadings have the same sign or not. The same sign indicates that the returns co-move with the corresponding factors, and *vice versa*. First, almost all individual stock returns are positively loaded on the global factor, suggesting that they co-move with the global factor. Next, turning to the local factor loadings, we find that the majority of the stock returns in NoDrl, Enrgy, Money, and Hlth are loaded wih the same sign. In Utils with two local factors, the majority of the returns are negatively loaded on the first factor while they are symmetric around 0 for the second factor. This confirms that the local/indistry factors are an important source of the within-industry co-movement.

[16] We download the weekly data of the Fama-French three factors from the Kenneth French Website.

Figure 1.5: Density plots of the global and local factor loadings



1.7 Conclusion

We have developed a novel procedure for identifying the number of the global factors and the number of the local factors jointly in a multilevel factor model. We first apply the principal component (PC) estimation to the data in each block and estimate the factors. We then evaluate the canonical correlations between factors in any two blocks and develop the canonical correlations difference (CCD) and the modified canonical correlations (MCC) criteria.

We show that both CCD and MCC are a consistent model selection criterion. Via Monte Carlo simulations, we demonstrate that CCD and MCC consistently select the number of global factors even in small samples. Further, they outperform other competing approaches even in the presence of serially correlated and weakly cross-sectionally correlated errors as well as the correlated local factors. We have also considered the simple modification by splitting the whole data with $R > 2$ groups into the two wide groups. We find that this modification can improve the estimation precision of r_0, especially if T is sufficiently large and the number of individuals in each group is much smaller than T.

We demonstrate the utility of our approach with an application to the multilevel asset pricing model for the weekly stock return data of twelve industries in the U.S. over the period, Jan. 2015 to Dec. 2016. By applying CCD and MCC, we find that there is only one global factor, which co-moves closely with the market factor. Next, by applying BIC_3, we find that the local factors explain non-trivial proportions of the return variations in 5 out of 12 industries.

We note in passing that the global factors can be common only to the blocks within a region, say emerging or advanced markets, e.g. Hallin & Liška (2011) and Chen (2012), which may be empirically more relevant. This factor structure can be regarded as the multilevel model with the regional factors rather than the global factors. This is similar to the three-level or overlapping factor models considered by Breitung & Eickmeier (2016) and Beck et al. (2016). Our approach can be easily extended to these cases given that the block membership within different layers is known.

In principle, if the (unknown) group membership as well as the number of the groups are (consistently) estimated using any exiting approaches (e.g. Su et al. (2016) and Ando & Bai (2017)), then we can apply our proposed section criteria to consistently estimate the number of global factors and the number of local factors in each group, jointly. Notice that there is a growing literature on weak factor model that is closely related to the multilevel factor model. In the 2-dimensional model, weak factors are harder to detect than strong factors. A number of recent papers have developed some novel but complex techniques, e.g. Lettau & Pelger (2020), Bailey et al. (2021) and Uematsu & Yamagata (2022). On the other hand, consistent estimation of both global and local factors and their loadings can be easily achieved in the the multilevel factor model, using the canonical-correlations-based approach as described in the paper. In this regard, we expect that the joint analysis of our proposed approach and the unknown group membership will shed further lights on enhancing our understanding of weak factor models, especially in relation to the recent asset pricing models following the factor zoo criticism raised by Cochrane (2011), see also Bailey et al. (2021) and Lettau & Pelger (2020).

Chapter 2

Generalised Canonical Correlation Estimation of the Multilevel Factor Model

Abstract We develop a novel approach based on the generalised canonical correlation (GCC) analysis to analyse the high dimensional panel data model with the multilevel factor structure. Importantly, our approach is shown to be valid even if some blocks share the common local/regional factors. We establish the consistency of the estimated factors and loadings, and derive their asymptotic normal distributions under fairly standard conditions. We also propose a GCC selection criterion for identifying the number of global factors. Via Monte Carlo simulations, we confirm the validity of our asymptotic theory, and also demonstrate the superior performance of the GCC selection criterion over 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.

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

JEL Classifications: C55, R31.

2.1 Introduction

In a data-rich environment with large cross-section units and time periods, the factor model provides a dimension reduction technique via a set of pervasive latent factors, e.g. Chamberlain & Rothschild (1983), Stock & Watson (2002) and Bai & Ng (2002). Recently, the multilevel factor models have gained increasing attention; some factors are pervasive (i.e. common to all individuals) whilst other are semi-pervasive (i.e. common only to a subset of individuals), 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 & 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. Applying a two-block factor model to a mixed-frequency data, Andreou et al. (2019) show that the industrial production is still the most important workhorse in the US.

The principal component (PC) estimation, a popular method in the single-level factor model, cannot be directly applied 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 the multilevel factor modelling a challenging issue. Wang (2008) proposes a sequential PC approach by updating the global and local factors iteratively, though this approach does not guarantee convergence to the global minimum unless the initial estimate is consistent. Breitung & 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.

In this paper we broaden the scope to encompass more generals scenarios in which certain blocks may share common regional factors, as analysed by Moench et al. (2013) and Beck et al. (2016). Alternatively, blocks share pairwise common local factors, as explored by Hallin & Liška (2011) and Rodríguez-Caballero & Caporin (2019). In such cases, the CCA approach fails to produce the consistent estimation of global factors, as the common local/regional factors can be mistakenly identified as global. As the main contribution, we propose the generalised canonical correlation analysis (GCC), which extends the standard CCA (using two blocks only) by conducting the simultaneous analysis of the factor spaces of all blocks collected in a system-wide matrix, $\mathbf{\Phi}$. By jointly dealing with the unit pairwise canonical correlation between any two blocks, GCC successfully addresses the aforementioned issues with common local/regional factors. Furthermore, GCC offers the main computational advantage, as it does not require any iterations, unlike many existing studies.

We establish the consistency of the estimated factors and loadings using matrix perturbation theory. Importantly, consistency can be achieved if the number of blocks, R is finite or tends to infinity. In this article, we focus on a finite R, a more practical case empirically, and derive the asymptotic normal distributions for the global and local factors and loadings under fairly standard assumptions.[1] To the best of our knowledge in the literature, only Andreou et al. (2019) has developed an asymptotic theory for factors and loadings under finite R. However, their approach is based on CCA and is confined to the case with two blocks ($R = 2$). Furthermore, their theory is subject to a few limitations, e.g. the asymptotic distribution of the estimated global factors is non-unique while the estimated local factors involve bias terms due to the sequential estimation. This suggests that it is unclear how to conduct inference using their results. By contrast, our theory is applicable for $R \geq 2$ while the asymptotic distribution of the estimated global factors remains unique.

The challenging issue is that the asymptotic covariance matrices of the estimated global factors and loadings involve a rotation matrix, that cannot be estimated consistently, rendering infeasible inference based on the asymptotic theory. Furthermore, if the number of blocks is finite, estimation errors from the estimated global factors and loadings are of the same magnitude as the local factors and loadings. To overcome these challenging issues, we propose a novel hybrid bootstrap method that is able to produce valid confidence intervals for the estimated global (local) factors and loadings. By resampling the global (local) factors toegether with the error terms, we can produce back-rotated factors and loadings. This procedure ensures that the bootstrap estimates have the correct limiting distributions across repetitions. It is worth mentioning that the GCC approach can be easily extended to more general cases, such as the multilevel factor model with more than two layers of classification considered by Ahn & Zhang (2023).

We also 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 global factors (r_0) in the single-level factor literature (e.g. Bai & Ng (2002), Onatski (2010) and Ahn & Horenstein (2013)), fail to generate reliable model selection in the multilevel case. Recently, a few approaches have been proposed to consistently estimate 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

[1] If $R \to \infty$, the estimation proceeds in an iterative manner. Wang (2008) developed a sequential estimation theory. Mao et al. (2021) provide an asymptotic theory for a bilateral trade flow model where both the number of importers and exporters tend to infinity. Jin et al. (2023) consider a generalised three-dimensional factor model and establish the asymptotic theory when the number of groups tends to infinity.

determining the number of the global factors based on the average pairwise canonical correlation among all blocks. Chen (2022) proposes a selection criterion based on the average residual sum of square from a regression of (estimated) global factors on the factor spaces in each block. Unlike most existing studies, 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 robust and general.

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 the circular projection estimation developed by Chen (2022) under all experiments. Next, we examine the performance of the hybrid bootstrap when evaluating confidence intervals for the estimated factors and loadings as well as the factor components. We find that the hybrid bootstrap can produce valid confidence intervals with correct coverage rates as the sample size increases.

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). Second, the national factor explains a considerable portion of the house 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 & 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-recovery 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 areal factor components display a heterogeneous but opposite 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 co-moves 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 2.2 introduces the multilevel factor model and provides a review of the related literature. Section 2.3 proposes the novel *GCC* approach and presents the main estimation algorithms. Section 2.4 develops the asymptotic theory for the *GCC* estimator, advances a new selection criterion for identifying the number of the global factors, and proposes a hybrid bootstrap approach for constructing confidence intervals. Section 2.5 reports Monte Carlo simulation results. Section 2.6 presents an empirical application to the house price inflation data in England and Wales. Section 2.7 offers concluding remarks. The detailed bootstrap algorithm, mathematical proofs, auxiliary lemmas, and additional simulation results are relegated to Appendix B.

2.2 The Multilevel Factor Model

Consider the multilevel factor model:

$$y_{ijt} = \boldsymbol{\gamma}_{ij}'\mathbf{G}_t + \boldsymbol{\lambda}_{ij}'\mathbf{F}_{it} + e_{ijt}, i = 1, ..., R, j = 1, ..., N_i, t = 1, ..., T \tag{2.2.1}$$

where $\mathbf{G}_t = \left[G_t^1, ..., G_t^{r_0}\right]'$ is the $r_0 \times 1$ vector of the global factors, $\mathbf{F}_{it} = \left[F_{it}^1, ..., F_{it}^{r_i}\right]'$ 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 heterogeneous factor loadings, and e_{ijt} is the idiosyncratic error. Stacking (2.2.1) across the N_i individuals in block i, we have:

$$\mathbf{Y}_{it} = \boldsymbol{\Gamma}_i\mathbf{G}_t + \boldsymbol{\Lambda}_i\mathbf{F}_{it} + \mathbf{e}_{it}, \tag{2.2.2}$$

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

$$\mathbf{Y}_t = \boldsymbol{\theta}^+\mathbf{K}_t^+ + \mathbf{e}_t, \tag{2.2.3}$$

where

$$\underset{N\times 1}{\mathbf{Y}_t} = \begin{bmatrix} \mathbf{Y}_{1t} \\ \vdots \\ \mathbf{Y}_{Rt} \end{bmatrix}, \underset{N\times 1}{\mathbf{e}_t} = \begin{bmatrix} \mathbf{e}_{1t} \\ \vdots \\ \mathbf{e}_{Rt} \end{bmatrix}, \underset{r^+\times 1}{\mathbf{K}_t^+} = \begin{bmatrix} \mathbf{G}_t \\ \mathbf{F}_{1t} \\ \vdots \\ \mathbf{F}_{Rt} \end{bmatrix}, \underset{N\times r^+}{\boldsymbol{\Theta}}^+ = \begin{bmatrix} \gamma_1 & \boldsymbol{\lambda}_1 & 0 & \cdots & 0 \\ \gamma_2 & 0 & \boldsymbol{\lambda}_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_R & 0 & 0 & \cdots & \boldsymbol{\lambda}_R \end{bmatrix}$$

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:

$$\mathbf{Y} = \mathbf{K}^+ \mathbf{\Theta}^{+\prime} + \mathbf{e}, \tag{2.2.4}$$

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

Alternatively, stacking (2.2.1) over time period t, we can rewrite the model as

$$\mathbf{Y}_{ij} = \mathbf{G}\boldsymbol{\gamma}_{ij} + \mathbf{F}_i \boldsymbol{\lambda}_{ij} + \mathbf{e}_{ij} = \mathbf{K}_i \boldsymbol{\theta}_{ij} + \mathbf{e}_{ij} \tag{2.2.5}$$

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

$$\mathbf{Y}_i = \mathbf{G}\mathbf{\Gamma}'_i + \mathbf{F}_i \mathbf{\Lambda}'_i + \mathbf{e}_i = \mathbf{K}_i \mathbf{\Theta}'_i + \mathbf{e}_i \tag{2.2.6}$$

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

The primary issue in the multilevel factor model is to identify the global and local factors, separately. We now express the model (2.2.2) as

$$\mathbf{Y}_{it} = \mathbf{\Gamma}_i \mathbf{G}_t + \breve{\mathbf{e}}_{it}, \;\; \breve{\mathbf{e}}_{it} = \mathbf{\Lambda}_i \mathbf{F}_{it} + \mathbf{e}_{it}, \tag{2.2.7}$$

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 $\mathbf{Y}_t = [\mathbf{Y}'_{1t}, \ldots, \mathbf{Y}'_{Rt}]'$, will be inconsistent estimates of $\mathbf{G}_t$ because the weak correlation condition among the error components in $\breve{\mathbf{e}}_t = [\breve{\mathbf{e}}'_{1t}, \ldots, \breve{\mathbf{e}}'_{Rt}]'$ is violated due to the presence of the local factors (see Breitung & Eickmeier (2016)). Alternatively, if we apply the *PC* estimation to each block $\mathbf{Y}_i$ in (2.2.6), the factor space spanned by $\mathbf{K}_i = [\mathbf{G}, \mathbf{F}_i]$ can be consistently estimated up to rotation, though the global and local factors cannot be separately identified.[2]

A number of alternative methods have been developed. Wang (2008) proposed an iterative sequential approach. Given the estimated global factors and loadings, denoted $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{\Gamma}}_i$, then the

[2]Moreover, the r^+ factors extracted from $\mathbf{Y}_t$ in (2.2.3) are not necessarily consistent estimates of $\mathbf{K}^+$. 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 $\sqrt{N}$.

local factors and loadings for each block i can be estimated by applying the PC estimation to

$$\mathbf{Y}_i - \widehat{\mathbf{G}}\widehat{\mathbf{\Gamma}}_i' = \mathbf{F}_i\mathbf{\Lambda}_i' + \mathbf{e}_i \tag{2.2.8}$$

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

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

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 & Eickmeier (2016) and Choi et al. (2018) propose the use of the canonical correlation analysis (CCA), where the canonical correlation between $\widehat{\mathbf{K}}_m$ and $\widehat{\mathbf{K}}_h$ is estimated for any two blocks, m and h. For convenience assume that r_0, r_m and r_h are known and set $r_0 + r_m = r_0 + r_h$. Consider the following characteristic equation:

$$\left(\widehat{\mathbf{S}}_{mh}\widehat{\mathbf{S}}_{hh}^{-1}\widehat{\mathbf{S}}_{hm} - \ell\widehat{\mathbf{S}}_{mm}\right)\mathbf{v}_m = \mathbf{0} \text{ or } \left(\widehat{\mathbf{S}}_{hm}\widehat{\mathbf{S}}_{mm}^{-1}\widehat{\mathbf{S}}_{mh} - \ell\widehat{\mathbf{S}}_{hh}\right)\mathbf{v}_h = \mathbf{0} \tag{2.2.9}$$

where $\widehat{\mathbf{S}}_{ab}$ $(a, b = m, h)$ denotes the covariance matrix between $\widehat{\mathbf{K}}_m$ and $\widehat{\mathbf{K}}_h$. We then obtain the solution ℓ by the (squared) canonical correlations between $\widehat{\mathbf{K}}_m$ and $\widehat{\mathbf{K}}_h$. Since they 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 $\widehat{\mathbf{G}} = \widehat{\mathbf{K}}_m\mathbf{V}_m^{r_0}$ or $\widehat{\mathbf{G}} = \widehat{\mathbf{K}}_h\mathbf{V}_h^{r_0}$, where $\mathbf{V}_m^{r_0}$ and $\mathbf{V}_h^{r_0}$ are the matrices consisting of the characteristic vectors corresponding to the r_0 largest characteristic roots. Next, after projecting $\widehat{\mathbf{G}}$ out, one can consistently estimate the local factors and loadings. In practice, this estimation proceeds iteratively until convergence. Breitung & 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. In a special case with $R = 2$, Andreou et al. (2019) suggests using CCA to estimate the global factors based on CCA between $\widehat{\mathbf{K}}_1$ and $\widehat{\mathbf{K}}_2$ by solving the characteristic equation (2.2.9) for $(m, h) = (1, 2)$ or $(m, h) = (2, 1)$. Namely, the global factors can be estimated by $\widehat{\mathbf{G}} = \widehat{\mathbf{K}}_1\mathbf{V}_1^{r_0}$ or $\widehat{\mathbf{G}} = \widehat{\mathbf{K}}_2\mathbf{V}_2^{r_0}$.

However, CCA does not always estimate the global factors consistently. Consider a three-block model ($R = 3$) with $r_0 = r_i = 1$ for $i = 1, 2, 3$. Suppose that the first and second blocks share the same local factor. When we obtain the largest canonical correlation between $\widehat{\mathbf{K}}_1$ and $\widehat{\mathbf{K}}_2$, we are no longer sure whether $\widehat{\mathbf{K}}_1\mathbf{V}_1^{r_0}$ yields the global 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 local factors.[3] Rodríguez-Caballero & Caporin (2019) consider the pairwise-common local factors by employing two parallel country classifications based on the Debt/GDP ratio and credit rating. In these cases, CCA fails to produce consistent estimates. See also Moench et al. (2013) and Beck et al. (2016).

Recently, Chen (2022) propose a circular projection estimation (CPE) approach. The circular projection matrix is a successive product of the factor spaces of $\mathbf{K}_i$, given by the product inside the bracket in $\left[\left(\prod_{i=1}^{R}\mathscr{P}\left(\mathbf{K}_i\right)\right)'\left(\prod_{i=1}^{R}\mathscr{P}\left(\mathbf{K}_i\right)\right)\right]\boldsymbol{\zeta} = \pi\boldsymbol{\zeta}$, where $\mathscr{P}(.)$ is the projection matrix, and π and $\boldsymbol{\zeta}$ are the eigenvalue and eigenvector. Only when $\pi = 1$ is $\boldsymbol{\zeta}$ 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 $\mathbf{K}_i$ by $\widehat{\mathbf{K}}_i$. CPE does not suffer from the issue related to the common local factors since it encompasses all blocks. Similarly, GCC incorporates the information from all blocks simultaneously via a linear combination of the whole factor spaces (see (2.3.19) below). However, our method yields a simpler asymptotic expansion of the global factors, enabling us to directly derive the asymptotic normal distribution of the estimator of the global factors whereas no such asymptotic theory is developed under CPE. Moreover, we show that GCC outperforms CPE via simulations (see Section 2.5).

2.3 The Generalised Canonical Correlation Analysis

We begin with the standard CCA and describe our approach via comparison between CCA and GCC. Select any two blocks, m and h and let $\mathbf{K}_m$ and $\mathbf{K}_h$ be $T \times (r_0 + r_m)$ and $T \times (r_0 + r_h)$ matrices consisting of the global and local factors. CCA aims to find the linear combinations $\mathbf{v}_{mj}$ and $\mathbf{v}_{hj}$ such that

$$(\mathbf{v}_{mj}, \mathbf{v}_{hj}) = \underset{\mathbf{v}_{mj}, \mathbf{v}_{hj}}{\arg\max}\, \mathrm{Corr}\left(\mathbf{K}_m\mathbf{v}_{mj}, \mathbf{K}_h\mathbf{v}_{hj}\right) \tag{2.3.10}$$

subject to the restrictions

$$\mathbf{V}_m'\mathbf{K}_m'\mathbf{K}_m\mathbf{V}_m = \mathbf{I}_{r_{\min}} \text{ and } \mathbf{V}_h'\mathbf{K}_h'\mathbf{K}_h\mathbf{V}_h = \mathbf{I}_{r_{\min}} \tag{2.3.11}$$

[3] Hallin & Liška (2011) find one common local factor between France and Germany in a three-country model for industrial production indices for France, Germany and Italy.

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

$$\mathbf{K}_m \mathbf{V}_m^{r_0} = \mathbf{K}_h \mathbf{V}_h^{r_0} \tag{2.3.12}$$

where $\mathbf{V}_m^{r_0} = [\mathbf{v}_{m1}, \ldots, \mathbf{v}_{mr_0}]$ and $\mathbf{V}_h^{r_0} = [\mathbf{v}_{h1}, \ldots, \mathbf{v}_{hr_0}]$ are the matrices collecting such linear combinations. We then solve the characteristic equation, (2.2.9) and obtain $\mathbf{V}_m^{r_0}$ that consists of characteristic vectors of $\mathbf{V}$ corresponding to the r_0 largest characteristic roots.

However, *CCA* does not always identify the global factors in the presence of common local factors as explained in Section 2.2. To address this important issue, we propose to construct the $T(R-1)R/2 \times \sum_{l=1}^{R}(r_0 + r_l)$ system-wide matrix as follows:

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{K}_1 & -\mathbf{K}_2 & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{K}_1 & \mathbf{0} & -\mathbf{K}_3 & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ & & & & \vdots & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{K}_{R-1} & -\mathbf{K}_R \end{bmatrix} \tag{2.3.13}$$

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

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

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

$$\mathbf{\Phi} = \mathbf{P \Delta Q}' \tag{2.3.14}$$

such that $\mathbf{\Phi Q} = \mathbf{P\Delta}$, where $\mathbf{P}$ and $\mathbf{Q}$ 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 $\mathbf{\Delta} = \operatorname{diag}\left\{\delta_1, \delta_2, \ldots, \delta_{\sum_{l=1}^{R}(r_0+r_l)}\right\}$ 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 $\mathbf{Q}$ and the singular values $\delta = 0$ such that $\mathbf{\Phi q} = \delta \mathbf{p} = \mathbf{0}$, then we obtain $\mathbf{Q}$ by the set of vectors, $\mathbf{q}$.

We establish the existence of the r_0 zero singular values and the corresponding eigenvectors, denoted $\mathbf{Q}^{r_0}$ in the following proposition.[4] A direct example of $\mathbf{Q}^{r_0}$ is such that each $\mathbf{Q}_i^{r_0} = [\mathbf{I}_{r_0}, \mathbf{0}]'$ 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 $\mathbf{G}\boldsymbol{\alpha}_0 = \mathbf{F}_1\boldsymbol{\alpha}_1 + \cdots + \mathbf{F}_R\boldsymbol{\alpha}_R$ if and only if $\boldsymbol{\alpha}_0 = \mathbf{0}$, $\boldsymbol{\alpha}_1 = \mathbf{0}$, $\ldots, \boldsymbol{\alpha}_R = \mathbf{0}$, which resembles the rank condition in Assumption A of Wang (2008).

Proposition 2.1. *There exists a $\sum_{l=1}^{R}(r_0 + r_l) \times r_0$ matrix, $\mathbf{Q}^{r_0} = [\mathbf{Q}_1^{r_0 \prime}, \mathbf{Q}_2^{r_0 \prime}, \ldots, \mathbf{Q}_R^{r_0 \prime}]'$ containing the right eigenvectors of $\boldsymbol{\Phi}$, such that $\boldsymbol{\Phi}\mathbf{Q}^{r_0} = \mathbf{0}$ with the r_0 zero singular values. Moreover, the remaining singular values of $\boldsymbol{\Phi}$ are larger than zero and of stochastic order $O_p\left(\sqrt{T}\right)$.*

From Proposition 2.1 we have:

$$\mathbf{K}_1\mathbf{Q}_1^{r_0} = \mathbf{K}_2\mathbf{Q}_2^{r_0} = \cdots = \mathbf{K}_R\mathbf{Q}_R^{r_0} \tag{2.3.15}$$

which shows that the pairwise canonical correlation in (2.3.12) is simultaneously satisfied for all pairs of the blocks. Importantly, this demonstrates that all $\mathbf{K}_i\mathbf{Q}_i^{r_0}$ for $i = 1, ..., R$, obtained by the *GCC* approach, can provide valid representation of $\mathbf{G}$ even in presence of the common local factors.

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

$$T^{-1}\boldsymbol{\Psi}\boldsymbol{\Psi}' = \mathbf{L}\boldsymbol{\Xi}\mathbf{L}', \tag{2.3.16}$$

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

Proposition 2.2. *The first r_0 columns of $\mathbf{L}$, denoted $\mathbf{L}^{r_0}$, consists of the factor space spanned by $\mathbf{G}$.*

Proposition 2.2 shows that the global factors can be identified by a linear combination of appropriately rotated block factor spaces. Importantly, the eigen-decomposition summarises the factor

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

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

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

spaces $\mathbf{K}_i\mathbf{Q}_i^{r_0}$, rendering us to develop a unique representation and asymptotic distribution of the global factors, differing from Proposition D.5 of Andreou et al. (2019).

The estimation algorithm proceeds as follows.

Estimation of global factors and loadings We first obtain the *PC* estimate of $\mathbf{K}_i$ for each block i, denoted $\widehat{\mathbf{K}}_i$, by $\sqrt{T}$ times the $r_{\max}$ eigenvectors of $\mathbf{Y}_i\mathbf{Y}_i'$ corresponding to the $r_{\max}$ largest eigenvalues, where $r_{\max} \geq \max_{i=1,\dots,R}\{r_0 + r_i\}$ is a common positive integer. We then construct the $TR(R-1)/2 \times Rr_{\max}$ matrix, $\widehat{\boldsymbol{\Phi}}$ by replacing $\mathbf{K}_i$ with $\widehat{\mathbf{K}}_i$ in (2.3.13), and evaluate the SDV of $\widehat{\boldsymbol{\Phi}}$ as

$$\widehat{\boldsymbol{\Phi}} = \widehat{\mathbf{P}}\widehat{\boldsymbol{\Delta}}\widehat{\mathbf{Q}}', \tag{2.3.17}$$

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

$$T^{-1}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Psi}}' = \widehat{\mathbf{L}}\widehat{\boldsymbol{\Xi}}\widehat{\mathbf{L}}' \tag{2.3.18}$$

where $\widehat{\mathbf{L}}$ 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*. From (2.3.18), we obtain the consistent estimator of the global factors, denoted $\widehat{\mathbf{G}}$, by the r_0 vectors of $\widehat{\mathbf{L}}$ corresponding to the r_0 largest eigenvalues multiplied by $\sqrt{T}$; namely,

$$\widehat{\mathbf{G}} = \frac{1}{\sqrt{T}}\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Psi}}'\widehat{\mathbf{J}}^{r_0} = \frac{1}{\sqrt{T}}\left(\sum_{i=1}^{R}\widehat{\mathbf{K}}_i\widehat{\mathbf{Q}}_i^{r_0}\widehat{\mathbf{Q}}_i^{r_0\prime}\widehat{\mathbf{K}}_i'\right)\widehat{\mathbf{J}}^{r_0} \tag{2.3.19}$$

where $\widehat{\mathbf{J}}^{r_0} = \widehat{\mathbf{L}}^{r_0}\left(\widehat{\boldsymbol{\Xi}}^{r_0}\right)^{-1}$, $\widehat{\mathbf{L}}^{r_0}$ collects the first r_0 columns of $\widehat{\mathbf{L}}$ 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 $\widehat{\boldsymbol{\Gamma}}_i = T^{-1}\mathbf{Y}_i'\widehat{\mathbf{G}}$.

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

2.4 Asymptotic Theory

Section 2.4.1 establishes the consistency of estimated factors and loadings based on the matrix perturbation theory. Section 2.4.2 develops a consistent selection criteria for determining the number of the global factors. In Section 2.4.3, we derive asymptotic normal distributions for the factors and loadings estimates. Section 2.4.4 discusses a hybrid bootstrap method for constructing valid confidence intervals.

2.4.1 Consistent estimation of factors and loadings

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

Assumption 2.A.

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

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

3. *Let* $\mathrm{E}(e_{ijt}e_{ikt}) = \tau_{i,(jk),t}$, *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* $\mathrm{E}(e_{ijt}e_{iks}) = \tau_{i,(jk),(ts)}$. *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
$$\mathrm{E}\left(\left|\frac{1}{\sqrt{N_i}}\sum_{j=1}^{N_i}[e_{ijs}e_{ijt} - \mathrm{E}(e_{ijs}e_{ijt})]\right|^4\right) \leq \mathcal{M}$$

Assumption 2.B.

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

2. *For each* m, h *and* t,
$$\mathrm{E}\left(\frac{1}{N_m}\sum_{j=1}^{N_m}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{K}_{ht}e_{mjt}\right\|^2\right) \leq \mathcal{M}$$

Assumption 2.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* $(\mathbf{\Theta}_i) = r_0 + r_i$ *where* $\mathbf{\Theta}_i = [\mathbf{\Gamma}_i, \mathbf{\Lambda}_i]$.

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

   (c) $\mathbf{\Sigma}_{\Theta_i}\mathbf{\Sigma}_{K_i}$ *has distinct eigenvalues.*

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

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

Assumption 2.A is an extended version of Assumption C in Bai & Ng (2002), which allows the idiosyncratic errors to be serially and (weakly) cross-sectionally correlated within blocks. This is less restrictive than the assumption in Choi et al. (2018). Assumptions 2.B and 2.C are standard. Assumption 2.B.2 allows weak correlation between global/local factors and idiosyncratic errors. Assumption 2.C requires the global (local) factors to have non-trivial contributions to the variance of all individuals within the corresponding block. Assumption 2.D is standard in multilevel factor models that ensures separate identification of global and local factors. 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 2.D for consistent estimation of $\boldsymbol{\gamma}_i$, $\boldsymbol{\lambda}_i$, $\mathbf{F}_i$ and r_i.

More importantly, in this paper, we allow the local factors to be correlated or even identical across some blocks unlike many existing studies that require the orthogonality among local factors, e.g. Choi et al. (2018) and Han (2021), and establish that the *GCC* estimator is valid even in the presence of the common local/regional factors. We focus on the more practical case with a fixed number of blocks R, although *GCC* is valid as $R \to \infty$.[5]

Lemma 2.1. *Under Assumptions 2.A–2.C, as $N_i, T \to \infty$, we have:*

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

where $\widehat{\mathbf{K}}_i$ 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, $\mathbf{K}_i = [\mathbf{G}, \mathbf{F}_i]$ is the $T \times (r_0 + r_i)$ factors, $\widehat{\mathbf{H}}_i$ is the $(r_0 + r_i) \times r_{\max}$ rotation matrix, $C_{N_i T} = \min\left\{\sqrt{N_i}, \sqrt{T}\right\}$, and

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

where $\mathbf{\Phi}$ is the $T(R-1)R/2 \times \sum_{l=1}^{R}(r_0 + r_l)$ matrix defined in (2.3.13), $\widehat{\mathbf{\Phi}}$ is the $T(R-1)R/2 \times Rr_{\max}$ matrix by replacing $\mathbf{K}_i$ with $\widehat{\mathbf{K}}_i$, $\widehat{\mathbf{H}} = diag\left\{\widehat{\mathbf{H}}_1, \widehat{\mathbf{H}}_2, \dots, \widehat{\mathbf{H}}_R\right\}$ is a $\sum_{l=1}^{R}(r_0 + r_l) \times Rr_{\max}$ block-diagonal rotation matrix and $C_{\underline{N},T} = \min\{\sqrt{\underline{N}}, \sqrt{T}\}$ with $\underline{N} = \min\{N_1, N_2, \dots, N_R\}$.

Lemma 2.1 establishes that as $N_i, T \to \infty$, $\widehat{\mathbf{K}}_i$ converges to their population counterpart up to a rotation. The rotation matrix, $\widehat{\mathbf{H}}_i$ is shown to exist in Bai & Ng (2002), but we do not need a specific form since any full rank rotation matrix yields the observationally equivalent model.

Lemma 2.2. *There exists an $Rr_{\max} \times r_0$ matrix $\bar{\mathbf{Q}}^{r_0}$ such that $\mathbf{\Phi}\widehat{\mathbf{H}}\bar{\mathbf{Q}}^{r_0} = \mathbf{0}$, where the r_0 singular values are zero. The remaining singular values of $\mathbf{\Phi}\widehat{\mathbf{H}}$ are larger than zero and of stochastic order $O_p\left(\sqrt{T}\right)$.*

Lemma 2.2 extends Proposition 2.1 to the case under the rotation incurred by the *PC* estimation, and enables us to apply Lemma 2.3 below to $\widehat{\mathbf{\Phi}}$ for deriving the convergence rate of the estimated

[5] If $R \to \infty$, the identification of global factors is simplified because each block is asymptotically negligible, and therefore, the *PC* estimation can be applied to the whole data matrix. See e.g. Wang (2008) and Jin et al. (2023).

eigenvectors under rotation. It helps to estimate the number of global factors r_0 by counting the number of zero singular values of $\widehat{\mathbf{\Phi}}$ (see Section 2.4.2).

While the consistency of the estimated eigenvalues are well-established, there are the two main issues when establishing the consistency of the estimated eigenvectors. First, it is acknowledged that the convergence of the eigenvectors may not be well-behaved under eigenvalue-multiplicity. Second, convergence rates of the eigenvectors associated with zero eigenvalues are unclear (see Theorem 3.4 of Stewart & Sun (1990)).

In Lemma 2.3 we state the perturbation theory developed by Yu et al. (2015), which is a variant of the Davis-Kahan Theorem, and necessary for deriving our consistency results.

Lemma 2.3. *Let $\mathbf{\Sigma}$ and $\widehat{\mathbf{\Sigma}}$ 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, $\mathbf{V} = [v_r, v_{r+1}, \ldots, v_s]$ and $\widehat{\mathbf{V}} = [\hat{v}_r, \hat{v}_{r+1}, \ldots, \hat{v}_s]$, have orthogonal columns, satisfying $\mathbf{\Sigma}\mathbf{V}_j = \lambda_j \mathbf{V}_j$ and $\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}_j = \hat{\lambda}_j \widehat{\mathbf{V}}_j$ for $j = r, r+1, \ldots, s$. Then, there exists a $d \times d$ orthogonal matrix $\widehat{\mathbf{O}}$ such that*

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

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 is not applicable to our case where the eigenvalues of interest are zero. Lemma 2.3 establishes that the convergence of the eigenvectors still holds up to an orthogonal rotation even if the population eigenvalues are zero.

With Lemmas 2.1–2.3, we establish the consistency of the estimated global factors and loadings (up to rotation) in Theorem 2.1.

Theorem 2.1. *Consistency of the global factors and loadings.*

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

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

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

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

where $\mathbb{H} = T^{-1/2}\mathbf{G}'\mathbf{J}^{r_0}\mathbf{U}$ is an $r_0 \times r_0$ rotation matrix, $\mathbf{J}^{r_0} = \mathbf{L}^{r_0}(\mathbf{\Xi}^{r_0})^{-1}$, $\mathbf{\Xi}^{r_0}$ is an $r_0 \times r_0$ diagonal matrix consisting of the r_0 non-zero eigenvalues of $T^{-1}\mathbf{G}\mathbf{G}'$ in descending order, $\mathbf{L}^{r_0}$ is a $T \times r_0$ matrix of the corresponding eigenvectors, $\mathbf{U}$ is an $r_0 \times r_0$ orthogonal matrix defined in (B.1.3), 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 & Otrok (2007)), then an orthogonality between global and local factors in Assumption 2.D is not required, unlike existing studies that impose an orthogonality, e.g. Wang (2008), Choi et al. (2018), Andreou et al. (2019) and Han (2021). Moreover, we allow the local factors to be correlated or identical across some blocks, unlike most studies that require orthogonality among the local factors, e.g. Choi et al. (2018) and Han (2021). Thus, the *GCC* is more general and robust.

Given consistent estimates of the global factors and loadings, we next establish the consistency of the estimated local factors and loadings in Theorem 2.2.

Theorem 2.2. *Consistency of the local factors and loadings.*

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

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

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

where $\widehat{\mathscr{H}_i} = (\mathbf{\Lambda}_i'\mathbf{\Lambda}_i/N_i)\left(\mathbf{F}_i'\widehat{\mathbf{F}}_i/T\right)\widehat{\mathbf{\Upsilon}}_i^{-1}$ 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, $\widehat{\mathbf{Y}}_i = \mathbf{Y}_i - \widehat{\mathbf{G}}\widehat{\mathbf{\Gamma}}_i'$, 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 2.2 establishes that the *GCC* estimator is still consistent even in the presence of the pairwise common local factors and the local/regional factors common across some blocks.

2.4.2 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 SVD of $\widehat{\mathbf{\Phi}}$ defined in (2.3.17). Then, we evaluate the ratio of adjacent (squared) singular values in a similar fashion as in Ahn & 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

$$\hat{r}_{0,GCC} = \underset{k=0,\ldots,r_{\max}}{\operatorname{argmax}} \frac{\hat{\delta}_{k+1}^2}{\hat{\delta}_k^2} \tag{2.4.20}$$

The main idea is that the ratio sharply separates the zero singular values from the positive ones. Using Lemma 2.2, 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 $k = 0, \ldots, r_0 - 1, r_0 + 1, \ldots, r_{\max}$ while 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, if $r_0 = 0$.

Theorem 2.3. *Under Assumptions 2.A–2.C, we have:*

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

where $\hat{r}_{0,GCC} = \underset{k=0,\ldots,r_{\max}}{\arg\max}\, \hat{\delta}_{k+1}^2/\hat{\delta}_k^2$, $\hat{\delta}_1 \leq \cdots \leq \hat{\delta}_{r_{\max}} \leq \cdots \leq \hat{\delta}_{Rr_{\max}}$ *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 2.3 lies in the sense of the matrix perturbation theory that the eigenvalues converge to their population counterparts under a small perturbation term (see Stewart & 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 makes the *GCC* criterion more general than existing ones 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 $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{\Gamma}}_i$. Then, the number of local factors, r_i can be consistently estimated by applying the existing approximate factor model to $\widehat{\mathbf{Y}}_i = \mathbf{Y}_i - \widehat{\mathbf{G}}\widehat{\mathbf{\Gamma}}_i'$ for $i = 1, ..., R$, e.g. Bai & Ng (2002), Onatski (2010) and Ahn & Horenstein (2013). See Choi & 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:

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

As described in Choi et al. (2021), however, it has 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 (2.2.9). 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 $r = 1$ such 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, their approach suffers from a few limitations. First, it can be applied to the model with only two blocks, suggesting that *CCA* does not guarantee a valid identification between the global and local factors under a multi-block setting. Second, their method requires orthogonality condition between the global and local factors, and rules out error correlation and heteroskedasticity in both cross-section and time.[6] Third, Choi et al. (2021) demonstrated via simulations that large sample sizes are required for the sequential test to work. On the other hand, the *GCC* criterion works for $R \geq 2$ even when the error terms are serially correlated and weakly cross-sectionally correlated. Furthermore, $\hat{r}_{0,GCC}$ remains consistent even when the global factors are correlated with local factors, and in the

[6] Andreou et al. (2019) claim that the strong assumption for the error terms (Assumption A.9) is made for simplifying the derivation of the feasible asymptotic distribution of the test statistic, and discuss how one can relax this assumption by using HAC-type estimators.

presence of common local factors (see Section 2.5 for simulation evidence).

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 (2.2.9) 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 two selection criteria, CCD and MCC, are proposed:

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

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

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.[7]

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

$$\left[\left(\prod_{i=1}^{R} \mathscr{P}\left(\widehat{\mathbf{K}}_i \right) \right)' \left(\prod_{i=1}^{R} \mathscr{P}\left(\widehat{\mathbf{K}}_i \right) \right) \right].$$

Chen suggests estimating r_0 by

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

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 $\tau = 14$. $ARSS$ 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 2.5).

[7] 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 robust to the coice of $r_{\max}$.

2.4.3 Asymptotic distributions of the estimated factors and loadings

Following Bai (2003), we make the additional assumptions, which are slightly stronger than those required for consistency in Section 2.4.1.

Assumption 2.E. *For each i, we have $\lim_{N_i, N\to\infty} N/N_i = \alpha_i \leq \mathcal{M}$*

Assumption 2.F.

1. *$\sum_{s=1}^{T} |\omega_{i,N_i}(s,t)| < \mathcal{M}$ for all i and t.*

2. *Let $\tau_{(mh),(kj),t} = \mathrm{E}\left(e_{mkt}e_{hjt}\right)$. 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 2.G.

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

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

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

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

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 finite covariance matrix with

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

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

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

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

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

Assumption 2.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 2.F and 2.G, corresponding to Assumptions E and F in Bai (2003), are standard in the literature. Assumption 2.F restricts the cross-sectional and serial dependence of the errors. Notice that Assumption 2.F.2 imposes limited cross-block dependence, which is not required in Assumption 2.A. Assumptions 2.G.1 and 2.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 sums of zero mean random variables. Assumptions 2.G.3 and 2.G.4 are the central limit theorems that can be applied to several mixing processes.

With Assumptions 2.F and 2.G, Lemma B.1.3 establishes that some parts in the asymptotic expansion of $\widehat{\mathbf{K}}_{it}$ achieve a convergence rate faster than $O_p \left(C_{N_i T}^{-1} \right)$, as shown in Lemma 2.1. This allows us to refine the convergence rates of $\widehat{\mathbf{Q}}^{r_0}$ and $\widehat{\mathbf{L}}^{r_0}$ in Lemma B.1.4 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 2.1. By applying these results, we are able to derive the asymptotic normal distributions of the estimated factors and loadings in Theorems 2.4 and 2.5.

Theorem 2.4. *Asymptotic distributions of the global factors and loadings.*

1. *Under Assumptions 2.A–2.C and 2.E–2.G, as $N_1,\ldots,N_R,T\to\infty$ and $\sqrt{N}/T\to 0$, we have for each t:*

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

where $\mathbb{H}$ is an $r_0\times r_0$ rotation matrix defined in Theorem 2.1, $\mathcal{I}=[\mathbf{I}_{r_0},\ldots,\mathbf{I}_{r_0}]'$ is an $Rr_0\times r_0$ matrix, $\mathbb{C}=\operatorname{diag}(\mathbb{C}_1,\ldots,\mathbb{C}_R)$ is an $Rr_0\times\sum_{i=1}^R(r_0+r_i)$ block diagonal matrix with diagonal elements being

$$\mathbb{C}_i=\sqrt{\frac{N}{N_i}}\mathbb{I}_i'\left(\frac{\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i}{N_i}\right)^{-1}$$

and $\mathbb{I}_i=[\mathbf{I}_{r_0},\mathbf{0}]'$ being an $(r_0+r_i)\times r_0$ matrix, $\mathbb{C}^0=\operatorname{plim}_{N_1,\ldots,N_R,T\to\infty}\mathbb{C}$, $\mathbb{E}_t$ and $\mathbb{D}_t^{(1)}$ are defined in Assumption 2.G.3, and $\mathbb{B}$ is an $r_0\times r_0$ matrix given by

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

where $\mathbf{J}^{r_0}$ is defined in Theorem 2.1 and $\mathbf{U}$ is an $r_0\times r_0$ unknown orthonormal matrix.

2. *Under Assumptions 2.A–2.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:*

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

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

Corollary 2.1. *Under Assumptions 2.A–2.G, we have:*

$$C_{NT}\left(\widehat{\boldsymbol{\gamma}}_{ij}'\widehat{\mathbf{G}}_t - \boldsymbol{\gamma}_{ij}'\mathbf{G}_t\right) \xrightarrow{d} N\left(0, \frac{1}{R^2}\frac{C_{NT}^2}{N}\boldsymbol{\gamma}_{ij}'\mathcal{I}'\mathbb{C}^0\mathbb{D}_t^{(1)}\mathbb{C}^{0\prime}\mathcal{I}\boldsymbol{\gamma}_{ij} + \frac{C_{NT}^2}{T}\mathbf{G}_t'\boldsymbol{\Sigma}_G^{-1}\mathbb{D}_{ij}^{(2)}\boldsymbol{\Sigma}_G^{-1}\mathbf{G}_t\right).$$

Theorem 2.4 establishes that the estimates of the global factors and loadings follow the asymptotic normal distributions. Unlike in Theorem 2.1, 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 matrix. To the best of our knowledge, the asymptotic distributions of the factors and loadings under finite R, have not been established in the literature. One exception is Andreou et al. (2019), but their theory applies only for $R = 2$. We note in passing 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 uncorrelated to each other, and the local factors are uncorrelated within blocks. Corollary 2.1 establishes the asymptotic normality of the global and local components in a similar fashion to Theorem 3 of Bai (2003).

Theorem 2.5. *Asymptotic distributions of the local factors and loadings.*

1. *Under Assumptions 2.A–2.G, as $N_1, N_2, \ldots, N_R, T \to \infty$, and if $\sqrt{N_i}/T \to 0$ and $N_i/T < \infty$, then*

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

where $\mathbb{D}_{ii,t}^{(4)} = \underset{N_1,\ldots,N_R,T\to\infty}{\text{plim}} \frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\boldsymbol{\lambda}_{ij}\boldsymbol{\lambda}_{ik}'\left(e_{ijt} + \widehat{S}_{ijt}\right)\left(e_{ikt} + \widehat{S}_{ikt}\right)$, $\widehat{\boldsymbol{\Upsilon}}_i$ is a diagonal matrix consisting of the r_i largest eigenvalues of $(N_iT)^{-1}\widehat{\mathbf{Y}}_i\widehat{\mathbf{Y}}_i'$ in descending order with $\widehat{\mathbf{Y}}_i = \mathbf{Y}_i - \widehat{\mathbf{G}}\widehat{\boldsymbol{\Gamma}}_i'$, $\boldsymbol{\Upsilon}_i$ is a diagonal matrix consisting of the eigenvalues of $\boldsymbol{\Sigma}_{\Lambda_i}\boldsymbol{\Sigma}_{F_i}$ in descending order, $\mathbb{W}_i = \text{plim}_{N_i,T\to\infty} T^{-1}\widehat{\mathbf{F}}_i'\mathbf{F}_i$, and $\widehat{\mathbf{S}}_{ijt}$ is the (t, j) element of $\widehat{\mathbf{S}}_i = \mathbf{G}\boldsymbol{\Gamma}_i' - \widehat{\mathbf{G}}\widehat{\boldsymbol{\Gamma}}_i'$.

2. *Under Assumptions 2.A–2.G, as $N_1, N_2, \ldots, N_R, T \to \infty$, and if $\sqrt{T}/N_i \to 0$, then for each i*

and j*:*

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

for $j=1,\ldots,N_i$ *where* $\mathbb{D}_{ij}^{(3)}$ *is define in Assumption 2.G.4.*

Corollary 2.2. *Under the Assumptions of Theorem 2.5, we have:*

$$C_{N_iT}\left(\widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\mathbf{F}}_{it}-\boldsymbol{\lambda}_{ij}'\mathbf{F}_{it}\right)\xrightarrow{d}N\left(0,\frac{C_{N_iT}^2}{N_i}\mathbf{F}_{it}'\boldsymbol{\Sigma}_{F_i}^{-1}\mathbb{D}_{ij}^{(3)}\boldsymbol{\Sigma}_{F_i}^{-1}\mathbf{F}_{it}+\frac{C_{N_iT}^2}{T}\boldsymbol{\lambda}_{ij}'\boldsymbol{\Sigma}_{\Lambda_i}^{-1}\mathbb{D}_{ii,t}^{(4)}\boldsymbol{\Sigma}_{\Lambda_i}^{-1}\boldsymbol{\lambda}_{ij}\right).$$

Theorem 2.5 establishes the asymptotic normal distributions of the local factor and loading estimates. In particular, Theorem 2.5.1 shows that the estimated local factors are well centered, whilst the local factors in Andreou et al. (2019) have a bias term that does not vanish if $0<N_i/T<\infty$ (see their Proposition D.5). We find that the estimation error incurred from the global factors and loadings contributes only to the variance of the local factors, since the cross-section average $N_i^{-1}\sum_{j=1}^{N_i}\boldsymbol{\lambda}_{ij}\sqrt{N_i}\widehat{S}_{ijt}$ has zero mean and a non-degenerate distribution as a result of Corollary 2.1 if $N_i/T<\infty$. In contrast, Theorem 2.5.2 does not require the condition $N_i/T<\infty$ because the time average $T^{-1}\sum_{t=1}^{T}\mathbf{F}_{it}\sqrt{T}\widehat{S}_{ijt}$ converges to zero due to Assumption 2.D and 2.G.2.

Theorem 2.5 establishes the asymptotic normal distributions of the local factor and loading estimates. Theorem Theorem 2.5.1 shows that the estimated local factors are well centered. On the contrary, the local factors estimated by Andreou et al. (2019) involve a bias term that does not vanish even if $0<N_i/T<\infty$ (see their Proposition D.5). We find that the estimation errors stemming from the global factors and loadings contribute only to the variance of the estimated local factors, since the cross-section average, $N_i^{-1}\sum_{j=1}^{N_i}\boldsymbol{\lambda}_{ij}\sqrt{N_i}\widehat{S}_{ijt}$ has zero mean and a non-degenerate distribution as a result of Corollary 2.1 if N_i and T are of comparable magnitude. But, Theorem 2.5.2 does not require the condition, $N_i/T=O(1)$ because the time average, $T^{-1}\sum_{t=1}^{T}\mathbf{F}_{it}\sqrt{T}\widehat{S}_{ijt}$ converges to zero under Assumption 2.D and 2.G.2. We also derive the asymptotic normal distribution for the estimated local components in Corollary 2.2.

2.4.4 Bootstrap-based confidence intervals

Although Theorem 2.4 and 2.5 establish the asymptotic distributions of the factor and loading estimates, they are not readily applicable in practice. The reasons are threefold. First, the sample analogue of the asymptotic covariance matrices in Theorem 2.4 are subject to the rotation matrix, which is unknown and cannot be estimated. For instance, for block i in Theorem 2.4.1, we have

$$\widehat{\mathbb{C}}_i\widehat{\mathbb{E}}_{it} = \mathbb{I}_i'\left(\frac{\widehat{\boldsymbol{\Theta}}_i'\widehat{\boldsymbol{\Theta}}_i}{N_i}\right)^{-1}\frac{1}{\sqrt{N_i}}\sum_{j=1}^{N_i}\widehat{\boldsymbol{\theta}}_{ij}\hat{e}_{ijt} \xrightarrow{p} \mathbb{I}_i'\begin{bmatrix}\mathbb{H}' & \mathbf{0}\\ \mathbf{0} & \mathscr{H}_i'\end{bmatrix}\left(\frac{\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i}{N_i}\right)^{-1}\frac{1}{\sqrt{N_i}}\sum_{j=1}^{N_i}\boldsymbol{\theta}_{ij}e_{ijt} \neq \mathbb{C}_i\mathbb{E}_{it}$$

where $\widehat{\boldsymbol{\Theta}}_i = \left[\widehat{\boldsymbol{\Gamma}}_i, \widehat{\boldsymbol{\Lambda}}_i\right]$ with $\widehat{\boldsymbol{\Gamma}}_i = T^{-1}\mathbf{Y}_i'\widehat{\mathbf{G}}$ and $\widehat{\boldsymbol{\Lambda}}_i = T^{-1}\left(\mathbf{Y_i} - \widehat{\mathbf{G}}\widehat{\boldsymbol{\Gamma}}_\mathbf{i}'\right)'\widehat{\mathbf{F}}_\mathbf{i}$. Consequently, the matrix $\mathbb{H}'$ prevents us from getting correct covariance estimates using sample analogues. Second, the bootstrapped factors and loadings will have different limiting distributions because the rotation matrices $\mathbb{H}^{*(b)}$ and $\widehat{\mathscr{H}}_i^{*(b)}$ vary in each bootstrap repetition b. Therefore, $\widehat{\mathbf{G}}_t^{*(b)}$ and $\widehat{\mathbf{F}}_{it}^{*(b)}$ do not necessarily align to $\widehat{\mathbf{G}}_t$ and $\widehat{\mathbf{F}}_{it}$ and the rotation matrices should be dealt with. Third, the covariance matrices of $\widehat{\gamma}_{ij}$, $\widehat{\mathbf{F}}_{it}$, $\widehat{\gamma}_{ij}'\widehat{\mathbf{G}}_t$, and $\widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\mathbf{F}}_{it}$ are subject to the correlation between the global and local factors and their autocorrelations, which cannot be accounted for by a standard bootstrap. For example, let $e_{ijt}^{*(b)}$ be the re-sampled error term and it follows that

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

where $\mathbb{D}_{ij}^{(2)*}$ is the asymptotic covariance matrix such that $\mathbb{D}_{ij}^{(2)*} \xrightarrow{p^*} \mathbb{H}'\mathbb{D}_{ij}^{(2)*}\mathbb{H}$, because $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{F}}_i$ are strictly orthogonal under the sequential estimation. Hence, a standard bootstrap cannot correctly recover the dependence in the limiting distributions.

Due to the aforementioned issues, it is necessary to advance a new hybrid bootstrap procedure to construct valid confidence intervals for the estimated factors, loadings and factor components. For convenience we assume that the error terms are cross-sectionally and serially independent.[8] We show that, in the bootstrap world, the rotation matrices can be consistently estimated since $\widehat{\mathbf{G}}_t$ and $\widehat{\mathbf{F}}_{it}$ are treated as known. Therefore, we use the estimated rotation matrices, $\widehat{\mathbb{H}}^{*(b)}$ and $\widetilde{\mathscr{H}}_i^{*(b)}$, to back-rotate $\widehat{\mathbf{G}}_t^{*(b)}$ and $\widehat{\mathbf{F}}_{it}^{*(b)}$ as follows:

$$\sqrt{N}\left[\left(\widehat{\mathbb{H}}^{*(b)\prime}\right)^{-1}\widehat{\mathbf{G}}_t^{*(b)} - \widehat{\mathbf{G}}_t\right] \text{ and } \sqrt{N_i}\left[\left(\widetilde{\mathscr{H}}_i^{*(b)\prime}\right)^{-1}\widehat{\mathbf{F}}_{it}^{*(b)} - \widehat{\mathbf{F}}_{it}\right].$$

As a result, the above statistics remain aligned to the ones in Theorem 2.4 and 2.5 across each bootstrap repetition (see also Gonçalves & Perron (2014)). To ensure that $\widehat{\gamma}_{ij}^{*(b)}$ and $\widehat{\mathbf{F}}_{it}^{*(b)}$ have correct asymptotic variances, we apply the dependent bootstrap by Shao (2010) and obtain a re-

[8] If the error terms are serially correlated or weakly cross-sectionally correlated, the dependent bootstrap can be adopted (e.g. Shao (2010) and Conley et al. (2023)). We maintain this assumption for computational tractability.

sampled global factors $\widetilde{\mathbf{G}}^{*(b)}$ or local factors $\mathbf{F}_i^{*(b)}$ such that

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

in order to preserve the dependent structure of the factors. Then, it follows that $\mathbb{D}_{ij}^{(2)*}\xrightarrow{p^*}\mathbb{H}'\mathbb{D}_{ij}^{(2)}\mathbb{H}$, suggesting that $\widehat{\boldsymbol{\gamma}}_{ij}^{*(b)}$ and $\widehat{\mathbf{F}}_{it}^{*(b)}$ will have correct asymptotic distributions which remain the same across each bootstrap repetition. We describe three detailed algorithms in Online Appendix B.2 for implementation.

Finally, the covariance matrix of $\widehat{\boldsymbol{\lambda}}_{ij}$, namely $\left(\mathbb{W}_i^{-1}\right)'\mathbb{D}_{ij}^{(3)}\mathbb{W}_i^{-1}$, can be estimated directly by the Newey-West HAC covariance estimator based on $\widehat{\mathbf{F}}_{it}\hat{e}_{ijt}$.[9]

We note in passing that Andreou et al. (2023) also propose a bootstrap procedure in a two-block setting. They only consider bootstrapping the canonical correlations between $\widehat{\mathbf{K}}_1$ and $\widehat{\mathbf{K}}_2$ (in our notation), which is invariant to any full rank rotation matrix. The rotation matrix in their covariance estimators of their bootstrap is offset by the trace operator. In such a case, the back-rotation and factor resampling is not required.

2.5 Monte Carlo Simulation

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

$$y_{ijt}=\boldsymbol{\gamma}_{ij}'\mathbf{G}_t+\sqrt{h_{i1}}\boldsymbol{\lambda}_{ij}'\mathbf{F}_{it}+\sqrt{\kappa h_{i2}}e_{ijt}=\sum_{z=1}^{r_0}\gamma_{ij}^zG_t^z+\sqrt{h_{i1}}\sum_{z=1}^{r_i}\lambda_{ij}^zF_{it}^z+\sqrt{\kappa h_{i2}}e_{ijt} \quad (2.5.21)$$

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

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

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

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

[9] From Theorem 2.5.1 and Lemma B.1.8, it follows that $T^{-1/2}\sum_{t=1}^{T}\widehat{\mathbf{F}}_{it}\hat{e}_{ijt}\xrightarrow{p}\widehat{\mathscr{H}}_i'T^{-1/2}\sum_{t=1}^{T}\mathbf{F}_{it}e_{ijt}$, which has asymptotic variance $\left(\mathbb{W}_i^{-1}\right)\mathbb{D}_{ij}^{(3)}\mathbb{W}_i^{-1}$. Therefore, the HAC estimator based on $\widehat{\mathbf{F}}_{it}\hat{e}_{ijt}$ is consistent. See also Section 5 of Bai (2003). In unreported simulations, we find that the bootstrap CI for $\boldsymbol{\lambda}_{ij}$ performs slightly worse than the asymptotic CI.

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

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

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

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

We then make the variance contribution of each component equalised for $\kappa = 1$. For $r_0 > 0$, we set:

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

while for $r_0 = 0$ we set:

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

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

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

where $\mathbf{F}_t = [\mathbf{F}_{1t}', \ldots, \mathbf{F}_{Rt}']'$ and $\mathbf{w}_t^F = \left[\mathbf{w}_{1t}^{F\prime}, \ldots, \mathbf{w}_{Rt}^{F\prime}\right]'$. 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 $\widehat{\mathbf{G}}$ and the number of the global factors $\hat{r}_0$.[10] 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

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

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 $TR\left(\widehat{\mathbf{G}}\right) = 1$. 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 $\widehat{\mathbf{F}}_i$ and $\hat{r}_i$ depend mainly on the precision of $\widehat{\mathbf{G}}$ 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 $\widehat{\mathbf{G}}$ and $\hat{r}_0$.

Table 2.1 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, 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 indicate 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.[11] Overall, we find that GCC dominates CCA and CPE in all cases we consider.

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).[12] Table 2.2 reports the average of $\hat{r}_0$ over 1,000

[10] The accuracy of $\widehat{\mathbf{F}}_i$ and $\hat{r}_i$ are contingent upon the accuracy of $\widehat{\mathbf{G}}$ and $\hat{r}_0$ (see Appendix A of Chapter 1). As the latter estimates become more precise, the local factors and $\hat{r}_i$ also become more precise. Therefore, we omit the results for the local factors in order to save space.

[11] 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 $R = 10$.

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

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 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 any tuning parameters, we conclude that GCC is the most robust and reliable criterion.

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 results in Table 2.3 to 2.6. The results are qualitative similar to those with $(r_0, r_i) = (2, 2)$. As the number of factors in the data increases, we find that the accuracy of the estimates becomes slightly lower.

We also propose a bootstrap approach to produce the valid confidence intervals for the estimated factors and loadings as well as the factor components. In Appendix B.2, we conduct a simulation study using the hybrid bootstrap approach, and find that the coverage rates of the bootstrap CIs are getting close to the nominal 95% as the sample size increases.

Table 2.1: 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 2.2: 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|0.8)	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 2.3: 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 2.4: Average estimates of the number of the global factors with $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (1, 1)$

			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 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 3)$			
R	N_i	T												
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)

			CCD	*MCC*	*ARSS*	*GCC*	*CCD*	*MCC*	*ARSS*	*GCC*
			DGP4 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ $\omega_F = 0.4$				DGP5 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ $\omega_F = 0.8$			
R	N_i	T								
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 2.5: 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 2.6: Average estimates of the number of the global factors with $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (3, 3)$

			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 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 3)$			
R	N_i	T												
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)

			CCD	*MCC*	*ARSS*	*GCC*	*CCD*	*MCC*	*ARSS*	*GCC*
			DGP4 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ $\omega_F = 0.4$				DGP5 $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ $\omega_F = 0.8$			
R	N_i	T								
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 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.

2.6 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. As the housing sector is directly related to employment, investment and consumption, it also plays 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, the 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 2.7 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 2.7: 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. Mean 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 $RIF_i = N_i^{-1} \sum_{j=1}^{N_i} \left(\widehat{\boldsymbol{\lambda}}_{ij}' \widehat{\boldsymbol{\lambda}}_{ij} / T^{-1} \tilde{\boldsymbol{\pi}}_{ij}' \tilde{\boldsymbol{\pi}}_{ij} \right)$.

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 (2.4.20), we detect one global (national) factor.[13] Next, by applying BIC_3 to each region,[14] 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 2.7). 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

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

[14] We have also applied alternative selection criteria, IC_{p2} by Bai & Ng (2002), ER by Ahn & 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 & Jeong (2019) and Choi et al. (2021).

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 2.7 show that the global factor explains a considerable proportion 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 & 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%.

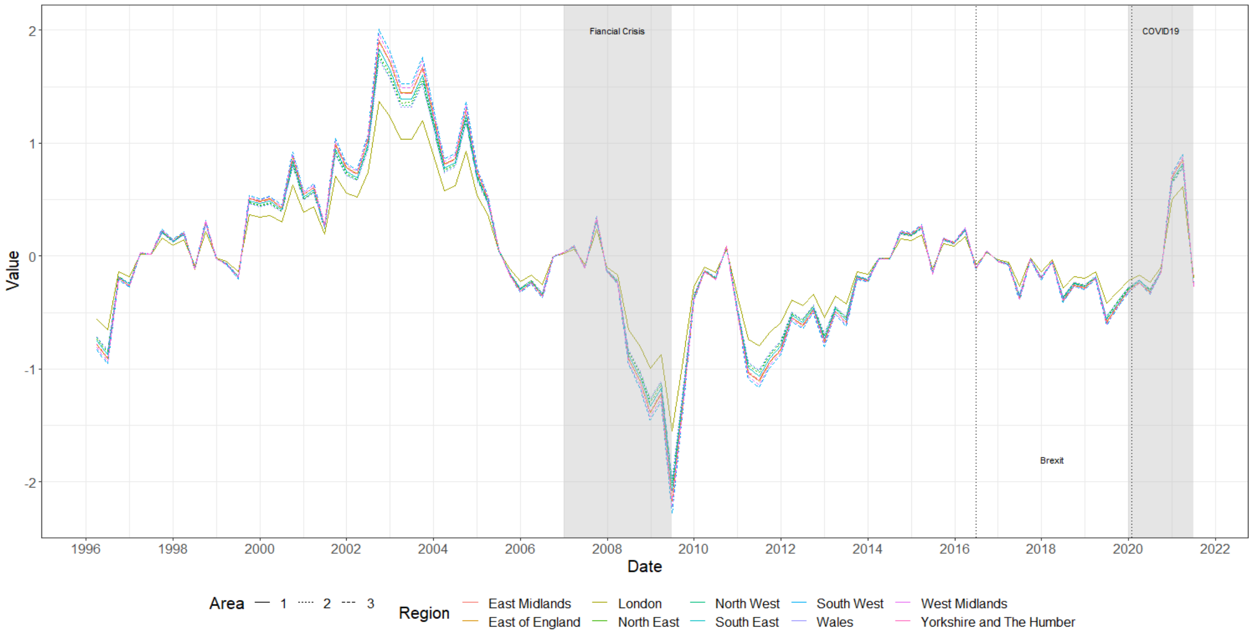
To avoid the issue that the estimated global and local factors are subject to rotation/sign indeterminacy, we report the time-varying behaviour of the average global (national) and local (regional) factor-components for each region i at time t that are constructed by $\widehat{\mathcal{G}}_{it} = \bar{\bar{\gamma}}_i'\widehat{\mathbf{G}}_t$ and $\widehat{\mathcal{F}}_{it} = \bar{\bar{\boldsymbol{\lambda}}}_i'\widehat{\mathbf{F}}_{it}$, where $\bar{\bar{\gamma}}_i = N_i^{-1}\sum_{j=1}^{N_i}\widehat{\gamma}_{ij}$ and $\bar{\bar{\boldsymbol{\lambda}}}_i = N_i^{-1}\sum_{j=1}^{N_i}\widehat{\boldsymbol{\lambda}}_{ij}$.[15] The trajectories of $\widehat{\mathcal{G}}_{it}$ plotted in Figure 2.1, are highly persistent but exhibit a typical "boom-bust-recovery" pattern of the (recent) housing market cycle.[16] 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 the COVID19 period, which was mainly prompted by a tax relief policy introduced by the UK government to boost the economy and improve liquidity.[17]

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

[16] 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).

[17] 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.

Figure 2.1: Estimated national components

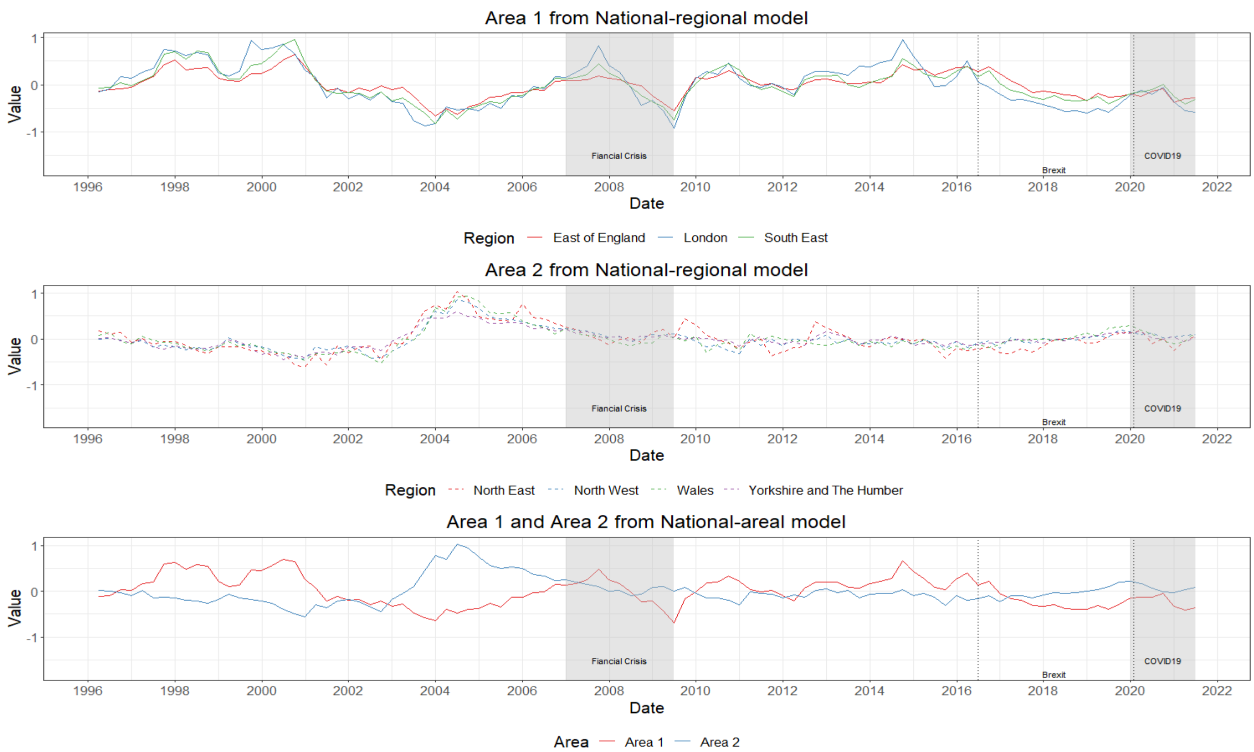


The first two figures in Figure 2.2 display the time-varying patterns of the regional factor-components $\widehat{\mathcal{F}}_{it}$, from which we can identify that the regional components of EE, LD and SE (solid lines) co-move closely (the upper panel) while those of NE, NW, YH and WA (dotted lines) tend to cluster together (the lower panel). These clustering patterns are corroborated by the correlation matrix among the estimated regional components in Table 2.8, 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 2.8: 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

Figure 2.2: Estimated regional components



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 2.9 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 2.3). Notice that the existence of an areal factor around London is not in line with the notion that the "London factor" is pervasive nationally,[18] 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 & Grimes (2008) and Montagnoli & Nagayasu (2015).

Figure 2.3: Map of regions in England and Wales

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

Table 2.9: Test of the number of common local factors from new blocks after $\widehat{\mathbf{G}}$ 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 2.10: 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

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 in which case the standard CCA-based estimates of the global and local factors may be inconsistent. The third panel in Figure 2.2 displays the areal factor components constructed by $\widehat{\mathcal{F}}_{at} = \left(N_a^{-1} \sum_{j=1}^{N_a} \widehat{\boldsymbol{\lambda}}'_{aj} \right) \widehat{\mathbf{F}}_{at}$ 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 2.2.

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 2.11. 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 plot of Figure 2.2. 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 2.11: 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.

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.[19] We also aggregate the areal factor components into the annual ones. The first two figures in Figure 2.4 display the areal factor components and the (lagged) population growth rate of in Area 1 and Area 2, respectively. We observe that they co-move closely 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 2.4, we observe

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

that the growth rate of the (lagged) population gap strongly co-moves 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.[20]

Figure 2.4: Areal components and population

2.7 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

[20] Howard & 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.

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 & 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 & Bai (2017), Coroneo et al. (2020) and Uematsu & Yamagata (2022) and apply GCC thereafter.

Chapter 3

Estimation and Inference for a Multi-dimensional Panel Data Model with Multilevel Factors

Abstract This paper considers a multi-dimensional panel data model with multilevel factors when the numbers of cross-sections and time observations are large. We develop a multilevel iterative principal component ($MIPC$) method for estimation by iteratively updating between the slope coefficients and factors, given one another. Under a finite number of blocks, our approach is able to produce consistent estimates of the slope coefficients, factors, and loadings. We also propose a model selection criteria based on the eigenvalue ratios to determine the numbers of factors. Given consistent factor estimates from each block, we apply the generalised canonical correlation (GCC) estimation to separately identifying the global and local factors. We show the consistency of our estimates and establish the asymptotic normality of the bias-corrected estimator for the slope coefficients. The Monte Carlo simulation demonstrates good finite sample performance of $MIPC$ compared to IPC in the presence of multilevel factor structure. In an empirical application, our model is applied to an analysis of the energy consumption and economic growth nexus using a cross-country panel data categorised by regions.

Keywords: Panel Data, Principal Components, Interactive Effects, Multilevel Factors.

JEL codes: C14, C23, C51.

3.1 Introduction

Panel data regression is an important tool for empirical studies as it can provide more efficient estimations than cross-sectional and time-series regressions. Moreover, it can also account for unobserved heterogeneity that may cause endogeneity when such variables are correlated with the regressors. However, when the additive effects of individual and time variables fail to capture the cross-sectional correlation properly, inconsistent estimations and invalid inferences can occur. To address this issue, interactive fixed effects (IFE) can be used, which are multiplicative terms of the individual and time effects. The iterative principal components (IPC) approach, which was pioneered by Bai (2009), has become a popular way to estimate such models, especially for large N and T. Another approach for IFE is the common correlated effects (CCE) approach developed by Pesaran (2006), where the interactive fixed effects are referred to as "factor structure".

However, in practice, researchers often encounter multi-dimensional panel data where individuals can be grouped into different categories (blocks). In such cases, local factors may be required in addition to the global factors to capture cross-sectional correlations within blocks. The use of local factors in combination with global factors can handle the heterogeneity across different blocks. Ando & Bai (2014) and Ando & Bai (2017) develop panel regression models with global and local factors by combining shrinkage estimation and PC. Rodríguez-Caballero (2022) extended the CCE approach to a three-dimensional setup with a hierarchical factor structure. Kapetanios et al. (2021) studied the three-dimensional CCE in a panel with origin-destination heterogeneity. Feng et al. (2023) extend Bai's approach to incorporate hierarchical multilevel factors when the number of blocks tends to infinity.

This article studies a panel regression model with unobserved global and local factors using a modified IPC estimation based on Bai (2009). The focus is on a finite number of blocks R, which is often the case in practice. We propose an estimation procedure that produces consistent slope parameters and the multilevel factors, and employ a wild dependent bootstrap, following Shao (2010) and Feng et al. (2023), for inference. Our approach is thereby called the multilevel IPC ($MIPC$). In addition, we propose a consistent selection criteria for estimating the numbers of global factors r_0 and local factors r_i for each block i. When R tends to infinity, Jin et al. (2023) show that the consistent estimation of r_0 and r_i can be achieved in a sequential manner because each block is asymptotically negligible. By contrast, when R is finite, consistent estimation of r_0 and r_i is difficult because the eigenvalues generated by the global and local factors are not well separated (see Han (2021)). While the information criteria in Ando & Bai (2014) and Ando & Bai (2017) are available for finite R, they can be computationally infeasible even when R is only mildly large, as pointed out by Choi et al. (2021). Our approach provides consistency and valid inference by requiring only the consistent estimation of $r_0 + r_i$ for each i, which can be easily done block-by-block.

We focus primarily on homogeneous slope coefficients instead of heterogeneous ones, which are considered in Ando & Bai (2014) and Ando & Bai (2017). In the heterogeneous case, the individual slope estimates are $\sqrt{T}$-consistent without bias terms.[1] By contrast, the optimal convergence rate $\sqrt{NT}$ can be achieved in our homogeneous model but comes with biased terms in the asymptotic

[1]We extend $MIPC$ to a heterogeneous model; see Appendix C.2 for details.

distribution. We develop consistent estimators of these bias terms and propose a bias-corrected estimator. These are new results in the literature. In the appendix, we show that $MIPC$ can be generalised to the heterogeneous coefficient model.

In the final step, we apply the generalized canonical correlation (GCC) estimation developed by Lin & Shin (2022) to disentangle the global and local factors, given consistent slope coefficients. Therefore, the results in Lin & Shin (2022) are directly applicable. This is useful in empirical studies when the global and local factors themselves are of interest rather than treated as nuisance parameters.

Via Monte Carlo simulation, we show the consistency of the estimated slope coefficients as well as the global and local factors. More importantly, $MIPC$ with the estimated numbers of factors has smaller size distortions than the infeasible version of IPC with the known total number of factors. This demonstrates the superiority of our approach.

The rest of the paper is organized as follows. We introduce the panel data model with multilevel factors and propose the $MIPC$ in Section 3.2. In Section 3.3, we analyze the asymptotic properties of the slope coefficients, factors, loadings, as well as the estimates of the numbers of factors. Section 3.4 provides a detailed algorithm for the implementation of $MIPC$. In Section 3.5, we demonstrate the finite sample performance of our approach using Monte Carlo simulations. We apply our approach to analysing the nexus between energy consumption and economic growth in Section 3.6 using cross-country multi-dimensional panel data. Section 3.7 is the concluding remarks. All the proofs are relegated to the Appendix.

3.2 The Model

Consider the following panel regression model with multilevel factors:

$$\begin{aligned} y_{ijt} &= \mathbf{X}'_{ijt}\boldsymbol{\beta} + u_{ijt},\ i = 1, ..., R,\ j = 1, ..., N_i,\ t = 1, ..., T \\ u_{ijt} &= \boldsymbol{\gamma}'_{ij}\mathbf{G}_t + \boldsymbol{\lambda}'_{ij}\mathbf{F}_{it} + e_{ijt} \end{aligned} \tag{3.2.1}$$

where $\mathbf{X}_{ijt} = \left[x^1_{ijt}, \ldots, x^p_{ijt}\right]'$ is the $p \times 1$ regressors and $\boldsymbol{\beta} = \left[\beta^1, \ldots, \beta^p\right]'$ is the corresponding slope parameters. The error term u_{ijt} consists of the global, local and idiosyncratic components. $\mathbf{G}_t = \left[G^1_t, ..., G^{r_0}_t\right]'$ is the $r_0 \times 1$ unobserved global factors, $\mathbf{F}_{it} = \left[F^1_{it}, ..., F^{r_i}_{it}\right]'$ is the $r_i \times 1$ unobserved 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. We assume that the number of blocks R, as well as the numbers of factors $r_0, r_1, \ldots, r_R$ are fixed, whilst the numbers of individuals in each block $N_1, \ldots, N_R$ and the time observations T tend to infinity. Let $\underline{N} = \min\{N_1, \ldots, N_R\}$ and $N = \sum_{i=1}^R N_i$, we assume that $\underline{N} \to \infty$ through out this paper. We also assume that the group membership is known a priori. In vector notation, the above model can be written as

$$\mathbf{Y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{u}_{ij} \text{ with } \mathbf{u}_{ij} = \mathbf{G}\boldsymbol{\gamma}_{ij} + \mathbf{F}_i\boldsymbol{\lambda}_{ij} + \mathbf{e}_{ij} = \mathbf{K}_i\boldsymbol{\theta}_{ij} + \mathbf{e}_{ij} \tag{3.2.2}$$

where $\mathbf{Y}_{ij}$, $\mathbf{X}_{ij}$, $\mathbf{G}$, $\mathbf{F}_i$, $\mathbf{u}_{ij}$ and $\mathbf{e}_{ij}$ are formed by stacking the corresponding identities across t and $\mathbf{K}_i = [\mathbf{G}, \mathbf{F}_i]$ and $\boldsymbol{\theta}_{ij} = \left[\boldsymbol{\gamma}'_{ij}, \boldsymbol{\lambda}'_{ij}\right]'$. Let $d_i = r_0 + r_i$ for all i. For convenience, we express the error components using the following matrix notation:

$$\mathbf{u}_i = \mathbf{G}\mathbf{\Gamma}_i' + \mathbf{F}_i\mathbf{\Lambda}_i' + \mathbf{e}_i = \mathbf{K}_i\mathbf{\Theta}_i' + \mathbf{e}_i$$

where $\mathbf{u}_i = [\mathbf{u}_{i1}, \ldots, \mathbf{u}_{iN_i}]$, $\mathbf{\Gamma}_i = [\boldsymbol{\gamma}_{i1}, \ldots, \boldsymbol{\gamma}_{iN_i}]'$, $\mathbf{\Lambda}_i = [\boldsymbol{\lambda}_{i1}, \ldots, \boldsymbol{\lambda}_{iN_i}]'$, and $\mathbf{\Theta}_i = [\mathbf{\Gamma}_i, \mathbf{\Lambda}_i]$.

As pointed out by Bai (2009) and Feng et al. (2023), the specification in (3.2.2) incorporates various widely applied fixed effect models as special cases, such as individual fixed effects and time effects. Moreover, the regressors are allowed to be arbitrarily correlated with the factors and loadings. Our model extends the interactive fixed effects by allowing the factors to be block specific (semi-pervasive) rather than pervasive. For example, y_{ijt} is the house price (or growth rate) for city j that belongs to region i and $\mathbf{X}_{ijt}$ is the income and population. The house prices across different regions are subject to the national business cycle captured by $\mathbf{G}_t$. In addition, the regional economic shocks, captured by the local factors $\mathbf{F}_{it}$, can also be an important source of the local house price co-movement. The omission of such local factors may lead to inconsistent estimation of the slope parameters if the regressors are correlated with the local components. It may also invalidate the inference due to the strong cross-section correlation induced by the local factors.

It is important to note that model (3.2.1) can be written as a two dimensional model as in Bai (2009). Consider the factor structure in u_{ijt}. Let $\mathbf{u}_{i.t} = [u_{i1t}, \ldots, u_{iN_it}]'$. The factor structure in the error component can be written as

$$\mathbf{u}_t = \mathbf{\Theta}^+\mathbf{K}_t^+ + \mathbf{e}_t, \tag{3.2.3}$$

where

$$\underset{N\times 1}{\mathbf{u}_t} = \begin{bmatrix} \mathbf{u}_{1.t} \\ \vdots \\ \mathbf{u}_{R.t} \end{bmatrix}, \underset{N\times 1}{\mathbf{e}_t} = \begin{bmatrix} \mathbf{e}_{1.t} \\ \vdots \\ \mathbf{e}_{R.t} \end{bmatrix}, \underset{r^+\times 1}{\mathbf{K}_t^+} = \begin{bmatrix} \mathbf{G}_t \\ \mathbf{F}_{1t} \\ \vdots \\ \mathbf{F}_{Rt} \end{bmatrix}, \underset{N\times r^+}{\mathbf{\Theta}^+} = \begin{bmatrix} \boldsymbol{\gamma}_1 & \boldsymbol{\lambda}_1 & 0 & \cdots & 0 \\ \boldsymbol{\gamma}_2 & 0 & \boldsymbol{\lambda}_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\gamma}_R & 0 & 0 & \cdots & \boldsymbol{\lambda}_R \end{bmatrix}.$$

Then, the model (3.2.1) can be written as

$$\mathbf{Y}_t = \mathbf{X}_t\boldsymbol{\beta} + \mathbf{\Theta}^+\mathbf{K}_t^+ + \mathbf{e}_t \tag{3.2.4}$$

where $\mathbf{Y}_t = [y_{11t}, \ldots, y_{1N_1t}, \ldots, y_{R1t}, \ldots, y_{RN_Rt}]'$ and $\mathbf{X}_t = [\mathbf{X}_{11t}, \ldots, \mathbf{X}_{1N_1t}, \ldots, \mathbf{X}_{R1t}, \ldots, \mathbf{X}_{RN_Rt}]'$. IPC can be valid when the total number of the factors r^+ is finite and known. However, it is well known that the determination of r^+ will be difficult as the local factors become too "weak" to be detected as the R increases. Additionally, the global and local factors cannot be separately identified by IPC. Feng et al. (2023) consider the multilevel model as in (3.2.1) in a setting that the number of blocks goes infinity. In such a case, the identification of the global and local factors from (3.2.3) can be done sequentially. Let $u_{ijt} = \boldsymbol{\gamma}_{ij}'\mathbf{G}_t + u_{ijt}^*$ where $u_{ijt}^* = \boldsymbol{\lambda}_{ij}'\mathbf{F}_{it} + e_{ijt}$. When $R \to \infty$, each block is asymptotically negligible so the local components $\boldsymbol{\lambda}_{ij}'\mathbf{F}_{it}$ only introduce weak correlations among the error terms u_{ijt}^*. As a result, the PC estimation applied to $\mathbf{u}_t$ can consistently estimate $\mathbf{G}_t$ and then $\mathbf{F}_{it}$ (as well as their dimension) in a sequential manner. However, if R is fixed, the local components $\boldsymbol{\lambda}_{ij}'\mathbf{F}_{it}$ violates the weak cross-section correlation assumption of u_{ijt}^* so PC is no longer applicable, as pointed out by Breitung & Eickmeier (2016). In many applications, R is finite rather than tends to infinity. Therefore, it is worth highlighting that the multilevel factor structure with finite R brings new challenges in estimation as well as model selection. In this article, we establish an estimation procedure that works under fixed R, providing valid model selection and inferential theory for the estimates.

We define $\mathbf{M}_A = \mathbf{I} - \mathbf{P}_A$ and $\mathbf{P}_A = \mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}'$ for any matrix $\mathbf{A}$ that has full column rank. Assuming that d_i's are known a priori, we minimise the objective function defined as follows:

$$\mathcal{Q}\left(\mathbf{b}, \boldsymbol{\mathcal{K}}_1, \ldots, \boldsymbol{\mathcal{K}}_R\right) = \sum_{i=1}^{R}\sum_{j=1}^{N_i}\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\mathbf{b}\right)'\mathbf{M}_{\mathcal{K}_i}\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\mathbf{b}\right) \tag{3.2.5}$$

where $\boldsymbol{\mathcal{K}}_i$ is a $T \times d_i$ matrix for all i. The estimators are given by

$$\left(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{K}}_1, \ldots, \widehat{\mathbf{K}}_R\right) = \underset{(\mathbf{b}, \boldsymbol{\mathcal{K}}_1, \ldots, \boldsymbol{\mathcal{K}}_R) \in \mathcal{D}}{\arg\min}\ \mathcal{Q}\left(\mathbf{b}, \boldsymbol{\mathcal{K}}_1, \ldots, \boldsymbol{\mathcal{K}}_R\right)$$

where $\mathcal{D} = \mathbb{R}^p \times \mathbb{K}^R$ and $\mathbb{K} = \left\{\mathbf{K} | T^{-1}\mathbf{K}'\mathbf{K} = \mathbf{I}_{d_i}\right\}$. The factors and loadings are subject to rotational indeterminacy, i.e. $\mathbf{K}_i\boldsymbol{\Theta}_i'$ is observationally equivalent to $\mathbf{K}_i\mathbf{A}\mathbf{A}^{-1}\boldsymbol{\Theta}_i'$ for any invertible matrix $\mathbf{A}$. Therefore, we impose $\widehat{\mathbf{K}}_i'\widehat{\mathbf{K}}_i/T = \mathbf{I}_{d_i}$ and that $\widehat{\boldsymbol{\Theta}}_i'\widehat{\boldsymbol{\Theta}}_i$ is diagonal for each block i as identification restrictions. The above objective function implies that the solutions $\left(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{K}}_1, \ldots, \widehat{\mathbf{K}}_R\right)$ are the following

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\boldsymbol{M}_{\widehat{K}_i}\mathbf{X}_{ij}\right)^{-1}\left(\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\boldsymbol{M}_{\widehat{K}_i}\mathbf{Y}_{ij}\right) \tag{3.2.6}$$

$$\widehat{\mathbf{K}}_i\widehat{\mathbf{V}}_i = \left[\frac{1}{N_iT}\sum_{j=1}^{N_i}\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}\right)\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}\right)'\right]\widehat{\mathbf{K}}_i \text{ for each } i \tag{3.2.7}$$

where $\widehat{\mathbf{V}}_i$ is a diagonal matrix consisting of the $d_i = r_0 + r_i$ largest eigenvalues of the term in the square bracket of (3.2.7).

Following Bai's IPC, $\left(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{K}}_1, \ldots, \widehat{\mathbf{K}}_R\right)$ are obtained in an iterative manner. Given $\widehat{\boldsymbol{\beta}}$, $\widehat{\mathbf{K}}_i$ is updated block by block as in (3.2.7). Given $\widehat{\mathbf{K}}_i$ for each i, $\widehat{\boldsymbol{\beta}}$ is updated using (3.2.6). We thus call this approach multilevel IPC (*MIPC*). The solutions (3.2.6) and (3.2.7) are the same as in Feng et al. (2023). However, it is important to note that, under finite R, the bias terms in $\widehat{\boldsymbol{\beta}}$ do not vanish as is the case when $R \to \infty$. Our theory suggests that the consistency of $\widehat{\boldsymbol{\beta}}$ only requires the consistency of $\widehat{\mathbf{K}}_i$'s without separate identification of $\mathbf{G}$ and $\mathbf{F}_i$'s, which is different from the approach developed by Ando & Bai (2014). The advantages of *MIPC* are twofold. From a theoretical perspective, the sequential projection of global and local factors in the $\sqrt{NT}$-consistency framework introduces asymptotic bias terms. However, these bias terms are significantly simplified by employing the *MIPC* method. On the practical side, the estimation of d_i's is substantially easier than the joint estimation of r_0 and r_i's, as we explained earlier. In the next section, we shall demonstrate the large sample properties of our estimators and propose a bias corrected estimator for the slope coefficients.

3.3 Asymptotic Analysis

3.3.1 Consistency and asymptotic distribution

In this section, we assume for the moment that the numbers of the factors d_i are known for each block i. The consistent estimation of the numbers of factors will be developed in Section 3.3.2. We make the following assumptions for the consistency of our estimators.

Let $\mathcal{M}$ be a finite constant.

Assumption 3.A. $\mathrm{E}\left(\|\mathbf{X}_{ijt}\|^4\right) \le \mathcal{M}$ *for all* i, j, *and* t. *Furthermore, we have*

$$\inf_{\mathcal{K}_1,\ldots,\mathcal{K}_R \in \mathbb{K}^R} \mathbf{D}\left(\mathcal{K}_1,\ldots,\mathcal{K}_R\right) > 0.$$

where $\mathbf{D}\left(\mathcal{K}_1,\ldots,\mathcal{K}_R\right)$ *is a* $p \times p$ *matrix such that*

$$\mathbf{D}\left(\mathcal{K}_1,\ldots,\mathcal{K}_R\right) = \frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}'_{ij}\left(\mathbf{M}_{\mathcal{K}_i}\mathbf{X}_{ij} - \frac{1}{N_i}\sum_{k=1}^{N_i} a_{i,kj}\mathbf{M}_{\mathcal{K}_i}\mathbf{X}_{ik}\right)$$

with $a_{i,kj} = \boldsymbol{\theta}'_{ik}\left(\boldsymbol{\Theta}'_i\boldsymbol{\Theta}_i/N_i\right)^{-1}\boldsymbol{\theta}_{ij}$.

Assumption 3.B.

1. $\mathrm{E}\left(\|\mathbf{K}_{it}\|^4\right) \le \mathcal{M}$ *and* $T^{-1}\mathbf{K}'_i\mathbf{K}_i \xrightarrow{p} \boldsymbol{\Sigma}_{K_i}$ *for each* i *as* $T \to \infty$ *where* $\boldsymbol{\Sigma}_{K_i}$ *is an* $d_i \times d_i$ *positive definite matrix.*
2. $\mathrm{E}\left(\|\boldsymbol{\theta}_{ij}\|^4\right) \le \mathcal{M}$ *and* $N_i^{-1}\boldsymbol{\Theta}'_i\boldsymbol{\Theta}_i \xrightarrow{p} \boldsymbol{\Sigma}_{\Theta_i}$ *for each* i *as* $N_i \to \infty$ *where* $\boldsymbol{\Sigma}_{\Theta_i}$ *is an* $d_i \times d_i$ *positive definite matrix.*

Assumption 3.C.

1. *For all* i, j *and* t, $\mathrm{E}(e_{ijt}) = 0$ *and* $\mathrm{E}\left(|e_{ijt}|^8\right) \le \mathcal{M}$.
2. *For all* i, $\mathrm{E}(e_{ijt}e_{iks}) = \sigma_{i,(jk),(ts)}$, $|\sigma_{i,(jk),(ts)}| \le \sigma_{i,(jk)}$ *for all* (t,s) *and* $|\sigma_{i,(jk),(ts)}| \le \tau_{i,(ts)}$ *for all* (j,k) *such that*

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

3. *For all* i, t *and* s

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

4. *For all* i, *we have*

$$\frac{1}{N_iT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{T}\sum_{v=1}^{T}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}|\operatorname{Cov}(e_{ijt}e_{ijs}, e_{iku}e_{ikv})| \le \mathcal{M}$$

$$\frac{1}{N_i^2 T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\sum_{c=1}^{N_i}\sum_{d=1}^{N_i}|\operatorname{Cov}(e_{ijt}e_{ikt}, e_{ics}e_{ids})| \leq \mathcal{M}$$

Assumption 3.D. *e_{mjt} is independent of $\mathbf{X}_{hks}$, $\boldsymbol{\theta}_{hk}$ and $\mathbf{K}_{hs}$ for all m, h, j, k, t, and s.*

Assumption 3.A plays an important role as an identification condition. This assumption rules out common regressors that do not vary across individuals and also time-invariant regressors that do not vary across time. Moreover, it is straightforward to show that the following equation holds

$$\mathbf{D}\left(\mathbf{K}_1, \ldots, \mathbf{K}_R\right) = \frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{Z}_{ij}'\mathbf{Z}_{ij}.$$

where $\mathbf{Z}_{ij}$ a $T \times p$ matrix such that

$$\mathbf{Z}_{ij} = \mathbf{M}_{K_i}\mathbf{X}_{ij} - \frac{1}{N_i}\sum_{k=1}^{N_i} a_{i,kj}\mathbf{M}_{K_i}\mathbf{X}_{ik}.$$

Assumption 3.B ensures the existence of the factors in each block i. Assumption 3.C allows cross-sectional and serial correlation in the error terms, but they are restricted by Assumption 3.C.2-3.C.4. Assumptions 3.A-3.D extend Assumptions A–D in Bai (2009) by requiring those conditions to be met in each block i. These assumptions are standard in the literature.

Lemma 3.1. *Let $\left(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{K}}_1, \ldots, \widehat{\mathbf{K}}_R\right)$ be the estimators from minimising (3.2.5). Under Assumption 3.A-3.D, as $N_1, \ldots, N_R, T \to \infty$, the following statements holds:*

1. *$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \xrightarrow{p} \mathbf{0}$.*
2. *For each i, the matrix $\mathbf{K}_i'\widehat{\mathbf{K}}_i/T$ is invertible and $\left\|\mathbf{P}_{\widehat{K}_i} - \mathbf{P}_{K_i}\right\| \xrightarrow{p} 0$.*

Assumption 3.E. *For each i, we have $\lim_{N_i, N\to\infty} N/N_i = \alpha_i \leq \mathcal{M}$.*

Lemma 3.2. *Under Assumption 3.A-3.E, as $N_1, \ldots, N_R, T \to \infty$ and $T/N_i \to \rho_i > 0$ for all i, then $\sqrt{NT}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = O_p(1)$.*

Lemma 3.1 establishes the consistency of the slope coefficients and the factor spaces for each block i based on the extreme estimation. At this stage, we are only able to claim consistency in Lemma 3.1.2 because the number of parameters grows with T. The consistency of $\widehat{\mathbf{K}}_i$ up to rotation is established by Lemma C.1.2 in the Appendix. Lemma 3.2 shows that $\widehat{\boldsymbol{\beta}}$ achieves a convergence rate of $\sqrt{NT}$ assuming that all N_i's are of the same order of magnitude and each N_i is comparable with T. Assumption 3.E is standard in the multilevel factor literature, see Ando & Bai (2014) and Andreou et al. (2019) for examples.

Assumption 3.F. *For some nonrandom positive definite matrix $\mathbb{D}_Z$, as $N, T \to \infty$, we have*

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{Z}_{ij}'\mathbf{e}_{ij} \xrightarrow{d} N\left(\mathbf{0}, \mathbb{D}_Z\right)$$

where

$$\mathbb{D}_Z = \text{plim} \frac{1}{NT} \sum_{m=1}^{R} \sum_{h=1}^{R} \sum_{t=1}^{T} \sum_{s=1}^{T} \sigma_{(mh),(jk),(ts)} \mathbf{Z}_{mjt} \mathbf{Z}'_{hks}.$$

Theorem 3.1. *Under Assumption 3.A-3.F, as $N_1, \ldots, N_R, T \to \infty$ and $T/N_i \to \rho_i > 0$ for all i, then*

$$\sqrt{NT}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left(\sum_{i=1}^{R}\left(\alpha_i^{-1/2}\rho_i^{1/2}\mathbb{B}_i^0 + \alpha_i^{-1/2}\rho_i^{-1/2}\mathbb{C}_i^0\right), \mathbb{D}_0^{-1}\mathbb{D}_Z\mathbb{D}_0^{-1}\right)$$

where $\mathbb{B}_i^0 = \text{plim}\, \mathbb{B}_i$ and $\mathbb{C}_i^0 = \text{plim}\, \mathbb{C}_i$ with

$$\mathbb{B}_i = -\mathbf{D}\left(\mathbf{K}_1, \ldots, \mathbf{K}_R\right)^{-1} \frac{1}{N_i} \sum_{j=1}^{N_i} \sum_{k=1}^{N_i} \left[\frac{\left(\mathbf{X}_{ij} - \mathbf{W}_{ij}\right)' \mathbf{K}_i}{T}\right] \mathbb{P}_i \boldsymbol{\theta}_{ik} \left(\frac{1}{T}\sum_{t=1}^{T} \sigma_{i,(jk),(tt)}\right),$$

$$\mathbb{C}_i = -\mathbf{D}\left(\mathbf{K}_1, \ldots, \mathbf{K}_R\right)^{-1} \frac{1}{N_i T} \sum_{j=1}^{N_i} \mathbf{X}'_{ij} \mathbf{M}_{K_i} \left[\frac{1}{N_i}\sum_{k=1}^{N_i} \mathrm{E}\left(\mathbf{e}_{ik}\mathbf{e}'_{ik}\right)\right] \mathbf{K}_i \mathbb{P}_i \boldsymbol{\theta}_{ij},$$

$$\mathbf{W}_{ij} = \frac{1}{N_i}\sum_{k=1}^{N_i} a_{i,jk}\mathbf{X}_{ik}, \ \ \textit{and}\ \mathbb{P}_i = \left(\frac{\mathbf{K}'_i\mathbf{K}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Theta}'_i\boldsymbol{\Theta}_i}{N_i}\right)^{-1}.$$

In addition, $\mathbb{D}_0 = \text{plim}_{N_1,\ldots,N_R,T\to\infty} \mathbf{D}\left(\mathbf{K}_1, \ldots, \mathbf{K}_R\right) = \text{plim}_{N_1,\ldots,N_R,T\to\infty} (NT)^{-1} \sum_{i=1}^{R}\sum_{j=1}^{N_i} \mathbf{Z}'_{ij}\mathbf{Z}_{ij}$ and $\mathbb{D}_Z$ is defined in Assumption 3.F.

Theorem 3.1 shows that the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ is not centered. The bias term $\mathbb{B}_i$ evolves due to the heteroskedasticity and cross-section correlation between contemporaneous error terms and $\mathbb{C}_i$ is a consequence of correlation and heteroskedasticity in the time dimension. If it is assumed that the error terms are *i.i.d*, we have $\mathbb{B}_i = \mathbf{0}$ and $\mathbb{C}_i = \mathbf{0}$ for all i so there is no bias. Furthermore, the *i.i.d* assumption leads to $\mathbb{D}_Z = \mathbb{D}_0$ and hence we have $\sqrt{NT}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbb{D}_0\right)$. However, if the error terms are not *i.i.d*, then a bias correction is required.

Let $\widehat{\mathbb{B}}_i$ and $\widehat{\mathbb{C}}_i$ be the estimators of $\mathbb{B}_i$ and $\mathbb{C}_i$, respectively, such that

$$\widehat{\mathbb{B}}_i = -\mathbf{D}\left(\widehat{\mathbf{K}}_1, \ldots, \widehat{\mathbf{K}}_R\right)^{-1} \frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\left[\frac{\left(\mathbf{X}_{ij} - \widehat{\mathbf{W}}_{ij}\right)'\widehat{\mathbf{K}}_i}{T}\right]\left(\frac{\widehat{\boldsymbol{\Theta}}'_i\widehat{\boldsymbol{\Theta}}_i}{N_i}\right)^{-1}\widehat{\boldsymbol{\theta}}_{ik}\hat{\sigma}_{i,(jk)}$$

with $\widehat{\boldsymbol{\theta}}_{ik} = T^{-1}\widehat{\mathbf{K}}'_i\left(\mathbf{Y}_{ik} - \mathbf{X}_{ik}\widehat{\boldsymbol{\beta}}\right)$, $\widehat{\boldsymbol{\Theta}}_i = \left[\widehat{\boldsymbol{\theta}}_{i1}, \ldots, \widehat{\boldsymbol{\theta}}_{iN_i}\right]'$, $\hat{\sigma}_{i,(jk)} = T^{-1}\widehat{\mathbf{e}}'_{ij}\widehat{\mathbf{e}}_{ik}$, $\widehat{\mathbf{e}}_{ij} = \mathbf{Y}_{ij} - \mathbf{X}_{ij}\widehat{\boldsymbol{\beta}} - \widehat{\mathbf{K}}_i\widehat{\boldsymbol{\theta}}_{ij}$, $\widehat{\mathbf{e}}_{ik} = \mathbf{Y}_{ik} - \mathbf{X}_{ik}\widehat{\boldsymbol{\beta}} - \widehat{\mathbf{K}}_i\widehat{\boldsymbol{\theta}}_{ik}$, and

$$\widehat{\mathbb{C}}_i = -\mathbf{D}\left(\widehat{\mathbf{K}}_1, \ldots, \widehat{\mathbf{K}}_R\right)^{-1}\frac{1}{N_i T}\sum_{j=1}^{N_i}\mathbf{X}'_{ij}\mathbf{M}_{\widehat{K}_i}\widehat{\mathbb{S}}_i\widehat{\mathbf{K}}_i\left(\frac{\widehat{\boldsymbol{\Theta}}'_i\widehat{\boldsymbol{\Theta}}_i}{N_i}\right)^{-1}\widehat{\boldsymbol{\theta}}_{ij},$$

where $\widehat{\mathbb{S}}_i$ is an estimator of $\mathbb{S}_i$ whose (t,s)-th element is $\widehat{\mathbb{S}}_{i,(ts)} = N_i^{-1}\sum_{k=1}^{N_i}\hat{e}_{ikt}\hat{e}_{iks}$. Assuming $\mathrm{E}(e_{mjt}e_{hkt}) = \sigma_{(mh),(jk)}$ for all m, h, j, k, and t, it can be shown that $\sqrt{T/N_i}\left(\widehat{\mathbb{B}}_i - \mathbb{B}_i\right) = o_p(1)$

and $\sqrt{N_i/T}\left(\widehat{\mathbb{C}}_i - \mathbb{C}_i\right) = o_p(1)$ (see Lemma C.1.7). Then, the bias-corrected estimator is given by

$$\widehat{\boldsymbol{\beta}}_{bc} = \widehat{\boldsymbol{\beta}} - \sum_{i=1}^{R}\left(\frac{1}{N}\widehat{\mathbb{B}}_i + \frac{N_i}{NT}\widehat{\mathbb{C}}_i\right).$$

Theorem 3.2. *Under Assumptions 3.A-3.F, and $E(e_{mjt}e_{hkt}) = \sigma_{(mh),(jk)}$ for all m, h, j, k, and t, as $N_1, \ldots, N_R, T \to \infty$, $T/N^2 \to 0$, and $N/T^2 \to 0$, then*

$$\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}_{bc} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbb{D}_0^{-1}\mathbb{D}_Z\mathbb{D}_0^{-1}\right).$$

Under the presence of weak cross-section correlation and serial correlation in the error term, the consistent estimation of the covariance matrix $\mathbb{D}_0^{-1}\mathbb{D}_Z\mathbb{D}_0^{-1}$ is infeasible in practice. Therefore, following Feng et al. (2023), we adopt a dependent wild bootstrap advanced by Shao (2010) to construct a valid confidence interval for $\widehat{\boldsymbol{\beta}}_{bc}$. The detailed algorithm for implementation is provided in Section 3.4.[2]

3.3.2 Determining the numbers of factors

So far we have assumed that the numbers of factors d_i's are known. In this section we develop a consistent selection criterion. We set a finite integer that is common to all blocks, $d_{\max}$, such that $d_{\max} \geq \max_i\{d_i\}$. We consider the objective function defined as follows:

$$\mathscr{Q}\left(\mathbf{b}, \mathscr{K}_1, \ldots, \mathscr{K}_R\right) = \sum_{i=1}^{R}\sum_{j=1}^{N_i}\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\mathbf{b}\right)' \mathbf{M}_{\mathscr{K}_i}\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\mathbf{b}\right) \tag{3.3.8}$$

where $\mathscr{K}_i$ is a $T \times d_{\max}$ matrix for all i. The initial estimators are given by

$$\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{K}}_1, \ldots, \widetilde{\mathbf{K}}_R\right) = \underset{(\boldsymbol{\beta}, \mathscr{K}_1, \ldots, \mathscr{K}_R) \in \mathscr{D}}{\arg\min}\ \mathscr{Q}\left(\mathbf{b}, \mathscr{K}_1, \ldots, \mathscr{K}_R\right)$$

where $\mathscr{D} = \mathbb{R}^p \times \widetilde{\mathbb{K}}^R$ and $\widetilde{\mathbb{K}} = \left\{\mathbf{K} | T^{-1}\mathbf{K}'\mathbf{K} = \mathbf{I}_{d_{\max}}\right\}$. Again, the above objective function implies the following solutions:

$$\widetilde{\boldsymbol{\beta}} = \left(\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\boldsymbol{M}_{\widetilde{K}_i}\mathbf{X}_{ij}\right)^{-1}\left(\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\boldsymbol{M}_{\widetilde{K}_i}\mathbf{Y}_{ij}\right) \tag{3.3.9}$$

$$\widetilde{\mathbf{K}}_i\widetilde{\mathbf{V}}_i = \left[\frac{1}{N_iT}\sum_{j=1}^{N_i}\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\widetilde{\boldsymbol{\beta}}\right)\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\widetilde{\boldsymbol{\beta}}\right)'\right]\widetilde{\mathbf{K}}_i \text{ for each } i. \tag{3.3.10}$$

Under Assumption 3.A-3.E, it can be shown that $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\mathbf{K}}_i$'s are consistent estimators. Define

[2]We validate the bootstrap procedure using simulations in Section 3.5 and the formal proof is beyond the scope of this paper.

the covariance matrix for each block i:

$$\widetilde{\mathbf{\Sigma}}_i = \frac{1}{N_i T} \sum_{j=1}^{N_i} \left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\widetilde{\boldsymbol{\beta}} \right) \left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\widetilde{\boldsymbol{\beta}} \right)'. \tag{3.3.11}$$

Let $\tilde{v}_{i,k}$ be the k-th largest eigenvalue of $\widetilde{\mathbf{\Sigma}}_i$. It can be shown that for $k = 1, \ldots, d_i$, $\tilde{v}_{i,k}$ converges to its population counterpart $v_{i,k}$, which is larger than zero. On the contrary, $\tilde{v}_{i,k}$ is asymptotically zero for $k = d_i + 1, \ldots, d_{\max}$. Therefore, we follow Lam & Yao (2012) and Ahn & Horenstein (2013) and develop estimators for d_i based on the ratios of the adjacent eigenvalues as follows:

$$\hat{d}_i = \underset{k=0,\ldots,d_{\max}-1}{\arg\max} \frac{\tilde{v}_{i,k}}{\tilde{v}_{i,k+1}} \text{ for } i = 1, \ldots, R$$

where $\tilde{v}_{i,0} = \sum_{\ell=1}^{C^2_{N_i T}} \tilde{v}_{i,\ell} / \log\left(C^2_{N_i T}\right)$ is a mock eigenvalue allowing the case where $d_i = 0$. Using the results of Lemma C.1.9, for each i, it is straightforward that for $k = 1, \ldots, d_i - 1, d_i + 1, \ldots, d_{\max} - 1$, $\tilde{v}_{i,k}/\tilde{v}_{i,k+1} = O_p(1)$, whilst for $k = d_i$ we have $\tilde{v}_{i,k}/\tilde{v}_{i,k+1} \to_p \infty$. Therefore, the ratio of the eigenvalues falls sharply at $k = d_i$. The mock eigenvalue $\tilde{v}_{i,0}$ is smaller in magnitude than the largest d_i eigenvalues but larger than the rest, which allows the case $d_i = 0$. The consistency of $\hat{d}_i$'s is summarised in the following proposition:

Proposition 3.1. *Under Assumptions 3.A–3.E, we have*

$$\lim_{N_i, T \to \infty} \Pr\left(\hat{d}_i = r_0 + r_i\right) = 1.$$

Alternatively, one may obtain $\hat{d}_i$ by implementing the selection criteria proposed by Bai & Ng (2002) to $\widetilde{\mathbf{Y}}_i = \left[\widetilde{\mathbf{Y}}_{i1}, \ldots, \widetilde{\mathbf{Y}}_{iN_i}\right]$ for each i separately where $\widetilde{\mathbf{Y}}_{ij} = \mathbf{Y}_{ij} - \mathbf{X}_{ij}\widetilde{\boldsymbol{\beta}}$. In the supplement appendix of Bai (2009), it is argued without rigorous proof that the estimated slope coefficients $\widetilde{\boldsymbol{\beta}}_i$ are $\sqrt{N_i T}$ consistent even when $d_{\max} > d_i$ factors are estimated. Moon & Weidner (2015) studies the asymptotic property of $\widetilde{\boldsymbol{\beta}}$ using a different set of assumptions other than Bai (2009).

3.3.3 Disentangling the global and local factors

While the estimation and inference do not necessarily require separate identification of global and local factors, practitioners may, in some instances, wish to estimate $\mathbf{G}$ and $\mathbf{F}_i$ individually, each with its economic interpretation. As a result, it becomes necessary to estimate the multilevel factor model as follows, after obtaining $\widehat{\boldsymbol{\beta}}_{bc}$:

$$\mathbf{Y}_{ij} - \mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}_{bc} = \mathbf{K}_i \boldsymbol{\theta}_{ij} + \mathbf{e}_{ij} = \mathbf{G}\boldsymbol{\gamma}_{ij} + \mathbf{F}_i \boldsymbol{\lambda}_{ij} + \mathbf{e}_{ij}$$

Different techniques have been devised to disentangle global and local factors. In a special case where $R = 2$, Andreou et al. (2019) propose using CCA to estimate the global factors based on the canonical correlation between $\widehat{\mathbf{K}}_1$ and $\widehat{\mathbf{K}}_2$ by solving the characteristic equations:

$$\left(\widehat{\mathbf{S}}_{12}\widehat{\mathbf{S}}_{22}^{-1}\widehat{\mathbf{S}}_{21} - \ell\widehat{\mathbf{S}}_{11}\right)\boldsymbol{\mathcal{V}}_1 = \mathbf{0} \text{ or } \left(\widehat{\mathbf{S}}_{21}\widehat{\mathbf{S}}_{11}^{-1}\widehat{\mathbf{S}}_{12} - \ell\widehat{\mathbf{S}}_{22}\right)\boldsymbol{\mathcal{V}}_2 = \mathbf{0}.$$

where $\widehat{\mathbf{S}}_{ab}$ $(a, b = 1, 2)$ represents the covariance matrix between $\widehat{\mathbf{K}}_1$ and $\widehat{\mathbf{K}}_2$. The global factors can be estimated as $\widehat{\mathbf{G}} = \widehat{\mathbf{K}}_1 \boldsymbol{\mathcal{V}}_1^{r_0}$ or $\widehat{\mathbf{G}} = \widehat{\mathbf{K}}_2 \boldsymbol{\mathcal{V}}_2^{r_0}$, where $\boldsymbol{\mathcal{V}}_m^{r_0}$ and $\boldsymbol{\mathcal{V}}_h^{r_0}$ are matrices comprising the characteristic vectors corresponding to the r_0 largest characteristic roots. Subsequently, by projecting out $\widehat{\mathbf{G}}$, one can obtain $\widehat{\mathbf{F}}_i$'s by applying PC to each block separately. In the case where $R \geq 2$, Breitung & Eickmeier (2016) and Choi et al. (2018) use the global factors generated by CCA as initial estimates and propose iterative estimation procedures where the global and local factors are sequentially updated given each other. However, as highlighted by Lin & Shin (2022), CCA may not always correctly identify the global factors when common local factors are present, as they might be misidentified as global factors. In such a scenario, CCA may lead to overestimation of the number of global factors and inconsistent factor estimates. Instead, Lin & Shin (2022) propose the generalised canonical correlation analysis (GCC), which extends the standard CCA by jointly dealing with the pairwise canonical correlation between any two blocks. They show that GCC outperforms existing approaches via simulation. In addition, GCC offers computational advantage as it does not involve iterations by achieving consistency in one sequential estimation. Therefore, in this article, we follow Lin & Shin (2022) to estimate the global and local factors using GCC.

Given the consistent estimates $\widehat{\mathbf{K}}_i$'s from the minimisation of (3.2.5), we follow the generalised canonical correlation (GCC) estimation developed by Lin & Shin (2022) for the global and local factors respectively. Specifically, we construct the following $T(R-1)R/2 \times \hat{d}^*$ system-wide matrix:

$$\widehat{\boldsymbol{\Phi}} = \begin{bmatrix} \widehat{\mathbf{K}}_1 & -\widehat{\mathbf{K}}_2 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \widehat{\mathbf{K}}_1 & \mathbf{0} & -\widehat{\mathbf{K}}_3 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ & & & & \vdots & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \widehat{\mathbf{K}}_{R-1} & -\widehat{\mathbf{K}}_R \end{bmatrix} \tag{3.3.12}$$

where $\hat{d}^* = \sum_{l=1}^R \hat{d}_l$ and perform a singular value decomposition (SVD) to $\widehat{\boldsymbol{\Phi}}$ as $\widehat{\boldsymbol{\Phi}} = \widehat{\mathbf{P}} \widehat{\boldsymbol{\Delta}} \widehat{\mathbf{Q}}'$. Let $\hat{\delta}_1, \dots, \hat{\delta}_{\hat{d}^*}$ be the diagonal elements of $\widehat{\boldsymbol{\Delta}}$ in *ascending order*. Then, the number of global factors r_0 can be obtained by

$$\hat{r}_0 = \underset{k=0,\dots,\hat{d}_{\min}}{\arg\max} \frac{\hat{\delta}_{k+1}^2}{\hat{\delta}_k^2}, \; \hat{d}_{\min} = \min\left\{\hat{d}_1, \dots, \hat{d}_R\right\}. \tag{3.3.13}$$

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

$$\hat{\delta}_0^2 = \frac{1}{C_{\underline{N}T} \hat{d}^*} \sum_{k=1}^{\hat{d}^*} \hat{\delta}_k^2.$$

The number of local factors can be simply obtained by $\hat{r}_i = \hat{d}_i - \hat{r}_0$ for each block i. The consistency of $\hat{r}_0$ and $\hat{r}_i$'s are summarised in the following proposition:

Proposition 3.2. *Under Assumptions 3.A–3.E, we have:*

$$\lim_{N_1,\dots,N_R,T\to\infty} \Pr(\hat{r}_0 = r_0) = 1 \;\textit{and}\; \lim_{N_1,\dots,N_R,T\to\infty} \Pr(\hat{r}_i = r_i) = 1$$

Let $\widehat{\mathbf{Q}}^{\hat{r}_0} = \left[\widehat{\mathbf{Q}}_1^{\hat{r}_0\prime}, \dots, \widehat{\mathbf{Q}}_R^{\hat{r}_0\prime}\right]'$ as the first $\hat{r}_0$ columns of $\widehat{\mathbf{Q}}$ from the above SVD, and construct the

$T \times R\hat{r}_0$ matrix, $\widehat{\mathbf{\Psi}} = \left[\widehat{\mathbf{K}}_1\widehat{\mathbf{Q}}_1^{\hat{r}_0}, \ldots, \widehat{\mathbf{K}}_R\widehat{\mathbf{Q}}_R^{\hat{r}_0}\right]$. We perform the eigen decomposition,

$$T^{-1}\widehat{\mathbf{\Psi}}\widehat{\mathbf{\Psi}}' = \widehat{\mathbf{L}}\widehat{\mathbf{\Xi}}\widehat{\mathbf{L}}' \tag{3.3.14}$$

where $\widehat{\mathbf{L}}$ is a $T \times R\hat{r}_0$ orthonormal matrix and $\widehat{\mathbf{\Xi}}$ is a $T \times T$ diagonal matrix consisting of the eigenvalues in *descending order*. Then, from (3.3.14), we obtain the consistent estimator of the global factors, denoted $\widehat{\mathbf{G}}$, by the $\hat{r}_0$ vectors of $\widehat{\mathbf{L}}$ corresponding to the $\hat{r}_0$ largest eigenvalues multiplied by $\sqrt{T}$. Let $\widehat{\mathbf{Y}}_{ij} = \mathbf{Y}_{ij} - \mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{Y}}_i = \left[\widehat{\mathbf{Y}}_{i1}, \ldots, \widehat{\mathbf{Y}}_{iN_i}\right]$. The global factor loadings can be estimated by $\widehat{\mathbf{\Gamma}}_i = T^{-1}\widehat{\mathbf{Y}}_i'\widehat{\mathbf{G}}$. Finally, let $\widehat{\mathbf{Y}}_{ij}^G = \mathbf{Y}_{ij} - \mathbf{X}_{ij}\widehat{\boldsymbol{\beta}} - \widehat{\mathbf{G}}\widehat{\gamma}_{ij}$, the local factors and loadings can be obtained by applying PC to each block $\widehat{\mathbf{Y}}_i^G = \left[\widehat{\mathbf{Y}}_{i1}^G, \ldots, \widehat{\mathbf{Y}}_{iN_i}^G\right]$. $\widehat{\mathbf{F}}_i$ is $\sqrt{T}$ multiply the $\hat{r}_i$ eigenvectors of $\widehat{\mathbf{Y}}_i^G\widehat{\mathbf{Y}}_i^{G\prime}$ corresponding to the $\hat{r}_i$ largest eigenvalues. The local factor loadings can be obtained by $\widehat{\mathbf{\Lambda}}_i = T^{-1}\widehat{\mathbf{Y}}_i^{G\prime}\widehat{\mathbf{F}}_i$.

To guarantee the consistency of $\widehat{\mathbf{\Gamma}}_i$, $\widehat{\mathbf{\Lambda}}_i$ and $\widehat{\mathbf{F}}_i$, we need the asymptotic orthogonality assumption stated as follows:

Assumption 3.G. *For every i, $T^{-1}\boldsymbol{K}_i'\boldsymbol{K}_i = \begin{bmatrix}\mathbf{\Sigma}_G & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{F_i}\end{bmatrix} + O_p\left(T^{-1/2}\right)$ where $\mathbf{\Sigma}_G$ and $\mathbf{\Sigma}_{F_i}$ are $r_0 \times r_0$ and $r_i \times r_i$ full rank matrices.*

Proposition 3.3.

Under Assumptions 3.A–3.E and 3.G, as $N_1, N_2, \ldots, N_R, T \to \infty$, we have:

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

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

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

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

where $\mathbb{H} = T^{-1/2}\mathbf{G}'\mathbf{J}^{r_0}\mathbf{U}$ is an $r_0 \times r_0$ rotation matrix, $\mathbf{J}^{r_0} = \mathbf{L}^{r_0}(\mathbf{\Xi}^{r_0})^{-1}$, $\mathbf{\Xi}^{r_0}$ is an $r_0 \times r_0$ diagonal matrix consisting of the r_0 non-zero eigenvalues of $T^{-1}\mathbf{G}\mathbf{G}'$ in descending order, $\mathbf{L}^{r_0}$ is a $T \times r_0$ matrix of the corresponding eigenvectors, and $\mathbf{U}$ is an $r_0 \times r_0$ orthogonal matrix. $\widehat{\mathscr{H}_i} = (\mathbf{\Lambda}_i'\mathbf{\Lambda}_i/N_i)\left(\widehat{\mathbf{F}}_i'\mathbf{F}_i/T\right)\widehat{\mathbf{\Upsilon}}_i^{-1}$ 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 $(N_iT)^{-1}\widehat{\mathbf{Y}}_i\widehat{\mathbf{Y}}_i'$ in descending order, $\widehat{\mathbf{Y}}_i = \mathbf{Y}_i - \widehat{\mathbf{G}}\widehat{\mathbf{\Gamma}}_i'$. Moreover, $C_{\underline{N},T} = \min\{\sqrt{\underline{N}}, \sqrt{T}\}$ with $\underline{N} = \min\{N_1, N_2, \ldots, N_R\}$.

3.4 Estimation Algorithm

In this section, we outline the detailed algorithm for estimation:

Estimating the number of factors d_i:

Step A1. Obtain the OLS estimates of $\boldsymbol{\beta}$ ignoring the factors, denoted $\widetilde{\boldsymbol{\beta}}^{(0)}$. Fix a sufficiently large integer $d_{\max}$ such that $d_{\max} \geq \max_i\{d_i\}$. For each i, obtain the factors $\widehat{\boldsymbol{K}}_i^{(0)}$ as $\sqrt{T}$ multiplied by the $d_{\max}$ eigenvectors of

$$\frac{1}{N_i T}\sum_{j=1}^{N_i}\left(\mathbf{Y}_{ij}-\mathbf{X}_{ij}\widetilde{\boldsymbol{\beta}}^{(0)}\right)\left(\mathbf{Y}_{ij}-\mathbf{X}_{ij}\widetilde{\boldsymbol{\beta}}^{(0)}\right)'.$$

$\widetilde{\boldsymbol{\beta}}^{(0)}$ and $\widetilde{\mathbf{K}}_i^{(0)}$'s are the initial estimates.
Step A2. Given $\widetilde{\mathbf{K}}_i^{(\ell-1)}$ for all i from the $(\ell-1)$-th step, update the slope coefficients as

$$\widetilde{\boldsymbol{\beta}}^{(\ell)} = \left(\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\boldsymbol{M}_{\widetilde{K}_i^{(\ell-1)}}\mathbf{X}_{ij}\right)^{-1}\left(\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\boldsymbol{M}_{\widetilde{K}_i^{(\ell-1)}}\mathbf{Y}_{ij}\right).$$

Step A3. Given $\widetilde{\boldsymbol{\beta}}^{(\ell)}$, update $\widetilde{\mathbf{K}}_i^{(\ell)}$ for each i as $\sqrt{T}$ multiplied by the $d_{\max}$ eigenvectors of

$$\frac{1}{N_i T}\sum_{j=1}^{N_i}\left(\mathbf{Y}_{ij}-\mathbf{X}_{ij}\widetilde{\boldsymbol{\beta}}^{(\ell)}\right)\left(\mathbf{Y}_{ij}-\mathbf{X}_{ij}\widetilde{\boldsymbol{\beta}}^{(\ell)}\right)'.$$

corresponding to the $d_{\max}$ largest eigenvalues in descending order.
Step A4. Repeat Step A2–A3 until convergence, and obtain the resulting $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\mathbf{K}}_i$ for each i.
Step A5. Construct the covariance matrix for each i as

$$\widetilde{\boldsymbol{\Sigma}}_i = \frac{1}{N_i T}\sum_{j=1}^{N_i}\left(\mathbf{Y}_{ij}-\mathbf{X}_{ij}\widetilde{\boldsymbol{\beta}}\right)\left(\mathbf{Y}_{ij}-\mathbf{X}_{ij}\widetilde{\boldsymbol{\beta}}\right)'$$

and obtain its $d_{\max}$ largest eigenvalues, denoted $\tilde{v}_{i,1},\ldots,\tilde{v}_{i,d_{\max}}$. For each i, estimate $\hat{d}_i$ as

$$\hat{d}_i = \underset{k=0,\ldots,d_{\max}-1}{\arg\max}\frac{\tilde{v}_{i,k}}{\tilde{v}_{i,k+1}}$$

where $\tilde{v}_{i,0} = \sum_{\ell=1}^{C_{N_iT}^2}\tilde{v}_{i,\ell}/\log\left(C_{N_iT}^2\right)$ is the mock eigenvalue.

Estimating the slope coefficients and factors:
Step B1. Let $\widehat{\boldsymbol{\beta}}^{(0)} = \widetilde{\boldsymbol{\beta}}$ and $\widehat{\mathbf{K}}_i^{(0)} = \widetilde{\mathbf{K}}_i$ for all i.
Step B2. Given $\widehat{\mathbf{K}}_i^{(\ell-1)}$ for all i from the $(\ell-1)$-th step, update the slope coefficients as

$$\widehat{\boldsymbol{\beta}}^{(\ell)} = \left(\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^{(\ell-1)}}\mathbf{X}_{ij}\right)^{-1}\left(\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^{(\ell-1)}}\mathbf{Y}_{ij}\right).$$

Step B3. Given $\widehat{\boldsymbol{\beta}}^{(\ell)}$, update $\widehat{\mathbf{K}}_i^{(\ell)}$ for each i as $\sqrt{T}$ multiplied by the $\hat{d}_i$ eigenvectors of

$$\frac{1}{N_i T}\sum_{j=1}^{N_i}\left(\mathbf{Y}_{ij}-\mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}^{(\ell)}\right)\left(\mathbf{Y}_{ij}-\mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}^{(\ell)}\right)'.$$

corresponding to the $\hat{d}_i$ largest eigenvalues in descending order.

Step B4. Repeat Step 2–3 until convergence, and obtain the resulting $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{K}}_i$ for each i. The bias-corrected estimator $\widehat{\boldsymbol{\beta}}_{bc}$ can be obtained following Theorem 3.2.

Step B5. (Optional) Apply *GCC* estimation following Section 3.3.3 with $r_0 + r_i$ replaced by $\hat{d}_i$ and obtain the global/local factors and loadings.

A bootstrap confidence interval:

The superscript $*(b)$ denotes the b-th realisation among B bootstrap repetitions. Following Feng et al. (2023), the bootstrap procedure proceeds as follows:

Step C1. Construct a random variable $\boldsymbol{\varepsilon}_t^{(b)} = \left[\varepsilon_1^{(b)}, \ldots, \varepsilon_T^{(b)}\right]'$ from a zero mean normal distribution with covariance

$$\mathrm{Cov}\left(\varepsilon_\tau^{(b)}, \varepsilon_s^{(b)}\right) = Bartlett\left(\frac{\tau - s}{l_T}\right) \text{ for } \tau, s = 1, \ldots, T$$

where l_T is a bandwidth parameter[3], and $Bartlett(x) = 1 - |x|$ if $|x| \leq 1$ or $Bartlett(x) = 0$ otherwise, is the Bartlett kernel function.

Step C2. For each i, j, and t, construct $e_{ijt}^{*(b)} = \hat{e}_{ijt}\varepsilon_t^{(b)}$ and $y_{ijt}^{*(b)} = \widehat{\mathbf{Z}}_{ijt}'\widehat{\boldsymbol{\beta}}_{bc} + \widehat{\boldsymbol{\theta}}_{ij}'\widehat{\mathbf{K}}_{it} + e_{ijt}^{*(b)}$ where $\widehat{\mathbf{Z}}_{ijt}$ is the t-th row vector of $\widehat{\mathbf{Z}}_{ij} = \mathbf{M}_{\widehat{K}_i}\mathbf{X}_{ij} - N_i^{-1}\sum_{k=1}^{N_i} a_{i,kj}\mathbf{M}_{\widehat{K}_i}\mathbf{X}_{ik}$.

Step C3. Obtain the bootstrap estimator of the slope parameters by

$$\widehat{\boldsymbol{\beta}}^{*(b)} = \left(\sum_{i=1}^{R}\sum_{j=1}^{N_i}\widehat{\mathbf{Z}}_{ij}'\boldsymbol{M}_{\widehat{K}_i}\widehat{\mathbf{Z}}_{ij}\right)^{-1}\left(\sum_{i=1}^{R}\sum_{j=1}^{N_i}\widehat{\mathbf{Z}}_{ij}'\boldsymbol{M}_{\widehat{K}_i}\mathbf{Y}_{ij}^{*(b)}\right).$$

Step C4. Repeat Step C1-C3 B times. Construct the empirical distribution function for the k-th slope coefficient

$$\widehat{\mathcal{D}}_{\beta^k}(\tau) = \frac{1}{B}\sum_{b=1}^{B}\mathbb{1}\left(\sqrt{NT}\left(\widehat{\beta}^{k,*(b)} - \widehat{\beta}^k\right) \leq \tau\right) \text{ for } k = 1, \ldots, p.$$

Then, the $1-\alpha$ CI is given by

$$\left[\widehat{\beta}^k - \frac{1}{\sqrt{NT}}\widehat{\mathcal{D}}_{\beta^k}^{-1}\left(\frac{\alpha}{2}\right), \widehat{\beta}^k - \frac{1}{\sqrt{NT}}\widehat{\mathcal{D}}_{\beta^k}^{-1}\left(1 - \frac{\alpha}{2}\right)\right] \tag{3.4.15}$$

where $\widehat{\mathcal{D}}_{\beta^k}^{-1}(\boldsymbol{\alpha}/2)$ and $\widehat{\mathcal{D}}_{\beta^k}^{-1}(1-\alpha/2)$ are the inverse functions of $\widehat{\mathcal{D}}_{\beta^k}$ evaluated at $\alpha/2$ and $1-\alpha/2$.

3.5 Monte Carlo Simulation

In this section, we conduct simulation studies to analyse the finite sample performance of the proposed estimators. The data is generated by

$$y_{ijt} = \mathbf{X}_{ijt}'\boldsymbol{\beta} + u_{ijt},\ u_{ijt} = \boldsymbol{\gamma}_{ij}'\mathbf{G}_t + \boldsymbol{\lambda}_{ij}'\mathbf{F}_{it} + e_{ijt}.$$

[3] We set $l_T = \lceil T^{1/3}\rceil$ at the order of $T^{1/3}$ following Feng et al. (2023) in our simulation and application.

The idiosyncratic errors are cross-sectionally and serially correlated as

$$\mathbf{e}_t = 0.2\mathbf{e}_{t-1} + \mathbf{w}_t^e,\ \mathbf{w}_t^e \sim i.i.d.N(\mathbf{0}, \boldsymbol{\Omega}_e)$$

where $\mathbf{e}_t = [e_{11t}, \ldots, e_{1N_1t}, \ldots, e_{R1t}, \ldots, e_{RN_Rt}]'$ and the (m,n) elements of $\boldsymbol{\Omega}_e$ are $0.2^{|m-n|}$. The factors follow $AR(1)$ processes

$$\mathbf{G}_t = 0.5\mathbf{G}_{t-1} + \mathbf{w}_t^G,\ \mathbf{w}_t^G \sim i.i.dN(\mathbf{0}, \mathbf{I}_{r_0})$$

$$\mathbf{F}_{it} = 0.5\mathbf{F}_{i,t-1} + \mathbf{w}_{i,t}^F,\ \mathbf{w}_{i,t}^F \sim i.i.dN(\mathbf{0}, \mathbf{I}_{r_i})$$

and the factor loadings are drawn from $i.i.d.$ standard normal distribution as $\boldsymbol{\gamma}_{ij} \sim i.i.d.N(\mathbf{0}, \mathbf{I}_{r_0})$ and $\boldsymbol{\lambda}_{ij} \sim i.i.d.\ N(\mathbf{0}, \mathbf{I}_{r_i})$. We consider $p = 2$ and allow the regressors to be correlated with the factor structure as

$$x_{ijt}^1 = \boldsymbol{\gamma}_{x,ij}'\mathbf{G}_t + \boldsymbol{\lambda}_{x,ij}'\mathbf{F}_{it} + \chi_{ijt}^1,\ \boldsymbol{\gamma}_{x,ij} \sim i.i.d.N(\mathbf{0}, \mathbf{I}_{r_0}),\ \boldsymbol{\lambda}_{x,ij} \sim i.i.d.N(\mathbf{0}, \mathbf{I}_{r_i})$$

and

$$x_{ijt}^2 = \left|\boldsymbol{\gamma}_{ij}'\mathbf{G}_t + \boldsymbol{\lambda}_{ij}'\mathbf{F}_{it}\right| + \chi_{ijt}^2$$

where χ_{ijt}^k's are cross-sectionally correlated such that $\boldsymbol{\chi}_t^k = \left[\chi_{11t}^k, \ldots, \chi_{1N_1t}^k, \ldots, \chi_{R1t}^k, \ldots, \chi_{RN_Rt}^k\right]'$ is generated by

$$\boldsymbol{\chi}_t^k = 0.5\boldsymbol{\iota}_N + 0.5\boldsymbol{\chi}_{t-1}^k + \boldsymbol{\epsilon}_t^k,\ \boldsymbol{\epsilon}_t^k \sim i.i.d.N(\mathbf{0}, \boldsymbol{\Omega}_\chi) \text{ for } k = 1, 2.$$

with covariance matrix $\boldsymbol{\Omega}_\chi$ whose (m,n) elements are $0.3^{|m-n|}$. We fix $\boldsymbol{\beta} = [1,1]'$, $R = 4$ and $r_0 = r_i = 2$ for all i. We set $N_1 = \cdots = N_R$ for convenience and consider the sample sizes $N_i \in \{30, 60, 90, 120, 150\}$ and $T \in \{40, 80, 120, 160, 200\}$. Each simulation experiment is repeated over 500 times.

We report the average bias for the bias-corrected estimators provided by $MIPC$, namely $\left\|\widehat{\boldsymbol{\beta}}_{bc} - \boldsymbol{\beta}\right\|$, and the size of the t-tests, namely the rejection rate of $H_0 : \hat{\beta}_{bc}^k = \beta^k$ against $H_1 : \hat{\beta}_{bc}^k \neq \beta^k$ for $k = 1, 2$ under 0.05 significance level. The precision of the estimated factors is evaluated by the trace ratio defined as

$$TR\left(\widehat{\mathbf{A}}\right) = \frac{\operatorname{tr}\left\{\mathbf{A}'\widehat{\mathbf{A}}\left(\widehat{\mathbf{A}}'\widehat{\mathbf{A}}\right)^{-1}\widehat{\mathbf{A}}'\mathbf{A}\right\}}{\operatorname{tr}\left\{\mathbf{A}'\mathbf{A}\right\}} \tag{3.5.16}$$

where $\widehat{\mathbf{A}}$ is the estimates of $\mathbf{A}$ and tr$\{.\}$ is the trace of a matrix. Apart from the global and local factors, we evaluate the precision of the estimated combined factors $\widehat{\mathbf{K}}^+ = \left[\widehat{\mathbf{G}}, \widehat{\mathbf{F}}_1, \ldots, \widehat{\mathbf{F}}_R\right]$ associated with the true factors $\mathbf{K}^+ = \left[\mathbf{K}_1^+, \ldots, \mathbf{K}_T^+\right]'$ where $\mathbf{K}_t^+$ is defined in (3.2.3). The closer the trace ratios are to one, the more precise the factors are estimated. For the estimated numbers of factors, we report the average $\hat{r}_0$ and $\hat{r}_i$ that are estimated by GCC and $\hat{r}_i = \hat{d}_i - \hat{r}_0$ respectively, where $\hat{d}_i$ is determined by our approach in Section 3.3.2 with $d_{\max} = 10$. To assess the overall selection precision, we also report the average estimated total number of factors $\hat{r}^+ = \hat{r}_0 + \sum_{l=1}^R \hat{r}_l$.

For comparison, we examine two different versions of IPC based on Bai (2009).[4] The first version is the bias-corrected IPC that is applied to the entire data. In this version, $\hat{r}^+$ is selected by BIC_3 with the maximum number of factors set at 15. We also consider an infeasible version of IPC, denoted by IPC^*, which is the same as IPC except that we enforce the true number of factors $r^+ = r_0 + \sum_{l=1}^{R} r_l$ in the estimation. We assess the bias of the slope coefficients and the size of the t-tests. Since these approaches cannot separately identify the global and local factors, the precision of the estimated combined factors $\widehat{\mathbf{K}}^+$ is evaluated.

The first panel of Table 3.1 shows the average bias of the slope coefficients. It is evident that $MIPC$ and IPC^* exhibit comparable performance, while IPC demonstrates a larger bias, particularly when N_i or T is small. IPC is only comparable with $MIPC$ and IPC^* when both N_i and T are sufficiently large. The second and third panels of Table 3.1 present the t-test sizes given by these approaches. IPC is severely oversized even when both N_i and T become large. Although IPC^* has a small bias, it demonstrates considerable size distortion. On the contrary, $MIPC$ achieves a smaller size distortion in general, as the sample size increases. Take $\widehat{\beta}^1$ as an example, when $N_i = 90$, the sizes of IPC^* are 13%, 27.8%, 20.2%, and 16.2%, and 13.8% for $T = 40, 80, 120, 160, 200$ respectively, whilst the corresponding figures for $MIPC$ are 13.2%, 9.2%, 8%, 7%, and 6.8%. This clearly shows that $MIPC$ is superior to IPC and IPC^* in the presence of multilevel factors.

We now focus on the estimation of the factor structure. In the first panel of Table 3.2, we present the trace ratios of the estimated global and local factors and their corresponding dimensions. The results reveal that the global factors are accurately estimated. The trace ratio of the global factors is not very high when $N_i = 30$ and $T = 40$. This is due to the underestimation of r_0 in small samples. As expected, the convergence of the local factors is slower than that of the global factors, but the trace ratio increases rapidly as the sample size increases. We also observe a similar pattern for the estimated dimensions of $\hat{r}_0$ and $\hat{r}_i$. This indicates that by combining $MIPC$ with GCC, we are able to successfully identify both global and local factors. In the second panel of Table 3.2, we report the estimated total number of factors obtained from the $MIPC$, and IPC with BIC_3. Our approach precisely estimates the true r^+, whereas IPC with BIC_3 fails, as discussed in Section 3.2. It tends to underestimate the true number of factors, unless both N_i and T are very large, leading to higher bias and size distortion, which we have demonstrated in Table 3.1. Finally, the third part of Table 3.2 presents the trace ratios of the estimated combined factors. It is observed that IPC performs poorly, as the dimension is always smaller. The precision of the estimated factors generated by $MIPC$ and IPC^* are comparable. This suggests that the performance of $MIPC$ and GCC in terms of estimating the combined factor space is reasonably good, and the slower convergence of the local factors does not impact the precision of the estimated combined factor space.

It is important to note that, IPC^* is implemented with the known number of factors, which is barely the case in practice. Nevertheless, in terms of inference, $MIPC$ performs better than IPC^* despite the uncertainty that arises from the estimation of the number of factors.

[4] We use the R package "phtt" for IPC estimation. See Bada & Liebl (2014) and `https://CRAN.R-project.org/package=phtt`.

Table 3.1: Simulation results for the slope coefficients ($R = 4$, $r_0 = r_i = 2$ for all i, $\boldsymbol{\beta} = [1,1]'$)

N_i	T	$MIPC$	IPC	IPC^*	$MIPC$	IPC	IPC^*	$MIPC$	IPC	IPC^*
		Bias $\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\| \times 100$			Size of $\widehat{\beta}^1$ $\times 100\%$			Size of $\widehat{\beta}^2$ $\times 100\%$		
30	40	1.796	3.777	1.878	13.2	33.8	17	14.6	21.6	23.4
30	80	1.159	1.68	1.225	8.8	26	16	9.4	16.4	11
30	120	0.968	1.066	0.989	10.4	16.8	12.4	8.4	29.6	31
30	160	0.804	0.837	0.799	9.4	14.8	12.8	11.6	22	21.8
30	200	0.804	0.837	0.799	7.4	14.8	13.4	7.4	19.4	19.2
60	40	1.17	2.613	1.298	12.2	31.6	18	12.8	20.6	11.4
60	80	0.77	0.866	0.797	7.6	9.2	7	9.2	26.8	25.4
60	120	0.635	0.653	0.653	7.8	27	26.8	8.6	17.8	18
60	160	0.556	0.563	0.563	7.4	24.6	24.6	8.8	17.8	17.8
60	200	0.484	0.456	0.455	8	20.2	20.2	7.4	11.2	11.2
90	40	0.939	2.191	1.04	13.2	33.4	13	14.4	21.8	25.8
90	80	0.64	0.697	0.652	9.2	27.6	27.8	9.8	21.4	21
90	120	0.501	0.501	0.501	8	20.2	20.2	8.8	17.6	17.6
90	160	0.435	0.379	0.379	7	16.2	16.2	8	9.6	9.6
90	200	0.389	0.329	0.328	6.8	13.8	13.8	7.4	12.4	12.4
120	40	0.777	1.98	0.887	10	36.8	10.2	12.6	22.2	27.4
120	80	0.534	0.553	0.536	10	20.6	21	10	18.6	16.4
120	120	0.434	0.394	0.395	6.6	17.4	17.6	7.4	11.2	11.2
120	160	0.396	0.354	0.354	9.4	18.2	18.2	10.2	14.6	14.6
120	200	0.33	0.257	0.257	6.4	9.8	9.8	7	7.6	7.6
150	40	0.704	1.967	0.803	13.6	37.2	24.4	13.2	38.2	24.2
150	80	0.487	0.498	0.48	9.4	21.2	21.2	9.6	16.2	14.8
150	120	0.384	0.336	0.338	9.4	16.6	16.8	9.8	10	10
150	160	0.327	0.26	0.26	6.4	10.2	10.2	8.4	6.6	6.6
150	200	0.314	0.236	0.236	8.6	10.6	10.6	8	10.6	10.6

r_0 is the true number of global factors. r_i is the true number of local factors in block i, $d_i = r_0 + r_i$, and $r^+ = r_0 + \sum_{l=1}^{R} r_l$ is the total number of factors in the model where R is the number of blocks. $MIPC$ is the result based on the numbers of factors estimated by our approach. IPC is Bai (2009)'s approach applied to the whole data with $\hat{r}^+$ selected by BIC_3 and IPC^* is IPC with the true number of factors r^+. The results of bias and size are obtained from the bias-corrected estimators for these three approaches. We set $N_1 = \cdots = N_R$ as the number of individuals in each block.

Table 3.2: Simulation results for the factors ($R = 4$, $r_0 = r_i = 2$ for all i, $\boldsymbol{\beta} = [1, 1]'$)

N_i	T	$\hat{r}_0$	$\hat{r}_i$	$TR(\widehat{\mathbf{G}})$	$TR(\widehat{\mathbf{F}}_i)$	$\hat{r}^+$		$TR(\widehat{\mathbf{K}}^+)$		
			$MIPC$			$MIPC$	IPC	$MIPC$	IPC	IPC^*
30	40	1.886(0|10.8)	2.071(10.6|2.2)	0.95	0.888	10.17(10|2.2)	3.018(0|99.8)	0.888	0.461	0.98
30	80	1.992(0|0.8)	2.006(0.8|0.2)	0.989	0.931	10.014(0.8|0.2)	7.998(0|80.2)	0.979	0.86	0.978
30	120	1.996(0|0.4)	2.003(0.4|0.2)	0.99	0.944	10.006(0.4|0.2)	9.642(0|28.2)	0.977	0.956	0.977
30	160	1.998(0|0.2)	2.001(0.2|0.2)	0.991	0.95	10(0.2|0.2)	9.886(0|10.4)	0.975	0.97	0.977
30	200	2(0|0)	2(0|0)	0.992	0.955	10(0|0)	9.948(0|5)	0.977	0.973	0.977
60	40	1.998(0|0.2)	2.002(0.2|0)	0.995	0.909	10.004(0.2|0)	3.142(0|100)	0.991	0.482	0.99
60	80	2(0|0)	2(0|0)	0.996	0.946	10(0|0)	9.45(0|42.8)	0.99	0.965	0.989
60	120	2(0|0)	2(0|0)	0.996	0.96	10(0|0)	9.992(0|0.8)	0.989	0.989	0.989
60	160	2(0|0)	2(0|0)	0.996	0.966	10(0|0)	10(0|0)	0.989	0.989	0.989
60	200	2(0|0)	2(0|0)	0.996	0.97	10(0|0)	10(0|0)	0.989	0.989	0.989
90	40	2(0|0)	2(0|0.2)	0.997	0.912	9.998(0|0.2)	3.112(0|100)	0.994	0.481	0.993
90	80	2(0|0)	2(0|0)	0.998	0.951	10(0|0)	9.696(0|25.6)	0.993	0.981	0.993
90	120	2(0|0)	2(0|0)	0.998	0.964	10(0|0)	10(0|0)	0.993	0.993	0.993
90	160	2(0|0)	2(0|0)	0.998	0.97	10(0|0)	10(0|0)	0.993	0.993	0.993
90	200	2(0|0)	2(0|0)	0.998	0.975	10(0|0)	10(0|0)	0.993	0.993	0.993
120	40	2(0|0)	2(0|0)	0.998	0.914	10(0|0)	2.862(0|100)	0.996	0.447	0.995
120	80	2(0|0)	2(0|0)	0.998	0.951	10(0|0)	9.774(0|19.6)	0.995	0.986	0.995
120	120	2(0|0)	2(0|0)	0.998	0.966	10(0|0)	9.998(0|0.2)	0.995	0.995	0.995
120	160	2(0|0)	2(0|0)	0.998	0.973	10(0|0)	10(0|0)	0.995	0.995	0.995
120	200	2(0|0)	2(0|0)	0.998	0.977	10(0|0)	10(0|0)	0.995	0.995	0.995
150	40	1.998(0|0.2)	2.002(0.2|0)	0.998	0.914	10.004(0.2|0)	2.696(0|100)	0.997	0.43	0.996
150	80	2(0|0)	2(0|0)	0.999	0.953	10(0|0)	9.816(0|16.8)	0.996	0.989	0.996
150	120	2(0|0)	2(0|0)	0.999	0.967	10(0|0)	10(0|0)	0.996	0.996	0.996
150	160	2(0|0)	2(0|0)	0.999	0.973	10(0|0)	10(0|0)	0.996	0.996	0.996
150	200	2(0|0)	2(0|0)	0.999	0.978	10(0|0)	10(0|0)	0.996	0.996	0.996

The average $\hat{r}_0$, $\hat{r}_i$, and $\hat{r}^+$ over 1000 replications are reported with the figures inside the parenthesis, $(O|U)$, indicating the percentage of overestimation and underestimation. r_0 is the true number of global factors. r_i is the true number of local factors in block i, $d_i = r_0 + r_i$, and $r^+ = r_0 + \sum_{l=1}^{R} r_l$ is the total number of factors in the model. $\hat{r}_0$ is estimated by GCC and $\hat{r}_i = \hat{d}_i - \hat{r}_0$ where $\hat{d}_i$ is determined by our approach. For $MIPC$, $\hat{r}^+ = \hat{r}_0 + \sum_{l=1}^{R} \hat{r}_l$. IPC is Bai (2009)'s approach applied to the whole data with $\hat{r}^+$ selected by BIC_3 and IPC^* is IPC with true number of factors $r^+ = \sum_{l=0}^{R} r_l$. We set $r_1 = \cdots = r_R$ where R is the number of blocks. We set $N_1 = \cdots = N_R$ as the number of individuals in each block. $TR(.)$ is the trace ratio defined in (3.5.16). The closer the trace ratio is to one, the more precise the estimated factors are. $\widehat{\mathbf{K}}^+$ is the estimated version of the combined factors $\mathbf{K}^+ = \left[\mathbf{K}_1^+, \ldots, \mathbf{K}_T^+\right]'$ where $\mathbf{K}_t^+$ is defined in (3.2.3).

3.6 Empirical Application

The literature has long debated the connection between energy use and economic expansion. For example, urbanisation and industrialisation are two economic processes that heavily rely on energy use. As a result, energy economists argue that classic growth theories neglect energy as an important economic driver and that energy should be incorporated into the production function. Meanwhile, the externality of energy use has grown to be a significant global concern. How much limiting energy use will have on economic growth may be of importance to policymakers. Therefore, assessing their relationship has significant policy-making ramifications.

Previous studies have employed techniques including cointegration analysis and the Granger Causality test to both time series and panel data. See Ozturk (2010) for a comprehensive review. Not until recently, the cross-section correlation between the error terms has drawn the attention of the researchers. In the panel ARDL context, a CD (cross-section dependence) test is often rejected, suggesting that the inference may be invalid for the panel data models if it is not appropriately dealt with (e.g. Damette & Seghir (2013) and Jalil (2014)). Numerous studies find that global and local business synchronisations have led to such cross-country correlation. Rodríguez-Caballero (2022) proposes a panel regression model with a multilevel factor structure to address this concern. Following his empirical specification, we estimate the following model:

$$\Delta \ln GDP_{ijt} = \beta^1 \Delta \ln EC_{ijt} + \beta^2 \Delta \ln K_{ijt} + \boldsymbol{\gamma}_{ij}'\mathbf{G}_t + \boldsymbol{\lambda}_{ij}'\mathbf{F}_{it} + e_{ijt} \tag{3.6.17}$$

where GDP is the per capita GDP (current US\$), K is the capital formation (current US\$), and EC is the energy consumption (kWh per capita). All the variables are log-differenced to achieve stationarity. Since the fixed effects and time effects are absorbed in the multilevel factors, practitioners do not have to choose which one should be used. The indices $i = 1, 2, 3, 4$ represents the four regions, namely Asia and Pacific, Europe, Africa, and The Americas, j represents the countries in the corresponding regions, and t is the time index covering the years from 1972 to 2014. We end up with a balanced panel data where $R = 4$, $\{N_i\}_{i=1}^R = \{21, 19, 18, 22\}$, $N = 80$, and $T = 43$.[5] As opposed to Rodríguez-Caballero (2022), we use a differenced version of the regression to circumvent the potential non-stationarity in the variables. However, the interpretation of the slope coefficients is the same in that β^1 and β^2 are the partial elasticity of the production on energy consumption and capital. For comparison, we also report the estimation results from a fixed effect model, a twoway fixed effect model, Bai's IPC estimation, and our (bias-corrected) $MIPC$ estimation.

The results of the estimation are presented in Table 3.3, where the first panel displays the estimated slope coefficients, and the second panel shows the number of factors. Initially, we performed a regression (3.6.17) using IPC with the number of factors determined by BIC3, as displayed in the third column of Table 3.3. However, no factor was detected so IPC reduces to OLS. Therefore, we proceeded to estimate a fixed effect model and a two-way fixed effect model, as shown in the first two columns. The inclusion of fixed effects resulted in a reduction of the partial elasticity of energy consumption from 0.209 to 0.121, and a slight decrease in the partial elasticity of capital. Furthermore, the coefficients became even smaller after adding time effects compared to the fixed effect model. We then moved on to MIPC estimation. No global factor is detected. Meanwhile, $\{\hat{r}_i\}_{i=1}^R = \{1, 1, 4, 2\}$ regional factors are selected for Asia Pacific, Europe, Africa, and The Americas respectively, as shown in Column four. The coefficient for energy consumption is 0.172, which is notably higher than that reported by the additive effect models, while the coefficient for capital formation is 0.392. To provide a comparison, column five displays the IPC estimates using the same total number of factors $\hat{r}^+ = 8$, and the slope coefficients are similar to those generated by MIPC.

[5]Our data sources from the World Bank: `https://databank.worldbank.org/source/world-development-indicators#`. We keep all the countries with less than 10 missing observations from 1971 to 2014. The missing values are imputed by the EM algorithm applied to each variable in the same region. We use the R package "mvdalab" to implement the EM algorithm.

Based on the above findings, it can be concluded that the energy effect is underestimated by FE models compared to the IFE models.

Table 3.3: Empirical results for the energy consumption and economic growth regression

	FE	Twoway FE	IPC	MIPC	IPC
	Slope coefficients				
$d\ln EC$	0.121	0.094	0.209	0.172	0.151
	(0.023)	(0.022)	(0.034)	(0.033)	(0.021)
$d\ln K$	0.451	0.399	0.474	0.392	0.370
	(0.008)	(0.008)	(0.019)	(0.012)	(0.012)
	Number of factors				
Global				0	
Asia Pacific				1	
Europe				1	
Africa				4	
The Americas				2	
Total			0	8	8
Method			BIC_3	$MIPC\&GCC$	fixed

The table shows the estimation results from the regression in (3.6.17). Each entry of the first panel shows the estimated coefficient with standard error in the parenthesis. The second panel shows the number of factors in each model and the corresponding selection criteria. For IPC, the total number of factors is either estimated by BIC_3 or fixed. For $MIPC$, $\hat{d}_i$ is estimated by our approach with $d_{\max} = 5$ and $\hat{r}_0$ is estimated by GCC. The number of local factors $\hat{r}_i$ is obtained by $\hat{r}_i = \hat{d}_i - \hat{r}_0$.

3.7 Conclusion

In this article, we consider a multi-dimensional panel data model with multilevel factors in the context of large numbers of cross-sections and time observations. The multilevel factor components capture the interactive fixed effects induced by the global factors that are pervasive across all individuals as well as the interactive fixed effects due to the local factors that are block-specific.

We develop a multilevel iterative principal component ($MIPC$) method for estimation by iteratively updating between the slope coefficients and factors, given one another. Under a finite number of blocks, our approach is able to produce consistent estimates of the slope coefficients, factors, and loadings. To determine the numbers of factors, we propose a model selection criteria based on the eigenvalue ratios. In addition, given consistent factor estimates from each block, we apply the generalised canonical correlation estimation developed by Lin & Shin (2022) to the to separately identify the global and local factors. In the asymptotic analysis, we show the consistency of our estimates and establish the asymptotic normality of the bias-corrected estimator for the slope coefficients. A dependent wild bootstrap is employed for inference which accounts for the cross-sectional and serial correlation in the idiosyncratic errors. The Monte Carlo simulation demonstrates good finite sample performance of $MIPC$ compared to Bai (2009)'s IPC in the presence of multilevel factor structure. We apply our model to study the energy consumption and economic growth nexus using a cross-country panel data categorised by regions.

There are several possible extensions. First, it would be interesting to incorporate spatial effects in the model. Such a model features both strong local dependence caused by local factors and weak local dependence evolving from spatial diffusion. Second, it is possible to extend the least squares estimation advanced by Moon & Weidner (2017) to our multilevel factor case which enables low-rank regressors such as time-invariant regressors and observed common factors.regressors such as time-invariant regressors and observed common factors.

Conclusions

This thesis aims to make contributions to modelling a high-dimensional panel data with a multilevel factor structure. In Chapter 1, we have developed two new selection criteria, namely the canonical correlation difference (CCD) and modified canonical correlation (MCC) to consitently estimate the number of global factors. In Chapter 2, we have advanced the generalized canonical correlation (GCC) estimation and established inferential theory for the multilevel factor model. Chapter 3 considers a panel regression model with an unobserved multilevel factor structure and proposes the bias corrected multilevel iterative principal component ($MIPC$) estimator and inferential theory. Monte Carlo experiments have demonstrated the good finite performance of the proposed methods. We have further showcased the utility of our approaches by applying them to several macroeconomics and finance applications. Thus, this thesis offers a comprehensive framework for theoretical and empirical analysis of the high-dimensional panel data with a multilevel factor structure.

In particular, we aim to apply/modify our approaches to address several important empirical research topics. One area of interest is asset pricing, where the ever-expanding "factor zoo" has sparked concerns about the presence of spurious pricing factors (Feng et al. (2020)). To tackle this issue, we can develop a formal test based on the GCC to examine whether a set of new factors spans the same space as the existing observable/unobserved factors. Additionally, another interesting application is to investigate the commonality among different groups of observable factors (Harvey et al. (2016)).

Furthermore, there are several prospect for methodological extensions. Firstly, the GCC approach can be readily applied to more complicated factor structures, such as block structures with multiple layers of classification. Secondly, one could consider non-stationary multilevel factors, although this would bring forth new challenges in terms of identification and inference. Lastly, it would be intriguing to combine spatial effects with multilevel factors, allowing for a joint consideration of both strong and weak local dependence.

Appendix A

Appendix to Chapter 1

A.1 Lemmas and Proofs

Lemma A.1.1. *Let* $\widehat{\mathbf{K}}_i = (M_i T)^{-1}\mathbf{Y}_i\mathbf{Y}_i'\widehat{\mathbf{K}}_i$. *Under Assumption A–D, as* $M_i, T \to \infty$, *we have:*

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

where $\mathbf{H}_i$ *is the* $r_{\max} \times (r_0 + r_i)$ *rotation matrix,* $\delta_{M_iT} = \min\left\{\sqrt{M_i}, \sqrt{T}\right\}$ *and* M_i *is the number of individuals in block* i.

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

Q.E.D

For any two blocks m and h, we apply the PC estimator to (1.3.5), and obtain consistent estimators of $\mathbf{K}_m = [\mathbf{G}, \mathbf{F}_m]$ and $\mathbf{K}_h = [\mathbf{G}, \mathbf{F}_h]$, denoted $\widehat{\mathbf{K}}_m$ and $\widehat{\mathbf{K}}_h$. Let $\ell_{mh,r}$ be the r-th largest squared canonical correlation between $\widehat{\mathbf{K}}_m$ and $\widehat{\mathbf{K}}_h$, which is given by the rth largest characteristic root of

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

where $\widehat{\mathbf{S}}_{ab}$ $(a, b = m, h)$ denotes the sample variance/covariance matrices for $\widehat{\mathbf{K}}_m$ and $\widehat{\mathbf{K}}_h$. Since $(1/\sqrt{T})\widehat{\mathbf{K}}_m$ is the eigenvector matrix corresponding to the $r_{\max}$ largest eigenvalues of $\mathbf{Y}_m\mathbf{Y}_m'$, we have:

$$\frac{1}{M_mT}\mathbf{Y}_m\mathbf{Y}_m'\frac{1}{\sqrt{T}}\widehat{\mathbf{K}}_m = \frac{1}{\sqrt{T}}\widehat{\mathbf{K}}_m\mathbf{V}_m$$

where $\mathbf{V}_m$ is an $r_{\max} \times r_{\max}$ diagonal matrix consisting of the $r_{\max}$ largest eigenvalues of $\mathbf{Y}_m\mathbf{Y}_m'$ in descending order divided by M_mT. This implies that $\widehat{\mathbf{K}}_m\mathbf{V}_m = \widetilde{\mathbf{K}}_m$. Similarly, we obtain $\widehat{\mathbf{K}}_h\mathbf{V}_h = \widetilde{\mathbf{K}}_h$ for block h. Since $r_{\max} < \min\{M_m, T\}$ $(r_{\max} < \min\{M_h, T\})$, the diagonal elements of $\mathbf{V}_m$ $(\mathbf{V}_h)$ are non-zero. This implies that $\mathbf{V}_m$ $(\mathbf{V}_h)$ is of full rank, though some diagonal elements may be very small. The canonical correlations between $\widehat{\mathbf{K}}_m$ and $\widehat{\mathbf{K}}_h$ are equal to those between $\widetilde{\mathbf{K}}_m$ and $\widetilde{\mathbf{K}}_h$, because the canonical correlations between two sets of variables are invariant to full rank transformations, see Theorem 12.2.2 in Anderson (2003). Therefore, we will study the limiting behaviour of the canonical correlations between $\widetilde{\mathbf{K}}_m$ and $\widetilde{\mathbf{K}}_h$ instead of those between $\widehat{\mathbf{K}}_m$ and $\widehat{\mathbf{K}}_h$. This enables us to employ Lemma A.1.1 subsequently.

Proof of Lemma 1.1. For any two blocks m and h, the population covariance between $\mathbf{K}_{mt}$ and $\mathbf{K}_{ht}$ can be expressed as

$$Var\begin{pmatrix}\mathbf{K}_{mt}\\ \mathbf{K}_{ht}\end{pmatrix} = \begin{bmatrix}\mathbf{\Sigma}_{mm} & \mathbf{\Sigma}_{mh}\\ \mathbf{\Sigma}_{hm} & \mathbf{\Sigma}_{hh}\end{bmatrix} = \begin{bmatrix}\mathbf{\Sigma}_G & \mathbf{0} & \mathbf{\Sigma}_G & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma}_{F_m} & \mathbf{0} & \mathbf{0}\\ \mathbf{\Sigma}_G & \mathbf{0} & \mathbf{\Sigma}_G & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_{F_h}\end{bmatrix} \tag{A.1.1}$$

where $\mathbf{\Sigma}_G$, $\mathbf{\Sigma}_{F_m}$ and $\mathbf{\Sigma}_{F_h}$ are defined in Assumption C. Without loss of generality, we assume $r_m \le r_h$. Using (A.1.1), we can rewrite the characteristic equation,

$$\left(\mathbf{\Sigma}_{mh}\mathbf{\Sigma}_{hh}^{-1}\mathbf{\Sigma}_{hm} - \rho\mathbf{\Sigma}_{mm}\right)\mathbf{v} = 0 \tag{A.1.2}$$

as

$$\begin{bmatrix}\mathbf{\Sigma}_G - \rho\mathbf{\Sigma}_G & \mathbf{0}\\ \mathbf{0} & -\rho\mathbf{\Sigma}_{F_m}\end{bmatrix}\mathbf{v} = \mathbf{0},$$

where $\rho_{mh,r}$ is the r-th largest squared canonical correlation between $\mathbf{K}_m$ and $\mathbf{K}_h$. It is clear that $\rho_{mh,1} = \cdots = \rho_{mh,r_0} = 1$ are the characteristic roots with multiplicity r_0, while $\rho_{mh,r_0+1} = \cdots = \rho_{mh,r_m} = 0$ are the characteristic roots with multiplicity, r_m. Since this holds for all m and h, we simply let $\rho_r = \rho_{mh,r}$. The characteristic vector corresponding to the rth eigenvalue is $\mathbf{v}_r = [0, \ldots, 0, 1, 0, \ldots, 0]$, which is the unit vector with the rth element being 1 and 0 otherwise.

Since $r_{\max} \ge r_0 + r_i$ for all i by construction, $\mathbf{H}_m$ and $\mathbf{H}_h$ are not of full column rank. This renders the variance-covariance matrices for the rotated factors $\mathbf{H}_m'\mathbf{K}_{mt}$ and $\mathbf{H}_h'\mathbf{K}_{ht}$, becoming singular as follows:

$$\text{Var}\left(\begin{bmatrix}\mathbf{H}_m'\mathbf{K}_{mt}\\ \mathbf{H}_h'\mathbf{K}_{ht}\end{bmatrix}\right) = \begin{bmatrix}\mathbf{H}_m'\mathbf{\Sigma}_{mm}\mathbf{H}_m & \mathbf{H}_m'\mathbf{\Sigma}_{mh}\mathbf{H}_h\\ \mathbf{H}_h'\mathbf{\Sigma}_{hm}\mathbf{H}_m & \mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h\end{bmatrix} \tag{A.1.3}$$

where both $\mathbf{H}_m'\mathbf{\Sigma}_{mm}\mathbf{H}_m$ and $\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h$ are the singular matrices. Consider the characteristic equation between the rotated factors as

$$\left[\mathbf{H}_m'\mathbf{\Sigma}_{mh}\mathbf{H}_h\left(\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h\right)^{-}\mathbf{H}_h'\mathbf{\Sigma}_{hm}\mathbf{H}_m - \rho\mathbf{H}_m'\mathbf{\Sigma}_{mm}\mathbf{H}_m\right]\mathbf{u} = \mathbf{0} \tag{A.1.4}$$

where $(\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h)^{-}$ is the Moore-Penrose inverse of $\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h$. Using the property of Moore-

Penrose inverse, we have:[1]

$$\mathbf{H}_h \left(\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h\right)^{-} \mathbf{H}_h' = \mathbf{\Sigma}_{hh}^{-1}$$

which holds if $\mathbf{H}_h$ has full row rank. Then, (A.1.4) becomes:

$$\mathbf{H}_m' \left(\mathbf{\Sigma}_{mh}\mathbf{\Sigma}_{hh}^{-1}\mathbf{\Sigma}_{hm} - \rho\mathbf{\Sigma}_{mm}\right)\mathbf{H}_m\mathbf{u} = \mathbf{0}. \tag{A.1.5}$$

Using (A.1.1), we rewrite (A.1.5) as

$$\mathbf{H}_m' \begin{bmatrix} \mathbf{\Sigma}_G - \rho\mathbf{\Sigma}_G & \mathbf{0} \\ \mathbf{0} & -\rho\mathbf{\Sigma}_{F_m} \end{bmatrix} \mathbf{H}_m\mathbf{u} = \mathbf{0}$$

which shows that both (A.1.2) and (A.1.4) will produce the same non-zero eigenvalues.

We now consider the following spectral decompositions:

$$\mathbf{H}_m'\mathbf{\Sigma}_{mm}\mathbf{H}_m = \mathbf{P}\mathbf{\Delta}_m\mathbf{P}' \text{ and } \mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h = \mathbf{Q}\mathbf{\Delta}_h\mathbf{Q}'$$

where $\mathbf{\Delta}_m(\mathbf{\Delta}_h)$ is a diagonal matrix of eignevalues of $\mathbf{H}_m'\mathbf{\Sigma}_{mm}\mathbf{H}_m(\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h)$, $\mathbf{P}(\mathbf{Q})$ is an orthogonal matrix whose columns are standardized eigenvectors associated with the diagonal entries of $\mathbf{\Delta}_m(\mathbf{\Delta}_h)$. As the rank of $\mathbf{H}_m'\mathbf{\Sigma}_{mm}\mathbf{H}_m(\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h)$ is $r_0+r_m \leq r_{\max}(r_0+r_h \leq r_{\max})$ asymptotically, we rewrite the above equation as

$$\begin{aligned} \mathbf{H}_m'\mathbf{\Sigma}_{mm}\mathbf{H}_m &= \begin{bmatrix}\mathbf{P}_1 & \mathbf{P}_2\end{bmatrix} \begin{bmatrix}\mathbf{\Delta}_1^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} \begin{bmatrix}\mathbf{P}_1 & \mathbf{P}_2\end{bmatrix}' \\ \mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h &= \begin{bmatrix}\mathbf{Q}_1 & \mathbf{Q}_2\end{bmatrix} \begin{bmatrix}\mathbf{\Delta}_2^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} \begin{bmatrix}\mathbf{Q}_1 & \mathbf{Q}_2\end{bmatrix}' \end{aligned} \tag{A.1.6}$$

where $\mathbf{P}_1$ and $\mathbf{P}_2$ are $r_{\max}\times(r_0+r_m)$ and $r_{\max}\times[r_{\max}-(r_0+r_m)]$ orthogonal matrices, and similarly for $\mathbf{Q}_1$ and $\mathbf{Q}_2$. Now, consider the $(r_0+r_m)\times(r_0+r_h)$ matrix, $\mathbf{\Delta}_1^{-1}\mathbf{P}_1'\left(\mathbf{H}_m'\mathbf{\Sigma}_{mh}\mathbf{H}_h\right)\mathbf{Q}_1\mathbf{\Delta}_2^{-1}$, whose singular value decomposition is given by (see Rao (1981))

$$\mathbf{\Delta}_1^{-1}\mathbf{P}_1'\left(\mathbf{H}_m'\mathbf{\Sigma}_{mh}\mathbf{H}_h\right)\mathbf{Q}_1\mathbf{\Delta}_2^{-1} = \mathbf{W}\begin{bmatrix}\mathbf{R}^{1/2} & \mathbf{0}\end{bmatrix}\mathbf{D}' \tag{A.1.7}$$

where $\mathbf{W}$ is an $(r_0+r_m)\times(r_0+r_m)$ orthonormal matrix, $\mathbf{D}$ an $(r_0+r_h)\times(r_0+r_h)$ orthonormal matrix and $\mathbf{R}$ the $(r_0+r_m)\times(r_0+r_m)$ diagonal matrix given by $\mathbf{R} = \text{diag}(\rho_1, ..., \rho_{r_0}, \rho_{r_0+1}, ..., \rho_{r_0+r_m}) = \text{diag}(1, ..., 1, 0, ..., 0)$.[2]

[1] We use two properties of the Moore-Penrose inverse. (1) Let $A \in R^{m\times n}$ and $B \in R^{n\times p}$. If A has full column rank and B has full row rank, then $(AB)^- = B^-A^-$. (2) If A has full column rank, then $A^-A = I$. If A has full row rank, then $AA^- = I$.

[2] Notice that $\mathbf{R}$ contains the same non-zero roots as in (A.1.4), see Rao (1981).

Define the full rank matrices,

$$\mathbf{A} = \left[\mathbf{P}_1\mathbf{\Delta}_1^{-1}\mathbf{W}, \mathbf{P}_2\right] \text{ and } \mathbf{B} = \left[\mathbf{Q}_1\mathbf{\Delta}_2^{-1}\mathbf{D}, \mathbf{Q}_2\right] \tag{A.1.8}$$

Combining (A.1.6), (A.1.7) and (A.1.8), it is straightforward to show that

$$\text{Var}\left(\begin{bmatrix}\mathbf{A}'\mathbf{H}_m'\mathbf{K}_{mt} \\ \mathbf{B}'\mathbf{H}_h'\mathbf{K}_{ht}\end{bmatrix}\right) = \begin{bmatrix}\mathbf{I}_{r_0+r_m} & \mathbf{0} & \mathbf{R}^{1/2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{R}^{1/2} & \mathbf{0} & \mathbf{I}_{r_0+r_m} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{r_h-r_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix} \tag{A.1.9}$$

From (A.1.9), we obtain the characteristic equation between $\mathbf{A}'\mathbf{H}_m'\mathbf{K}_{mt}$ and $\mathbf{B}'\mathbf{H}_h'\mathbf{K}_{ht}$ by

$$\left[\mathbf{A}'\mathbf{H}_m'\mathbf{\Sigma}_{mh}\mathbf{H}_h\mathbf{B}\left(\mathbf{B}'\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h\mathbf{B}\right)^{-}\mathbf{B}'\mathbf{H}_h'\mathbf{\Sigma}_{hm}\mathbf{H}_m\mathbf{A} - \rho\mathbf{A}'\mathbf{H}_m'\mathbf{\Sigma}_{mm}\mathbf{H}_m\mathbf{A}\right]\mathbf{u} = \mathbf{0}$$

which can be simplified as

$$\left(\begin{bmatrix}\mathbf{R}^{1/2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{I}_{r_0+r_m} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{r_h-r_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{R}^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} - \rho\begin{bmatrix}\mathbf{I}_{r_0+r_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\right)\mathbf{u} = \mathbf{0}$$

Hence,

$$\left(\begin{bmatrix}\mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} - \rho\begin{bmatrix}\mathbf{I}_{r_0+r_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\right)\mathbf{u} = \mathbf{0} \tag{A.1.10}$$

Obviously, (A.1.10) has the same characteristic roots from (A.1.4) and the same non-zero characteristic roots from (A.1.2), consequently.

Now, we consider the sample covariance matrix for $\widetilde{\mathbf{K}}_m$ and $\widetilde{\mathbf{K}}_h$ given by

$$Var\left(\begin{array}{c}\widetilde{\mathbf{K}}_m \\ \widetilde{\mathbf{K}}_h\end{array}\right) = \frac{1}{T}\begin{bmatrix}\widetilde{\mathbf{K}}_m'\widetilde{\mathbf{K}}_m & \widetilde{\mathbf{K}}_m'\widetilde{\mathbf{K}}_h \\ \widetilde{\mathbf{K}}_h'\widetilde{\mathbf{K}}_m & \widetilde{\mathbf{K}}_h'\widetilde{\mathbf{K}}_h\end{bmatrix} = \begin{bmatrix}\widetilde{\mathbf{S}}_{mm} & \widetilde{\mathbf{S}}_{mh} \\ \widetilde{\mathbf{S}}_{hm} & \widetilde{\mathbf{S}}_{hh}\end{bmatrix}$$

Consider the full rank transformation $\widetilde{\mathbf{K}}_m\mathbf{A}$ and $\widetilde{\mathbf{K}}_h\mathbf{B}$, where $\mathbf{A}$ and $\mathbf{B}$ are defined in (A.1.8). The canonical correlations between them are equivalent to those between $\widetilde{\mathbf{K}}_m$ and $\widetilde{\mathbf{K}}_h$. By Lemma A.1.1, we obtain: $\mathbf{A}'\widetilde{\mathbf{S}}_{mm}\mathbf{A} \xrightarrow{p} \mathbf{A}'\mathbf{H}_m'\mathbf{\Sigma}_{mm}\mathbf{H}_m\mathbf{A}$, $\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B} \xrightarrow{p} \mathbf{B}'\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h\mathbf{B}$ and $\mathbf{A}'\widetilde{\mathbf{S}}_{mh}\mathbf{B} \xrightarrow{p} \mathbf{A}'\mathbf{H}_m'\mathbf{\Sigma}_{mh}\mathbf{H}_h\mathbf{B}$. Let $\underline{M} = \min\{M_m, M_h\}$ and $\delta_{\underline{M}T} = \min\{\sqrt{\underline{M}}, \sqrt{T}\}$. Applying (A.1.9) and Lemma A.1.1, we can rewrite these transformed variance/covariance matrices as

$$\mathbf{A}'\widetilde{\mathbf{S}}_{mm}\mathbf{A} = \begin{bmatrix}\mathbf{I}_{r_0+r_m} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right)\end{bmatrix}$$

$$\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B} = \begin{bmatrix} \mathbf{I}_{r_0+r_m} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & \mathbf{I}_{r_h-r_m} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \end{bmatrix}$$

and

$$\mathbf{A}'\widetilde{\mathbf{S}}_{mh}\mathbf{B} = \begin{bmatrix} \mathbf{R}^{1/2} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \end{bmatrix}$$

Notice that the Moore-Penrose inverse of the lower $[r_{\max} - (r_0 + r_m)] \times [r_{\max} - (r_0 + r_m)]$ block of $\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B}$ does not converge to $\begin{bmatrix} \mathbf{I}_{r_h-r_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, because

$$\operatorname{rank}\left(\begin{bmatrix} \mathbf{I}_{r_h-r_m} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \end{bmatrix}\right) \neq \operatorname{rank}\left(\begin{bmatrix} \mathbf{I}_{r_h-r_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\right)$$

Similarly, $\left(\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B}\right)^{-}$ does not converge to $\begin{bmatrix} \mathbf{I}_{r_0+r_m} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{r_h-r_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$. See Theorem 1 in Karabiyik et al. (2017). But, the Moore-Penrose inverse follows the Banachiewicz-Schur form.[3] Thus,

$$\begin{bmatrix} \mathbf{I}_{r_h-r_m} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \end{bmatrix}^{-} = \begin{bmatrix} \mathbf{I}_{r_h-r_m} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & -O_p(1) \\ -O_p(1) & O_p\left(\delta_{\underline{M}T}^{2}\right) \end{bmatrix} = O_p\left(\delta_{\underline{M}T}^{2}\right) \quad \text{(A.1.11)}$$

Also, $\left(\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B}\right)^{-}$ follows the Banachiewicz-Schur form, from which we obtain:

$$\left(\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B}\right)^{-} = \begin{bmatrix} \mathbf{I}_{r_0+r_m} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & -O_p(1) \\ -O_p(1) & O_p\left(\delta_{\underline{M}T}^{2}\right) \end{bmatrix}. \quad \text{(A.1.12)}$$

Using the above results, we obtain:

$$\mathbf{A}'\widetilde{\mathbf{S}}_{mh}\mathbf{B}\left(\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B}\right)^{-}\mathbf{B}'\widetilde{\mathbf{S}}_{hm}\mathbf{A} =$$

$$\begin{bmatrix} \mathbf{R}^{1/2} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \end{bmatrix} \begin{bmatrix} \mathbf{I}_{r_0+r_m} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & -O_p(1) \\ -O_p(1) & O_p\left(\delta_{\underline{M}T}^{2}\right) \end{bmatrix} \begin{bmatrix} \mathbf{R}^{1/2} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{R} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \end{bmatrix}$$

[3] Let $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$. Under some conditions, the MP inverse of M is given as $M^{-} = \begin{bmatrix} A^{-} + A^{-}CS^{-}BA^{-} & -A^{-}CS^{-} \\ -S^{-}BA^{-} & S^{-} \end{bmatrix}$, where $S = D - BA^{-}C$. We check that the required conditions hold in our case. See Tian & Takane (2009) and Castro-González et al. (2015)

Therefore, the characteristic equation between $\widetilde{\mathbf{K}}_m\mathbf{A}$ and $\widetilde{\mathbf{K}}_h\mathbf{B}$,

$$\left[\mathbf{A}'\widetilde{\mathbf{S}}_{mh}\mathbf{B}\left(\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B}\right)^{-}\mathbf{B}'\widetilde{\mathbf{S}}_{hm}\mathbf{A}-\ell\mathbf{A}'\widetilde{\mathbf{S}}_{mm}\mathbf{A}\right]\boldsymbol{\xi}=\mathbf{0}$$

can be rewritten as

$$\left(\begin{bmatrix}\mathbf{R} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}-\ell\begin{bmatrix}\mathbf{I}_{r_0+r_m} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}+O_p\left(\delta_{\underline{M}T}^{-2}\right)\right)\boldsymbol{\xi}=\mathbf{0}$$

which is analogous to (A.1.10) with a small perturbation term.

Finally, by the continuity of the characteristic roots, we have $\ell_{mh,r}\xrightarrow{p}1$ for $r=1,\ldots,r_0$ and $\ell_{mh,r}\xrightarrow{p}0$ for $r=r_0+1,\ldots,r_{\max}$ as $T,M_m,M_h\to\infty$.

Q.E.D

Proof of Lemma 1.2 Using Lemma 1.1, it is straightforward to show that $\xi(r)\xrightarrow{p}1$ for $0\leq r\leq r_0$ and $\xi(r)\xrightarrow{p}0$ otherwise.

Q.E.D

Proof of Lemma 1.3. By applying Lemma 1 to the definition of $CCD(r)$, it is straightforward to derive the main results in Lemmas 1.3.

Q.E.D

Proof of Theorem 1.1. We need to show that

$$\Pr(CCD(r)<CCD(r_0))\longrightarrow 1 \text{ as } M_1,\ldots,M_R,T\longrightarrow\infty$$

for $r\neq r_0$ and $r\leq r_{\max}$. By Lemma 1.3, it is easily seen that for $r_0<r\leq r_{\max}$ we have:

$$CCD(r)-CCD(r_0)\xrightarrow{p}-1<0$$

while for $0\leq r<r_0$:

$$CCD(r)-CCD(r_0)\xrightarrow{p}-1<0.$$

Next, consider the case with $r_0=0$. Then, for $r_0<r\leq r_{\max}$, it is straightforward to show that

$$CCD(r)-CCD(r_0)\xrightarrow{p}-1<0.$$

Q.E.D

A.2 The Performance of Existing 2D Model Selection Criteria

We investigate the finite sample performance of the existing 2D model selection criteria when we apply them directly to the whole data matrix, $\mathbf{YY}'$ by ignoring the multilevel factors structure.

Let r_0 (r_i) be the number of global (local) factors and R be the number of groups. In the case where the number of blocks, R is finite, it is well-established that the existing information criteria mainly developed for the single level panel data, fail to consistently estimate the number of global factors r_0 because the weak cross-section correlation condition is violated in the presence of the local factors. But, as $R \to \infty$, the impacts of the local factors would be asymptotically negligible. In this case, we may rewrite the multilevel factor model as the single level counterpart:

$$y_{ijt} = \boldsymbol{\gamma}_{ij}' \mathbf{G}_t + u_{ijt}$$

where

$$u_{ijt} = \boldsymbol{\lambda}_{ij}' \mathbf{F}_{it} + e_{ijt}$$

First, we assume that the local factors are orthogonal to each other. Then, the errors u_{ijt} would satisfy the weak cross-sectional correlation condition in the sense of Assumption A3. In this case, Han (2021) conjectures that the number of global factors can be consistently estimated asymptotically by the existing selection criteria (see Remark 4). Next, in the general case where the local factors are mutually correlated, Assumption 1.A.3 may be violated because u_{ijt}'s share stronger cross-section correlations. In this situation we show via simulations that the existing criteria will be inconsistent.

To conduct additional Monte Carlo simulations we consider the same benchmark design as used in Section 1.5.[4] We fix $r_0 = 2$ and $r_i = 2$ for all $i = 1, ..., R$ and set $T \in \{50, 100\}$. We set $r_{\max} = 10$ for the four criteria, IC_{p2}, BIC_3 by Bai & Ng (2002), ER by Ahn & Horenstein (2013) and ED by Onatski (2010) whilst we apply our practical guide in selecting $r^*_{\max}$ for CCD (see Section II).

Table A.1 shows the simulation results for $T = 100$. Panel A of Table A.1 reports the results for the case with $(\beta, \phi_e, \kappa) = (0, 0, 1)$ and $\omega_F = 0$ in which u_{ijt}'s are *iid* and the local factors are mutually uncorrelated. For $R = 2$, all of the existing criteria fail to distinguish between the global and local factors, and tend to select the total number of factors, $r_0 + \sum_{i=1}^{R} r_i = 6$, not r_0. For sufficiently large R, all of the four existing criteria tend to select the number of global factors as M and T increase. If the value of R is moderate, e.g. $R = 5$ or 10, then they tend to select the intermediate value between r_0 and $r_0 + \sum_{i=1}^{R} r_i$, suggesting that these approaches are rather unreliable in practice.[5] Notice, however, that CCD selects the true number of global factors in all cases, regardless of whether R is small or large.

[4] We have also conducted more simulations by using the different numbers for r_0 and r_i together with the different sample sizes. These simulation results, available upon request, are qualitatively similar.

[5] Their performance is also very sensitive to the values of $r_{\max}$. For instance, for $R = 5$, the average estimates of r_0 by ER and ED are 1.96 and 1.99 if $r_{\max} = 10$ while they become 8.96 and 12.0 if $r_{\max} = 15$. In principle, it is unclear how to set the value of $r_{\max}$ sufficiently large for large R.

Panel B presents the results for the case with $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ and $\omega_F = 0$ in which we allow u_{ijt}'s to be cross-sectionally and serially correlated. The performances of the four existing criteria are all adversely affected in small samples. Except for IC_{p2}, they can select r_0 only if $R > 30$. Still for the moderate value of R, they tend to select the intermediate value between r_0 and $r_0 + \sum_{i=1}^{R} r_i$. CCD remains insensitive to the presence of the cross-sectionally and serially correlated errors, and selects the number of global factors correctly in all cases.

Panels C and D display the results for the case with $(\beta, \phi_e, \kappa) = (0, 0, 1)$ and $\omega_F = \{0.2, 0.4\}$ in which we allow the local factors to be mutually correlated. The performances of the four existing criteria are all unreliable when R is relatively small. Even for large R, say $R = 50$, they tend to overestimate the number of global factors significantly. By contrast, CCD still selects r_0 precisely as M rises.

To further examine the impact of the small T, we repeat the above experiments with $T = 50$, and report the results in Table A.2. Though the performance of all selection criteria are slightly adversely affected by the small T, overall patterns are qualitatively similar to those reported in Table A.1. Still remarkably, only CCD selects the number of global factors correctly in almost all cases.

In sum we demonstrate that the existing selection criteria will produce unreliable inference in finite samples by ignoring the multilevel structure. This result also implies that the sequential procedure proposed by Wang (2008) and applied by Hallin & Liška (2011) will be quite unreliable. In particular, for large R, it will severely overestimate by selecting $r_0 + \sum_{i=1}^{R} r_i/R$ instead of r_0. Furthermore, Han (2021) provides the simulation evidence that this approach can lead to even negative estimates of both the number of global factors and the number of local factors in small samples, for $R = 3$.

Table A.1: The performance of 2D criteria applied to 3D data ($r_0 = 2, r_i = 2$, $r_{\max} = 10$, $T = 100$)

R	M	T	CCD	IC_{p2}	BIC_3	ER	ED
			Panel A: $(\beta, \phi_e, \kappa) = (0, 0, 1)$, $\omega_F = 0$				
2	20	100	2(0|0.1)	5.81(100|0)	5.02(100|0)	5.45(91.2|4.8)	5.99(99.9|0)
2	100	100	2(0|0)	6(100|0)	5.97(100|0)	6(100|0)	6(100|0)
5	20	100	2(0|0)	6.35(99.8|0)	4.15(99.5|0)	1.89(2.9|16.9)	1.92(8|13.4)
5	100	100	2(0|0)	10(100|0)	5.1(100|0)	1.96(0.6|4.6)	1.99(1.2|1.6)
10	20	100	2(0|0)	4.07(92.3|0)	2.11(11.1|0)	1.99(0|1.4)	2.08(6.4|0)
10	100	100	2(0|0)	8.73(100|0)	2.02(2|0)	2(0|0.3)	2.06(4.4|0)
30	20	100	2(0|0)	2.01(1.4|0)	2(0|0)	2(0|0)	2.06(3.8|0)
30	100	100	2(0|0)	2.01(1|0)	2(0|0)	2(0|0)	2.04(2.7|0)
50	20	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.03(2|0)
50	100	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.02(2|0)
			Panel B: $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$, $\omega_F = 0$				
2	20	100	2(0|0.4)	8.03(100|0)	6.27(100|0)	4.21(66.8|17.5)	2.37(32.5|58.7)
2	100	100	2(0|0)	6.54(100|0)	6(100|0)	5.97(99.4|0.2)	6.01(100|0)
5	20	100	2(0|0)	9.02(100|0)	5.34(100|0)	1.88(4|17.9)	1.93(9.6|13.9)
5	100	100	2(0|0)	10(100|0)	6.32(100|0)	1.94(0.5|6.3)	2.01(3.9|2.9)
10	20	100	2(0|0)	7.98(100|0)	2.77(66.4|0)	1.97(0|2.9)	2.1(6.9|0)
10	100	100	2(0|0)	10(100|0)	2.43(40.4|0)	2(0|0.2)	2.06(4.7|0)
30	20	100	2(0|0)	2.83(58.1|0)	2(0|0)	2(0|0.1)	2.04(2.6|0)
30	100	100	2(0|0)	3.55(82|0)	2(0|0)	2(0|0)	2.02(1.7|0)
50	20	100	2(0|0)	2.06(5.4|0)	2(0|0)	2(0|0)	2.02(1.6|0)
50	100	100	2(0|0)	2.05(4.8|0)	2(0|0)	2(0|0)	2.02(1.5|0)
			Panel C: $(\beta, \phi_e, \kappa) = (0, 0, 1)$, $\omega_F = 0.2$				
2	20	100	2(0.2|0.4)	5.69(100|0)	4.76(100|0)	5.13(88.9|6.1)	5.96(99.9|0)
2	100	100	2(0|0)	6(100|0)	5.87(100|0)	6(100|0)	6(100|0)
5	20	100	2(0|0)	5.45(100|0)	3.8(100|0)	2.44(53.4|13.5)	2.53(68.4|14.2)
5	100	100	2(0|0)	9.89(100|0)	4.4(100|0)	2.61(66|5.1)	2.77(84.5|4.7)
10	20	100	2(0|0)	3.9(100|0)	2.99(99.2|0)	2.64(68|4.3)	3.02(95.7|0.3)
10	100	100	2(0|0)	6.65(100|0)	3(99.9|0)	2.78(78.4|0.8)	3.03(99.6|0)
30	20	100	2(0|0)	3(100|0)	2.95(94.6|0)	2.9(89.6|0)	3.04(100|0)
30	100	100	2(0|0)	3(100|0)	2.92(91.8|0)	2.93(93.3|0)	3.02(100|0)
50	20	100	2(0|0)	3(100|0)	2.92(91.5|0)	2.96(95.9|0)	3.02(100|0)
50	100	100	2(0|0)	3(100|0)	2.95(94.6|0)	2.9(89.6|0)	3.04(100|0)
			Panel D: $(\beta, \phi_e, \kappa) = (0, 0, 1)$, $\omega_F = 0.4$				
2	20	100	2.18(19.3|0.8)	5.25(100|0)	4.19(100|0)	4.03(83.7|8.7)	5.87(99.5|0.3)
2	100	100	2.09(8.6|0)	6(100|0)	5.25(100|0)	5.97(100|0)	6.01(100|0)
5	20	100	2.1(9.6|0)	3.92(100|0)	3.19(100|0)	2.96(96.4|0.9)	3.13(98.9|0.3)
5	100	100	2.02(1.7|0)	8.29(100|0)	3.28(100|0)	3(99.6|0.1)	3.02(100|0)
10	20	100	2.06(6.4|0)	3.2(100|0)	3(100|0)	3(100|0)	3.07(100|0)
10	100	100	2.01(0.5|0)	4.12(100|0)	3(100|0)	3(100|0)	3.03(100|0)
30	20	100	2.06(5.8|0.1)	3(100|0)	3(100|0)	3(100|0)	3.04(100|0)
30	100	100	2.01(0.6|0)	3(100|0)	3(100|0)	3(100|0)	3.02(100|0)
50	20	100	2.05(5.4|0)	3(100|0)	3(100|0)	3(100|0)	3.03(100|0)
50	100	100	2(0|0)	3(100|0)	3(100|0)	3(100|0)	3.02(100|0)

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

Table A.2: The performance of 2D criteria applied to 3D data ($r_0 = 2, r_i = 2$, $r_{\max} = 10$, $T = 50$)

R	M	T	CCD	IC_{p2}	BIC_3	ER	ED
			Panel A: $(\beta, \phi_e, \kappa) = (0, 0, 1)$, $\omega_F = 0$				
2	20	50	1.98(0.3|2.6)	5.06(99.5|0)	4.96(100|0)	3.96(66.7|19.8)	5.35(91.5|6.5)
2	100	50	2(0|0)	5.99(100|0)	5.1(100|0)	5.96(99.5|0.4)	6(100|0)
5	20	50	2(0|0)	4.6(96.7|0)	3.63(96.8|0)	1.77(6.7|30.7)	1.9(14.4|27.9)
5	100	50	2(0|0)	7.58(100|0)	3.25(90.1|0)	1.78(2.9|25.5)	1.87(9.5|19.3)
10	20	50	2(0|0)	2.93(66.5|0)	2.09(9.2|0.2)	1.84(0.8|15.8)	2.1(10.9|4)
10	100	50	2(0|0)	3.85(91|0)	2(0.1|0.1)	1.9(0.2|9.8)	2.13(8.3|1)
30	20	50	2(0|0)	2.01(0.6|0)	1.98(0|1.9)	1.97(0|3.1)	2.11(6.9|0)
30	100	50	2(0|0)	2(0.4|0)	1.97(0|2.9)	1.98(0|1.8)	2.07(5|0)
50	20	50	2(0|0.1)	2(0|0)	1.98(0|2.1)	1.98(0|2.4)	2.05(3.9|0)
50	100	50	2(0|0)	2(0|0)	1.92(0|7.8)	1.99(0|1)	2.04(3.1|0)
			Panel B: $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$, $\omega_F = 0$				
2	20	50	2.17(14.7|1.6)	7.31(100|0)	6.73(100|0)	3.36(54.3|26.5)	2.11(30.2|57.7)
2	100	50	2.02(1.5|0)	7.49(100|0)	5.9(100|0)	4.9(81.9|9.8)	4.98(82.9|14.8)
5	20	50	2(0.2|0.1)	8.69(100|0)	5.49(100|0)	1.79(8.3|35.3)	1.66(10.7|37)
5	100	50	2(0|0)	9.97(100|0)	4.85(100|0)	1.78(3.5|27.3)	1.75(6.2|24.5)
10	20	50	2(0|0)	7.65(100|0)	3.33(93|0)	1.8(1.1|20.3)	2.02(8.6|8)
10	100	50	2(0|0)	9.64(100|0)	2.48(45|0)	1.85(0.3|14.8)	2.06(6.2|3.1)
30	20	50	2(0|0)	3.95(80.2|0)	2(0|0)	1.92(0|8.3)	2.04(2.3|0.2)
30	100	50	2(0|0)	4.77(89.1|0)	2(0|0.1)	1.95(0|4.6)	2.02(1.9|0)
50	20	50	2(0|0)	2.75(46.4|0)	2(0|0)	1.95(0|5.4)	2.02(1.2|0)
50	100	50	2(0|0)	2.77(45.4|0)	2(0|0.4)	1.96(0|4.3)	2.01(0.9|0)
			Panel C: $(\beta, \phi_e, \kappa) = (0, 0, 1)$, $\omega_F = 0.2$				
2	20	50	2.01(3.6|2.7)	4.82(99.9|0)	4.77(100|0)	3.65(65|18.1)	5.01(87.3|9.1)
2	100	50	2.01(0.8|0)	5.97(100|0)	4.78(100|0)	5.9(98.9|1)	6(100|0)
5	20	50	2(0.1|0.6)	4.31(99.7|0)	3.54(99.4|0)	2.16(38.3|27.6)	2.27(47.8|28.5)
5	100	50	2(0|0)	6.48(100|0)	3.22(98.4|0)	2.23(41.6|20.6)	2.29(53.3|21)
10	20	50	2(0|0.4)	3.37(99.2|0)	2.85(84.7|0)	2.26(43.4|18)	2.78(74.8|5.3)
10	100	50	2(0|0)	3.77(99.9|0)	2.72(72|0)	2.4(52.1|11.7)	2.87(82.9|2.1)
30	20	50	2(0|0.1)	2.99(98.8|0)	2.37(37.1|0.2)	2.57(61.3|4.8)	3.01(98|0)
30	100	50	2(0|0)	2.99(99.2|0)	2.14(14.6|0.9)	2.63(67|4.2)	3.04(98.8|0)
50	20	50	2(0|0)	2.99(98.6|0)	2.19(20.1|1.2)	2.65(68.1|3.6)	3.03(99.9|0)
50	100	50	2(0|0)	2.99(99.1|0)	2.06(7.9|1.6)	2.69(72.2|3.2)	3.02(99.7|0)
			Panel D: $(\beta, \phi_e, \kappa) = (0, 0, 1)$, $\omega_F = 0.4$				
2	20	50	2.27(30.8|3.9)	4.26(99.5|0.1)	4.34(100|0)	2.94(59.7|20.5)	4.52(85.5|9.7)
2	100	50	2.21(21.4|0)	5.76(100|0)	4.14(100|0)	5.2(94.8|2.4)	6(100|0)
5	20	50	2.23(23.1|0.6)	3.55(100|0)	3.14(99.9|0)	2.68(77.2|10.5)	3.12(90.5|5)
5	100	50	2.09(8.9|0)	4.74(100|0)	3.01(99.9|0)	2.9(91.9|2.4)	3.09(97.8|1.3)
10	20	50	2.17(16.8|0.1)	3.09(99.9|0)	2.99(99.3|0)	2.9(92.7|2.4)	3.1(99.4|0.2)
10	100	50	2.05(4.6|0)	3.14(100|0)	2.99(98.7|0)	2.96(96.3|0.8)	3.12(99.9|0)
30	20	50	2.15(14.8|0)	3(100|0)	2.96(96.1|0)	2.99(99.2|0)	3.08(100|0)
30	100	50	2.04(4.1|0)	3(100|0)	2.9(89.9|0)	2.99(99.4|0.1)	3.05(100|0)
50	20	50	2.12(12.4|0)	3(100|0)	2.93(93.3|0)	2.99(99.3|0)	3.03(100|0)
50	100	50	2.05(5|0)	3(100|0)	2.82(81.7|0.1)	2.99(99.6|0.2)	3.03(100|0)

See footnotes to Table A.1.

A.3 Practical Guidelines for Selecting the Maximum Number of Factors

It is well-established in the literature that if the maximum number of factors is set too high, the redundant factors are likely to be selected, see Ahn & Horenstein (2013). Hence, we propose a practical selection guideline. We first apply BIC_3 to the data $\mathbf{Y}_i$ in each block with a sufficiently large $r_{\max}$, and obtain the consistent estimate of $r_0 + r_i$, denoted $\widehat{r_0 + r_i}$ for $i = 1, ..., R$. Then, we select the common maximum number of factors by $r^*_{\max} = \max\left\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_R}\right\}$. This procedure selects $r^*_{\max} \leq r_{\max}$, while ensuring that $\Pr(r^*_{\max} \geq r_0 + r_i) \xrightarrow{p} 1$ for all $i = 1, ..., R$.

We generate the data following the similar simulation design used in Section 1.5, but fix $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (2, 2)$ and $\kappa = 1$. We consider the four cases with $(\beta, \phi_e) \in \{(0, 0), (0.1, 0), (0, 0.5), (0.1, 0.5)\}$.

In Tables A.3 and A.4 we report the simulation results for CCD and MCC using $r^*_{\max}$ together with the fixed $r_{\max} = 10$.[6] In particular, if idiosyncratic errors are serially correlated, then CCD with the large $r_{\max}$ tends to overestimate r_0 for small T whereas its performance seems to be intact in the presence of the cross-sectional correlations. The performance of MCC is adversely affected by the presence of both cross-section and serial correlation in the errors, especially if T is small. On the other hand, we find that both CCD and MCC with $r^*_{\max}$, select the number of global factors correctly even in small samples. These results confirm our claim that the practical guide outlined in Section 1.5 improves the selection precision of CCD and MCC in finite samples.

[6]Here, we run BIC_3 R times using $r_{\max} = 10$ for all $i = 1, ..., R$, and then obtain the common maximum number of factors by $r^*_{\max} = \max\left\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_R}\right\}$.

Table A.3: Average *CCD* estimates of the number of global factors for experiments with $r_{\max} = 10$ and $r^*_{\max} = \max\left\{\widehat{r_0+r_1}, \ldots, \widehat{r_0+r_R}\right\}$

			$(\beta, \phi_e, \kappa) = (0, 0, 1)$		$(\beta, \phi_e, \kappa) = (0.1, 0, 1)$		$(\beta, \phi_e, \kappa) = (0, 0.5, 1)$		$(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$	
R	M	T	$r_{\max}$	$r^*_{\max}$	$r_{\max}$	$r^*_{\max}$	$r_{\max}$	$r^*_{\max}$	$r_{\max}$	$r^*_{\max}$
2	20	50	2.07(9.9|5.4)	1.98(0.6|2.4)	2.05(6.7|2)	1.99(0.7|1.5)	3(44.5|2.1)	2.18(16.2|2.2)	2.71(35.3|0.8)	2.01(5|4.3)
2	50	50	2.02(2|0)	2(0.2|0)	2.02(1.5|0)	2(0|0)	4.51(70.5|0)	2(0.2|0)	2.98(36.6|0)	2(0.5|0.4)
2	100	50	2.01(1.3|0)	2(0|0)	2(0.3|0)	2(0.1|0)	6.36(97.4|0)	2(0.1|0)	3.8(54.4|0)	2(0|0)
2	200	50	2.01(1|0)	2(0|0)	2.01(0.8|0)	2(0|0)	7.01(100|0)	2(0|0)	5.94(92|0)	2(0|0.2)
2	20	100	2(0.1|0.4)	2(0|0.5)	2(0|0)	1.98(0|1.6)	2(1.5|1.4)	2(0|0.3)	2(0.4|0.3)	2(0|0.4)
2	50	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.01(1.1|0)	2(0|0)	2(0.1|0)	2(0|0)
2	100	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.03(1.9|0)	2(0|0)	2.04(2.5|0)	2(0|0)
2	200	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.3(12.8|0)	2(0|0)	2(0.3|0)	2(0|0)
2	20	200	2(0|0)	1.99(0|0.5)	2(0|0)	2(0|0.5)	2(0|0)	2(0|0.3)	2(0|0.1)	2(0|0.9)
2	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)
2	100	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	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)
5	20	50	2(0|0.4)	2(0|0.1)	2(0|0)	2(0|0.2)	2.15(8.6|0.2)	2.01(0.9|0.1)	2.14(8.3|0.1)	2(0|0.4)
5	50	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.87(23.1|0)	2(0|0)	2.88(23|0)	2(0|0)
5	100	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	6.49(97.8|0)	2(0|0)	2.09(2.2|0)	2(0|0)
5	200	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	7.13(100|0)	2(0|0)	5.61(80.4|0)	2(0|0)
5	20	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.98(0|2.3)
5	50	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
5	100	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
5	200	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
5	20	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)
5	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)
5	100	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)
5	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)
10	20	50	2(0|0)	2(0|0)	2(0|0.1)	2(0|0)	2(0.5|0.2)	2(0|0.1)	2(0.1|0.1)	2(0|0)
10	50	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2.12(3|0)	2(0|0)	2.13(3.2|0)	2(0|0)
10	100	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	6.67(99.8|0)	2(0|0)	6.62(99.6|0)	2(0|0)
10	200	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	7.14(100|0)	2(0|0)	5.38(74|0)	2(0|0)
10	20	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0.5)
10	50	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
10	100	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
10	200	100	2(0|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)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	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)
10	100	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)
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)

The average of $\hat{r}_0$ over 1,000 replications are reported together with the figures inside the parenthesis, $(O|U)$, indicating the percentage of overestimation and underestimation, respectively. For example, (0|0) implies that r_0 is perfectly correctly estimated. β, ϕ_e and κ control the cross-section correlation, serial correlation and noise to signal ratio, respectively.

Table A.4: Average MCC estimates of the number of global factors for experiments with $r_{\max} = 10$ and $r^*_{\max} = \max\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_R}\}$

			$(\beta, \phi_e, \kappa) = (0, 0, 1)$		$(\beta, \phi_e, \kappa) = (0.1, 0, 1)$		$(\beta, \phi_e, \kappa) = (0, 0.5, 1)$		$(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$	
R	M	T	$r_{\max}$	$r^*_{\max}$	$r_{\max}$	$r^*_{\max}$	$r_{\max}$	$r^*_{\max}$	$r_{\max}$	$r^*_{\max}$
2	20	50	2.18(18.3|0.2)	1.99(0|1.3)	2.15(15.1|0)	1.99(0.1|0.7)	2.16(16.1|0)	1.99(0.1|0.8)	3.23(91.7|0)	2.26(25.1|0)
2	50	50	2.04(3.8|0)	2(0|0.1)	2.02(2.3|0)	2(0|0.2)	2.04(3.5|0)	2(0|0.1)	3.13(88.9|0)	2.01(1|0)
2	100	50	2.01(1.1|0)	2(0|0)	2(0.3|0)	2(0|0)	2.01(0.9|0)	2(0|0)	3.14(89.1|0)	2(0.2|0)
2	200	50	2(0|0)	2(0|0)	2(0.1|0)	2(0|0)	2(0|0)	2(0|0)	3.43(96.8|0)	2(0.1|0)
2	20	100	1.98(0|1.7)	1.98(0|2.4)	1.99(0|0.6)	1.96(0|3.7)	2(0|0)	1.97(0|2.8)	2(0|0.4)	1.98(0.1|2.2)
2	50	100	2(0|0)	2(0|0)	2(0|0)	2(0|0.1)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
2	100	100	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	200	100	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	20	200	1.92(0|8.4)	1.94(0|5.7)	1.97(0|2.9)	1.86(0|13.7)	1.97(0|3.3)	1.9(0|10.5)	1.97(0|2.8)	1.88(0|12.2)
2	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)
2	100	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	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)
5	20	50	2.01(0.7|0)	2(0|0.2)	2(0.1|0)	2(0|0)	2(0.1|0)	2(0|0.2)	3.07(100|0)	2.2(20.4|0)
5	50	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	3.03(99.9|0)	2(0|0)
5	100	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	3.03(99.8|0)	2(0|0)
5	200	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	3.36(100|0)	2(0|0)
5	20	100	2(0|0.1)	2(0|0.2)	2(0|0)	2(0|0.3)	2(0|0)	2(0|0.3)	2(0|0)	2(0|0)
5	50	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
5	100	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
5	200	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
5	20	200	1.97(0|2.9)	1.99(0|0.8)	2(0|0.4)	1.93(0|7.2)	2(0|0.3)	1.94(0|6.5)	2(0|0.3)	1.96(0|4.3)
5	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)
5	100	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)
5	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)
10	20	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	3(100|0)	2.19(18.5|0)
10	50	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	3(100|0)	2(0|0)
10	100	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	3(100|0)	2(0|0)
10	200	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	3.29(100|0)	2(0|0)
10	20	100	2(0|0)	2(0|0)	1.98(0|0)	1.98(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
10	50	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
10	100	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
10	200	100	2(0|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	1.99(0|0.8)	2(0|0.1)	2(0|0)	1.97(0|3)	2(0|0.1)	1.97(0|2.8)	2(0|0)	1.99(0|1.1)
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)
10	100	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)
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)

See footnotes to Table A.3.

A.4 The Performance of MCC with BIC_3

In the main text, we propose using the following penalty function in MCC:

$$P(\underline{M}, T) = \frac{\ln \underline{M} + \ln T}{\sqrt{\underline{M}T}} \ln \ln(\underline{M}T)$$

Since the different penalty functions may lead to the different performance (e.g. Bai & Ng (2002)), we also consider the popular penalty function in BIC_3:

$$BIC_3 = \frac{\underline{M} + T}{\underline{M}T} \ln(\underline{M}T)$$

We find that $P(\underline{M}, T)$ and BIC_3 produce values close to each other under the normal panel sizes. For example, for $\bar{M} = 50$ and $T = 100$, $BIC_3 = 0.23$ and $P(\underline{M}, T) = 0.21$. Notice, however, that BIC_3 does not always guarantee consistency. For example, if $\underline{M} = \exp(T)$, then $BIC_3 \to 1$. On the other hand, $P(\underline{M}, T)$ is not subject to this issue.

Table A.5 reports the relative performance of MCC with BIC_3 and $P(\underline{M}, T)$. When cross-sectional and serial correlation are absent in the error terms, both approaches are able to produce precise estimates. Under the presence of cross-sectional and serial correlation, MCC with BIC_3 tends to severely overestimate r_0 in small samples. For the noisier data with $\kappa = 3$, we have the mixed results. Overall, we may conclude that MCC using $P(\underline{M}, T)$ outperforms.

Table A.5: Average estimates of the number of global factors by BIC_3 and $P(\underline{M}, T)$ for experiments with $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (2, 2)$ and $r^*_{\max} = \max\left\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_R}\right\}$

			$(\beta, \phi_e, \kappa) = (0, 0, 1)$		$(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$		$(\beta, \phi_e, \kappa) = (0.1, 0.5, 3)$	
R	M	T	BIC_3	$P(\underline{M}, T)$	BIC_3	$P(\underline{M}, T)$	BIC_3	$P(\underline{M}, T)$
2	20	50	2(0.1|0.3)	1.98(0|1.8)	2.53(46.8|2.8)	2.24(22.9|0)	2.57(52.7|1.6)	2.18(23.3|5.8)
2	50	50	2(0|0)	2(0|0)	2.01(0.8|0)	2.01(0.9|0)	1.88(1.9|13.8)	1.89(2.3|13.3)
2	100	50	2(0|0)	2(0|0)	2(0|0)	2(0.3|0)	1.91(0.5|9.7)	1.92(0.5|8.4)
2	200	50	2(0|0.1)	2(0|0)	2(0|0)	2(0|0)	1.96(0|4.3)	1.95(0|5.4)
2	20	100	2(0|0.3)	1.97(0|2.7)	2(0|0.1)	1.99(0|1.5)	1.71(0|28.9)	1.22(0|71.9)
2	50	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.57(0|42.3)	1.6(0|39)
2	100	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.88(0|12.5)	1.93(0|6.6)
2	200	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0.1)	2(0|0.1)
2	20	200	2(0|0.4)	1.95(0|5.5)	1.99(0|0.7)	1.86(0|13.5)	1.47(0|50.3)	0.63(0|97.8)
2	50	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.36(0|61.6)	1.21(0|70.9)
2	100	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.92(0|8.1)	1.95(0|5)
2	200	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
5	20	50	2(0|0)	2(0|0.1)	2.69(68.2|0)	2.2(19.6|0)	2.7(69.6|0)	2.22(22.4|0)
5	50	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.95(0|5.1)	1.97(0|2.9)
5	100	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.95(0|5.2)	1.93(0|6.7)
5	200	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.99(0|1.5)	1.98(0|2.2)
5	20	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.85(0|15.1)	1.26(0|26.2)
5	50	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.66(0|34.3)	1.67(0|33.1)
5	100	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.93(0|7.4)	1.97(3.4|0)
5	200	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0.1)
5	20	200	2(0|0)	1.99(0|1.3)	2(0|0)	1.97(0|3)	1.56(0|43.7)	0.7(0|99.7)
5	50	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.43(0|56.7)	1.25(0|74.6)
5	100	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.97(0|2.9)	1.99(0|1.4)
5	200	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
10	20	50	2(0|0)	2(0|0)	2.88(87.4|0)	2.21(21|0)	2.86(84.9|0)	2.28(28.4|0)
10	50	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.98(0|1.9)	1.98(0.1|1.7)
10	100	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.97(0|3.2)	1.98(0|2.3)
10	200	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.99(0|1.4)	1.98(0|1.8)
10	20	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.95(0|4.9)	1.28(0|72)
10	50	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.71(0|29.5)	1.73(0|27.1)
10	100	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.95(0|4.8)	1.97(0|3.2)
10	200	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
10	20	200	2(0|0)	2(0|0.4)	2(0|0)	2(0|0.3)	1.61(0|38.6)	0.78(0|99.9)
10	50	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.44(0|56.4)	1.21(0|78.7)
10	100	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	1.98(0|2.1)	1.99(0|0.8)
10	200	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)

See footnotes to Table A.3.

A.5 Uneven Block Sizes and Heterogeneous Correlations among Local Factors

We investigate the performance of CCD and MCC with uneven block sizes and correlated local factors. We use the similar DGP in Section 1.5, but generate the local factors by

$$\mathbf{F}_{it} = \mathbf{\Phi}_F \mathbf{F}_{i,t-1} + \mathbf{w}_t,\ \mathbf{w}_t \sim iidN(0, \mathbf{\Omega}_F)$$

where $\mathbf{F}_t = [\mathbf{F}'_{1t}, \ldots, \mathbf{F}'_{Rt}]'$ and $\mathbf{\Phi}_F$ is a diagonal matrix with the common elements, 0.5. We fix $R = 4$ and $(r_0, r_i) = (2, 1)$ for all $i = 1, ..., 4$. We set $(M_1, M_2, M_3, M_4) \in \{(20, 50, 70, 100), (50, 70, 100, 150), (100, 150, 200, 250)\}$. We allow the correlations among local factors to be heterogeneous. We consider the three different covariance matrices of $\mathbf{\Omega}_F$. The first covariance matrix is simply set at $\mathbf{\Omega}_F = \mathbf{I}_4$ such that the local factors are uncorrelated. We also consider the following covariance matrices:

$$\mathbf{\Omega}_F = \mathbf{\Omega}_F^{(1)} = \begin{bmatrix} 1 & 0.8 & 0.4 & 0.2 \\ 0.8 & 1 & 0.6 & 0.3 \\ 0.4 & 0.6 & 1 & 0.5 \\ 0.2 & 0.3 & 0.5 & 1 \end{bmatrix} \text{ and } \mathbf{\Omega}_F = \mathbf{\Omega}_F^{(2)} = \begin{bmatrix} 1 & 0.9 & 0.8 & 0.8 \\ 0.9 & 1 & 0.7 & 0.8 \\ 0.8 & 0.7 & 1 & 0.4 \\ 0.8 & 0.8 & 0.4 & 1 \end{bmatrix}$$

The average off-diagonal elements of $\mathbf{\Omega}_F^{(1)}$ and $\mathbf{\Omega}_F^{(2)}$ are 0.47 and 0.73 respectively, which represent a moderate and high level of correlation among the local factors.

The results for uncorrelated local factors reported in Panel A of Table A.6 show that CCD performs well whilst MCC is less satisfactory when the block sizes are small and $T = 50$ but its performance improves sharply as block sizes increase. Panel B shows the results for $\mathbf{\Omega}_F = \mathbf{\Omega}_F^{(1)}$. Now that the upper bound condition for CCD is not violated, we find that the performance of both CCD and MCC are satisfactory and qualitatively similar to that in Panel A. In Panel C, we present the results for $\mathbf{\Omega}_F = \mathbf{\Omega}_F^{(2)}$, where the local factors are strongly correlated. CCD now overestimates r_0 significantly mainly due to the violation of the upper bound condition. The performance of MCC remains satisfactory unless T is too small. When the data is noisier, we find that its performance improves sharply with the large block sizes.

Table A.6: Average estimates of the number of global factors with uneven block sizes and heterogeneous local correlations, $R = 4$, $(\phi_G, \phi_F) = (0.5, 0.5)$ and $r^*_{\max} = \max\left\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_R}\right\}$

				$M_i = (20, 50, 70, 100)$		$M_i = (50, 70, 100, 150)$		$M_i = (100, 150, 200, 250)$	
Panel A: $\Omega_F = I_4$									
β	ϕ_e	κ	T	CCD	MCC	CCD	MCC	CCD	MCC
0	0	1	50	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
0	0	1	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
0	0	1	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
0.1	0.5	1	50	2.01(3|0)	2.45(44.2|0)	2(0|0)	2(0.1|0)	2(0|0)	2(0|0)
0.1	0.5	1	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
0.1	0.5	1	200	2(0|0)	2(0|0.2)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
0.1	0.5	3	50	2.03(7.2|5)	2.98(88.1|0)	1.96(0.9|4.5)	2.05(5.4|0.2)	2(0|0)	2(0|0)
0.1	0.5	3	100	1.95(0|5.2)	1.99(0|0.6)	1.99(1|0)	1.99(0|0.8)	2(0|0)	2(0|0)
0.1	0.5	3	200	1.92(0|6.7)	1.9(0|9.6)	1.99(0|0.8)	1.96(0|3.9)	2(0|0)	2(0|0)
Panel B: $\Omega_F = \Omega_F^{(1)}$									
0	0	1	50	2.02(2.1|0)	2.01(1|0)	2.01(0.7|0)	2(0|0)	2.01(0.6|0)	2(0|0)
0	0	1	100	2(0.1|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
0	0	1	200	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
0.1	0.5	1	50	2.05(4.8|0)	2.78(75.9|0)	2.03(3.3|0)	2.03(2.6|0)	2.02(2|0)	2(0|0)
0.1	0.5	1	100	2(0|0)	2(0|0)	2(0.1|0)	2(0|0)	2(0|0)	2(0|0)
0.1	0.5	1	200	2(0|0)	2(0|0.1)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
0.1	0.5	3	50	2.27(23.2|3)	2.97(10|0)	2.12(13.9|2)	2.1(10.4|0.3)	2.05(5.8|0.5)	2(0.2|0.5)
0.1	0.5	3	100	1.98(2.5|3.5)	1.99(0|0.8)	2(0.5|0.3)	1.97(0|2.9)	2(0.1|0)	2(0|0)
0.1	0.5	3	200	1.96(0.1|4.1)	1.85(14.7|0)	2(0.1|0.3)	1.93(0|7)	2(0|0)	2(0|0)
Panel C: $\Omega_F = \Omega_F^{(2)}$									
0	0	1	50	2.85(84.9|0)	2.85(84.9|0)	2.83(82.5|0)	2.42(41.9|0)	2.79(78.5|0)	2.11(11.2|0)
0	0	1	100	2.91(91.2|0)	2.44(43.7|0)	2.9(89.5|0)	2.03(2.5|0)	2.9(89.5|0)	2(0.1|0)
0	0	1	200	2.98(97.5|0)	2(0.3|0)	2.97(96.7|0)	2(0|0)	2.96(95.8|0)	2(0|0)
0.1	0.5	1	50	2.87(81.2|0)	3.08(99.5|0)	2.85(84.6|0)	2.81(80.5|0)	2.86(85.9|0)	2.23(22.6|0)
0.1	0.5	1	100	2.92(92|0)	2.52(52.1|0)	2.93(92.7|0)	2.05(4.8|0)	2.89(88.5|0)	2(0|0)
0.1	0.5	1	200	2.98(97.8|0)	2.01(0.5|0)	2.97(96.9|0)	2(0|0)	2.95(94.8|0)	2(0|0)
0.1	0.5	3	50	2.94(79.5|0.3)	3.19(99.3|0)	2.88(85.2|0.3)	2.77(76.6|0)	2.87(86.4|0)	2.42(41.7|0)
0.1	0.5	3	100	2.8(80.5|0.2)	2.46(45.9|0)	2.91(90.7|0)	2.09(8.6|0)	2.93(93.3|0)	2.01(1.4|0)
0.1	0.5	3	200	2.76(76.3|0)	2.01(1.4|0.9)	2.95(95.3|0)	1.99(0|0.6)	2.98(97.9|0)	2(0|0)

ω_F is the common off-diagonal element of Ω_F that is the covariance matrix of the innovations in the local factors. See footnotes to Table A.3.

A.6 Estimating the Number of Local Factors

In this section we provide the additional simulation results for the estimation of local factors. We apply the estimation procedure suggested in Section 1.4.2 to the same experiments conducted in Section 1.5. We consider the four estimators, IC_{p2}, BIC_3 by Bai & Ng (2002), ER by Ahn & Horenstein (2013) and ED by Onatski (2010). Before applying these estimators to the data within each block ($i = 1, \cdots, R$), we need to concentrate out the $T \times \hat{r}_0$ matrix of estimated global factors, $\widehat{\mathbf{G}}$, where $\hat{r}_0$ is estimated by CCD or MCC.

We iterate each simulation experiment 1000 times, yielding $R \times 1000$ $\hat{r}_i$ in total. The maximum number of local factors, $r^*_{i,\max}$ is set to $r^*_{\max} - \hat{r}_0$. We report the average estimates together with the percentage of overestimates and underestimates inside the parenthesis as $(O|U)$ in Table A.7-A.12.

Table A.7 presents the simulation results for the benchmark case with $(r_0, r_i) = (2, 2)$ and $(\beta, \phi_e, \kappa) = (0, 0, 1)$. The four approaches can produce accurate estimates of $\hat{r}_i$, unless M is too small. If the idiosyncratic errors are cross-sectionally and serially correlated, IC_{p2} always overestimates r_i. BIC_3 and ER perform slightly better than ED, see Table A.8. Qualitatively similar patterns can be observed in Table A.9 and Table A.10.

The results for $(r_0, r_i) = (3, 3)$ reported in Table A.11, show that it is more difficult to precisely estimate r_i in small sample, although the number of global factors is still accurately estimated. As M and T become large, BIC_3 and ER can provide precise estimates. When working with the very noisy data with $(\beta, \phi_e, \kappa) = (0.1, 0.5, 3)$, we find that all the estimators perform poorly, see Table A.12. This is a small sample issue as the biases will vanish as M and T increases.

Combining these results with the main simulation results in Section 1.5, we may conclude that the sequential procedure using BIC or ER in conjunction with CCD or MCC will provide reliable inference for jointly selecting the number of global factors and the number of local factors in multilevel factor model.

Table A.7: Average estimates of the number of local factors for experiments with $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (2, 2)$, $(\beta, \phi_e, \kappa) = (0, 0, 1)$ and $r^*_{i,\max} = r^*_{\max} - \hat{r}_0$

R	M	T	IC_{p2}	BIC_3	ER	ED
2	20	50	2.02(5|3)	1.59(0|40)	1.75(2|27)	2.13(17|3)
2	50	50	2(0|0)	1.9(0|10)	1.97(0|3)	2.06(6|0)
2	100	50	2(0|0)	1.98(0|2)	2(0|0)	2.03(3|0)
2	200	50	2(0|0)	2(0|0)	2(0|0)	2.02(2|0)
2	20	100	2.06(6|0)	1.63(0|36)	1.88(1|13)	2.19(19|1)
2	50	100	2(0|0)	1.99(0|1)	2(0|0)	2.1(10|0)
2	100	100	2(0|0)	2(0|0)	2(0|0)	2.04(4|0)
2	200	100	2(0|0)	2(0|0)	2(0|0)	2.02(2|0)
2	20	200	2.07(7|0)	1.62(0|37)	1.93(1|9)	2.25(25|0)
2	50	200	2(0|0)	2(0|0)	2(0|0)	2.17(17|0)
2	100	200	2(0|0)	2(0|0)	2(0|0)	2.08(8|0)
2	200	200	2(0|0)	2(0|0)	2(0|0)	2.04(4|0)
5	20	50	2.05(9|4)	1.52(0|46)	1.73(2|29)	2.13(18|5)
5	50	50	2(0|0)	1.9(0|10)	1.97(0|3)	2.06(6|0)
5	100	50	2(0|0)	1.98(0|2)	2(0|0)	2.03(3|0)
5	200	50	2(0|0)	1.99(0|1)	2(0|0)	2.02(2|0)
5	20	100	2.09(10|1)	1.58(0|41)	1.85(2|17)	2.28(29|1)
5	50	100	2(0|0)	1.99(0|1)	2(0|0)	2.15(15|0)
5	100	100	2(0|0)	2(0|0)	2(0|0)	2.05(5|0)
5	200	100	2(0|0)	2(0|0)	2(0|0)	2.02(2|0)
5	20	200	2.11(11|0)	1.56(0|43)	1.89(2|12)	2.38(39|1)
5	50	200	2(0|0)	2(0|0)	2(0|0)	2.26(26|0)
5	100	200	2(0|0)	2(0|0)	2(0|0)	2.12(12|0)
5	200	200	2(0|0)	2(0|0)	2(0|0)	2.04(4|0)
10	20	50	2.06(10|4)	1.49(0|49)	1.72(2|30)	2.14(19|5)
10	50	50	2(0|0)	1.9(0|10)	1.96(0|4)	2.07(7|0)
10	100	50	2(0|0)	1.98(0|2)	2(0|0)	2.03(3|0)
10	200	50	2(0|0)	1.99(0|1)	2(0|0)	2.02(2|0)
10	20	100	2.1(11|1)	1.55(0|44)	1.84(2|18)	2.29(30|1)
10	50	100	2(0|0)	1.99(0|1)	2(0|0)	2.15(15|0)
10	100	100	2(0|0)	2(0|0)	2(0|0)	2.05(5|0)
10	200	100	2(0|0)	2(0|0)	2(0|0)	2.02(2|0)
10	20	200	2.12(12|0)	1.53(0|45)	1.89(2|13)	2.41(42|1)
10	50	200	2(0|0)	2(0|0)	2(0|0)	2.29(29|0)
10	100	200	2(0|0)	2(0|0)	2(0|0)	2.14(14|0)
10	200	200	2(0|0)	2(0|0)	2(0|0)	2.04(4|0)

The average of $\hat{r}_i$ across R blocks over 1,000 replications are reported together with the figures inside the parenthesis, $(O|U)$, indicating the percentage of overestimation and underestimation, respectively. ϕ_G and ϕ_F are the AR coefficients of the global and local factors respectively. β, ϕ_e and κ control the cross-section correlation, serial correlation and noise to signal ratio, respectively. $\hat{r}_0$ is estimated by CCD with $r^*_{\max} = \max\left\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_R}\right\}$.

Table A.8: Average estimates of the number of local factors for experiments with $(\phi_G, \phi_F) = (0.5, 0.5), (r_0, r_i) = (2, 2), (\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ and $r^*_{i,\max} = r^*_{\max} - \hat{r}_0$

R	M	T	IC_{p2}	BIC_3	ER	ED
2	20	50	2.95(92|0)	2.06(17|11)	1.77(5|28)	2.23(46|18)
2	50	50	2.81(81|0)	1.99(1|3)	1.88(2|14)	2.12(18|5)
2	100	50	2.81(81|0)	1.99(0|1)	1.94(0|6)	2.09(10|1)
2	200	50	2.58(60|2)	1.96(0|4)	1.95(0|5)	2.01(5|3)
2	20	100	2.99(99|0)	1.99(9|10)	1.82(2|20)	2.33(55|16)
2	50	100	2.99(99|0)	2(0|0)	1.96(0|4)	2.07(8|1)
2	100	100	2.74(74|0)	2(0|0)	2(0|0)	2.04(4|0)
2	200	100	2.45(45|0)	2(0|0)	2(0|0)	2.04(4|0)
2	20	200	3(100|0)	1.91(3|12)	1.87(2|15)	2.4(58|13)
2	50	200	3(100|0)	2(0|0)	1.99(0|1)	2.02(2|0)
2	100	200	2.98(98|0)	2(0|0)	2(0|0)	2.02(2|0)
2	200	200	2.26(26|0)	2(0|0)	2(0|0)	2.02(2|0)
5	20	50	2.96(96|0)	1.97(11|13)	1.75(6|31)	2.02(37|25)
5	50	50	2.84(84|0)	1.99(1|2)	1.88(1|14)	2.08(15|6)
5	100	50	2.84(84|0)	2(0|0)	1.95(0|5)	2.08(8|0)
5	200	50	2.73(73|0)	2(0|0)	1.99(0|1)	2.05(5|0)
5	20	100	2.99(99|0)	1.9(5|15)	1.81(4|23)	2.13(43|22)
5	50	100	2.98(98|0)	2(0|0)	1.96(0|4)	2.06(7|2)
5	100	100	2.71(71|0)	2(0|0)	2(0|0)	2.04(4|0)
5	200	100	2.44(44|0)	2(0|0)	2(0|0)	2.04(4|0)
5	20	200	3(100|0)	1.83(2|18)	1.85(3|18)	2.19(46|20)
5	50	200	3(100|0)	2(0|0)	1.99(0|1)	2.01(1|0)
5	100	200	2.98(98|0)	2(0|0)	2(0|0)	2.02(2|0)
5	200	200	2.24(24|0)	2(0|0)	2(0|0)	2.03(3|0)
10	20	50	2.96(96|0)	1.93(8|15)	1.73(6|33)	1.95(33|28)
10	50	50	2.82(82|0)	1.99(1|2)	1.88(1|13)	2.07(14|5)
10	100	50	2.82(82|0)	2(0|0)	1.96(0|5)	2.07(7|0)
10	200	50	2.72(72|0)	2(0|0)	1.99(0|1)	2.05(5|0)
10	20	100	2.99(99|0)	1.87(4|17)	1.8(4|24)	2.09(41|23)
10	50	100	2.98(98|0)	2(0|0)	1.96(0|4)	2.05(7|1)
10	100	100	2.7(70|0)	2(0|0)	2(0|0)	2.05(5|0)
10	200	100	2.43(43|0)	2(0|0)	2(0|0)	2.04(4|0)
10	20	200	3(100|0)	1.81(1|20)	1.84(3|19)	2.1(42|22)
10	50	200	3(100|0)	2(0|0)	1.99(0|1)	2.01(1|0)
10	100	200	2.98(98|0)	2(0|0)	2(0|0)	2.02(2|0)
10	200	200	2.26(24|0)	2(0|0)	2(0|0)	2.03(2|0)

See footnotes to Table A.7.

Table A.9: Average estimates of the number of local factors for experiments with $(\phi_G, \phi_F) = (0.5, 0.5), (r_0, r_i) = (0, 2), (\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ and $r^*_{i,\max} = r^*_{\max} - \hat{r}_0$

R	M	T	IC_{p2}	BIC_3	ER	ED
2	20	50	3.83(98|0)	2.77(74|0)	1.77(4|27)	2.74(63|12)
2	50	50	3.46(85|0)	2.25(25|0)	1.89(1|13)	2.35(24|4)
2	100	50	3.44(82|0)	2(1|1)	1.94(0|6)	2.12(10|1)
2	200	50	2.82(57|4)	1.93(0|7)	1.91(0|8)	2.01(6|4)
2	20	100	3.79(100|0)	2.57(57|0)	1.84(3|19)	3.33(85|3)
2	50	100	4.35(100|0)	2.12(12|0)	1.97(0|4)	2.49(22|1)
2	100	100	3.13(74|0)	2(0|0)	2(0|0)	2.06(4|0)
2	200	100	2.49(42|0)	2(0|0)	2(0|0)	2.05(4|0)
2	20	200	3.86(100|0)	2.38(38|0)	1.86(3|17)	3.83(98|0)
2	50	200	3.86(100|0)	2.38(39|1)	1.88(4|16)	3.82(98|0)
2	100	200	4.44(99|0)	2(0|0)	2(0|0)	2.02(1|0)
2	200	200	2.26(24|0)	2(0|0)	2(0|0)	2.02(1|0)
5	20	50	3.84(97|0)	2.76(73|0)	1.78(4|26)	2.72(61|11)
5	50	50	3.48(85|0)	2.27(26|0)	1.88(1|13)	2.36(25|4)
5	100	50	3.55(85|0)	2.01(1|0)	1.95(0|5)	2.13(10|0)
5	200	50	3.12(70|0)	2(0|0)	1.99(0|1)	2.06(4|0)
5	20	100	3.79(100|0)	2.58(58|0)	1.85(3|19)	3.35(86|3)
5	50	100	4.32(99|0)	2.12(12|0)	1.96(0|4)	2.48(22|1)
5	100	100	3.08(72|0)	2(0|0)	2(0|0)	2.07(5|0)
5	200	100	2.5(42|0)	2(0|0)	2(0|0)	2.04(3|0)
5	20	200	3.86(100|0)	2.38(39|1)	1.89(3|14)	3.83(98|0)
5	50	200	4.82(100|0)	2.03(3|0)	1.98(0|2)	2.38(14|0)
5	100	200	4.46(99|0)	2(0|0)	2(0|0)	2.02(1|0)
5	200	200	2.26(24|0)	2(0|0)	2(0|0)	2.02(2|0)
10	20	50	3.85(98|0)	2.77(74|0)	1.76(4|28)	2.72(61|11)
10	50	50	3.48(86|0)	2.26(26|0)	1.89(1|13)	2.38(26|4)
10	100	50	3.54(84|0)	2.01(1|0)	1.95(0|5)	2.14(10|0)
10	200	50	3.16(71|0)	2(0|0)	1.99(0|1)	2.07(5|0)
10	20	100	3.79(100|0)	2.58(58|0)	1.84(3|19)	3.33(85|3)
10	50	100	4.33(99|0)	2.12(12|0)	1.97(0|4)	2.48(22|1)
10	100	100	3.1(74|0)	2(0|0)	2(0|0)	2.07(5|0)
10	200	100	2.5(42|0)	2(0|0)	2(0|0)	2.05(3|0)
10	20	200	3.86(100|0)	2.38(39|1)	1.89(3|14)	3.83(98|0)
10	50	200	4.83(100|0)	2.03(3|0)	1.99(0|1)	2.37(14|0)
10	100	200	4.46(99|0)	2(0|0)	2(0|0)	2.01(1|0)
10	200	200	2.26(24|0)	2(0|0)	2(0|0)	2.02(2|0)

See footnotes to Table A.7.

Table A.10: Average estimates of the number of local factors for experiments with $(\phi_G, \phi_F) = (0.5, 0.5), (r_0, r_i) = (1, 1), (\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ and $r^*_{i,\max} = r^*_{\max} - \hat{r}_0$

R	M	T	IC_{p2}	BIC_3	ER	ED
2	20	50	2.93(99|0)	1.82(78|0)	1.01(1|0)	2.11(75|0)
2	50	50	2.58(88|0)	1.26(26|0)	1(0|0)	1.6(34|0)
2	100	50	2.52(85|0)	1.01(1|0)	1(0|0)	1.2(14|0)
2	200	50	1.86(59|2)	0.97(0|3)	0.98(0|2)	1.1(9|1)
2	20	100	2.88(100|0)	1.66(66|0)	1(0|0)	2.51(90|0)
2	50	100	3.34(99|0)	1.11(11|0)	1(0|0)	1.62(27|0)
2	100	100	2.09(73|0)	1(0|0)	1(0|0)	1.11(7|0)
2	200	100	1.53(45|0)	1(0|0)	1(0|0)	1.07(5|0)
2	20	200	2.94(100|0)	1.47(47|0)	1(0|0)	2.93(98|0)
2	50	200	3.83(100|0)	1.02(2|0)	1(0|0)	1.44(17|0)
2	100	200	3.48(99|0)	1(0|0)	1(0|0)	1.02(2|0)
2	200	200	1.27(25|0)	1(0|0)	1(0|0)	1.04(3|0)
5	20	50	2.93(98|0)	1.78(75|0)	1.01(1|0)	2.04(70|0)
5	50	50	2.5(87|0)	1.23(23|0)	1(0|0)	1.53(31|0)
5	100	50	2.55(85|0)	1.01(1|0)	1(0|0)	1.2(14|0)
5	200	50	2.11(69|0)	1(0|0)	1(0|0)	1.09(6|0)
5	20	100	2.89(100|0)	1.61(61|0)	1(0|0)	2.47(88|0)
5	50	100	3.33(99|0)	1.1(10|0)	1(0|0)	1.59(27|0)
5	100	100	2.08(72|0)	1(0|0)	1(0|0)	1.1(7|0)
5	200	100	1.48(40|0)	1(0|0)	1(0|0)	1.07(5|0)
5	20	200	2.95(100|0)	1.41(41|0)	1(0|0)	2.9(97|0)
5	50	200	3.81(100|0)	1.02(2|0)	1(0|0)	1.46(17|0)
5	100	200	3.4(99|0)	1(0|0)	1(0|0)	1.03(2|0)
5	200	200	1.26(25|0)	1(0|0)	1(0|0)	1.04(3|0)
10	20	50	2.91(98|0)	1.78(75|0)	1(0|0)	2.03(70|0)
10	50	50	2.47(86|0)	1.23(22|0)	1(0|0)	1.51(31|0)
10	100	50	2.54(85|0)	1.01(1|0)	1(0|0)	1.19(13|0)
10	200	50	2.12(70|0)	1(0|0)	1(0|0)	1.09(6|0)
10	20	100	2.91(100|0)	1.59(59|0)	1(0|0)	2.45(86|0)
10	50	100	3.31(99|0)	1.09(9|0)	1(0|0)	1.58(26|0)
10	100	100	2.05(71|0)	1(0|0)	1(0|0)	1.1(7|0)
10	200	100	1.49(41|0)	1(0|0)	1(0|0)	1.07(5|0)
10	20	200	2.97(100|0)	1.39(39|0)	1(0|0)	2.9(96|0)
10	50	200	3.81(100|0)	1.02(2|0)	1(0|0)	1.42(16|0)
10	100	200	3.4(98|0)	1(0|0)	1(0|0)	1.02(1|0)
10	200	200	1.25(23|0)	1(0|0)	1(0|0)	1.04(3|0)

See footnotes to Table A.7.

Table A.11: Average estimates of the number of local factors for experiments with $(\phi_G, \phi_F) = (0.5, 0.5), (r_0, r_i) = (3, 3), (\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$ and $r^*_{i,\max} = r^*_{\max} - \hat{r}_0$

R	M	T	IC_{p2}	BIC_3	ER	ED
2	20	50	4.41(82|4)	3.11(36|24)	2.18(11|63)	2.4(31|50)
2	50	50	4.19(80|2)	2.9(9|17)	2.47(8|43)	2.39(20|42)
2	100	50	4.17(77|4)	2.84(0|15)	2.65(1|25)	2.87(9|15)
2	200	50	3.67(57|12)	2.72(0|24)	2.68(0|23)	2.79(5|15)
2	20	100	4.89(99|0)	3.13(28|16)	2.48(13|45)	3.01(53|35)
2	50	100	4.84(99|0)	3.02(5|3)	2.78(4|19)	2.57(13|28)
2	100	100	4.12(76|0)	3(0|0)	2.98(0|2)	3.05(4|0)
2	200	100	3.65(52|0)	3(0|0)	3(0|0)	3.05(4|0)
2	20	200	4.89(99|0)	3.14(30|16)	2.48(13|46)	2.93(51|37)
2	50	200	4.99(100|0)	3.01(1|1)	2.9(2|9)	2.58(4|21)
2	100	200	4.86(98|0)	3(0|0)	3(0|0)	3.02(1|0)
2	200	200	3.33(30|0)	3(0|0)	3(0|0)	3.01(1|0)
5	20	50	4.69(95|1)	3.15(30|17)	2.24(13|60)	2.11(27|58)
5	50	50	4.28(83|0)	2.97(7|10)	2.5(7|40)	2.34(16|43)
5	100	50	4.38(86|0)	2.94(0|6)	2.72(1|21)	2.91(8|11)
5	200	50	4.21(79|0)	2.95(0|5)	2.88(0|9)	3.03(4|1)
5	20	100	4.81(98|0)	2.87(14|27)	2.36(11|51)	2.42(37|49)
5	50	100	4.79(98|0)	2.99(2|4)	2.78(3|19)	2.53(11|28)
5	100	100	4.01(73|0)	3(0|0)	2.97(0|2)	3.04(4|0)
5	200	100	3.57(47|0)	3(0|0)	3(0|0)	3.04(3|0)
5	20	200	4.86(99|0)	2.65(8|41)	2.46(11|45)	2.68(46|43)
5	50	200	4.98(100|0)	2.99(0|1)	2.91(1|8)	2.63(4|18)
5	100	200	4.81(98|0)	3(0|0)	3(0|0)	3.01(1|0)
5	200	200	3.28(25|0)	3(0|0)	3(0|0)	3.02(2|0)
10	20	50	4.7(94|1)	3.09(25|16)	2.2(12|61)	1.95(23|62)
10	50	50	4.23(81|0)	2.94(6|12)	2.48(6|41)	2.29(14|43)
10	100	50	4.37(86|0)	2.94(0|6)	2.72(1|21)	2.9(8|11)
10	200	50	4.21(79|0)	2.96(0|4)	2.9(0|8)	3.04(4|1)
10	20	100	4.8(98|0)	2.8(10|30)	2.32(10|53)	2.22(32|54)
10	50	100	4.78(98|0)	2.97(2|4)	2.78(3|19)	2.53(10|28)
10	100	100	3.97(71|0)	3(0|0)	2.97(0|2)	3.04(4|0)
10	200	100	3.54(45|0)	3(0|0)	3(0|0)	3.04(3|0)
10	20	200	4.86(99|0)	2.57(4|44)	2.44(10|46)	2.36(39|50)
10	50	200	4.98(100|0)	2.99(0|2)	2.9(1|9)	2.64(4|18)
10	100	200	4.79(97|0)	3(0|0)	3(0|0)	3.01(1|0)
10	200	200	3.26(24|0)	3(0|0)	3(0|0)	3.02(1|0)

See footnotes to Table A.7.

Table A.12: Average estimates of the number of local factors for Experiments with $(\phi_G, \phi_F) = (0.5, 0.5), (r_0, r_i) = (2, 2), (\beta, \phi_e, \kappa) = (0.1, 0.5, 3)$ and $r^*_{i,\max} = r^*_{\max} - \hat{r}_0$

R	M	T	IC_{p2}	BIC_3	ER	ED
2	20	50	3.13(76|5)	1.51(12|51)	1.56(17|55)	1.34(23|57)
2	50	50	2.58(62|7)	1.02(0|80)	1.5(11|53)	1.02(12|68)
2	100	50	2.6(65|5)	1.03(0|81)	1.4(4|50)	1.34(8|49)
2	200	50	2.4(50|9)	1.02(0|81)	1.47(0|42)	1.7(3|25)
2	20	100	3.4(94|0)	1.25(5|66)	1.53(16|52)	1.45(32|56)
2	50	100	3.01(97|0)	1.13(0|76)	1.47(9|48)	0.78(7|75)
2	100	100	2.68(68|0)	1.57(0|42)	1.79(1|20)	1.79(5|18)
2	200	100	2.44(44|0)	1.88(0|12)	1.96(0|4)	2.04(4|0)
2	20	200	3.29(96|0)	0.91(1|82)	1.48(16|52)	1.45(35|55)
2	50	200	3.02(100|0)	1.16(0|75)	1.47(6|43)	0.6(2|80)
2	100	200	2.98(98|0)	1.88(0|12)	1.93(0|6)	1.98(2|3)
2	200	200	2.26(26|0)	2(0|0)	2(0|0)	2.03(3|0)
5	20	50	3.29(84|3)	1.54(15|52)	1.59(18|54)	1.14(19|66)
5	50	50	2.62(65|6)	0.99(0|84)	1.48(12|55)	0.87(9|74)
5	100	50	2.77(77|1)	1.07(0|81)	1.41(3|49)	1.35(8|48)
5	200	50	2.67(67|0)	1.12(0|79)	1.56(0|35)	1.82(4|17)
5	20	100	3.36(92|1)	1.11(5|72)	1.42(16|59)	1.17(22|65)
5	50	100	2.96(95|0)	1.05(0|81)	1.42(9|51)	0.76(6|76)
5	100	100	2.66(66|0)	1.53(0|45)	1.76(1|22)	1.76(5|19)
5	200	100	2.43(43|0)	1.85(0|15)	1.96(0|4)	2.04(4|0)
5	20	200	3.14(95|0)	0.68(1|90)	1.34(15|60)	1.14(24|64)
5	50	200	3(100|0)	1.08(0|79)	1.38(5|47)	0.61(2|79)
5	100	200	2.97(97|0)	1.87(0|13)	1.92(0|7)	1.98(2|3)
5	200	200	2.23(23|0)	2(0|0)	2(0|0)	2.02(2|0)
10	20	50	3.24(84|3)	1.48(14|56)	1.54(18|56)	1.12(18|67)
10	50	50	2.62(66|6)	0.97(0|84)	1.46(12|55)	0.86(9|74)
10	100	50	2.76(76|1)	1.05(0|82)	1.39(3|50)	1.33(7|49)
10	200	50	2.67(67|0)	1.13(0|78)	1.57(0|35)	1.81(5|17)
10	20	100	3.27(91|1)	1.04(4|76)	1.38(15|60)	1.09(20|67)
10	50	100	2.96(96|0)	1.01(0|83)	1.38(8|53)	0.71(5|78)
10	100	100	2.65(65|0)	1.53(0|46)	1.77(1|21)	1.78(4|18)
10	200	100	2.4(40|0)	1.85(0|15)	1.95(0|4)	2.04(4|0)
10	20	200	3.06(95|0)	0.58(0|94)	1.25(13|62)	0.98(20|69)
10	50	200	3(100|0)	1.05(0|82)	1.32(5|50)	0.6(2|80)
10	100	200	2.97(97|0)	1.86(0|14)	1.92(0|7)	1.98(2|3)
10	200	200	2.22(22|0)	2(0|0)	2(0|0)	2.02(2|0)

See footnotes to Table A.7.

A.7 Alternative Selection Criteria under Two Wide-groups Division when $R > 2$

In this Section we follow the anonymous referee's suggestion and split the whole data with $R > 2$ groups into the two wide groups. This simple modification enables us to apply the *AGGR*'s procedure for estimating the number of global factors even if $R > 2$, without developing complex analytic counterparts. Furthermore, this scheme may improve the finite sample performance of CCD and MCC estimators by increasing the number of cross-section observations used in the estimation of the number of global factors, r_0 and the global factors, $\mathbf{G}$. This implementation requires to calculate only one pair of sample squared canonical correlations, though we need to adjust $r_{\max}$ when estimating the two wide-group specific factors.

Let $\lfloor R/2 \rfloor$ be the integer part of $R/2$. Without loss of generality, we suppose that the first (the second) "w" group contains the blocks, $i = 1, \ldots, \lfloor R/2 \rfloor$ ($i = \lfloor R/2 \rfloor + 1, \ldots, R$). We use the superscript, w to denote the data constructed by this strategy. The numbers of cross-sectional observations in each "w" group are $M_1^w = M_1 + \cdots + M_{\lfloor R/2 \rfloor}$ and $M_2^w = M_{\lfloor R/2 \rfloor + 1} + \cdots + M_R$, respectively. Denote the corresponding $T \times M_1^w$ and $T \times M_2^w$ data matrices by $\mathbf{Y}_1^w$ and $\mathbf{Y}_2^w$. The first "w" group contains $r_1^w = r_0 + r_1 + \cdots + r_{\lfloor R/2 \rfloor}$ factors, denoted by $\mathbf{K}_1^w$ while the second "w" group contains $r_2^w = r_0 + r_{\lfloor R/2 \rfloor + 1} + \cdots + r_R$ factors, denoted by $\mathbf{K}_2^w$. Then, we can estimate r_0 through evaluating the canonical correlations between $\mathbf{K}_1^w$ and $\mathbf{K}_2^w$.

We explore the performance of $AGGR$, CCD and MCC with the two wide-group division, denoted respectively by $AGGR^w$, CCD^w and MCC^w,[7] via additional Monte Carlo experiments. We consider the same DGP employed under Experiments 1 and 3 as in Section 1.5.

To consistently estimate factors and canonical correlations by CCD^w and MCC^w, we need to select a sufficiently large $r_{\max} \geq \max\{r_1^w, r_2^w\}$. Following a practical selection guideline described in Section 1.5, we select the common $r_{\max}^*$ for the two-wide groups by $r_{\max}^* = \max\{\hat{r}_1^w, \hat{r}_2^w\}$, where $\hat{r}_1^w$ and $\hat{r}_2^w$ are the number of factors estimated by BIC_3. Here we set $r_{\max} = r_0 + r_i \times \lceil R/2 \rceil + 5$ for each w-groups when applying BIC_3, where $\lceil R/2 \rceil$ is the smallest integer that is larger than or equal to $R/2$. For example, if $(r_0, r_i) = (2, 2)$, then $r_{\max} = 9, 13, 17$ for $R = 2, 5, 10$. This ensures the consistency of BIC_3. When implementing $AGGR^w$, we assume that the true numbers of factors, r_1^w and r_2^w, are known, and set $r_{\max} = \max\{r_1^w, r_2^w\}$. But, we still allow the estimation uncertainty in implementing CCD^w and MCC^w.

Table A.13 shows the simulation results for Experiment 1. For $R = 2$, the performances of CCD^w, MCC^w and $AGGR^w$ are qualitatively similar to those in Table 1. As R increases, however, all three criteria tend to overestimate r_0 if T is small ($T = 50$). In particular, the performance of $AGGR^w$ becomes satisfactory only if M and T are sufficiently large. Overall, we find that CCD^w and MCC^w appear to be outperformed by CCD and MCC. First, we need to set $r_{\max}^*$ larger as R grows. This may adversely affect the precision of canonical correlations. Second, if $M_i \geq T$, the two wide-group division does not yield more precise factor estimates. By Lemma 4, the convergence

[7]We do not report the results fo the IC approach by Chen (2012), since it is computationally too burdensome and its performance is dominated by CCD and MCC as reported in the main text.

rate of $\widehat{\mathbf{K}}_1^w$ is $O_p(1/\delta_{M_1^w T})$ where $\delta_{M_1^w T} = \min\left\{\sqrt{M_1^w}, \sqrt{T}\right\} = \sqrt{T}$. In this case the use of more cross-section observations does not necessarily result in a faster convergence rate.

However, if $T > M_i$, then the larger M_1^w and M_2^w will improve the precision of factor estimates by CCD^w and MCC^w. If T is also sufficiently large ($T \geq 100$) with $M_i << T$, then the advantage of using the larger M_1^w and M_2^w can outweigh the disadvantage caused by assigning the larger r_{max}^*. This gain is more noticeable when the data is noisier. For example, in Panel C of Table A.13, the average estimates of CCD^w and MCC^w with $(M, T) = (20, 200)$ are 1.68 and 1.21 (2 and 1.93) for $R = 5$ ($R = 10$) whilst the corresponding figures for CCD and MCC are only 0.92 and 0.7 (0.96 and 0.78) for $R = 5$ ($R = 10$) (see Table 1.1).

Next, Table A.14 shows the simulation results for Experiment 3. We allow the number of global factors to vary from 0 to 3 by setting $(r_0, r_i) \in \{(0,2), (1,1), (3,3)\}$ for $i = 1, ..., R$ and $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$. First, the results for the case with $(r_0, r_i) = (0, 2)$, are reported in Panel A. CCD^w tends to select zero global factor correctly for $R = 2, 5$. But, if $R = 10$, CCD^w slightly overestimate r_0 if T is small, mainly due to the adverse effect of the larger $r_{\max}^{*w}$. Such overestimation becomes more severe for MCC^w and $AGGR^w$. Second, turning to the case with $(r_0, r_i) = (1, 1)$ in Panel B, we find that CCD^w and MCC^w estimate $r_0 = 1$ correctly in most cases. $AGGR^w$ overestimates r_0 when M or T is small. Finally, the results for $(r_0, r_i) = (3, 3)$ presented in Panel C, display that CCD^w, MCC^w and $AGGR^w$ are all subject to severe overestimation biases if T is small. The performance of CCD^w and MCC^w improves sharply with T. By contrast, the performance of $AGGR$ becomes satisfactory only if both M and T are large. Overall, CCD and MCC still outperform CCD^w, MCC^w and $AGGR^w$ in most cases.

We may draw three conclusions from these findings. First, we are able to apply the $AGGR$ approach to the multilevel panel with $R > 2$ by forming the two wide-groups, though its performance becomes satisfactory only if both M and T are substantially large. Its performance becomes rather unreliable, especially if T is small. Second, CCD and MCC still outperform CCD^w, MCC^w and $AGGR^w$ in most cases. Third, there is a trade-off between a selection of the larger $r_{\max}^*$ and the use of more cross-section observations. CCD^w and MCC^w significantly improves the estimation precision of r_0 for the multilevel panel with $R > 2$, if T is sufficiently large and M is much smaller than T. On the other hand, if T is small, then the performance of CCD^w and MCC^w appears to be unreliable, and MCC^w tends to overestimate r_0. Hence, we may recommend to apply this 2-wide groups modification in practice, only when T is sufficiently large and M is much smaller than T.

Table A.13: Average estimates of the number of global factors for $w-$groups for Experiment 1. $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (2, 2)$, $(\beta, \phi_e, \kappa) \in \{(0, 0, 1), (0.1, 0.5, 1), (0.1, 0.5, 3)\}$

		CCD^w	MCC^w	$AGGR^w$	CCD^w	MCC^w	$AGGR^w$	CCD^w	MCC^w	$AGGR^w$
Panel A: $(\beta, \phi_e, \kappa) = (0, 0, 1)$										
M	T		$R = 2$			$R = 5$			$R = 10$	
20	50	1.95(0.3|4.5)	1.96(0|4.2)	2.82(65.9|3.3)	1.99(0.9|1.5)	1.99(0.2|1.1)	7.55(98.7|0)	2.02(2.1|0)	2.01(0.8|0)	5.5(45.5|0.1)
50	50	1.98(0|1.7)	1.98(0|1.8)	1.98(0|2)	2(0.6|0.4)	2(0|0.4)	2.28(6.4|0.3)	2.01(1.4|0)	2(0|0)	5.25(38.7|0)
100	50	1.99(0|0.6)	1.99(0|0.7)	1.99(0|0.7)	2(0.3|0.1)	1.99(0|1)	2.3(7.4|0)	2.01(0.7|0)	2(0|0.1)	5.04(36.5|0)
200	50	2(0|0.4)	2(0|0.5)	2(0|0.3)	2(0.2|0.3)	1.98(0|1.6)	2.21(5.7|0.1)	2.01(0.6|0)	2(0|0.3)	5.09(37.3|0)
20	100	1.92(0|6.3)	1.91(0|8.3)	2.54(52.3|4.2)	2(0|0.5)	1.99(0|1.3)	8(100|0)	2(0|0)	2(0|0)	2.13(1.7|0.3)
50	100	2(0|0.1)	2(0|0.1)	2.28(28.1|0)	2(0|0)	2(0|0)	2(0.1|0.1)	2(0|0)	2(0|0)	2.14(1.8|0)
100	100	2(0|0.1)	2(0|0.1)	2(0|0)	2(0|0)	2(0|0)	2.01(0.3|0)	2(0|0)	2(0|0)	2.1(1.3|0)
200	100	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0.2|0)	2(0|0)	2(0|0)	2.05(0.6|0)
20	200	1.83(0|13.1)	1.78(0|18.9)	2.37(46.5|8.6)	2(0|0.3)	1.99(0|1.3)	8(100|0)	2(0|0)	2(0|0)	11.95(99.9|0)
50	200	2(0|0)	2(0|0)	2.19(20.1|0.6)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
100	200	2(0|0)	2(0|0)	2.11(11.2|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
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)
Panel B: $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$										
M	T		$R = 2$			$R = 5$			$R = 10$	
20	50	2.06(6.8|2.4)	2.05(5.9|0.6)	2.56(61.6|16.8)	2.24(14|0)	2.45(41.7|0)	5.69(88.5|0.5)	3.42(42.1|0)	3.06(81.9|0)	9.4(87.7|0.1)
50	50	2.02(2|0.2)	2(0.1|0.2)	1.66(0|31.4)	2.07(5.7|0)	2.04(4.4|0)	2.52(11.7|2)	2.67(24.3|0)	2.19(18.5|0)	9.47(84.6|0)
100	50	2(0.2|0)	2(0|0.1)	1.86(0|13.2)	2.1(7.5|0)	2.01(1|0)	2.56(12.6|0.4)	2.43(17.8|0)	2.02(1.8|0)	9.4(83.5|0)
200	50	2(0|0)	2(0|0)	1.94(0|5.6)	2.07(5|0)	2(0.1|0)	2.64(14.5|0.2)	2.26(13.1|0)	2(0.1|0)	8.81(77.7|0.3)
20	100	1.99(0|1.5)	1.95(0|5.2)	2.26(52|24.7)	2(0|0)	2(0|0)	5.62(86.8|1.5)	2(0|0)	2(0|0)	2.27(3.8|4.1)
50	100	2(0|0)	2(0|0.1)	2.18(30.8|11.4)	2(0|0)	2(0|0)	1.99(0.2|2.3)	2(0|0)	2(0|0)	2.31(3.7|0)
100	100	2(0|0)	2(0|0)	1.91(0|8.7)	2(0|0)	2(0|0)	2(0.1|0.6)	2(0|0)	2(0|0)	2.29(3.4|0)
200	100	2(0|0)	2(0|0)	1.98(0|1.6)	2(0|0)	2(0|0)	2.01(0.5|0.1)	2(0.1|0)	2(0|0)	2.24(2.9|0)
20	200	1.96(0|3.6)	1.8(0|19.8)	2.04(45.2|31.6)	2(0|0)	2(0|0)	4.58(69.7|6.4)	2(0|0)	2(0|0)	11.63(99.8|0)
50	200	2(0|0)	2(0|0.2)	1.98(17.8|17)	2(0|0)	2(0|0)	4.64(80|1.1)	2(0|0)	2(0|0)	1.99(0|0.6)
100	200	2(0|0)	2(0|0)	2.06(11.1|4.8)	2(0|0)	2(0|0)	2(0|0.4)	2(0|0)	2(0|0)	2(0|0)
200	200	2(0|0)	2(0|0)	1.99(0|1.5)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
Panel C: $(\beta, \phi_e, \kappa) = (0.1, 0.5, 3)$										
20	50	1.38(9.8|48.3)	1.79(5.5|26.1)	3.08(96.2|1.1)	2.31(29.7|12.3)	2.3(31.5|2.2)	5.41(80|5.4)	3.99(68.6|0.9)	3.03(78.4|0)	10.34(99.5|0)
50	50	1.67(4.7|31.8)	1.69(1|31.2)	0.7(0|93.6)	2.18(18|4.5)	2.02(6.1|4.7)	2.88(22|14.5)	3.03(52.4|0)	2.27(26.4|0.4)	10.69(99|0.1)
100	50	1.83(1.8|17.6)	1.77(0.1|23.1)	0.96(0|83.8)	2.02(7.3|5.9)	1.92(0.2|8.5)	3.13(26.3|12.4)	2.62(39.9|0.1)	2.04(4.4|0.1)	10.59(97.6|0)
200	50	1.78(0.3|22.2)	1.71(0|28.4)	1.22(0|69.4)	1.98(3.8|5.8)	1.82(0|17.5)	3.43(32.2|13.2)	2.3(22.7|0.1)	1.98(0.2|1.8)	10.53(95.9|0.1)
20	100	0.93(0|67.9)	1.02(0|83)	2.88(91.8|4)	1.76(0|19.2)	1.68(0|31.6)	5.05(74.4|14.9)	2(0.3|0.3)	1.98(0|2.5)	2.4(6.2|10.7)
50	100	1.59(0|32.6)	1.38(0|58.6)	2.2(60.8|26.7)	1.96(0|4.3)	1.86(0|14)	1.61(0.3|37.3)	2(0.1|0.1)	2(0|0.5)	3.09(12.9|3.1)
100	100	1.95(0|4.8)	1.89(0|11)	1.1(0|74.7)	1.98(0|2.5)	1.9(0|10.4)	1.8(0.1|19.5)	2(0|0)	2(0|0.2)	3.72(19.8|1)
200	100	1.96(0|3.5)	1.96(0|4.2)	1.55(0|41.5)	1.99(0|1.3)	1.9(0|9.9)	1.93(0.1|7.1)	2(0|0)	2(0|0.2)	3.99(23.1|1.1)
20	200	0.55(0|86.4)	0.44(0|98.8)	2.72(85.4|8.4)	1.68(0|22.3)	1.21(0|72)	4.35(65|25.2)	2(0|0)	1.93(0|6.6)	7.75(87|7)
50	200	1.62(0|30.7)	1.06(0|81.4)	1.76(43.4|42)	1.99(0|1.3)	1.86(0|13.6)	4.02(58.3|21.7)	2(0|0)	2(0|0)	1.9(0|10.1)
100	200	1.99(0|1.1)	1.95(0|5.3)	1.4(19.7|55.8)	2(0|0.1)	1.98(0|2)	1.87(0|12.5)	2(0|0)	2(0|0)	1.99(0|0.8)
200	200	2(0|0)	2(0|0)	1.73(0|26.6)	2(0|0)	2(0|0.4)	1.98(0|2.3)	2(0|0)	2(0|0)	2(0|0)

The average of $\hat{r}_0$ over 1,000 replications is reported together with the figures inside the parenthesis, $(O|U)$, indicating the percentage of overestimation and underestimation. 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. For $AGGR$, we assume that the true number of factors, $r_0 + r_i$ is known whereas we allow the estimation uncertainty in implementing CCD^w and MCC^w. We select the common $r^*_{\max}$ for the two-wide groups by $r^*_{\max} = \max\{\hat{r}^w_1, \hat{r}^w_2\}$, where $\hat{r}^w_1$ and $\hat{r}^w_2$ are the number of factors estimated by BIC_3 with $r_{\max} = r_0 + r_i \times \lceil R/2 \rceil + 5$ for each of w-groups, where $\lceil R/2 \rceil$ is the smallest integer that is larger than or equal to $R/2$.

Table A.14: Average estimates of the number of global factors for $w-$groups for Experiment 3. $(\phi_G, \phi_F) = (0.5, 0.5)$, $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$, $(r_0, r_i) \in \{(0, 2), (1, 1), (3, 3)\}$

		CCD^w	MCC^w	$AGGR^w$	CCD^w	MCC^w	$AGGR^w$	CCD^w	MCC^w	$AGGR^w$
					Panel A: $(r_0, r_i) = (0, 2)$					
M	T		$R = 2$			$R = 5$			$R = 10$	
20	50	0.02(1.9|0)	0.3(28.7|0)	1.39(65.4|0)	0.13(6.9|0)	1.08(83.3|0)	7.62(99.9|0)	0.57(18|0)	1.28(89.6|0)	3.51(51.5|0)
50	50	0(0.2|0)	0.01(1|0)	0(0|0)	0.1(6.6|0)	0.25(24.2|0)	0.62(11.8|0)	0.3(13|0)	0.31(29.6|0)	3.44(39.9|0)
100	50	0(0.3|0)	0(0.2|0)	0(0|0)	0.06(4.7|0)	0.03(3.2|0)	0.64(10.8|0)	0.24(12|0)	0.04(4.1|0)	3.46(39.7|0)
200	50	0(0|0)	0(0|0)	0(0|0)	0.11(7.7|0)	0.01(0.8|0)	0.78(12.8|0)	0.34(14.3|0)	0(0.4|0)	8.58(79.9|0)
20	100	0(0|0)	0(0|0)	0.74(41.7|0)	0(0|0)	0(0|0)	0.68(40|0)	0(0|0)	0(0|0)	0.13(1.2|0)
50	100	0(0|0)	0(0|0)	0.44(30.8|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0.02(0.2|0)
100	100	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)
200	100	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0.02(0.3|0)	0(0|0)	0(0|0)	0.01(0.1|0)
20	200	0(0|0)	0(0|0)	0.43(24.5|0)	0(0|0)	0(0|0)	7.66(99.9|0)	0(0|0)	0(0|0)	7.87(98|0)
50	200	0(0|0)	0(0|0)	0.08(8.3|0)	0(0|0)	0(0|0)	6.19(100|0)	0(0|0)	0(0|0)	0(0|0)
100	200	0(0|0)	0(0|0)	1.33(69.6|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)
200	200	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)	0(0|0)
					Panel B: $(r_0, r_i) = (1, 1)$					
M	T		$R = 2$			$R = 5$			$R = 10$	
20	50	1(0.2|0)	1(0.4|0)	0.93(4.2|11.2)	1.01(1|0)	1.01(1.4|0)	1.96(56.8|1)	1.04(3.1|0)	1.02(1.7|0)	1.1(2.9|0.7)
50	50	1(0.3|0)	1(0.2|0)	0.92(0|7.8)	1(0.4|0)	1(0|0)	0.98(0|1.6)	1.02(1.8|0)	1(0|0)	1.06(1.7|0.2)
100	50	1(0.2|0)	1(0|0)	0.97(0|3)	1.01(0.7|0)	1(0|0)	0.99(0|0.6)	1.01(1.3|0)	1(0|0)	1.06(1.5|0.1)
200	50	1(0|0)	1(0|0)	0.98(0|1.6)	1.01(1|0)	1(0|0)	1(0|0.4)	1.01(0.9|0)	1(0|0)	1.02(0.8|0.1)
20	100	1(0|0)	1(0|0)	0.88(1.3|13.6)	1(0|0)	1(0|0)	2.15(64.8|1.8)	1(0|0)	1(0|0)	0.98(0|1.6)
50	100	1(0|0)	1(0|0)	0.97(0|3.5)	1(0|0)	1(0|0)	0.99(0|1.3)	1(0|0)	1(0|0)	1(0|0)
100	100	1(0|0)	1(0|0)	0.98(0|2.2)	1(0|0)	1(0|0)	1(0|0.5)	1(0|0)	1(0|0)	1(0|0)
200	100	1(0|0)	1(0|0)	0.99(0|0.9)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
20	200	1(0|0)	1(0|0)	0.84(0.4|16.4)	1(0|0)	1(0|0)	1.83(53.6|4.4)	1(0|0)	1(0|0)	3.47(76|0.5)
50	200	1(0|0)	1(0|0)	0.94(0|5.9)	1(0|0)	1(0|0)	2(70.9|0.5)	1(0|0)	1(0|0)	1(0|0)
100	200	1(0|0)	1(0|0)	0.99(0|1.3)	1(0|0)	1(0|0)	1(0|0.1)	1(0|0)	1(0|0)	1(0|0)
200	200	1(0|0)	1(0|0)	1(0|0.4)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
					Panel C: $(r_0, r_i) = (3, 3)$					
20	50	3.33(31.8|12.9)	3.24(26.9|4.1)	4.93(88.7|7.1)	5.17(64.2|0.5)	4.72(96.9|0)	11.7(100|0)	10.87(99|0)	7.08(100|0)	17(100|0)
50	50	3.07(8.7|2.5)	3(1.9|2.4)	2.1(0.2|73.9)	3.88(36.5|0.2)	3.52(47.9|0)	8.9(77.5|0.2)	9.01(95.8|0)	5.09(98.6|0)	17(100|0)
100	50	3.02(2.1|0.2)	2.99(0.1|0.7)	2.55(0.3|42.5)	3.48(22.2|0)	3.13(12.4|0)	9.08(78.4|0.1)	7.85(89.7|0)	4(79.3|0)	17(100|0)
200	50	3(0.3|0.1)	3(0|0.4)	2.8(0|19.4)	3.26(15.6|0)	3.01(1.3|0)	9.01(77.5|0.2)	6.88(80.3|0)	3.27(26.8|0)	17(100|0)
20	100	2.81(0.4|16.8)	2.47(0|49.7)	4.21(76.6|16.2)	3(0.4|0.5)	2.99(0.1|1)	10.91(98.5|0.5)	3.01(0.6|0)	3(0|0)	9.56(49.7|2.2)
50	100	2.99(0|1)	2.93(0|7.2)	3.79(59.5|19.5)	3(0|0)	3(0|0)	3.15(3.2|6)	3(0.2|0)	3(0|0)	10.51(55.5|0.1)
100	100	3(0|0)	3(0|0)	2.67(0|30.3)	3(0|0)	3(0|0)	3.23(3.3|0.3)	3(0.2|0)	3(0|0)	10.43(55.5|0)
200	100	3(0|0)	3(0|0)	2.91(0|9.1)	3(0|0)	3(0|0)	3.22(3.3|0.2)	3(0|0)	3(0|0)	9.71(50.5|0)
20	200	2.72(0|22.2)	1.85(0|87.9)	3.94(71.7|19.5)	3(0|0.1)	2.86(0|13.8)	8.78(89.2|5.1)	3(0|0)	3(0|0.1)	18(100|0)
50	200	3(0|0.1)	2.85(0|15.3)	3.39(44.1|26.3)	3(0|0)	3(0|0)	8.02(86.3|1.4)	3(0|0)	3(0|0)	3(0.1|0.9)
100	200	3(0|0)	3(0|0)	3.24(23.8|11.9)	3(0|0)	3(0|0)	2.98(0|1.6)	3(0|0)	3(0|0)	3.07(0.5|0)
200	200	3(0|0)	3(0|0)	2.95(0|5.1)	3(0|0)	3(0|0)	3(0|0.3)	3(0|0)	3(0|0)	3.01(0.1|0)

See footnotes to Table A.13.

A.8 Alternative Selection Criteria for the Number of Global Factors

A few studies have attempted to develop consistent selection criteria for the number of global factors under the multilevel setting. Here we provide the detailed estimation algorithms for the alternative approaches proposed by Chen (2012), Andreou et al. (2019) and Han (2021).

First, Chen (2012) proposes the modified information criteria as follows:

$$IC_{Chen}(k_0, k_1, \ldots, k_R) = \sum_{i=1}^{R} \frac{M_i}{N} V_i \left(\widehat{\mathbf{G}}^{k_0}, \widehat{\mathbf{F}}_i^{k_i}, k_0, k_i \right) + \hat{\sigma} \left(\sum_{i=1}^{R} k_i / R + \bar{h} + \bar{\alpha} k_0 \right) g(N, T)$$

$$(\hat{r}_0, \hat{r}_1, \ldots, \hat{r}_R) = \underset{0 \le k_0, k_1, \ldots, k_R \le r_{\max}}{\arg\min} IC_{Chen}(k_0, k_1, \ldots, k_R)$$

where M_i is the number of the individuals in each block i, V_i is the sum of squared residuals for each block i for $i = 1, \ldots, R$, $\widehat{\mathbf{G}}^{k_0}$ are the k_0 estimated global factors, $\widehat{\mathbf{F}}_i^{k_i}$ are the k_i estimated local factors, $\hat{\sigma}$ is the average variance of the idiosyncratic errors, $\bar{h}$ and $\bar{\alpha} < 1$ are fixed scaling parameters given by[8]

$$\bar{h} = \sum_{i=1}^{R} h(\alpha_i), \;\; h(\alpha_i) = \frac{\alpha_i g(N, T)}{\underline{\alpha} g(N, T)}, \;\; \alpha_i = \frac{M_i}{N} \text{ and } \underline{\alpha} = \min\{\alpha_1, \ldots, \alpha_R\}.$$

and $g(N, T)$ is a penalty function given by

$$g(N, T) = \frac{N + T}{NT} \ln \frac{NT}{N + T}$$

When implementing IC_{chen} in the simulations, we simply assume that the true number of factors $r_0 + r_i$ is known in order to avoid estimating too many candidate models.[9]

Next, Andreou et al. (2019) (AGGR) apply the canonical correlation analysis to estimate global and local factors in a two-group factor model with mixed frequency data. They then develop a novel inference on the numbers of global and group-specific factors. AGGR apply the existing information criteria to the data in each block and obtain $\widehat{r_0 + r_i}$ for $i = 1, 2$. Set $r_{\max} = \min \left\{ \widehat{r_0 + r_1}, \widehat{r_0 + r_2} \right\}$ and $M = \min\{M_1, M_2\}$. Then, we extract the $r_{\max}$ PCs from the data matrix $\mathbf{Y}_i$, that is denoted $\widehat{\mathbf{K}}_i$ for $i = 1, 2$. Let $\widehat{\mathbf{V}}_{ab} = \widehat{\mathbf{K}}_a' \widehat{\mathbf{K}}_b / T$ be the covariance matrix between $\widehat{\mathbf{K}}_1$ and $\widehat{\mathbf{K}}_2$ for $a, b = 1, 2$. Construct $\widehat{\mathbf{R}} = \widehat{\mathbf{V}}_{11}^{-1} \widehat{\mathbf{V}}_{12} \widehat{\mathbf{V}}_{22}^{-1} \widehat{\mathbf{V}}_{21}$ and $\widehat{\mathbf{R}}^* = \widehat{\mathbf{V}}_{22}^{-1} \widehat{\mathbf{V}}_{21} \widehat{\mathbf{V}}_{11}^{-1} \widehat{\mathbf{V}}_{12}$. Evaluate the $r_{\max} \times k$ matrix $\widehat{\mathbf{W}}_1$ (resp. $\widehat{\mathbf{W}}_2$) which collects the eigenvectors corresponding to k largest eigenvalues of $\widehat{\mathbf{R}}$ in descending

[8] Chen (2012) uses this criterion to jointly determine the (unknown) block memberships. Then, $\bar{h}$ is a function of M_i, N and T controlling the dispersion of blocks. If the membership is known, $\bar{h}$ is a fixed constant.

[9] For example, in the case where $(r_0, r_i) = (2, 2)$, the candidate models are restricted to $\{(0, 4), (1, 3), (2, 2), (3, 1), (4, 0)\}$.

order (resp. $\widehat{\mathbf{R}}^*$). Define the $r_{\max} \times (r_{\max} - k)$ matrix $\widehat{\mathbf{W}}_1^-$ (resp. $\widehat{\mathbf{W}}_2^-$), which collects the $r_{\max} - k$ eigenvectors corresponding to the remaining eigenvalues of $\widehat{\mathbf{R}}$ (resp. $\widehat{\mathbf{R}}^*$) in descending order. The global factors can be estimated by $\widehat{\mathbf{G}} = \widehat{\mathbf{K}}_1 \widehat{\mathbf{W}}_1$ or $\widehat{\mathbf{G}} = \widehat{\mathbf{K}}_2 \widehat{\mathbf{W}}_2$. The local factors can be estimated by $\widehat{\mathbf{F}}_1 = \widehat{\mathbf{K}}_1 \widehat{\mathbf{W}}_1^-$ and $\widehat{\mathbf{F}}_2 = \widehat{\mathbf{K}}_2 \widehat{\mathbf{W}}_2^-$. The factor loadings can be estimated by $\widehat{\mathbf{\Gamma}}_i = \mathbf{Y}_i' \widehat{\mathbf{G}}/T$ and $\widehat{\mathbf{\Lambda}}_i = \mathbf{Y}_i' \widehat{\mathbf{F}}_i / T$ for $i = 1, 2$. Let $\widehat{\mathbf{D}}_i = \widehat{\mathbf{E}}_i' \widehat{\mathbf{E}}_i / T$ and $\widehat{\mathbf{\Theta}}_i = \left[\widehat{\mathbf{\Gamma}}_i, \widehat{\mathbf{\Lambda}}_i \right]$ for $i = 1, 2$, where $\widehat{\mathbf{E}}_i = \mathbf{Y}_i - \widehat{\mathbf{G}} \widehat{\mathbf{\Gamma}}_i' - \widehat{\mathbf{F}}_i \widehat{\mathbf{\Lambda}}_i'$. Define

$$\widehat{\mathbf{\Sigma}}_{u,i} = \left(\frac{\widehat{\mathbf{\Theta}}_i' \widehat{\mathbf{\Theta}}_i}{M_i} \right)^{-1} \left(\frac{\widehat{\mathbf{\Theta}}_i' \widehat{\mathbf{C}}_i \widehat{\mathbf{\Theta}}_i}{M_i} \right) \left(\frac{\widehat{\mathbf{\Theta}}_i' \widehat{\mathbf{\Theta}}_i}{M_i} \right)^{-1}, \; i = 1, 2$$

where $\widehat{\mathbf{C}}_i$ is an $M_i \times M_i$ diagonal matrix with diagonal elements equal to those in $\widehat{\mathbf{D}}_i$.

Then, they construct the following (feasible) scaled and centered test statistic:

$$\tilde{\xi}(k) = M\sqrt{T} \left(\frac{1}{2} \operatorname{tr} \left\{ \widehat{\mathbf{\Sigma}}_U^2 \right\} \right)^{-1/2} \left[\hat{\xi}(k) - k + \frac{1}{2M} \operatorname{tr} \left\{ \widehat{\mathbf{\Sigma}}_U^2 \right\} \right]$$

where $\hat{\xi}(k)$ is the sum of the k largest eigenvalues of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{\Sigma}}_U = (M_2/M_1) \widehat{\mathbf{\Sigma}}_{u,1}^{(kk)} + \widehat{\mathbf{\Sigma}}_{u,2}^{(kk)}$ with (kk) indicating the upper-left (k, k) blocks of the matrix. By imposing the strong assumption that idiosyncratic errors have neither serial nor cross-sectional correlations, AGGR can derive that $\tilde{\xi}(r)$ follows the standard normal distribution asymptotically under the null hypothesis, $r_0 = r$.

The number of global factors can be estimated by applying the $\tilde{\xi}(r)$ test sequentially for $r = r_{\max}, r_{\max} - 1, \ldots, 1$ backwards:

$$\hat{r}_0 = \max \left\{ r : 1 \leq r \leq r_{\max} : \tilde{\xi}(r) > z_{\alpha_{MT}} \right\}$$

where α_{MT} is a sequence of real scalars defined in the interval $(0, 1)$ such that $\alpha_{MT} \to 0$ and $\left(M\sqrt{T} \right)^{-1} z_{\alpha_{MT}} \to 0$ as $M, T \to \infty$. $z_{\alpha_{MT}}$ is a threshold value given by $z_{\alpha_{MT}} = -c \left(M\sqrt{T} \right)^{\gamma}$ for $c > 0$ and $0 < \gamma < 1$.

The above procedure can be used for model selection if the critical value diverges at a certain rate, γ. Following AGGR (see Footnote 7 on p.1277), we set $c = 0.95$ and $\gamma = 0.1$ in simulations. We also assume that the number of factors $r_0 + r_i$ is known. If we use the estimates of $r_0 + r_i$, then we find that the finite sample performance of *AGGR* becomes worse.

Han (2021) proposes an alternative method to identify the spaces spanned by global and local factors separately, and develops a shrinkage estimator that can consistently estimate the factor loadings and determine the number of factors, simultaneously. We describe Han's algorithms in details.

1. Let $\widetilde{\mathbf{F}}$ be $\sqrt{T}$ times the eigenvectors corresponding to the $\bar{k}$ largest eigenvalues of the $T \times T$

matrix $\mathbf{YY}'$ in descending order. Compute $\tilde{\mathbf{\Lambda}} = \mathbf{Y}'\widetilde{\mathbf{F}}/T$. Let $\widetilde{\mathbf{\Lambda}}_i$ ($M_i \times \bar{k}$) be the i-th block of $\widetilde{\mathbf{\Lambda}}$. Let $\widetilde{\mathbf{S}}_1 = \widetilde{\mathbf{\Lambda}}_1'\widetilde{\mathbf{\Lambda}}_1/M_1$ and then compute the spectral decomposition:

$$\widetilde{\mathbf{S}}_1 = \widetilde{\mathbf{A}}_1\widetilde{\mathbf{V}}_1\widetilde{\mathbf{A}}_1' \tag{A.8.13}$$

where $\widetilde{\mathbf{V}}_1$ is a diagonal matrix consisting of the eigenvalues of $\widetilde{\mathbf{S}}_1$ in descending order and $\widetilde{\mathbf{A}}_1$ denotes the corresponding eigenvectors. Let $\breve{\mathbf{F}} = \widetilde{\mathbf{F}}\widetilde{\mathbf{A}}_1$ and $\breve{\mathbf{\Lambda}} = \widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{A}}_1$.

2. The estimate of $\breve{\mathbf{\Lambda}}_1$, denoted $\widehat{\mathbf{\Lambda}}_1$, can be obtained by minimising the LASSO function:

$$\mathbb{Q}_1(\mathbf{\Lambda}_1) = \frac{1}{M_1T}\sum_{t=1}^{T}\left\|\mathbf{Y}_{1t} - \mathbf{\Lambda}_1\breve{\mathbf{f}}_t\right\|^2 + \frac{\gamma_1}{\delta_{NT}^2}\sum_{k=1}^{\bar{k}} w_{1j}\|\mathbf{\Lambda}_{1k}\| \tag{A.8.14}$$

where $\breve{\mathbf{f}}_t$ is the transpose of the t-th row of $\breve{\mathbf{F}}$, $\mathbf{\Lambda}_{1k}$ is an $M_1 \times 1$ vector of the k-th column of $\mathbf{\Lambda}_{1k}$, $\delta_{NT}^2 = \min\{N, T\}$, $\gamma_{1,NT}$ is a positive tuning parameter depending on N and T, and w_{1k} is an adaptive weight. Let $\hat{k}_1$ be the number of nonzero columns in $\widehat{\mathbf{\Lambda}}_1$.

3. Partition $\breve{\mathbf{\Lambda}}$ as $\left[\breve{\mathbf{\Lambda}}_1', \ldots, \breve{\mathbf{\Lambda}}_R'\right]'$ where $\breve{\mathbf{\Lambda}}_i'$ is an $M_i \times \bar{k}$ loading matrix of the i-th block for $i = 1, \ldots, R$. Let $\breve{\mathbf{\Lambda}}_{i,1:\hat{k}_1}$ and $\breve{\mathbf{\Lambda}}_{i,\hat{k}_1+1:\bar{k}}$ be the first $\hat{k}_1$ columns and the last $\bar{k} - \hat{k}_1$ columns of $\breve{\mathbf{\Lambda}}_i$, respectively. Define

$$\breve{\mathbf{S}}_0 = (N - M_1)^{-1}\sum_{i=2}^{R}\breve{\mathbf{\Lambda}}_{i,1:\hat{k}_1}'\breve{\mathbf{\Lambda}}_{i,1:\hat{k}_1}$$

$$\widetilde{\mathbf{S}}_i = \frac{1}{M_i}\breve{\mathbf{\Lambda}}_{i,\hat{k}_1+1:\bar{k}}'\breve{\mathbf{\Lambda}}_{i,\hat{k}_1+1:\bar{k}},\ i = 2, \ldots, R.$$

Evaluate the spectral decompositions, $\breve{\mathbf{S}}_0 = \breve{\mathbf{A}}_0\breve{\mathbf{V}}_0\breve{\mathbf{S}}_0'$ and $\widetilde{\mathbf{S}}_i = \widetilde{\mathbf{A}}_i\widetilde{\mathbf{V}}_i\widetilde{\mathbf{A}}_i'$ for $i = 2, \ldots, R$. Let $\widehat{\mathbf{G}} = \breve{\mathbf{F}}_{1:\hat{k}}\breve{\mathbf{A}}_0$ and $\widehat{\mathbf{F}}_i = \breve{\mathbf{F}}_{\hat{k}_1+1:\bar{k}}\widetilde{\mathbf{A}}_i$ for $i = 2, \ldots, R$.

4. Let $\mathbf{\Lambda}^{(-1)} = [\mathbf{\Lambda}_2', \ldots, \mathbf{\Lambda}_R']'$. This factor loading matrix can be estimated by minimising:

$$\begin{aligned}\mathbb{Q}\left(\mathbf{\Lambda}^{(-1)}\right) = & \frac{1}{(N - M_1)T}\sum_{i=2}^{R}\sum_{t=1}^{T}\left\|\mathbf{Y}_{it} - \mathbf{\Lambda}_{i,1:\hat{k}_1}\widehat{\mathbf{G}}_t - \mathbf{\Lambda}_{i,\hat{k}_1+1:\bar{k}}\widehat{\mathbf{F}}_{it}\right\|^2 \\ & + \frac{\gamma_0}{\delta_{NT}^2}\sum_{k=1}^{\hat{k}_1} w_{0k}\left\|\mathbf{\Lambda}_k^{(-1)}\right\| + \frac{1}{\delta_{NT}^2}\sum_{i=2}^{R}\gamma_i\sum_{k=\hat{k}+1}^{\bar{k}} w_{ik}\|\mathbf{\Lambda}_{ik}\|\end{aligned} \tag{A.8.15}$$

where $\widehat{\mathbf{G}}_t$ and $\widehat{\mathbf{F}}_{it}$ denote the transpose of the t-th rows of $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{F}}$, $\mathbf{\Lambda}_k^{(-1)}$ is the k-th column of $\mathbf{\Lambda}^{(-1)}$, $\mathbf{\Lambda}_{ik}$ is the k-th column of the $M_i \times \hat{k}$ matrix $\mathbf{\Lambda}_i$ for $i = 2, \ldots, R$, $\gamma_{0,NT}$ and $\gamma_{i,NT}$ are positive tuning parameters, and w_{0k} and w_{ik} are adaptive weights.

Let $\widehat{\mathbf{\Lambda}}^{(-1)} = [\widehat{\mathbf{\Lambda}}_2', \ldots, \widehat{\mathbf{\Lambda}}_R']'$, where $\widehat{\mathbf{\Lambda}}_i$ denotes the estimate of the $M_i \times \bar{k}$ loadings for the i-th block ($i = 2, \ldots, R$). Then, we can estimate the number of global factors by the number of nonzero columns in $\mathbf{\Lambda}^{(-1)}_{1:\hat{k}_1}$, denoted $\hat{r}_0$. Similarly, we estimate the number of local factors by the number of nonzero columns in $\widehat{\mathbf{\Lambda}}_{i,\hat{k}+1:\bar{k}}$, denoted $\hat{r}_i$, for $i = 2, \ldots, R$. Finally, we obtain $\hat{r}_1 = \hat{k}_1 - \hat{r}_0$.

Notice that the above shrinkage estimator requires us to select tuning parameters by some information criteria. The tuning parameter for the first block, γ_1 can be selected by minimising the following information criterion:

$$\mathcal{T}_1(\gamma_1) = \ln\left(\frac{1}{M_1T}\sum_{t=1}^{T}\left\|\mathbf{Y}_{1t} - \widehat{\mathbf{\Lambda}}_1\breve{\mathbf{f}}_t\right\|^2\right) + \hat{k}_1(\gamma_1)\frac{M_1+T}{M_1T}\ln\left(\frac{M_1T}{M_1+T}\right)\cdot\ln[\ln(M_1)].$$

Let γ_1^* be the selected tuning parameter and denote the estimate of k_1 based on γ_1^* by $\hat{k}_1(\gamma_1^*)$. To select the tuning parameters for rest of the blocks, Han proposes the following information criterion:

$$\begin{aligned}\mathcal{T}_2(\gamma_0, \gamma_2, \ldots, \gamma_R) = \ln&\left(\frac{1}{(N-M_1)T}\sum_{i=2}^{R}\sum_{t=1}^{T}\left\|\mathbf{Y}_{it} - \widehat{\mathbf{\Lambda}}_{i,1:\hat{k}_1}\widehat{\mathbf{G}}_t - \widehat{\mathbf{\Lambda}}_{i,\hat{k}_1+1:\bar{k}}\widehat{\mathbf{F}}_{it}\right\|^2\right)\\ &+\left(\hat{r}_0(\gamma_0,\gamma_i,\gamma_1^*) + \sum_{i=2}^{R}\hat{\gamma}_i(\gamma_0,\gamma_i,\gamma_1^*)\right)\left(\frac{\bar{N}+T}{\bar{N}T}\right)\ln\left(\frac{\bar{N}T}{\bar{N}+T}\right)\ln[\ln(\bar{N})].\end{aligned}$$

Han's approach imposes different penalty terms for different blocks. Consequently, even when the number of blocks, R is (mildly) large, there will be a large number of candidate tuning parameters to be selected coherently.

Moreover, the shrinkage estimation results are not invariant to the order of the blocks. Hence, the selection of the appropriate first group is crucially important in practice (see the empirical applications in Han (2021)). To deal with this issue, he proposes an additional information criteria to determine which block is ordered first as follows:

$$IC(s) = \ln(SSR_s) + \left(\hat{r}_0 + \sum_{i=1}^{R}\hat{r}_i\right)\left(\frac{\bar{N}+T}{\bar{N}T}\right)\ln\left(\frac{\bar{N}T}{\bar{N}+T}\right)\ln(\ln(\bar{N})), \text{ for } s = 1, \ldots, R$$

where SSR_s is the sum of squared residual when the s-th block is ordered first and $\bar{N} = \sum_i M_i/R$. This further increases the computational burden.

In the simulation study, Han only considers the sample sizes by combining $R \in \{2, 3, 4, 5\}$, $M \in \{50, 100, 150, 200\}$ and $T \in \{100, 200\}$. However, if R is sufficiently large,[10] there are too many candidate models to be estimated along with too many combinations of the tuning parameters selection, which is computationally too heavy.

More importantly, to derive the consistency of the shrinkage estimator in Theorem 2, he has to assume that the local factors are mutually uncorrelated,[11] though idiosyncratic errors are still

[10] Indeed, Han (2021) focus the model with a fixed number of groups.

[11] He acknowledges that it is challenging to develop a shrinkage estimator fully robust to between-group correlations among group-specific factors.

allowed to be cross sectionally and serially correlated. He addressed this issue by conducting the additional simulations (see Table 5), from which we find that his approach severely overestimates (underestimates) the number of global (local) factors, if the correlation between the local factors is as small as 0.1.

A.9 Proofs of Lemmas 1∗–3∗

Proof of Lemma 1∗. For any two groups m and h, the population covariance between $\mathbf{K}_{mt}$ and $\mathbf{K}_{ht}$ can be expressed as

$$Var\begin{pmatrix}\mathbf{K}_{mt}\\ \mathbf{K}_{ht}\end{pmatrix}=\begin{bmatrix}\mathbf{\Sigma}_{mm} & \mathbf{\Sigma}_{mh}\\ \mathbf{\Sigma}_{hm} & \mathbf{\Sigma}_{hh}\end{bmatrix}=\begin{bmatrix}\mathbf{\Sigma}_G & \mathbf{0} & \mathbf{\Sigma}_G & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma}_{F_m} & \mathbf{0} & \mathbf{\Sigma}_{F_{mh}}\\ \mathbf{\Sigma}_G & \mathbf{0} & \mathbf{\Sigma}_G & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma}_{F_{hm}} & \mathbf{0} & \mathbf{\Sigma}_{F_h}\end{bmatrix} \tag{A.9.16}$$

where $\mathbf{\Sigma}_G$, $\mathbf{\Sigma}_{F_m}$ and $\mathbf{\Sigma}_{F_h}$ are defined in Assumption C, $\mathbf{\Sigma}_{F_{mh}}=E\left(\mathbf{F}_{mt}\mathbf{F}'_{ht}\right)$ and $\mathbf{\Sigma}_{F_{hm}}=\mathbf{\Sigma}'_{F_{mh}}$. We assume $r_m \le r_h$. Using (A.9.16), we can express the characteristic equation,

$$\left(\mathbf{\Sigma}_{mh}\mathbf{\Sigma}_{hh}^{-1}\mathbf{\Sigma}_{hm}-\rho\mathbf{\Sigma}_{mm}\right)\mathbf{v}=0 \tag{A.9.17}$$

as

$$\begin{bmatrix}\mathbf{\Sigma}_G-\rho\mathbf{\Sigma}_G & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma}_{F_{mh}}\mathbf{\Sigma}_{F_h}^{-1}\mathbf{\Sigma}_{F_{hm}}-\rho\mathbf{\Sigma}_{F_m}\end{bmatrix}\mathbf{v}=\mathbf{0},$$

where $\rho_{mh,r}$ is the r-th largest squared canonical correlation between $\mathbf{K}_m$ and $\mathbf{K}_h$. Here, $\rho_{mh,1}=\cdots=\rho_{mh,r_0}=1$ are the characteristic roots with multiplicity r_0. From the lower block,

$$\mathbf{\Sigma}_{F_{mh}}\mathbf{\Sigma}_{F_h}^{-1}\mathbf{\Sigma}_{F_{hm}}-\rho\mathbf{\Sigma}_{F_m},$$

we obtain $1>\rho_{mh,r_0+1}\ge ... \ge \rho_{mh,r_0+r_m}\ge 0$. The characteristic vector corresponding to the r_0 largest eigenvalue is $\mathbf{v}_r=[0,\dots,0,1,0,\dots,0]$, which is the unit vector with the rth element being 1 and zeros otherwise.

In practice, we should deal with the rotation matrix and the selection of r_{max} in the PC estimation. As $r_{max}\ge r_0+r_i$ for all i by construction, the rotation matrices, $\mathbf{H}_m$ and $\mathbf{H}_h$, makes the standard CCA inapplicable to the rotated factors, $\mathbf{H}'_m\mathbf{K}_{mt}$ and $\mathbf{H}'_h\mathbf{K}_{ht}$, because their variance/covariance matrices will be singular. The variance-covariance matrix for the rotated factors $\mathbf{H}'_m\mathbf{K}_{mt}$ and $\mathbf{H}'_h\mathbf{K}_{ht}$, becomes:

$$\text{Var}\left(\begin{bmatrix}\mathbf{H}'_m\mathbf{K}_{mt}\\ \mathbf{H}'_h\mathbf{K}_{ht}\end{bmatrix}\right)=\begin{bmatrix}\mathbf{H}'_m\mathbf{\Sigma}_{mm}\mathbf{H}_m & \mathbf{H}'_m\mathbf{\Sigma}_{mh}\mathbf{H}_h\\ \mathbf{H}'_h\mathbf{\Sigma}_{hm}\mathbf{H}_m & \mathbf{H}'_h\mathbf{\Sigma}_{hh}\mathbf{H}_h\end{bmatrix} \tag{A.9.18}$$

where both $\mathbf{H}_m'\mathbf{\Sigma}_{mm}\mathbf{H}_m$ and $\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h$ are singular. Consider the characteristic equation between the rotated factors as

$$\left[\mathbf{H}_m'\mathbf{\Sigma}_{mh}\mathbf{H}_h\left(\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h\right)^{-}\mathbf{H}_h'\mathbf{\Sigma}_{hm}\mathbf{H}_m - \rho\mathbf{H}_m'\mathbf{\Sigma}_{mm}\mathbf{H}_m\right]\mathbf{u} = \mathbf{0} \tag{A.9.19}$$

where $(\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h)^{-}$ is the Moore-Penrose inverse of $\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h$. Using the property of Moore-Penrose inverse, we have:

$$\mathbf{H}_h\left(\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h\right)^{-}\mathbf{H}_h' = \mathbf{\Sigma}_{hh}^{-1}$$

which holds if $\mathbf{H}_h$ has full row rank. Then (A.9.19) becomes

$$\mathbf{H}_m'\left(\mathbf{\Sigma}_{mh}\mathbf{\Sigma}_{hh}^{-1}\mathbf{\Sigma}_{hm} - \rho\mathbf{\Sigma}_{mm}\right)\mathbf{H}_m\mathbf{u} = \mathbf{0}. \tag{A.9.20}$$

Using (A.9.16), we rewrite (A.9.20) as

$$\mathbf{H}_m'\begin{bmatrix}\mathbf{\Sigma}_G - \rho\mathbf{\Sigma}_G & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma}_{F_{mh}}\mathbf{\Sigma}_{F_h}^{-1}\mathbf{\Sigma}_{F_{hm}} - \rho\mathbf{\Sigma}_{F_m}\end{bmatrix}\mathbf{H}_m\mathbf{u} = \mathbf{0}$$

which shows that both (A.9.17) and (A.9.19) will produce the same non-zero eigenvalues. This implies that the rotation matrices do not alter the non-zero canonical correlation.

We now consider the following spectral decompositions:

$$\mathbf{H}_m'\mathbf{\Sigma}_{mm}\mathbf{H}_m = \mathbf{P}\mathbf{\Delta}_m\mathbf{P}' \text{ and } \mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h = \mathbf{Q}\mathbf{\Delta}_h\mathbf{Q}'$$

where $\mathbf{\Delta}_m(\mathbf{\Delta}_h)$ is a diagonal matrix of eignevalues of $\mathbf{H}_m'\mathbf{\Sigma}_{mm}\mathbf{H}_m(\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h)$, $\mathbf{P}(\mathbf{Q})$ is an orthogonal matrix whose columns are standardized eigenvectors associated with the diagonal entries of $\mathbf{\Delta}_m(\mathbf{\Delta}_h)$. As the rank of $\mathbf{H}_m'\mathbf{\Sigma}_{mm}\mathbf{H}_m(\mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h)$ is $r_0+r_m \leq r_{\max}(r_0+r_h \leq r_{\max})$ asymptotically, we rewrite the above equation as:

$$\begin{aligned}\mathbf{H}_m'\mathbf{\Sigma}_{mm}\mathbf{H}_m &= \begin{bmatrix}\mathbf{P}_1 & \mathbf{P}_2\end{bmatrix}\begin{bmatrix}\mathbf{\Delta}_1^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{P}_1 & \mathbf{P}_2\end{bmatrix}'\\ \mathbf{H}_h'\mathbf{\Sigma}_{hh}\mathbf{H}_h &= \begin{bmatrix}\mathbf{Q}_1 & \mathbf{Q}_2\end{bmatrix}\begin{bmatrix}\mathbf{\Delta}_2^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{Q}_1 & \mathbf{Q}_2\end{bmatrix}'\end{aligned} \tag{A.9.21}$$

where $\mathbf{P}_1$ and $\mathbf{P}_2$ are $r_{\max} \times (r_0 + r_m)$ and $r_{\max} \times [r_{\max} - (r_0 + r_m)]$ orthogonal matrices, and similarly for $\mathbf{Q}_1$ and $\mathbf{Q}_2$. Consider the $(r_0 + r_m) \times (r_0 + r_h)$ matrix, $\mathbf{\Delta}_1^{-1}\mathbf{P}_1'\left(\mathbf{H}_m'\mathbf{\Sigma}_{mh}\mathbf{H}_h\right)\mathbf{Q}_1\mathbf{\Delta}_2^{-1}$, whose singular value decomposition is given by (see Rao (1981))

$$\mathbf{\Delta}_1^{-1}\mathbf{P}_1'\left(\mathbf{H}_m'\mathbf{\Sigma}_{mh}\mathbf{H}_h\right)\mathbf{Q}_1\mathbf{\Delta}_2^{-1} = \mathbf{W}\begin{bmatrix}\mathbf{R}^{1/2} & \mathbf{0}\end{bmatrix}\mathbf{D}' \tag{A.9.22}$$

where $\mathbf{W}$ is an $(r_0+r_m)\times(r_0+r_m)$ orthonormal matrix, $\mathbf{D}$ an $(r_0+r_h)\times(r_0+r_h)$ orthonormal matrix and $\mathbf{R}$ the $(r_0 + r_m) \times (r_0 + r_m)$ diagonal matrix given by $\mathbf{R} = \text{diag}(\rho_1, ..., \rho_{r_0}, \rho_{r_0+1}, ..., \rho_{r_0+r_m}) = \text{diag}(1, ..., 1, \rho_{mh,r_0+1}, ..., \rho_{mh,r_0+r_m})$.[12]

Define the full rank matrices,

$$\mathbf{A} = \left[\mathbf{P}_1\boldsymbol{\Delta}_1^{-1}\mathbf{W}, \mathbf{P}_2\right] \text{ and } \mathbf{B} = \left[\mathbf{Q}_1\boldsymbol{\Delta}_2^{-1}\mathbf{D}, \mathbf{Q}_2\right] \tag{A.9.23}$$

Combining (A.9.21), (A.9.22) and (A.9.23), it is straightforward to show that

$$\text{Var}\left(\begin{bmatrix}\mathbf{A}'\mathbf{H}_m'\mathbf{K}_{mt}\\ \mathbf{B}'\mathbf{H}_h'\mathbf{K}_{ht}\end{bmatrix}\right) = \begin{bmatrix}\mathbf{I}_{r_0+r_m} & \mathbf{0} & \mathbf{R}^{1/2} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{R}^{1/2} & \mathbf{0} & \mathbf{I}_{r_0+r_m} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{r_h-r_m} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix} \tag{A.9.24}$$

From (A.9.24), the characteristic equation between $\mathbf{A}'\mathbf{H}_m'\mathbf{K}_{mt}$ and $\mathbf{B}'\mathbf{H}_h'\mathbf{K}_{ht}$ is

$$\left[\mathbf{A}'\mathbf{H}_m'\boldsymbol{\Sigma}_{mh}\mathbf{H}_h\mathbf{B}\left(\mathbf{B}'\mathbf{H}_h'\boldsymbol{\Sigma}_{hh}\mathbf{H}_h\mathbf{B}\right)^{-}\mathbf{B}'\mathbf{H}_h'\boldsymbol{\Sigma}_{hm}\mathbf{H}_m\mathbf{A} - \rho\mathbf{A}'\mathbf{H}_m'\boldsymbol{\Sigma}_{mm}\mathbf{H}_m\mathbf{A}\right]\mathbf{u} = \mathbf{0}$$

which can be simplified as

$$\left(\begin{bmatrix}\mathbf{R}^{1/2} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{I}_{r_0+r_m} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{r_h-r_m} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{R}^{1/2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} - \rho\begin{bmatrix}\mathbf{I}_{r_0+r_m} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\right)\mathbf{u} = \mathbf{0}$$

Hence,

$$\left(\begin{bmatrix}\mathbf{R} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} - \rho\begin{bmatrix}\mathbf{I}_{r_0+r_m} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\right)\mathbf{u} = \mathbf{0} \tag{A.9.25}$$

Obviously, (A.9.25) has the same characteristic roots as (A.9.19) and the same non-zero characteristic roots in (A.9.17) consequently.

Now, we consider the sample covariance matrix for $\widetilde{\mathbf{K}}_m$ and $\widetilde{\mathbf{K}}_h$ given by

$$\text{Var}\begin{pmatrix}\widetilde{\mathbf{K}}_m\\ \widetilde{\mathbf{K}}_h\end{pmatrix} = \frac{1}{T}\begin{bmatrix}\widetilde{\mathbf{K}}_m'\widetilde{\mathbf{K}}_m & \widetilde{\mathbf{K}}_m'\widetilde{\mathbf{K}}_h\\ \widetilde{\mathbf{K}}_h'\widetilde{\mathbf{K}}_m & \widetilde{\mathbf{K}}_h'\widetilde{\mathbf{K}}_h\end{bmatrix} = \begin{bmatrix}\widetilde{\mathbf{S}}_{mm} & \widetilde{\mathbf{S}}_{mh}\\ \widetilde{\mathbf{S}}_{hm} & \widetilde{\mathbf{S}}_{hh}\end{bmatrix}$$

Consider the full rank transformation $\widetilde{\mathbf{K}}_m\mathbf{A}$ and $\widetilde{\mathbf{K}}_h\mathbf{B}$, where $\mathbf{A}$ and $\mathbf{B}$ are defined in (A.9.23). The canonical correlations between them are equivalent to those between $\widetilde{\mathbf{K}}_m$ and $\widetilde{\mathbf{K}}_h$. From Lemma 5, we obtain: $\mathbf{A}'\widetilde{\mathbf{S}}_{mm}\mathbf{A} \xrightarrow{p} \mathbf{A}'\mathbf{H}_m'\boldsymbol{\Sigma}_{mm}\mathbf{H}_m\mathbf{A}$, $\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B} \xrightarrow{p} \mathbf{B}'\mathbf{H}_h'\boldsymbol{\Sigma}_{hh}\mathbf{H}_h\mathbf{B}$ and $\mathbf{A}'\widetilde{\mathbf{S}}_{mh}\mathbf{B} \xrightarrow{p} \mathbf{A}'\mathbf{H}_m'\boldsymbol{\Sigma}_{mh}\mathbf{H}_h\mathbf{B}$. Let $\underline{M}$ be $\min\{M_m, M_h\}$. Applying (A.9.24) and Lemma 5, we can rewrite these

[12] Notice that $\mathbf{R}$ contains the same non-zero roots as in (A.9.19), see Rao (1981).

transformed variance/covariance matrices as:

$$\mathbf{A}'\widetilde{\mathbf{S}}_{mm}\mathbf{A} = \begin{bmatrix} \mathbf{I}_{r_0+r_m} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \end{bmatrix}$$

$$\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B} = \begin{bmatrix} \mathbf{I}_{r_0+r_m} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & \mathbf{I}_{r_h-r_m} + O_p(\delta_{\underline{M}T}^{-2}) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \end{bmatrix}$$

and

$$\mathbf{A}'\widetilde{\mathbf{S}}_{mh}\mathbf{B} = \begin{bmatrix} \mathbf{R}^{1/2} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \end{bmatrix}$$

Notice that the Moore-Penrose inverse of the lower $[r_{\max} - (r_0 + r_m)] \times [r_{\max} - (r_0 + r_m)]$ block of $\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B}$ does not converge to $\begin{bmatrix} \mathbf{I}_{r_h-r_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ as a result of

$$\text{rank}\left(\begin{bmatrix} \mathbf{I}_{r_h-r_m} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \end{bmatrix}\right) \neq \text{rank}\left(\begin{bmatrix} \mathbf{I}_{r_h-r_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\right)$$

For the same reason, $\left(\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B}\right)^{-}$ do not converge to $\begin{bmatrix} \mathbf{I}_{r_0+r_m} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{r_h-r_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$. See Theorem 1 in Karabiyik et al. (2017). But, the Moore-Penrose inverse follows the Banachiewicz-Schur form and thus we have

$$\begin{bmatrix} \mathbf{I}_{r_h-r_m} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \end{bmatrix}^{-} = \begin{bmatrix} \mathbf{I}_{r_h-r_m} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & -O_p(1) \\ -O_p(1) & O_p\left(\delta_{\underline{M}T}^{2}\right) \end{bmatrix} = O_p\left(\delta_{\underline{M}T}^{2}\right) \quad \text{(A.9.26)}$$

as shown in Tian & Takane (2009) and Castro-González et al. (2015). Again, $\left(\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B}\right)^{-}$ also follows the Banachiewicz-Schur form which leads to

$$\left(\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B}\right)^{-} = \begin{bmatrix} \mathbf{I}_{r_0+r_m} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & -O_p(1) \\ -O_p(1) & O_p\left(\delta_{\underline{M}T}^{2}\right) \end{bmatrix}. \quad \text{(A.9.27)}$$

Using the above results, we obtain:

$$\mathbf{A}'\widetilde{\mathbf{S}}_{mh}\mathbf{B}\left(\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B}\right)^{-}\mathbf{B}'\widetilde{\mathbf{S}}_{hm}\mathbf{A} =$$

$$\begin{bmatrix} \mathbf{R}^{1/2} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \end{bmatrix} \begin{bmatrix} \mathbf{I}_{r_0+r_m} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & -O_p(1) \\ -O_p(1) & O_p\left(\delta_{\underline{M}T}^{2}\right) \end{bmatrix} \begin{bmatrix} \mathbf{R}^{1/2} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{R} + O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p(\delta_{\underline{M}T}^{-2}) \\ O_p\left(\delta_{\underline{M}T}^{-2}\right) & O_p\left(\delta_{\underline{M}T}^{-2}\right) \end{bmatrix}$$

Therefore, the characteristic equation between $\widetilde{\mathbf{K}}_m\mathbf{A}$ and $\widetilde{\mathbf{K}}_h\mathbf{B}$

$$\left[\mathbf{A}'\widetilde{\mathbf{S}}_{mh}\mathbf{B}\left(\mathbf{B}'\widetilde{\mathbf{S}}_{hh}\mathbf{B}\right)^{-}\mathbf{B}'\widetilde{\mathbf{S}}_{hm}\mathbf{A} - \ell\mathbf{A}'\widetilde{\mathbf{S}}_{mm}\mathbf{A}\right]\boldsymbol{\xi} = \mathbf{0}$$

can be rewritten as

$$\left(\begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} - \ell \begin{bmatrix} \mathbf{I}_{r_0+r_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + O_p\left(\delta_{\underline{M}T}^{-2}\right)\right)\boldsymbol{\xi} = \mathbf{0}$$

which is analogous to (A.9.25) with a small perturbation term. Finally, by the continuity of the characteristic roots, we have $\ell_{mh,r} \xrightarrow{p} 1$ for $r = 1, \ldots, r_0$, $\ell_{mh,r} \xrightarrow{p} \rho_{mh,r}$ for $r = r_0+1, \ldots, r_0+r_m$ and $\ell_{mh,r} \xrightarrow{p} 0$ for $r = r_0 + r_m + 1, \ldots, r_{\max}$ as $T, M_m, M_h \to \infty$.

Q.E.D

Proof of Lemma 2∗. This lemma follows directly by Lemma 1∗ and the definition of $\xi(r)$ and $\bar{\rho}_r$.

Q.E.D

Proof of Lemma 3∗. This lemma follows directly from Lemma 2∗ and the definition of $CCD(r)$.

Q.E.D

Appendix B

Appendix to Chapter 2

B.1 Lemmas and Proofs

In Section B.1.1, we state some auxiliary lemmas and show the proofs of the results established in Section 2.4. Section B.1.2 presents the proofs of the auxiliary lemmas. We use the following facts throughout the proofs. By Assumption 2.B.1, we have: $\left\|T^{-1/2}\mathbf{G}\right\| = O_p(1)$ and $\left\|T^{-1/2}\mathbf{F}_i\right\| = O_p(1)$ for all $i = 1, ..., R$. By Assumptions 2.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 $a\boldsymbol{\Sigma}\mathbf{V} = a\lambda\mathbf{V}$ where $\mathbf{V}$ is the eigenvector associated with the eigenvalue λ and a is a non-zero real number.

B.1.1 Proofs of the main results

Proof of Proposition 2.1.

Using $\mathbf{K}_i = [\mathbf{G}, \mathbf{F}_i]$ for $i = 1, ..., R$, we can be express the matrix $\boldsymbol{\Phi}$ in (2.3.13) as

$$\boldsymbol{\Phi} = \begin{bmatrix} \mathbf{G} & \mathbf{F}_1 & -\mathbf{G} & -\mathbf{F}_2 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{G} & \mathbf{F}_1 & \mathbf{0} & \mathbf{0} & -\mathbf{G} & -\mathbf{F}_3 & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & & & & & & \vdots & & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{G} & \mathbf{F}_{R-1} & -\mathbf{G} & -\mathbf{F}_R \end{bmatrix}$$

Let

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

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

$$\mathbf{K}_i\mathbf{Q}_i^{r_0} = [\mathbf{G}, \mathbf{F}_i] \begin{bmatrix} \frac{1}{\sqrt{R}}\mathbf{A} \\ \mathbf{0} \end{bmatrix} = \mathbf{G}\mathbf{B} \tag{B.1.1}$$

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

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

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

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

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

It follows that

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

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

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

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

Q.E.D

Proof of Proposition 2.2.

Using (B.1.1) we obtain:

$$\frac{1}{\sqrt{T}}\mathbf{\Psi} = \frac{1}{\sqrt{T}}\left[\mathbf{K}_1\mathbf{Q}_1^{r_0}, \ldots, \mathbf{K}_R\mathbf{Q}_R^{r_0}\right] = \frac{1}{\sqrt{T}}\left[\mathbf{GB}, \ldots, \mathbf{GB}\right] \tag{B.1.2}$$

which yields

$$\frac{\mathbf{\Psi\Psi}'}{T} = \frac{\mathbf{GG}'}{T} = \mathbf{L\Xi L}'$$

where $\mathbf{\Xi}$ is a diagonal matrix with the first r_0 non-zero elements and the remaining zero elements.

Finally, it follows that

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

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 2.1.

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

Q.E.D

Proof of Lemma 2.2.

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

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

where $\widehat{\mathbf{H}} = diag\left\{\widehat{\mathbf{H}}_1, \widehat{\mathbf{H}}_2, \ldots, \widehat{\mathbf{H}}_R\right\}$. Along the same arguments in Proof of Proposition 2.1, we obtain the desired result.

Q.E.D

Proof of Lemma 2.3.

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

Q.E.D

Lemma B.1.1. *Under Assumption 2.A–2.C, as* $N_1, N_2, \ldots, N_R, T \longrightarrow \infty$, *we have:*

1. *For every* m *and* h,

$$\frac{1}{T\sqrt{N_h}}\left\|\left(\widehat{\mathbf{K}}_m - \mathbf{K}_m\widehat{\mathbf{H}}_m\right)'\mathbf{e}_h\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right)$$

2. *For each* i,

$$\frac{1}{T\sqrt{N_i}}\left\|\widehat{\mathbf{G}}'\mathbf{e}_i\right\| = O_p\left(\frac{1}{C_{\underline{N}T}}\right)$$

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

Proof of Theorem 2.1.

By Lemma 2.1, we have:

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

Furthermore, by Lemma 2.2 and Lemma 2.3, we obtain:

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

where $\mathbf{D}$ is an $r_0\times r_0$ orthogonal matrix. Then, using the definition $\bar{\mathbf{Q}}^{r_0}=\widehat{\mathbf{H}}_i^{-}\mathbf{Q}^{r_0}$ and (B.1.1), it follows for each i that

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

where the inequalities hold due to the Cauchy-Schwarz inequality, and the last equality follows from Lemma 2.1 and the fact that $\left\|\widehat{\mathbf{Q}}_i^{r_0}\right\|=O_p(1)$ and $\left\|\widehat{\mathbf{H}}_i\right\|=O_p(1)$. Using this convergence rate, we obtain:

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

where the inequality follows from the Cauchy-Schwarz inequality. Applying Lemma 2.3 to the above equation, we obtain:

$$\left\|\widehat{\mathbf{L}}^{r_0}-\mathbf{L}^{r_0}\mathbf{U}\right\|=O_p\left(\frac{1}{C_{\underline{N}T}}\right)\tag{B.1.3}$$

where $\mathbf{U}$ is an $r_0\times r_0$ orthogonal matrix.[1] Finally, by definition of $\widehat{\mathbf{G}}$ and Proposition 2.2, we conclude:

$$\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}-\mathbf{G}\mathbb{H}\right\|=O_p\left(\frac{1}{C_{\underline{N}T}}\right)\tag{B.1.4}$$

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

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

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

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

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

$$\frac{1}{\sqrt{N_i}}\left(\widehat{\mathbf{\Gamma}}_i' - \mathbb{H}^{-1}\mathbf{\Gamma}_i'\right) = \frac{1}{T\sqrt{N_i}}\widehat{\mathbf{G}}'\left(\mathbf{G} - \widehat{\mathbf{G}}\mathbb{H}^{-1}\right)\mathbf{\Gamma}_i' + \frac{1}{T\sqrt{N_i}}\widehat{\mathbf{G}}'\mathbf{F}_i\mathbf{\Lambda}_i' + \frac{1}{T\sqrt{N_i}}\widehat{\mathbf{G}}'\mathbf{e}_i \tag{B.1.5}$$

The first term of RHS is bounded by $O_p(C_{\underline{N}T}^{-1})$ due to (B.1.4). The second term is bounded as

$$\begin{aligned}\left\|\frac{1}{T\sqrt{N_i}}\widehat{\mathbf{G}}'\mathbf{F}_i\mathbf{\Lambda}_i'\right\| &= \left\|\frac{1}{T\sqrt{N_i}}\left(\widehat{\mathbf{G}}' - \mathbf{G}\mathbb{H} + \mathbf{G}\mathbb{H}\right)'\mathbf{F}_i\mathbf{\Lambda}_i'\right\| \\ &\leq \left\|\frac{1}{T\sqrt{N_i}}\left(\widehat{\mathbf{G}}' - \mathbf{G}\mathbb{H}\right)'\mathbf{F}_i\mathbf{\Lambda}_i'\right\| + \left\|\frac{1}{T\sqrt{N_i}}\mathbb{H}'\mathbf{G}'\mathbf{F}_i\mathbf{\Lambda}_i'\right\| \\ &= O_p\left(\frac{1}{C_{\underline{N}T}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) = O_p\left(\frac{1}{C_{\underline{N}T}}\right) \end{aligned} \tag{B.1.6}$$

where the inequality follows from the Cauchy-Schwarz inequality and the second to last equalities use Lemma 2.1 and Assumptioin 2.D. The last term of (B.1.5) is bounded by $O_p\left(C_{\underline{N}T}^{-1}\right)$ due to Lemma B.1.1.2. Then,

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

Q.E.D

Lemma B.1.2. *Under Assumptions 2.A–2.C, as $N_1, N_2, \ldots, N_R, T \longrightarrow \infty$, we have for each $i = 1, \ldots, R$:*

1.

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

2.

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

3.

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

Proof of Theorem 2.2.

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

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

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

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

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

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

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

Next, we study the second term of the above equation. Using the relation that

$$\begin{aligned}\widehat{\mathbf{S}}_i = \mathbf{G}\boldsymbol{\Gamma}_i' - \widehat{\mathbf{G}}\widehat{\boldsymbol{\Gamma}}_i' &= \mathbf{G}\boldsymbol{\Gamma}_i' - \left(\widehat{\mathbf{G}} - \mathbf{G}\mathbb{H} + \mathbf{G}\mathbb{H}\right)\left(\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i' + \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right)\\ &= -\left(\widehat{\mathbf{G}} - \mathbf{G}\mathbb{H}\right)\left(\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right) - \left(\widehat{\mathbf{G}} - \mathbf{G}\mathbb{H}\right)\mathbb{H}^{-1}\boldsymbol{\Gamma}_i' - \mathbf{G}\mathbb{H}\left(\widehat{\boldsymbol{\Gamma}}_i' - \mathbb{H}^{-1}\boldsymbol{\Gamma}_i'\right),\end{aligned} \quad \text{(B.1.7)}$$

we obtain:

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

By Theorem 2.1, it follows that

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

Using (B.1.7), it follows that

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

Therefore, by Theorem 2.1,

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

From (B.1.7) we obtain:

$$\begin{aligned} \frac{1}{\sqrt{T}} \frac{1}{N_i T} \widehat{\mathbf{S}}_i \mathbf{e}_i' \widehat{\mathbf{F}}_i = & - \frac{1}{\sqrt{T}} \frac{1}{N_i T} \left(\widehat{\mathbf{G}} - \mathbf{G}\mathbb{H} \right) \left(\widehat{\mathbf{\Gamma}}_i' - \mathbb{H}^{-1} \mathbf{\Gamma}_i' \right) \mathbf{e}_i' \widehat{\mathbf{F}}_i \\ & - \frac{1}{\sqrt{T}} \frac{1}{N_i T} \left(\widehat{\mathbf{G}} - \mathbf{G}\mathbb{H} \right) \mathbb{H}^{-1} \mathbf{\Gamma}_i' \mathbf{e}_i' \widehat{\mathbf{F}}_i - \frac{1}{\sqrt{T}} \frac{1}{N_i T} \mathbf{G}\mathbb{H} \left(\widehat{\mathbf{\Gamma}}_i' - \mathbb{H}^{-1} \mathbf{\Gamma}_i' \right) \mathbf{e}_i' \widehat{\mathbf{F}}_i \end{aligned}$$

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 2.1 and Lemma 3. The second term is bounded by $N_i^{-1/2} O_p \left(C_{\underline{N}T}^{-1} \right)$ due to Theorem 2.1 and Lemma 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:

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

It is straightforward to show that $\frac{1}{\sqrt{T}} \frac{1}{N_i T} \left\| \mathbf{e}_i \widehat{\mathbf{S}}_i' \widehat{\mathbf{F}}_i \right\|$ has the same stochastic order. Again using (B.1.7):

$$\begin{aligned} \frac{1}{\sqrt{T}} \frac{1}{N_i T} \mathbf{F}_i' \mathbf{\Lambda}_i' \widehat{\mathbf{S}}_i' \widehat{\mathbf{F}}_i = & - \frac{1}{\sqrt{T}} \frac{1}{N_i T} \mathbf{F}_i' \mathbf{\Lambda}_i' \left(\widehat{\mathbf{\Gamma}}_i' - \mathbb{H}^{-1} \mathbf{\Gamma}_i' \right)' \left(\widehat{\mathbf{G}} - \mathbf{G}\mathbb{H} \right)' \widehat{\mathbf{F}}_i \\ & - \frac{1}{\sqrt{T}} \frac{1}{N_i T} \mathbf{F}_i' \mathbf{\Lambda}_i' \mathbf{\Gamma}_i \left(\mathbb{H}^{-1} \right)' \left(\widehat{\mathbf{G}} - \mathbf{G}\mathbb{H} \right)' \widehat{\mathbf{F}}_i - \frac{1}{\sqrt{T}} \frac{1}{N_i T} \mathbf{F}_i' \mathbf{\Lambda}_i' \left(\widehat{\mathbf{\Gamma}}_i' - \mathbb{H}^{-1} \mathbf{\Gamma}_i' \right)' \mathbb{H}' \mathbf{G}' \widehat{\mathbf{F}}_i \end{aligned}$$

Using Theorem 2.1, we obtain:

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

Combining all the results, we conclude that

$$\frac{1}{\sqrt{T}} \left\| \widehat{\mathbf{F}}_i - \mathbf{F}_i \widehat{\mathscr{H}_i} \right\| = O_p \left(\frac{1}{C_{\underline{N}T}} \right). \tag{B.1.8}$$

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

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

Plugging $\mathbf{Y}_i = \mathbf{G}\mathbf{\Gamma}_i' + \mathbf{F}_i \mathbf{\Lambda}_i' + \mathbf{e}_i$, $\mathbf{F}_i = \mathbf{F}_i - \widehat{\mathbf{F}}_i \widehat{\mathscr{H}_i}^{-1} + \widehat{\mathbf{F}}_i \widehat{\mathscr{H}_i}^{-1}$ and (B.1.7) into the above equation, we obtain:

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

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

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 2.1. The fourth term is bounded by $O_p\left(C_{\underline{N}T}^{-1}\right)$ from (B.1.8). The last term can be written as

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

The first term of the above equation 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 $(N_iT)^{-1/2}\left\|\mathbf{F}_i\mathbf{e}_i\right\|=O_p(1)$ under Assumption 2.B.2. Collecting all the results, we conclude that

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

Q.E.D

Proof of Theorem 2.3.

By Lemmas 2.1 and 2.2 and using the continuity of the singular values, we have:

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

First, if $r_0>0$, we have:

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

Next, 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 $C_{\underline{N}T}\to\infty$, we have the desired result that the ratio $\hat{\delta}_{k+1}/\hat{\delta}_k$ attains maximum at $k=r_0$.

Q.E.D

Lemma B.1.3. *Let $C_{N_i,T}=\min\{\sqrt{N_i},\sqrt{T}\}$. Under Assumptions 2.A-2.C and 2.F–2.G, we have:*

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

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

where $\widehat{\mathbf{H}}_i = (\mathbf{\Theta}_i'\mathbf{\Theta}_i/N_i)\left(\mathbf{K}_i'\widehat{\mathbf{K}}_i/T\right)\widehat{\mathbf{V}}_i^{-1}$ *is an* $(r_0+r_i)\times(r_0+r_i)$ *matrix with* $\widehat{\mathbf{V}}_i$ *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. Furthermore,*

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

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

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

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

where $\mathbf{e}_{i.t} = (e_{i1t}, e_{i2t}..., e_{iN_it})'$.

2. *Let* $\widehat{\boldsymbol{\mathcal{R}}}_i = T^{-1/2}\left(\widehat{\mathbf{K}}_i - \mathbf{K}_i\widehat{\mathbf{H}}_i\right)$. *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:* $T^{-1/2}\widehat{\boldsymbol{\mathcal{R}}}_m'\mathbf{K}_h = O_p\left(C_{N_mT}^{-2}\right)$.

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

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

Lemma B.1.4. *Under Assumptions 2.A–2.C and 2.E–2.G, as* $N_1,\ldots,N_R,T\to\infty$, *we have:*

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

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

where $\mathbf{D}$ *is an* $r_0\times r_0$ *orthogonal matrix,* $\mathbf{U}$ *is an* $r_0\times r_0$ *orthogonal matrix defined in (B.1.3), and* $C_{\underline{N},T} = \min\left\{\sqrt{\underline{N}},\sqrt{T}\right\}$ *with* $\underline{N} = \min\{N_1, N_2,\ldots,N_R\}$.

Lemma B.1.5. *Under Assumptions 2.A–2.C and 2.F–2.G, as* $N_i, T\to\infty$, *we have for each* i:

1.

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

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

2.

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

3.

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

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

4.

$$\operatorname*{plim}_{N_i,T\to\infty}\widehat{\mathbf{H}}_i=\mathbf{H}_i$$

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

Proof of Theorem 2.4.

1. From (2.3.19), we have for each t:

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

Using the asymptotic expansions in Lemma B.1.4.1 and Lemma B.1.4.2:

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

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

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

where we use that $(\mathbf{\Xi}^{r_0})^{-1}\mathbf{U}'=\mathbf{U}(\mathbf{\Xi}^{r_0})^{-1}$ because $\mathbf{\Xi}$ is diagonal. Replacing $T^{-1/2}\widehat{\mathbf{K}}_i$ with $T^{-1/2}\mathbf{K}_i\widehat{\mathbf{H}}_i+$

$\widehat{\mathcal{R}}_i$, the above equation can be written as

$$\widehat{\mathbf{G}}_t = \mathbb{H}'\frac{1}{R}\sum_{i=1}^{R}\mathbb{I}_i'\left(\widehat{\mathbf{H}}_i^{-1}\right)'\widehat{\mathbf{K}}_{it} + \mathbf{U}(\boldsymbol{\Xi}^{r_0})^{-1}\mathbf{L}^{r_0\prime}\left[\sum_{i=1}^{R}\widehat{\mathcal{R}}_i\widehat{\mathbf{H}}_i^{-1}\mathbf{Q}_i^{r_0}\mathbf{Q}_i^{r_0\prime}\left(\widehat{\mathbf{H}}_i^{-1}\right)'\widehat{\mathbf{K}}_{it}\right] + O_p\left(\frac{1}{C_{\underline{N}T}^2}\right) \tag{B.1.9}$$

where we use $\mathbf{K}_i\mathbf{Q}_i^{r_0} = \mathbf{GB}$, $\mathbf{B} = R^{-1/2}\mathbf{A}$, $\mathbf{Q}_i^{r_0} = \left[R^{-1/2}\mathbf{A}',\mathbf{0}\right]'$, $\mathbf{BQ}_i^{r_0} = R^{-1}\left[\mathbf{I}_{r_0},\mathbf{0}\right] = R^{-1}\mathbb{I}_i'$ and $\mathbf{A}$ is an orthogonal matrix. From the asymptotic expansion in Lemma 1, it follows that $T^{-1}\sum_{s=1}^{T}\widehat{\mathbf{K}}_{is}\eta_{i,st}$ and $(N_iT)^{-1}\mathbf{e}_i\boldsymbol{\Theta}_i\mathbf{K}_i'\widehat{\mathbf{K}}_i\widehat{\mathbf{V}}_i^{-1}$ are dominant terms in $\widehat{\mathbf{K}}_{it} - \widehat{\mathbf{H}}_i'\mathbf{K}_{it}$ and $\widehat{\mathcal{R}}_i$, respectively. So we have:

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

and

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

Plugging these expressions into (B.1.9) and multiplying both sides by $\sqrt{N}$, we can show that

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

Using $\widehat{\mathbf{H}}_i = (\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i)\left(\mathbf{K}_i'\widehat{\mathbf{K}}_i/T\right)\widehat{\mathbf{V}}_i^{-1}$ from Lemma B.1.3.1 and rearranging terms, the above equation can be reduced to

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

where

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

By Lemmas B.1.2 and B.1.5, it is easily seen that $\mathbb{B} = O_p\left(\underline{N}^{-1/2}\right)$.

Finally, we achieve the desired result that

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

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

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

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

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

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

$$\mathbb{C} \xrightarrow{p} \mathbb{C}^0 = \begin{bmatrix} \alpha_1^{1/2}\mathbb{I}_1'\boldsymbol{\Sigma}_{\Theta_1} & & \\ & \ddots & \\ & & \alpha_R^{1/2}\mathbb{I}_R'\boldsymbol{\Sigma}_{\Theta_R} \end{bmatrix}$$

Therefore,

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

2. For each i and j, we have $\widehat{\boldsymbol{\gamma}}_{ij} = T^{-1}\widehat{\mathbf{G}}'\mathbf{Y}_{ij}$. Using (2.2.5) and $\mathbf{G} = \mathbf{G} - \widehat{\mathbf{G}}\left(\mathbb{H}+\mathbb{B}\right)^{-1} + \widehat{\mathbf{G}}\left(\mathbb{H}+\mathbb{B}\right)^{-1}$, we have

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

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

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

The first term is bounded by $O_p\left(\underline{N}^{-1}\right)$ by Theorem 2.4.1. The second to fourth terms are $O_p\left(C_{\underline{N}T}^{-2}\right)$ by Lemma B.1.6. Then, we obtain:

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

Next, consider the eigen-decomposition, $T^{-1}\mathbf{G}'\mathbf{G} = \mathbf{X}\boldsymbol{\Xi}^{r_0}\mathbf{X}'$. Let ξ_k and $\mathbf{X}_k$ be the k-th eigenvalue and eigenvector pair of the matrix $T^{-1}\mathbf{G}'\mathbf{G}$. Pre-multiplying $\mathbf{G}$ by $T^{-1}\mathbf{G}'\mathbf{G}\mathbf{X}_k = \xi_k\mathbf{X}_k$

yields $T^{-1}\mathbf{G}\mathbf{G}'(\mathbf{G}\mathbf{X}_k) = \xi_k(\mathbf{G}\mathbf{X}_k)$, which shows that $\mathbf{G}\mathbf{X}_k$ is the (non-normalised) eigenvector of $T^{-1}\mathbf{G}\mathbf{G}'$ for $k = 1, \dots, r_0$, where $T^{-1}\mathbf{G}'\mathbf{G}$ and $T^{-1}\mathbf{G}\mathbf{G}'$ have the same nonzero eigenvalues. Moreover, the normalised eigenvector becomes $(\mathbf{X}_k'\mathbf{G}'\mathbf{G}\mathbf{X}_k)^{-1/2}\mathbf{G}\mathbf{X}_k = (\xi_k T)^{-1/2}\mathbf{G}\mathbf{X}_k$. By definition $\mathbf{L}^{r_0}$ is the normalised eigenvectors of $T^{-1}\mathbf{G}\mathbf{G}'$, and it is easily seen that $\mathbf{L}^{r_0} = T^{-1/2}\mathbf{G}\mathbf{X}(\mathbf{\Xi}^{r_0})^{-1/2}$. Thus, we have:

$$\mathbb{H}\mathbb{H}' = T^{-1}\mathbf{G}'\mathbf{L}^{r_0}(\mathbf{\Xi}^{r_0})^{-1}(\mathbf{\Xi}^{r_0})^{-1}\mathbf{L}^{r_0\prime}\mathbf{G}.$$

Plugging the expression of $\mathbf{L}^{r_0}$ into the above equation, we obtain: $\mathbb{H}\mathbb{H}' = \left(T^{-1}\mathbf{G}'\mathbf{G}\right)^{-1}$. Since $T^{-1}\mathbf{G}'\mathbf{G} \xrightarrow{p} \mathbf{\Sigma}_G$, we have $\mathbb{H}\mathbb{H}' \xrightarrow{p} \mathbf{\Sigma}_G^{-1}$.

Therefore, under Assumption 2.G.4, we have:

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

where we use the fact that $\mathbb{B} = O_p\left(\underline{N}^{-1/2}\right)$.

Q.E.D

Proof of Corollary 2.1.

Using Theorem 2.4, $\widehat{\mathbf{G}}_t = \widehat{\mathbf{G}}_t - (\mathbb{H}+\mathbb{B})'\mathbf{G}_t + (\mathbb{H}+\mathbb{B})'\mathbf{G}_t$ and $\widehat{\gamma}_{ij} = \widehat{\gamma}_{ij} - (\mathbb{H}+\mathbb{B})^{-1}\gamma_{ij} + (\mathbb{H}+\mathbb{B})^{-1}\gamma_{ij}$, we have:

$$C_{\underline{N}T}\left(\widehat{\gamma}_{ij}'\widehat{\mathbf{G}}_t - \gamma_{ij}'\mathbf{G}_t\right) = \frac{C_{\underline{N}T}}{\sqrt{\underline{N}}}\frac{1}{R}\gamma_{ij}'\mathcal{I}'\mathbb{C}\mathbb{E}_t + \frac{C_{\underline{N}T}}{\sqrt{T}}\mathbf{G}_t'\mathbb{H}\mathbb{H}'\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\mathbf{G}_s\left(\boldsymbol{\lambda}_{ij}'\mathbf{F}_{is} + e_{ijs}\right) + o_p(1). \tag{B.1.10}$$

where we use the fact that $\mathbb{B} = o_p(1)$ from Theorem 2.4. Using the same arguments in the proof of Theorem 3 in Bai (2003), the leading terms in (B.1.10) are asymptotically independent. Under Assumption 2.G.3 and 2.G.4, we finally obtain the desired result.

Q.E.D

Lemma B.1.6. *Under Assumptions 2.A–2.G, as $N_1, \dots, N_R, T \to \infty$, we have:*

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

Lemma B.1.7. *Under Assumptions 2.A–2.G, as $N_i, T \to \infty$, we have for each i, j and t:*

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

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

Lemma B.1.8. *Under Assumptions 2.A–2.G, for each i, as $N_i, T \to \infty$, we have:*

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

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

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

3.
$$\operatorname*{plim}_{N_i,T\to\infty} \frac{\widehat{\mathbf{F}}_i'\mathbf{F}_i}{T} = \mathbb{W}_i$$

The $r_i \times r_i$ *matrix* $\mathbb{W}_i$ *is given by* $\mathbb{W}_i = \mathbf{\Upsilon}_i^{1/2}\boldsymbol{\mathcal{L}}_i'\mathbf{\Sigma}_{\Lambda_i}^{-1/2}$ *and invertible, where* $\mathbf{\Upsilon}_i$ *is also an* $r_i \times r_i$ *diagonal matrix consisting of the eigenvalues of* $\mathbf{\Sigma}_{\Lambda_i}^{1/2}\mathbf{\Sigma}_{F_i}\mathbf{\Sigma}_{\Lambda_i}^{1/2}$, *and* $\boldsymbol{\mathcal{L}}_i$ *is the corresponding eigenvector matrix such that* $\boldsymbol{\mathcal{L}}_i'\boldsymbol{\mathcal{L}}_i = \mathbf{I}_{r_i}$.

4.
$$\operatorname*{plim}_{N_i,T\to\infty} \widehat{\mathscr{H}_i} = \mathscr{H}_i$$

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

Lemma B.1.9. *Under the assumptions of Theorem 2.5.1, we have for each i and j:*

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

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

3.
$$\frac{1}{T}\left(\widehat{\mathbf{F}}_i - \mathbf{F}_i\widehat{\mathscr{H}_i}\right)'\widehat{\mathbf{S}}_{ij} = O_p\left(\frac{1}{C_{\underline{N}T}^2}\right)$$

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

Proof of Theorem 2.5.

1. To analyse the first term in (B.1.12), we let:

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

Using $\widehat{\mathbf{F}}_{is}=\widehat{\mathbf{F}}_{is}-\widehat{\mathscr{H}}_i'\mathbf{F}_{is}+\widehat{\mathscr{H}}_i'\mathbf{F}_{is}$ and by Theorem 2.1.2, we obtain:

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

Similarly,

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

The first term is $O_p\left(C_{\underline{N}T}^{-2}\right)$ by Theorem 2.2 and Lemma B.1.7. Using Lemma B.1.7, we can express the second term as

$$\begin{aligned}\widehat{\mathscr{H}}_i'\frac{1}{N_iT}\sum_{s=1}^{T}\sum_{j=1}^{N_i}\mathbf{F}_{is}\widehat{S}_{ijs}\boldsymbol{\lambda}_{ij}'\mathbf{F}_{it}=\\
&-\widehat{\mathscr{H}}_i'\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{1}{T}\sum_{s=1}^{T}\mathbf{F}_{is}\left(\widehat{\mathbf{G}}_s-\left(\mathbb{H}'+\mathbb{B}'\right)\mathbf{G}_s\right)'\left[\widehat{\gamma}_{ij}-\left(\mathbb{H}+\mathbb{B}\right)^{-1}\gamma_{ij}\right]\boldsymbol{\lambda}_{ij}'\mathbf{F}_{it}\\
&-\widehat{\mathscr{H}}_i'\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{1}{T}\sum_{s=1}^{T}\mathbf{F}_{is}\left(\widehat{\mathbf{G}}_s-\left(\mathbb{H}'+\mathbb{B}'\right)\mathbf{G}_s\right)'\left(\mathbb{H}+\mathbb{B}\right)^{-1}\gamma_{ij}\boldsymbol{\lambda}_{ij}'\mathbf{F}_{it}\\
&-\widehat{\mathscr{H}}_i'\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{1}{T}\sum_{s=1}^{T}\mathbf{F}_{is}\mathbf{G}_s'\left(\mathbb{H}+\mathbb{B}\right)'\left[\widehat{\gamma}_{ij}-\left(\mathbb{H}+\mathbb{B}\right)^{-1}\gamma_{ij}\right]\boldsymbol{\lambda}_{ij}'\mathbf{F}_{it}\end{aligned}$$

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 B.1.6.1 and Theorem 2.5. The second term is $O_p\left(C_{\underline{N}T}^{-2}\right)$ by Lemma B.1.6.1 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 2.D. Therefore, we obtain: $\mathscr{X}2=O_p\left(C_{\underline{N}T}^{-2}\right)$. Using $\widehat{\mathbf{F}}_{is}=\widehat{\mathbf{F}}_{is}-\widehat{\mathscr{H}}_i'\mathbf{F}_{is}+\widehat{\mathscr{H}}_i'\mathbf{F}_{is}$, we have:

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

The first term is bounded by $O_p\left(C_{\underline{N}T}^{-2}\right)$ by Theorem 2.2 and Lemma B.1.7. The second term is written as

$$\begin{aligned}
&\widehat{\mathscr{H}}_i' \frac{1}{N_i T}\sum_{s=1}^{T}\sum_{j=1}^{N_i}\mathbf{F}_{is}\widehat{S}_{ijs}e_{ijt} = \\
&\qquad -\widehat{\mathscr{H}}_i' \frac{1}{N_i}\sum_{j=1}^{N_i}\frac{1}{T}\sum_{s=1}^{T}\mathbf{F}_{is}\left(\widehat{\mathbf{G}}_s - \left(\mathbb{H}'+\mathbb{B}'\right)\mathbf{G}_s\right)'\left[\widehat{\gamma}_{ij} - \left(\mathbb{H}+\mathbb{B}\right)^{-1}\gamma_{ij}\right]e_{ijt} \\
&\qquad -\widehat{\mathscr{H}}_i' \frac{1}{N_i}\sum_{j=1}^{N_i}\frac{1}{T}\sum_{s=1}^{T}\mathbf{F}_{is}\left(\widehat{\mathbf{G}}_s - \left(\mathbb{H}'+\mathbb{B}'\right)\mathbf{G}_s\right)'\left(\mathbb{H}+\mathbb{B}\right)^{-1}\gamma_{ij}e_{ijt} \\
&\qquad -\widehat{\mathscr{H}}_i' \frac{1}{N_i}\sum_{j=1}^{N_i}\frac{1}{T}\sum_{s=1}^{T}\mathbf{F}_{is}\mathbf{G}_s'\left(\mathbb{H}+\mathbb{B}\right)'\left[\widehat{\gamma}_{ij} - \left(\mathbb{H}+\mathbb{B}\right)^{-1}\gamma_{ij}\right]e_{ijt}
\end{aligned}$$

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 B.1.6.1 and Theorem 2.5. The second term is $O_p\left(C_{\underline{N}T}^{-2}\right)$ by Lemma B.1.6.1 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 2.D. Combining these results, we have: $\mathscr{X}3 = O_p\left(C_{\underline{N}T}^{-2}\right)$.

Next, consider

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

The first term of the above equation is of order $O_p\left(C_{\underline{N}T}^{-2}\right)$ by Theorem 2.2 and Lemma B.1.7. Therefore, we have:

$$\begin{aligned}
\|\mathscr{X}5\| &\le O_p\left(\frac{1}{C_{\underline{N}T}^2}\right) + \left\|\widehat{\mathscr{H}}_i\right\|\frac{1}{\sqrt{T}}\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\mathbf{F}_{is}e_{ijs}\right\|^2\right)^{1/2}\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\left|\widehat{S}_{ijt}\right|^2\right)^{1/2} \\
&= O_p\left(\frac{1}{\sqrt{T}C_{\underline{N}T}}\right)
\end{aligned}$$

where the last equality follows from Assumption 2.B.2 and Lemma B.1.7.

Collecting the results above, (B.1.12) becomes

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

It then follows that

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

Under Assumption 2.G.3, we have

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

where

$$\mathbb{D}_{ii,t}^{(1)(r_i,r_i)} = \operatorname*{plim}_{N_i \to \infty} N_i^{-1} \sum_{j=1}^{N_i} \sum_{k=1}^{N_i} \boldsymbol{\lambda}_{ij} \boldsymbol{\lambda}_{ik}' E(e_{ijt} e_{ikt})$$

is a lower-right $r_i \times r_i$ matrix of $\mathbb{D}_{ii,t}^{(1)}$. As a result, it follows that

$$\begin{aligned}\sqrt{N_i} \left(\widehat{\mathbf{F}}_{it} - \widehat{\mathscr{H}}_i' \mathbf{F}_{it} \right) &= \widehat{\mathbf{\Upsilon}}_i^{-1} \left(\frac{\widehat{\mathbf{F}}_i' \mathbf{F}_i}{T} \right) \frac{1}{N_i} \sum_{j=1}^{N_i} \boldsymbol{\lambda}_{ij} \sqrt{N_i} \widehat{S}_{ijt} + \widehat{\mathbf{\Upsilon}}_i^{-1} \left(\frac{\widehat{\mathbf{F}}_i' \mathbf{F}_i}{T} \right) \frac{1}{\sqrt{N_i}} \sum_{j=1}^{N_i} \boldsymbol{\lambda}_{ij} e_{ijt} + o_p(1) \\ &= \widehat{\mathbf{\Upsilon}}_i^{-1} \left(\frac{\widehat{\mathbf{F}}_i' \mathbf{F}_i}{T} \right) \frac{1}{\sqrt{N_i}} \sum_{j=1}^{N_i} \boldsymbol{\lambda}_{ij} \left(e_{ijt} + \widehat{S}_{ijt} \right) + o_p(1) \xrightarrow{d} N \left(\mathbf{0}, \mathbf{\Upsilon}_i^{-1} \mathbb{W}_i \mathbb{D}_{ii,t}^{(4)} \mathbb{W}_i' \mathbf{\Upsilon}_i^{-1} \right).\end{aligned}$$

since the first leading term also have a non-degenerate distribution by Corollary 2.1.

2. Using $\widehat{\mathbf{\Lambda}}_i = T^{-1} \widehat{\mathbf{F}}_i' \widehat{\mathbf{Y}}_{ij}$, $\widehat{\mathbf{Y}}_{ij} = \widehat{\mathbf{S}}_{ij} + \mathbf{F}_i \lambda_{ij} + \mathbf{e}_{ij}$ and $\mathbf{F}_i = \mathbf{F}_i - \widehat{\mathbf{F}}_i \widehat{\mathscr{H}}_i^{-1} + \widehat{\mathbf{F}}_i \widehat{\mathscr{H}}_i^{-1}$, we obtain:

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

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

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

Notice that, using Corollary 2.1, we obtain

$$\begin{aligned}&\widehat{\mathscr{H}}_i' \frac{1}{T} \mathbf{F}_i' \widehat{\mathbf{S}}_{ij} = \widehat{\mathscr{H}}_i' \frac{1}{T} \sum_{t=1}^{T} \mathbf{F}_{it} \widehat{\mathbf{S}}_{ijt} \\ &= \widehat{\mathscr{H}}_i' \frac{1}{T} \sum_{t=1}^{T} \mathbf{F}_{it} \left(\frac{1}{\sqrt{N}} \frac{1}{R} \gamma_{ij}' \mathcal{I}' \mathbb{C} \mathbb{E}_t + \frac{1}{\sqrt{T}} \mathbf{G}_t' \left(\frac{\mathbf{G}'\mathbf{G}}{T} \right)^{-1} \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \mathbf{G}_s \left(\boldsymbol{\lambda}_{ij}' \mathbf{F}_{is} + e_{ijs} \right) + o_p \left(\frac{1}{C_{\underline{N}T}} \right) \right) \\ &= \frac{1}{\sqrt{N}} \widehat{\mathscr{H}}_i' \frac{1}{T} \sum_{t=1}^{T} \mathbf{F}_{it} \frac{1}{R} \gamma_{ij}' \mathcal{I}' \mathbb{C} \mathbb{E}_t + \frac{1}{\sqrt{T}} \widehat{\mathscr{H}}_i' \left(\frac{\mathbf{F}_i' \mathbf{G}}{T} \right) \left(\frac{\mathbf{G}'\mathbf{G}}{T} \right)^{-1} \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \mathbf{G}_s \left(\boldsymbol{\lambda}_{ij}' \mathbf{F}_{is} + e_{ijs} \right) + o_p \left(\frac{1}{C_{\underline{N}T}} \right)\end{aligned}$$

$$= \frac{1}{\sqrt{N}} \widehat{\mathscr{H}}_i' \frac{1}{T} \sum_{t=1}^{T} \mathbf{F}_{it} \frac{1}{R} \gamma_{ij}' \mathcal{I}' \mathbb{C}\mathbb{E}_t + O_p\left(\frac{1}{T}\right) + o_p\left(\frac{1}{C_{\underline{N}T}}\right)$$

where the last equality is due to Assumption 2.D and 2.G.4. The first term of the RHS in the above equation is bounded by

$$\frac{1}{\sqrt{N}} \widehat{\mathscr{H}}_i' \frac{1}{T} \sum_{t=1}^{T} \mathbf{F}_{it} \frac{1}{R} \gamma_{ij}' \mathcal{I}' \mathbb{C}\mathbb{E}_t$$
$$= \frac{1}{\sqrt{N}} \widehat{\mathscr{H}}_i' \frac{1}{T} \sum_{t=1}^{T} \mathbf{F}_{it} \frac{1}{R} \gamma_{ij}' \sum_{m=1}^{R} \sqrt{\frac{N}{N_m}} \mathbb{I}_m' \left(\frac{\boldsymbol{\Theta}_m' \boldsymbol{\Theta}_m}{N_m}\right)^{-1} \frac{1}{\sqrt{N_m}} \sum_{k=1}^{N_m} \boldsymbol{\theta}_{mk} e_{mkt}$$
$$= \frac{1}{\sqrt{TN_m}} \widehat{\mathscr{H}}_i' \frac{1}{R} \sum_{m=1}^{R} \sqrt{\frac{1}{TN_m} \sum_{k=1}^{N_m} \sum_{t=1}^{T} \mathbf{F}_{it} \boldsymbol{\theta}_{mk}' e_{mkt}} \left(\frac{\boldsymbol{\Theta}_m' \boldsymbol{\Theta}_m}{N_m}\right)^{-1} \mathbb{I}_m \gamma_{ij} = O_p\left(\frac{1}{\sqrt{TN_m}}\right)$$

where the last equality follows from Assumption 2.G.2. Then, by the above development, part 1 of this theorem, Lemma B.1.9, and Assumption 2.D, it follows that

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

By Assumption 2.G.4, we have

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

since the leading terms have non-degenerate distributions.

Q.E.D

B.1.2 Proofs of the auxiliary lemmas

Proof of Lemma B.1.1.

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

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

The first term is of stochastic order $O_p\left(C_{N_m T}^{-1}\right)$ by Lemma 2.1. For the second term, we have:

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

Since $E(e_{hjt}^2) = O(1)$, the above term is $O_p(1)$. Combining these, we obtain the required result.

2. Using equation (2.3.19) and $\widehat{\mathbf{K}}_m = \widehat{\mathbf{K}}_m - \mathbf{K}_m \widehat{\mathbf{H}}_m + \mathbf{K}_m \widehat{\mathbf{H}}_m$, we have:

$$\frac{1}{T\sqrt{N_i}}\left\|\widehat{\mathbf{G}}'\mathbf{e}_i\right\| = \frac{1}{T\sqrt{N_iT}}\left\|\sum_{m=1}^{R}\left\{\widehat{\mathbf{J}}^{r_0\prime}\left(\widehat{\mathbf{K}}_m - \mathbf{K}_m\widehat{\mathbf{H}}_m\right)\widetilde{\mathbf{Q}}_m^{r_0}\left(\widehat{\mathbf{K}}_m - \mathbf{K}_m\widehat{\mathbf{H}}_m\right)'\mathbf{e}_i\right.\right.$$
$$+\widehat{\mathbf{J}}^{r_0\prime}\left(\widehat{\mathbf{K}}_m - \mathbf{K}_m\widehat{\mathbf{H}}_m\right)\widetilde{\mathbf{Q}}_m^{r_0}\widehat{\mathbf{H}}_m'\mathbf{K}_m'\mathbf{e}_i + \widehat{\mathbf{J}}^{r_0\prime}\mathbf{K}_m\widehat{\mathbf{H}}_m\widetilde{\mathbf{Q}}_m^{r_0}\left(\widehat{\mathbf{K}}_m - \mathbf{K}_m\widehat{\mathbf{H}}_m\right)'\mathbf{e}_i$$
$$\left.\left.+\widehat{\mathbf{J}}^{r_0\prime}\mathbf{K}_m\widehat{\mathbf{H}}_m\widetilde{\mathbf{Q}}_m^{r_0}\widehat{\mathbf{H}}_m'\mathbf{K}_m'\mathbf{e}_i\right\}\right\|$$

where $\widetilde{\mathbf{Q}}_i^{r_0} = \widehat{\mathbf{Q}}_i^{r_0}\widehat{\mathbf{Q}}_i^{r_0\prime}$. We note that $\left\|\widehat{\mathbf{J}}^{r_0}\right\| = O_p(1)$ since $\widehat{\mathbf{L}}^{r_0\prime}\widehat{\mathbf{L}}^{r_0} = \mathbf{I}_{r_0}$ 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 B.1.1.1 and Lemma 1. The second term is bounded by $O_p\left(T^{-1/2}C_{\underline{N}T}^{-1}\right)$ by Lemma 2.1 and the fact that $(N_mT)^{-1/2}\|\mathbf{K}_m'\mathbf{e}_i\| = O_p(1)$ under Assumption 2.B.2. The third term is bounded by $O_p\left(C_{\underline{N}T}^{-1}\right)$ by Lemma 1. The last term is bounded by $O_p(T^{-1/2})$ since $(N_mT)^{-1/2}\|\mathbf{K}_m'\mathbf{e}_i\| = O_p(1)$ under Assumption 2.B.2. The proof completes by combining all these results.

Q.E.D

Proof of Lemma B.1.2.

1.

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

Taking expectations of the term inside the bracket, by Assumption 2.A.3 and 2.C.1, we have:

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

2. The proof is similar to part 1 and thus omitted.

3. From (B.1.5) we have:

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

The first term is bounded by $O_p\left(C_{\underline{N}T}^{-1}N_i^{-1/2}\right)$ by Theorem 2.1 and Lemma 1. The second term is bounded by $O_p\left(C_{\underline{N}T}^{-1}N_i^{-1/2}\right)$ due to (B.1.6) and Lemma 2. Using (2.3.19), the third term can be written as

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

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

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

Therefore, it follows that

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

The proof completes by combining the above results.

Q.E.D

Proof of Lemma B.1.3.

1. For each i, by definition, we have: $\widehat{\mathbf{K}}_i\widehat{\mathbf{V}}_i = (N_iT)^{-1}\mathbf{Y}_i\mathbf{Y}_i'\widehat{\mathbf{K}}_i$. By plugging (2.2.6) into this equation, we obtain:

$$\widehat{\mathbf{K}}_i - \mathbf{K}_i\widehat{\mathbf{H}}_i = \left(\frac{1}{N_iT}\mathbf{e}_i\mathbf{\Theta}_i\mathbf{K}_i'\widehat{\mathbf{K}}_i + \frac{1}{N_iT}\mathbf{K}_i\mathbf{\Theta}_i'\mathbf{e}_i'\widehat{\mathbf{K}}_i + \frac{1}{N_iT}\mathbf{e}_i\mathbf{e}_i'\widehat{\mathbf{K}}_i\right)\widehat{\mathbf{V}}_i^{-1} \tag{B.1.11}$$

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

2. For each i, we have:

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

Combining (a)-(d) in Lemma B.1.3.1, the result follows.

3. Consider the term,

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

where $\left\|\widehat{\mathbf{V}}_m^{-1}\right\| = O_p(1)$ by Lemma B.1.5. Let $T^{-1/2}\widehat{\mathcal{R}}_m'\mathbf{K}_h = \widehat{\mathbf{V}}_m^{-1}(\mathbf{X}1 + \mathbf{X}2 + \mathbf{X}3 + \mathbf{X}4)$.

First, $\mathbf{X}1$ can be written as

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widehat{\mathbf{K}}_{ms} - \widehat{\mathbf{H}}_m'\mathbf{K}_{ms}\right)\omega_m(s,t)\mathbf{K}_{ht}' + \widehat{\mathbf{H}}_m'\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{K}_{ms}\omega_m(s,t)\mathbf{K}_{ht}' = \mathbf{X}1.1 + \mathbf{X}1.2$$

By the Cauchy-Schwarz inequality, we have:

$$\|\mathbf{X}1.1\| \leq \frac{1}{\sqrt{T}}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{T}\sum_{s=1}^{T}\left\|\widehat{\mathbf{K}}_{ms} - \widehat{\mathbf{H}}_m'\mathbf{K}_{ms}\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}|\omega_m(s,t)|^2\|\mathbf{K}_{ht}\|^2\right)^{1/2} = \left[O_p\left(\frac{1}{\sqrt{N_m}}\right) + O_p\left(\frac{1}{\sqrt{T}C_{N_mT}}\right)\right]\frac{1}{\sqrt{T}} = O_p\left(\frac{1}{\sqrt{N_mT}}\right) + O_p\left(\frac{1}{TC_{N_mT}}\right)$$

where we use Lemma 1, Assumption 2.B.1 and the fact that $T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}|\omega_m(s,t)|^2 = O(1)$ (see Lemma 1.(i) in Bai & Ng (2002)). The expected value of $\mathbf{X}1.2$ without $\widehat{\mathbf{H}}'_m$, is bounded by

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

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

By the Cauchy-Schwarz inequality, $\mathbf{X}2$ is bounded by

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

under Assumptions 2.B.1 and 2.G.1.

Next, $\mathbf{X}3$ can be expressed as

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

Applying the Cauchy-Schwarz inequality to $\mathbf{X}3.1$, we obtain:

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

The second term can be expressed as

$$\begin{aligned}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{K}_{ht}\eta_{m,st}\right\|^2\right)^{1/2} &= \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{N_mT}\sum_{t=1}^{T}\mathbf{K}_{ht}\mathbf{K}'_{ms}\boldsymbol{\theta}_{mj}e_{mjt}\right\|^2\right)^{1/2} \\ &\leq \left(\frac{1}{T}\sum_{s=1}^{T}\|\mathbf{K}_{ms}\|^2\left\|\frac{1}{N_mT}\sum_{t=1}^{T}\sum_{j=1}^{N_m}\mathbf{K}'_{ht}\boldsymbol{\theta}_{mj}e_{mjt}\right\|^2\right)^{1/2} = O_p\left(\frac{1}{\sqrt{N_mT}}\right)\end{aligned}$$

under Assumptions 2.B.1 and 2.G.2. Hence, $\|\mathbf{X}3.1\| = O_p\left(C_{N_mT}^{-1}\right)O_p\left(N_m^{-1/2}T^{-1/2}\right)$. For $\mathbf{X}3.2$, we have:

$$\mathbf{X}3.2 = \frac{1}{T}\sum_{s=1}^{T}\mathbf{K}_{ms}\mathbf{K}'_{ms}\frac{1}{N_mT}\sum_{t=1}^{T}\sum_{j=1}^{N_m}\mathbf{K}'_{ht}\boldsymbol{\theta}_{mj}e_{mjt} = O_p\left(\frac{1}{\sqrt{N_mT}}\right)$$

by Assumption 2.G.2. Therefore, $\|\mathbf{X}3\| = O_p\left(N_m^{-1/2}T^{-1/2}\right)$.

Following similar steps, we also obtain: $\mathbf{X}4 = O_p\left(N_m^{-1/2}T^{-1/2}\right)$. Collecting all these results, we have: $T^{-1/2}\widehat{\boldsymbol{\mathcal{R}}}'_m\mathbf{K}_h = O_p\left(C_{N_mT}^{-2}\right)$.

4.

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

By Lemmas B.1.3.2 and B.1.3.3, it follows that $T^{-1/2}\widehat{\mathcal{R}}'_m\widehat{\mathbf{K}}_h = O_p\left(C^{-2}_{N_mT}\right)$.

5. Consider

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

where $\left\|\widehat{\mathbf{V}}^{-1}_m\right\| = O_p(1)$ by Lemma B.1.5. Let $T^{-1/2}\widehat{\mathcal{R}}'_m\mathbf{e}_{hj} = \widehat{\mathbf{V}}^{-1}_m\left(\mathcal{X}1 + \mathcal{X}2 + \mathcal{X}3 + \mathcal{X}4\right)$.

As $\mathcal{X}1$ is of order $O_p\left(C^{-2}_{N_mT}\right)$, the proof is the same as that of $\mathbf{X}1$ in Lemma B.1.3. Next,

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

Using the Cauchy-Schwarz inequality, we have:

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

Notice that by Assumption 2.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 B.1.3.2, we show that

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

By Assumption 2.G.1,

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

Combining these two terms, we have $\mathcal{X}2 = O_p(C^{-2}_{N_mT})$.

Next, we can write $\mathcal{X}3$ as

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

By the Cauchy-Schwarz inequality, we have:

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

Notice that

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

Using Lemma B.1.3.2, we have:

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

By Assumption 2.F.2, we have:

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

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

Q.E.D

Proof of Lemma B.1.4.

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

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

Using (B.1.11) and the definition of $\widehat{\boldsymbol{\mathcal{R}}}_i$ in Lemma B.1.3.2, we obtain:

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

where

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

and

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

Using Lemma B.1.3.3 and the fact that $\widehat{\mathbf{H}}_i$ 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 B.1.3.2, we have $\widehat{\mathbb{A}}_3 = O_p\left(C_{\underline{N}T}^{-2}\right)$.

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

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

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

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

The first two terms are bounded by $O_p\left(C_{N_iT}^{-2}\right)$ by Lemmas B.1.3.3 and B.1.3.4. Using $\widehat{\mathbf{Q}}_i = \widehat{\mathbf{H}}_i^{-1}\mathbf{Q}_i\mathbf{D} + O_p\left(C_{N_iT}^{-2}\right)$, the last two terms can be expressed as

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

Notice that

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

since $\mathbf{D}$ and $\mathbf{B}$ are orthogonal matrices. Therefore,

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

By Lemma 2.3 and Assumption 2.B.1, we have: $\left\|\widehat{\mathbf{L}}^{r_0} - \mathbf{L}^{r_0}\mathbf{U}\right\| = O_p\left(C_{\underline{N}T}^{-2}\right)$ where $\mathbf{U}$ is an $r_0 \times r_0$ orthogonal matrix defined in (B.1.3).

Q.E.D

Proof of Lemma B.1.5.

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 Lemma B.1.6.

Using $\widehat{\mathbf{K}}_{it} = \widehat{\mathbf{K}}_{it} - \widehat{\mathbf{H}}_i'\mathbf{K}_{it} + \widehat{\mathbf{H}}_i'\mathbf{K}_{it}$, we can write (B.1.9) as

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

Then, we have:

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

By Lemma B.1.3.4, $T^{-1}\sum_{t=1}^{T}\left(\widehat{\mathbf{K}}_{mt} - \widehat{\mathbf{H}}_m'\mathbf{K}_{mt}\right)\mathbf{K}_{it}' = O_p\left(C_{N_mT}^{-2}\right)$. Then, the required result follows.

We can also prove Lemma B.1.6.2 along similar arguments using Lemma B.1.3.4.

Q.E.D

Proof of Lemma B.1.7.

Using the expansions, $\widehat{\mathbf{G}} = \widehat{\mathbf{G}} - \mathbf{G}\left(\mathbb{H}+\mathbb{B}\right) + \mathbf{G}\left(\mathbb{H}+\mathbb{B}\right)$ and $\widehat{\boldsymbol{\Gamma}}_i' = \widehat{\boldsymbol{\Gamma}}_i' - \left(\mathbb{H}+\mathbb{B}\right)^{-1}\boldsymbol{\Gamma}_i' + \left(\mathbb{H}+\mathbb{B}\right)^{-1}\boldsymbol{\Gamma}_i'$, the result follows from Theorems 2.4.

Q.E.D

Proof of Lemma B.1.8.

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

Q.E.D

Proof of Lemma B.1.9.

1. By construction we have:

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

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

$$\widehat{\mathbf{F}}_{it} - \widehat{\mathscr{H}}_i'\mathbf{F}_{it} =$$
$$\widehat{\boldsymbol{\Upsilon}}_i^{-1}\frac{1}{N_iT}\left(\sum_{s=1}^{T}\widehat{\mathbf{F}}_{is}\widehat{\mathbf{S}}_{i.s}'\widehat{\mathbf{S}}_{i.t} + \sum_{s=1}^{T}\widehat{\mathbf{F}}_{is}\widehat{\mathbf{S}}_{i.s}'\boldsymbol{\Lambda}_i\mathbf{F}_{it} + \sum_{s=1}^{T}\widehat{\mathbf{F}}_{is}\widehat{\mathbf{S}}_{i.s}'\mathbf{e}_{i.t} + \sum_{s=1}^{T}\widehat{\mathbf{F}}_{is}\mathbf{F}_{is}'\boldsymbol{\Lambda}_i'\widehat{\mathbf{S}}_{i.t} + \sum_{s=1}^{T}\widehat{\mathbf{F}}_{is}\mathbf{e}_{i.s}'\widehat{\mathbf{S}}_{i.t}\right)$$
$$+\widehat{\boldsymbol{\Upsilon}}_i^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\widehat{\mathbf{F}}_{is}\omega_i(s,t) + \frac{1}{T}\sum_{s=1}^{T}\widehat{\mathbf{F}}_{is}\zeta_{i,st} + \frac{1}{T}\sum_{s=1}^{T}\widehat{\mathbf{F}}_{is}\eta_{i,st}^* + \frac{1}{T}\sum_{s=1}^{T}\widehat{\mathbf{F}}_{is}\mu_{i,st}^*\right) \quad \text{(B.1.12)}$$

where $\widehat{\mathscr{H}_i} = (\mathbf{\Lambda}_i'\mathbf{\Lambda}_i/N_i)(\mathbf{F}_i'\mathbf{F}_i/T)\,\widehat{\mathbf{\Upsilon}}_i^{-1}$, $\widehat{\mathbf{S}}_{i.t}$ is the $N_i \times 1$ vector of $\widehat{\mathbf{S}}_i$ (the t-th row vector), $\eta_{i,st}^* = N_i^{-1}\mathbf{F}_{is}'\mathbf{\Lambda}_i'\mathbf{e}_{i.t}$ and $\mu_{i,st}^* = N_i^{-1}\mathbf{F}_{it}'\mathbf{\Lambda}_i'\mathbf{e}_{i.s}$. $\omega_i(s,t)$ and $\zeta_{i,st}$ are defined in Lemma B.1.3.1.

It then follows that:

$$\frac{1}{T}\left(\widehat{\mathbf{F}}_i - \mathbf{F}_i\widehat{\mathscr{H}_i}\right)'\mathbf{F}_i = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\mathbf{F}}_{it} - \widehat{\mathscr{H}_i}'\mathbf{F}_{it}\right)\mathbf{F}_{it}' =$$

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

By Lemma B.1.8, we have $\widehat{\mathbf{\Upsilon}}_i = O_p(1)$. It is straightforward to show that the second term is of order $O_p\left(C_{\underline{N}T}^{-2}\right)$ using the similar lines in the proof of Lemma B.1.3.3.

We now focus on the first term, which is written as $\widehat{\mathbf{\Upsilon}}_i^{-1}(\mathcal{Q}1 + \mathcal{Q}2 + \mathcal{Q}3 + \mathcal{Q}4 + \mathcal{Q}5)$. By Lemma B.1.7, $\mathcal{Q}1 = O_p\left(C_{\underline{N}T}^{-2}\right)$. Next, using $\widehat{\mathbf{F}}_{is} = \widehat{\mathbf{F}}_{is} - \widehat{\mathscr{H}_i}'\mathbf{F}_{is} + \widehat{\mathscr{H}_i}'\mathbf{F}_{is}$, we have:

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

Notice that $\mathbf{F}_i'\mathbf{F}_i/T = O_p(1)$ under Assumption 2.B.1. Combining Lemmas B.1.6 and B.1.7, Theorems 2.4 and Assumption 2.D, we have: $T^{-1}\sum_{s=1}^{T}\mathbf{F}_{is}\widehat{S}_{ijs} = O_p\left(C_{\underline{N}T}^{-2}\right)$ such that the second term of the above equation is $O_p\left(C_{\underline{N}T}^{-2}\right)$. Hence, $\mathcal{Q}2 = \left(C_{\underline{N}T}^{-2}\right)$. Along similar arguments, it is easily seen that $\mathcal{Q}3$ to $\mathcal{Q}5$ have stochastic order $O_p\left(C_{\underline{N}T}^{-2}\right)$.

2. The proof is similar to part 1 of Lemma B.1.9 and thus omitted.

3. The result follows from Theorem 2.2 and Lemma B.1.7.

Q.E.D

B.2 Bootstrap Confidence Intervals

In this section, we outline the bootstrap procedure for estimating the confidence intervals for $(\mathbb{H}+\mathbb{B})'\mathbf{G}_t$, $(\mathbb{H}+\mathbb{B})^{-1}\boldsymbol{\gamma}_{ij}$, $\boldsymbol{\gamma}_{ij}'\mathbf{G}_t$, $\widehat{\mathscr{H}_i}'\mathbf{F}_{it}$, and $\boldsymbol{\lambda}_{ij}'\mathbf{F}_{it}$, respectively. For computational tractability, we assume that the error terms are cross-sectionally and serially uncorrelated. If the error terms are serially correlated or weakly cross-sectionally correlated, the dependent bootstrap can be adopted (e.g. Shao (2010) and Conley et al. (2023)). The validity of the bootstrap is corroborated by simulations whilst the formal proof is beyond the scope of this paper.

We introduce some notations in line with the bootstrap literature. The superscript $*(b)$ denotes the b-th realisation among B bootstrap repetitions. Let $\Pr^*$ be the bootstrap probability measure conditional on the original sample. Though $\Pr^*$ depends on the sample size (N, T) and the realisation (b), we omit them for notational simplicity. For a bootstrapped sequence $\mathbf{X}_n^{*(b)}$ for $b = 1, .., B$, we denote $\mathbf{X}_n^{*(b)} = o_{p^*}(1)$ if $\lim_{n\to\infty} \Pr^*\left(\left|\mathbf{X}_n^{*(b)}\right| > \epsilon\right) = o_p(1)$ for all $\epsilon > 0$. Furthermore, $\mathbf{X}_n^{*(b)} \xrightarrow{d^*} D$ if $\mathbf{X}_n^{*(b)}$ weakly converges in distribution to D conditional on the original sample. $Bartlett(x) = 1-|x|$ if $|x| \leq 1$ and $Bartlett(x) = 0$ if $|x| > 0$ is the Bartlett kernel function, see Andrews (1991). Finally, let $\boldsymbol{\iota}_z = [0, \ldots, 1, \ldots, 0]'$ be an $r_0 \times 1$ vector such that the z-th element is one and zeros elsewhere and $\boldsymbol{\iota}_{i,z}$ be an $r_i \times 1$ vector defined similarly for $i = 1, \ldots, R$.

We provide three algorithms to generate the confidence intervals.

Algorithm 1: Bootstrap the global factors

Step 1.1 *For each i, j and s, construct $e_{ijs}^{*(b)} = \hat{e}_{ijs}\varepsilon_{ijs}^{*(b)}$ where $\hat{e}_{ijs} = y_{ijs} - \widehat{\gamma}_{ij}'\widehat{\mathbf{G}}_s - \widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\mathbf{F}}_{is}$ and $\varepsilon_{ijs}^{*(b)} \sim$ i.i.d. $N(0,1)$, $i = 1, ..., R$, $j = 1, ..., N_m$ and $s = 1, ..., T$.*

Step 1.2 *Re-sample the data by $y_{ijs}^{*(b)} = \widehat{\gamma}_{ij}'\widehat{\mathbf{G}}_s + \widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\mathbf{F}}_{is} + e_{ijs}^{*(b)}$.*

Step 1.3 *Apply the estimation procedure developed in Section 2.3 to the re-sampled data, and obtain the bootstrap estimates of global factors, denoted $\widehat{\mathbf{G}}^{*(b)}$. We then construct the rotation matrix by $\widehat{\mathbb{H}}^{*(b)} = T^{-1}\widehat{\mathbf{G}}'\widehat{\mathbf{G}}^{*(b)}$.*

Step 1.4 *Repeat Steps 1.1–1.3 B times, and construct the empirical distribution function as*

$$\widehat{\mathcal{D}}_{G_t^z}(\tau) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\left(\sqrt{N}\left[\boldsymbol{\iota}_z'\left(\widehat{\mathbb{H}}^{*(b)\prime}\right)^{-1}\widehat{\mathbf{G}}_t^{*(b)} - \widehat{G}_t^z\right] \leq \tau\right) \text{ for } z = 1, \ldots, r_0$$

where $\mathbb{1}$ is the indicator function. The $1-\alpha$ CI for the z-th (rotated) global factor is given by

$$\left[\widehat{G}_t^z - \frac{1}{\sqrt{N}}\widehat{\mathcal{D}}_{G_t^z}^{-1}\left(\frac{\alpha}{2}\right), \widehat{G}_t^z - \frac{1}{\sqrt{N}}\widehat{\mathcal{D}}_{G_t^z}^{-1}\left(1 - \frac{\alpha}{2}\right)\right] \tag{B.2.13}$$

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

The consistency and asymptotic normality of the *GCC* estimator can be achieved for the re-sampled data since the validity of Assumptions 2.A–2.G is still maintained in Step 1.1. Hence, for each $b = 1, \ldots, B$, we have:

$$\sqrt{N}\left[\left(\mathbb{H}^{*(b)\prime} + \mathbb{B}^{*(b)\prime}\right)^{-1}\widehat{\mathbf{G}}_t^{*(b)} - \widehat{\mathbf{G}}_t\right] = \frac{1}{R}\mathcal{I}'\widehat{\mathbb{C}}\mathbb{E}_t^{*(b)} + o_{p^*}(1)$$

where $\mathbb{H}^{*(b)} = T^{-1/2}\widehat{\mathbf{G}}'\widehat{\mathbf{L}}^{r_0}\mathbf{U}^{*(b)}$ and $\mathbb{B}^{*(b)} = \frac{1}{R}\sum_{i=1}^{R}\sqrt{\frac{1}{N_i}}\mathbb{I}_i'\left(\frac{\widehat{\mathbf{\Theta}}_i'\widehat{\mathbf{\Theta}}_i}{N_i}\right)^{-1}\frac{\widehat{\mathbf{\Theta}}_i'\widehat{\mathbf{e}}_i'}{\sqrt{N_i T}}\widehat{\mathbf{J}}^{r_0}\mathbf{U}^{*(b)}$ with $\mathbf{U}^{*(b)}$ being an orthonormal matrix induced by Lemma 2.3, $\widehat{\mathbf{\Theta}}_i = \left[\widehat{\mathbf{\Gamma}}_i, \widehat{\mathbf{\Lambda}}_i\right]$,

$$\widehat{\mathbb{C}} = \text{diag}\left(\sqrt{\frac{N}{N_1}}\mathbb{I}_1'\left(\frac{\widehat{\mathbf{\Theta}}_1'\widehat{\mathbf{\Theta}}_1}{N_1}\right)^{-1}, \ldots, \sqrt{\frac{N}{N_R}}\mathbb{I}_R'\left(\frac{\widehat{\mathbf{\Theta}}_R'\widehat{\mathbf{\Theta}}_R}{N_R}\right)^{-1}\right),$$

and

$$\mathbb{E}_t^{*(b)} = \begin{bmatrix} \mathbb{E}_{1t}^{*(b)} \\ \mathbb{E}_{2t}^{*(b)} \\ \vdots \\ \mathbb{E}_{Rt}^{*(b)} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{N_1}}\sum_{j=1}^{N_1}\widehat{\boldsymbol{\theta}}_{1j}e_{1jt}^{*(b)} \\ \frac{1}{\sqrt{N_2}}\sum_{j=1}^{N_2}\widehat{\boldsymbol{\theta}}_{2j}e_{2jt}^{*(b)} \\ \vdots \\ \frac{1}{\sqrt{N_R}}\sum_{j=1}^{N_R}\widehat{\boldsymbol{\theta}}_{Rj}e_{Rjt}^{*(b)} \end{bmatrix}.$$

Notice that $\mathbb{H}^{*(b)} + \mathbb{B}^{*(b)}$ can be evaluated directly under the bootstrap world. By applying Theorem 2.4, we have

$$\left[\widehat{\mathbf{G}}^{*(b)} - \widehat{\mathbf{G}}\left(\mathbb{H}^{*(b)} + \mathbb{B}^{*(b)}\right)\right] = \frac{1}{R\sqrt{N}}\mathbb{E}^{*(b)}\widehat{\mathbb{C}}'\mathcal{I}$$

where $\mathbb{E}^{*(b)} = \left[\mathbb{E}_1^{*(b)}, \ldots, \mathbb{E}_T^{*(b)}\right]'$. Pre-multiplying the above equation by $T^{-1}\widehat{\mathbf{G}}'$ and using the fact that $T^{-1}\widehat{\mathbf{G}}'\widehat{\mathbf{G}} = \mathbf{I}_{r_0}$, we obtain:

$$\left(\frac{\widehat{\mathbf{G}}'\widehat{\mathbf{G}}^{*(b)}}{T}\right) - \left(\mathbb{H}^{*(b)} + \mathbb{B}^{*(b)}\right) = \frac{1}{RT\sqrt{N}}\widehat{\mathbf{G}}'\mathbb{E}^{*(b)}\widehat{\mathbb{C}}'\mathcal{I} = o_{p^*}(1)$$

which implies that $\left(\mathbb{H}^{*(b)} + \mathbb{B}^{*(b)}\right)$ can be consistently estimated by $\widehat{\mathbb{H}}^{*(b)} = T^{-1}\widehat{\mathbf{G}}'\widehat{\mathbf{G}}^{*(b)}$. As a result, we have

$$\sqrt{N}\left[\left(\widehat{\mathbb{H}}^{*(b)\prime}\right)^{-1}\widehat{\mathbf{G}}_t^{*(b)} - \widehat{\mathbf{G}}_t\right] = \frac{1}{R}\mathcal{I}'\widehat{\mathbb{C}}\mathbb{E}_t^{*(b)} + o_{p^*}(1) \xrightarrow{d^*} N\left(\mathbf{0}, \frac{1}{R^2}\mathcal{I}'\widehat{\mathbb{C}}\left(\mathbb{D}_t^{(1)*}\right)\widehat{\mathbb{C}}'\mathcal{I}\right) \quad \text{(B.2.14)}$$

where $\mathbb{D}_t^{(1)*}$ is the (conditional) asymptotic covariance matrix of $\mathbb{E}_t^{*(b)}$. Under the assumptions of Theorem 2.4.1, it follows that

$$\sup_{x\in\mathbb{R}}\left|\Pr{}^*\left(\sqrt{N}\left[\boldsymbol{\iota}_z'\left(\widehat{\mathbb{H}}^{*(b)\prime}\right)^{-1}\widehat{\mathbf{G}}_t^{*(b)} - \widehat{G}_t^z\right] \le x\right) - \Pr\left(\sqrt{N}\left[\widehat{G}_t^z - \boldsymbol{\iota}_z'\left(\mathbb{H}+\mathbb{B}\right)'\mathbf{G}_t\right] \le x\right)\right| \xrightarrow{p} 0,$$

suggesting that the confidence interval (B.2.13) has correct coverage rate for $\iota_z'\left(\mathbb{H}+\mathbb{B}\right)'\mathbf{G}_t$, namely the z-th (rotated) global factor.

Algorithm 2: Bootstrap the global factor loadings and global component

Step 2.1 *For each m, k and s, let $e_{mks}^{*(b)} = \hat{e}_{mks}\varepsilon_{mks}^{*(b)}$ where $\hat{e}_{mks} = y_{mks} - \widehat{\gamma}_{mk}'\widehat{\mathbf{G}}_s - \widehat{\boldsymbol{\lambda}}_{mk}'\widehat{\mathbf{F}}_{ms}$ and $\varepsilon_{mks}^{*(b)} \sim$ i.i.d. $N(0,1)$, $m = 1, ..., R$, $k = 1, ..., N_m$ and $s = 1, ..., T$.*

Step 2.2 *Apply the dependent wild bootstrap by Shao (2010) and re-sample the local factors for* *block i as*

$$F_{is}^{z,*(b)} = \widehat{F}_{is}^{z} \cdot \omega_{is}^{z,*(b)} \text{ for } z = 1, \ldots, r_i,\ s = 1, \ldots, T,$$

where $\omega_{is}^{z,*(b)}$ is drawn from a zero mean normal distribution independent across i and z with covariance,

$$\operatorname{Cov}\left(\omega_{ic}^{z,*(b)}, \omega_{id}^{z,*(b)}\right) = Bartlett\left(\frac{c-d}{l_i^z}\right) \text{ for } c, d = 1, \ldots, T$$

where l_i^z is a bandwidth parameter.[2]

Step 2.3 *Re-sample the data for the i-th block as $y_{iks}^{*(b)} = \widehat{\gamma}_{ik}' \widehat{\mathbf{G}}_s + \widehat{\boldsymbol{\lambda}}_{ik}' \mathbf{F}_{is}^{*(b)} + e_{iks}^{*(b)}$, while $y_{mks}^{*(b)} = \widehat{\gamma}_{mk}' \widehat{\mathbf{G}}_s + \widehat{\boldsymbol{\lambda}}_{mk}' \widehat{\mathbf{F}}_{ms} + e_{mks}^{*(b)}$ for other blocks.*

Step 2.4 *Apply the procedure developed in Section 2.3 to the re-sampled data, and obtain the bootstrap estimates of global factors and loadings, denoted $\widehat{\mathbf{G}}^{*(b)}$ and $\widehat{\boldsymbol{\Gamma}}_i^{*(b)} = T^{-1}\mathbf{Y}_i^{*(b)\prime}\widehat{\mathbf{G}}^{*(b)}$. Construct the estimated rotation matrix $\widehat{\mathbb{H}}^{*(b)} = T^{-1}\widehat{\mathbf{G}}'\widehat{\mathbf{G}}^{*(b)}$.*

Step 2.5 *Repeat Steps 2.1–2.4 for B times. Construct the empirical distribution functions*

$$\widehat{\mathcal{D}}_{\gamma_{ij}^z}(\tau) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\left(\sqrt{T}\left[\boldsymbol{\iota}_z'\widehat{\mathbb{H}}^{*(b)}\widehat{\boldsymbol{\gamma}}_{ij}^{*(b)} - \widehat{\gamma}_{ij}^{z}\right] \leq \tau\right) \text{ for } z = 1, \ldots, r_0$$

and

$$\widehat{\mathcal{D}}_{\gamma_{ij}'G_t}(\tau) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\left(C_{NT}\left(\widehat{\gamma}_{ij}^{*(b)\prime}\widehat{\mathbf{G}}_t^{*(b)} - \widehat{\gamma}_{ij}'\widehat{\mathbf{G}}_t\right) \leq \tau\right).$$

Then, the $1-\alpha$ CI for the z-th (rotated) global factor loading is given by

$$\left[\widehat{\gamma}_{ij} - \frac{1}{\sqrt{T}}\widehat{\mathcal{D}}_{\gamma_{ij}^z}^{-1}\left(\frac{\alpha}{2}\right), \widehat{\gamma}_{ij} - \frac{1}{\sqrt{T}}\widehat{\mathcal{D}}_{\gamma_{ij}^z}^{-1}\left(1 - \frac{\alpha}{2}\right)\right] \tag{B.2.15}$$

where $\widehat{\mathcal{D}}_{\gamma_{ij}^z}^{-1}(\alpha/2)$ and $\widehat{\mathcal{D}}_{\gamma_{ij}^z}^{-1}(1-\alpha/2)$ are the inverse functions of $\widehat{\mathcal{D}}_{\gamma_{ij}^z}$ evaluated at $\alpha/2$ and $1-\alpha/2$, respectively. The $1-\alpha$ CI for the global component is given by

$$\left[\widehat{\gamma}_{ij}'\widehat{\mathbf{G}}_t - \frac{1}{C_{NT}}\widehat{\mathcal{D}}_{\gamma_{ij}'G_t}^{-1}\left(\frac{\alpha}{2}\right), \widehat{\gamma}_{ij}'\widehat{\mathbf{G}}_t - \frac{1}{C_{NT}}\widehat{\mathcal{D}}_{\gamma_{ij}'G_t}^{-1}\left(1 - \frac{\alpha}{2}\right)\right] \tag{B.2.16}$$

where $\widehat{\mathcal{D}}_{\gamma_{ij}'G_t}^{-1}(\alpha/2)$ and $\widehat{\mathcal{D}}_{\gamma_{ij}'G_t}^{-1}(1-\alpha/2)$ are the inverse function of $\widehat{\mathcal{D}}_{\gamma_{ij}'G_t}$ evaluated at $\alpha/2$ and $1-\alpha/2$.

For each $b = 1, \ldots, B$, by Theorem 2.4.2 we have:

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

[2] The bandwidth parameter can be chosen following the data dependent approach developed by Andrews (1991). In simulation, we use the R package "cointReg" to perform the bandwidth selection. See https://cran.rstudio.com/web/packages/cointReg/index.html.

In Step 2.2 we follow the dependent wild bootstrap by Shao (2010), which accounts for times series dependence of the local factors. This ensures that the covariance matrix of $\left(\mathbb{H}^{*(b)}+\mathbb{B}^{*(b)}\right)\widehat{\gamma}_{ij}^{*(b)}$ matches that of $\widehat{\gamma}_{ij}$, i.e. $\mathbb{D}_{ij}^{(2)*}\xrightarrow{p^*}\mathbb{H}'\mathbb{D}_{ij}^{(2)}\mathbb{H}$. If we do not resample $\widehat{\mathbf{F}}_i$, then it is easily seen that

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

because $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{F}}_i$ are orthogonal. Consequently, the covariance matrix will be incorrect.

Under the assumptions of Theorem 2.4, it follows that

$$\sup_{x\in\mathbb{R}}\left|\Pr^*\left(\sqrt{T}\left[\boldsymbol{\iota}_z'\widehat{\mathbb{H}}^{*(b)}\widehat{\gamma}_{ij}^{*(b)}-\widehat{\gamma}_{ij}^{z}\right]\le x\right)-\Pr\left(\sqrt{T}\left[\widehat{\gamma}_{ij}^{z}-\boldsymbol{\iota}_z'\left(\mathbb{H}+\mathbb{B}\right)^{-1}\boldsymbol{\gamma}_{ij}\right]\le x\right)\right|\xrightarrow{p}0$$

and

$$\sup_{x\in\mathbb{R}}\left|\Pr^*\left(C_{NT}\left[\widehat{\gamma}_{ij}^{*(b)\prime}\widehat{\mathbf{G}}_t^{*(b)}-\widehat{\gamma}_{ij}'\widehat{\mathbf{G}}_t\right]\le x\right)-\Pr\left(C_{NT}\left[\widehat{\gamma}_{ij}'\widehat{\mathbf{G}}_t-\boldsymbol{\gamma}_{ij}'\mathbf{G}_t\right]\le x\right)\right|\xrightarrow{p}0$$

suggesting that the confidence intervals (B.2.15) and (B.2.16) have correct coverage rates for the z-th (rotated) global factor loading $\boldsymbol{\iota}_z'\left(\mathbb{H}+\mathbb{B}\right)^{-1}\boldsymbol{\gamma}_{ij}$ and $\boldsymbol{\gamma}_{ij}'\mathbf{G}_t$, respectively.

Algorithm 3: Bootstrap the local factors and local components

Step 3.1 *For each m, k and s, let $e_{mks}^{*(b)}=\hat{e}_{mks}\varepsilon_{mks}^{*(b)}$ where $\hat{e}_{mks}=y_{mks}-\widehat{\gamma}_{mk}'\widehat{\mathbf{G}}_s-\widehat{\boldsymbol{\lambda}}_{mk}'\widehat{\mathbf{F}}_{ms}$ and $\varepsilon_{mks}^{*(b)}\sim$ i.i.d. $N(0,1)$, $m=1,...,R$, $k=1,...,N_m$ and $s=1,...,T$.*

Step 3.2 *Apply the dependent wild bootstrap and re-sample the global factors as*

$$\widetilde{G}_s^{z,*(b)}=\widehat{G}_s^{z}\cdot\omega_s^{z,*(b)}\ \textit{for}\ z=1,\dots,r_0,\ s=1,\dots,T,$$

where $\omega_s^{z,*(b)}$ is drawn from a zero mean normal distribution independent across z with covariance

$$\mathrm{Cov}\left(\omega_c^{z,*(b)},\omega_d^{z,*(b)}\right)=\textit{Bartlett}\left(\frac{c-d}{l^z}\right)\ \textit{for}\ c,d=1,\dots,T$$

where l^z is a bandwidth parameter.

Step 3.3 *Re-sample the data as $y_{mks}^{*(b)}=\widehat{\gamma}_{mk}'\widetilde{\mathbf{G}}_s^{*(b)}+\widehat{\boldsymbol{\lambda}}_{mk}'\widehat{\mathbf{F}}_{ms}+e_{mks}^{*(b)}$ where $\widetilde{\mathbf{G}}_s^{*(b)}=(\widetilde{G}_s^{1,*(b)},\dots,\widetilde{G}_s^{r_0,*(b)})'$.*

Step 3.4 *Apply the procedure in Section 2.3 to the re-sampled data, and obtain the bootstrap estimates, denoted $\widehat{\mathbf{G}}^{*(b)}$, $\widehat{\boldsymbol{\Gamma}}_i^{*(b)}=T^{-1}\mathbf{Y}_i^{*(b)\prime}\widehat{\mathbf{G}}^{*(b)}$, $\widehat{\mathbf{F}}_i^{*(b)}$, and $\widehat{\boldsymbol{\Lambda}}_i^{*(b)}=T^{-1}\widetilde{\mathbf{Y}}_i^{*(b)\prime}\widehat{\mathbf{F}}_i^{*(b)}$ where $\widetilde{\mathbf{Y}}_i^{*(b)}=\mathbf{Y}_i^{*(b)}-\widehat{\mathbf{G}}^{*(b)}\widehat{\boldsymbol{\Gamma}}_i^{*(b)}$. Following Gonçalves & Perron (2014) we construct the estimated rotation matrix $\widetilde{\mathscr{H}}_i^{*(b)}=\mathrm{diag}(\pm1)$ where the signs are determined by the signs of the diagonal elements of $\widehat{\mathbf{F}}_i^{*(b)\prime}\widehat{\mathbf{F}}_i$.*

Step 3.5 *Repeat Steps 3.1–3.4 for B times. Construct the empirical distribution functions as*

$$\widehat{\mathcal{D}}_{F_{it}^z}(\tau) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\left(\sqrt{N_i}\left[\boldsymbol{\iota}_{i,z}'\left(\widetilde{\mathscr{H}}_i^{*(b)\prime}\right)^{-1}\widehat{\mathbf{F}}_{it}^{*(b)} - \widehat{F}_{it}^z\right] \le \tau\right) \text{ for } z = 1,\ldots,r_i$$

and

$$\widehat{\mathcal{D}}_{\lambda_{ij}' F_{it}}(\tau) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\left(C_{N_iT}\left(\widehat{\boldsymbol{\lambda}}_{ij}^{*(b)\prime}\widehat{\mathbf{F}}_{it}^{*(b)} - \widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\mathbf{F}}_{it}\right) \le \tau\right).$$

The $1-\alpha$ CI for the z-th (rotated) local factor is given by

$$\left[\widehat{F}_{it}^z - \frac{1}{\sqrt{N_i}}\widehat{\mathcal{D}}_{F_{it}^z}^{-1}\left(\frac{\alpha}{2}\right), \widehat{F}_{it}^z - \frac{1}{\sqrt{N_i}}\widehat{\mathcal{D}}_{F_{it}^z}^{-1}\left(1-\frac{\alpha}{2}\right)\right] \tag{B.2.17}$$

where $\widehat{\mathcal{D}}_{F_{it}^z}^{-1}(\alpha/2)$ and $\widehat{\mathcal{D}}_{F_{it}^z}^{-1}(1-\alpha/2)$ are the inverse function of $\widehat{\mathcal{D}}_{F_{it}}$ evaluated at $\alpha/2$ and $1-\alpha/2$, respectively. The $1-\alpha$ CI for the local component is given by

$$\left[\widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\mathbf{F}}_{it} - \frac{1}{C_{N_iT}}\widehat{\mathcal{D}}_{\lambda_{ij}'F_{it}}^{-1}\left(\frac{\alpha}{2}\right), \widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\mathbf{F}}_{it} - \frac{1}{C_{N_iT}}\widehat{\mathcal{D}}_{\lambda_{ij}'F_{it}}^{-1}\left(1-\frac{\alpha}{2}\right)\right] \tag{B.2.18}$$

where $\widehat{\mathcal{D}}_{\lambda_{ij}'F_{it}}^{-1}(\alpha/2)$ and $\widehat{\mathcal{D}}_{\lambda_{ij}'F_{it}}^{-1}(1-\alpha/2)$ are the inverse function of $\widehat{\mathcal{D}}_{\lambda_{ij}'F_{it}}$ evaluated at $\alpha/2$ and $1-\alpha/2$.

Using Theorem 2.5.1, we have

$$\sqrt{N_i}\left[\left(\widehat{\mathscr{H}}_i^{*(b)\prime}\right)^{-1}\widehat{\mathbf{F}}_{it}^{*(b)} - \widehat{\mathbf{F}}_{it}\right] \xrightarrow{d^*} N\left(\mathbf{0}, \mathbb{D}_{ii,t}^{(4)*}\right) \tag{B.2.19}$$

where

$$\mathbb{D}_{ii,t}^{(4)*} = \underset{N_1,\ldots,N_R,T\to\infty}{\operatorname{plim}} \frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\widehat{\boldsymbol{\lambda}}_{ij}\widehat{\boldsymbol{\lambda}}_{ik}'\left(\mathbf{e}_{ijt}^{*(b)} + \widehat{S}_{ijt}^{*(b)}\right)\left(\mathbf{e}_{ikt}^{*(b)} + \widehat{S}_{ikt}^{*(b)}\right)$$

with $\widehat{S}_{ikt}^{*(b)}$ being the (t,k) element of $\widehat{\mathbf{S}}_i^{*(b)} = \widehat{\mathbf{G}}\widehat{\boldsymbol{\Gamma}}_i' - \widehat{\mathbf{G}}^{*(b)}\widehat{\boldsymbol{\Gamma}}_i^{*(b)\prime}$. From Lemma B.1.8 and the normalisation $T^{-1}\widehat{\mathbf{F}}_i'\widehat{\mathbf{F}}_i = \mathbf{I}_{r_i}$, it can be easily seen that $\widehat{\mathscr{H}}_i^{*(b)} = \operatorname{diag}(\pm 1) + o_{p^*}(1)$. Therefore, we can replace $\widehat{\mathscr{H}}_i^{*(b)}$ by $\widetilde{\mathscr{H}}_i^{*(b)}$ which is a diagonal matrix with all elements equal to ± 1 and the signs are determined by the signs of the diagonal elements of $\widehat{\mathbf{F}}_i^{*(b)\prime}\widehat{\mathbf{F}}_i$.

Under the assumptions of Theorem 2.5, it follows that

$$\sup_{x\in\mathbb{R}}\left|\Pr{}^*\left(\sqrt{N_i}\left[\boldsymbol{\iota}_{i,z}'\left(\widetilde{\mathscr{H}}_i^{*(b)}\right)^{-1}\widehat{\mathbf{F}}_{it}^{*(b)} - \widehat{F}_{it}^z\right] \le x\right) - \Pr\left(\sqrt{N_i}\left[\widehat{F}_{it}^z - \boldsymbol{\iota}_{i,z}'\widehat{\mathscr{H}}_i'\mathbf{F}_{it}\right] \le x\right)\right| \xrightarrow{p} 0$$

and

$$\sup_{x\in\mathbb{R}}\left|\Pr{}^*\left(C_{N_iT}\left[\widehat{\boldsymbol{\lambda}}_{ij}^{*(b)\prime}\widehat{\mathbf{F}}_{it}^{*(b)} - \widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\mathbf{F}}_{it}\right] \le x\right) - \Pr\left(C_{N_iT}\left[\widehat{\boldsymbol{\lambda}}_{ij}'\widehat{\mathbf{F}}_{it} - \boldsymbol{\lambda}_{ij}'\mathbf{F}_{it}\right] \le x\right)\right| \xrightarrow{p} 0$$

suggesting that the confidence intervals (B.2.17) and (B.2.18) have correct coverage rates for the z-th (rotated) local factor, namely $\iota'_{i,z}\widehat{\mathscr{H}}_i'\mathbf{F}_{it}$, and the local component $\boldsymbol{\lambda}'_{ij}\mathbf{F}_{it}$, for block i.

A simulation is conducted to examine the validity of the hybrid bootstrap procedure. We use the same DGP as in Section 2.5. We fix $R = 3$, $(r_0, r_i) = (2, 2)$, $(\phi_G, \phi_F) = (0.5, 0.5)$, and $(\beta, \phi_e, \kappa) = (0, 0, 1)$. The sample size varies as $N_i \in \{20, 50, 100, 200, 300\}$ with $N_1 = \cdots = N_R$ and $T \in \{50, 100, 200, 300\}$. We focus on the CIs for the first element of the global and local factors as well as their loadings since the corresponding second elements have the same statistical properties. We also investigate the CIs for the global and local components. The bootstrap CIs for $(\mathbb{H} + \mathbb{B})'\mathbf{G}_t$, $(\mathbb{H} + \mathbb{B})^{-1}\boldsymbol{\gamma}_{ij}$, $\boldsymbol{\gamma}'_{ij}\mathbf{G}_t$, $\widehat{\mathscr{H}}_i'\mathbf{F}_{it}$, and $\boldsymbol{\lambda}'_{ij}\mathbf{F}_{it}$ are generated by the by (B.2.13), (B.2.15), (B.2.16), (B.2.17), and (B.2.18) respectively. The CI for $\widehat{\mathscr{H}}_i^{-1}\boldsymbol{\lambda}_{ij}$ is constructed using the estimated covariance matrix. For comparison, the CIs generated by theoretical (infeasible) variances of are also reported. All the CIs are We set the number of bootstrap repetition $B = 599$ and the significance level $\alpha = 0.05$ throughout the study.

Each entry of Table B.1 is the coverage rate calculated as the ratios of CIs that contain the true values over 1000 repetitions. The top panel of Table B.1 shows that the infeasible CIs for the global factors and loadings have coverage rates around 0.95 whilst the coverage rates of the bootstrapped CIs tend to the nominal value as either N_i or T increases. Moreover, the infeasible CI for the global component is always larger than 0.95, whereas the bootstrap CI approach to 0.95 as the sample size increases. The bottom panel of Table B.1 presents the results for the local factor, factor loading, and local component. On one hand, the coverage rates of infeasible and bootstrap CIs for the local factors converge to the nominal value as sample size becomes large. On the other hand, the coverage rates of infeasible and asymptotic CIs for the local factor loadings exhibit a slower speed of convergence, but they tend to 0.95 quickly unless T is small. Additionally, the infeasible CI for the local component converges very slowly, whilst the bootstrap counterpart tends to 0.95 for large samples. To conclude, the above results confirm the validity of our hybrid bootstrap procedure.

Table B.1: Coverage rates for the CIs with $R = 3$, $(r_0, r_i) = (2, 2)$, $(\phi_G, \phi_F) = (0.5, 0.5)$, and $(\beta, \phi_e, \kappa) = (0, 0, 1)$

		$\widehat{\mathbf{G}}_t$		$\widehat{\gamma}_{ij}$		$\widehat{\gamma}'_{ij}\widehat{\mathbf{G}}_t$	
N_i	T	Infeasible	Bootstrap	Infeasible	Bootstrap	Infeasible	Bootstrap
20	50	0.942	0.820	0.962	0.909	0.960	0.905
50	50	0.950	0.914	0.945	0.884	0.968	0.909
100	50	0.934	0.911	0.956	0.880	0.971	0.917
200	50	0.927	0.914	0.955	0.895	0.981	0.901
300	50	0.938	0.918	0.963	0.897	0.975	0.909
20	100	0.971	0.917	0.928	0.900	0.962	0.908
50	100	0.939	0.918	0.953	0.924	0.983	0.921
100	100	0.951	0.936	0.939	0.906	0.984	0.936
200	100	0.938	0.928	0.942	0.914	0.983	0.925
300	100	0.949	0.945	0.943	0.902	0.984	0.933
20	200	0.955	0.912	0.927	0.907	0.973	0.922
50	200	0.946	0.916	0.940	0.927	0.971	0.946
100	200	0.947	0.930	0.937	0.923	0.981	0.940
200	200	0.953	0.942	0.944	0.933	0.986	0.946
300	200	0.953	0.952	0.955	0.932	0.982	0.931
20	300	0.959	0.902	0.934	0.928	0.961	0.911
50	300	0.944	0.918	0.932	0.923	0.976	0.938
100	300	0.953	0.941	0.957	0.941	0.974	0.936
200	300	0.945	0.943	0.948	0.938	0.980	0.931
300	300	0.953	0.948	0.941	0.925	0.983	0.935
		$\widehat{\mathbf{F}}_{it}$		$\widehat{\boldsymbol{\lambda}}_{ij}$		$\widehat{\boldsymbol{\lambda}}'_{ij}\widehat{\mathbf{F}}_{it}$	
N_i	T	Infeasible	Bootstrap	Infeasible	Asymptotic	Infeasible	Bootstrap
20	50	0.910	0.893	0.839	0.745	0.895	0.895
50	50	0.928	0.924	0.920	0.865	0.904	0.908
100	50	0.930	0.920	0.954	0.889	0.880	0.909
200	50	0.926	0.924	0.958	0.890	0.858	0.908
300	50	0.918	0.905	0.952	0.893	0.834	0.893
20	100	0.938	0.886	0.786	0.751	0.877	0.903
50	100	0.926	0.918	0.904	0.853	0.901	0.927
100	100	0.944	0.929	0.945	0.901	0.875	0.916
200	100	0.949	0.944	0.945	0.902	0.843	0.916
300	100	0.953	0.945	0.962	0.914	0.840	0.922
20	200	0.930	0.873	0.780	0.762	0.867	0.902
50	200	0.945	0.929	0.897	0.881	0.871	0.924
100	200	0.947	0.930	0.921	0.906	0.890	0.929
200	200	0.954	0.934	0.949	0.924	0.877	0.937
300	200	0.948	0.928	0.953	0.931	0.882	0.935
20	300	0.933	0.868	0.721	0.693	0.837	0.892
50	300	0.947	0.919	0.879	0.865	0.880	0.932
100	300	0.944	0.932	0.929	0.917	0.881	0.940
200	300	0.969	0.941	0.938	0.924	0.887	0.933
300	300	0.950	0.931	0.949	0.934	0.872	0.946

Each entry shows the coverage rate calculated as the ratios of CIs that contains the true factors or loadings over 1000 repetitions. We set the bootstrap repetition as $B = 599$. The infeasible CIs are generated by the theoretical asymptotic distributions in Theorem 2.4 or 2.5, and the bootstrap CIs for $(\mathbb{H}+\mathbb{B})'\mathbf{G}_t$, $(\mathbb{H}+\mathbb{B})^{-1}\gamma_{ij}$, $\gamma'_{ij}\mathbf{G}_t$, $\widehat{\mathscr{H}}_i'\mathbf{F}_{it}$, and $\boldsymbol{\lambda}'_{ij}\mathbf{F}_{it}$ are generated by the by (B.2.13), (B.2.15), (B.2.16), (B.2.17), and (B.2.18) respectively. The CI for$\widehat{\mathscr{H}}_i^{-1}\boldsymbol{\lambda}_{ij}$ is generated by the estimated covariance matrix. We report the CIs 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 $r_1 = \cdots = r_R$ 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 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.

Appendix C

Appendix to Chapter 3

C.1 Lemmas and Proofs

In Section C.1.1, we establish some useful auxiliary lemmas. Based on these lemmas, Section C.1.2 presents the proofs of the main theoretical results in Section 3.3. Finally, the proofs of the auxiliary lemmas are provided in Section C.1.3. We use the following facts throughout: $T^{-1}\left\|\mathbf{X}_{ij}\right\|^2 = T^{-1}\sum_{t=1}^{T}\left\|\mathbf{X}_{ijt}\right\|^2 = O_p(1)$ due to Assumption 3.A. Therefore, $(N_iT)^{-1}\left\|\mathbf{X}_{ij}\right\|^2 = O_p(1)$. We also have $T^{-1/2}\left\|\mathbf{K}_i\right\| = O_p(1)$ by Assumption 3.B.1 and $T^{-1/2}\left\|\widehat{\mathbf{K}}_i\right\| = O_p(1)$ due to the normalisation $T^{-1}\widehat{\mathbf{K}}_i'\widehat{\mathbf{K}}_i = \mathbf{I}_{r_0+r_i}$.

C.1.1 Auxiliary lemmas

Lemma C.1.1. *Under Assumptions 3.A–3.E, we have:*

$$\sup_{\boldsymbol{\mathcal{K}}_1,\ldots,\boldsymbol{\mathcal{K}}_R\in\mathbb{K}^R}\left\|\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{e}_{ij}\right\| = o_p(1)$$

$$\sup_{\boldsymbol{\mathcal{K}}_1,\ldots,\boldsymbol{\mathcal{K}}_R\in\mathbb{K}^R}\left\|\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\boldsymbol{\theta}_{ij}'\mathbf{K}_i'\mathbf{M}_{\mathcal{K}_i}\mathbf{e}_{ij}\right\| = o_p(1)$$

$$\sup_{\boldsymbol{\mathcal{K}}_1,\ldots,\boldsymbol{\mathcal{K}}_R\in\mathbb{K}^R}\left\|\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\mathbf{P}_{\mathcal{K}_i}\mathbf{e}_{ij}\right\| = o_p(1)$$

where $\mathbb{K} = \left\{\mathbf{K}|T^{-1}\mathbf{K}'\mathbf{K} = \mathbf{I}_{r_0+r_i}\right\}$.

Lemma C.1.2. *Let* $\widehat{\mathbf{H}}_i = (\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i)\left(\mathbf{K}_i'\widehat{\mathbf{K}}_i/T\right)\widehat{\mathbf{V}}_i^{-1}$*. Under Assumptions 3.A–3.E, for all i,* $\widehat{\mathbf{H}}_i$ *is an* $(r_0+r_i)\times(r_0+r_i)$ *invertible matrix,* $\widehat{\mathbf{V}}_i \xrightarrow{p} \mathbf{V}_i$ *where* $\mathbf{V}_i$ *is a* $(r_0+r_i)\times(r_0+r_i)$ *diagonal matrix consisting of the eigenvalues of* $\boldsymbol{\Sigma}_{\mathbf{K}_i}\boldsymbol{\Sigma}_{\boldsymbol{\Theta}_i}$ *which are defined in Assumption 3.B, and*

$$\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{K}}_i - \mathbf{K}_i\widehat{\mathbf{H}}_i\right\| = O_p\left(\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|\right) + O_p\left(\frac{1}{C_{N_iT}}\right).$$

Lemma C.1.3. *Under Assumptions 3.A–3.E, for all i and j, we have*

1. $T^{-1}\mathbf{K}_i'\left(\widehat{\mathbf{K}}_i-\mathbf{K}_i\widehat{\mathbf{H}}_i\right)=O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+O_p\left(C_{N_iT}^{-2}\right)$.
2. $T^{-1}\widehat{\mathbf{K}}_i'\left(\widehat{\mathbf{K}}_i-\mathbf{K}_i\widehat{\mathbf{H}}_i\right)=O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+O_p\left(C_{N_iT}^{-2}\right)$.
3. $T^{-1}\mathbf{X}_{ij}'\left(\widehat{\mathbf{K}}_i-\mathbf{K}_i\widehat{\mathbf{H}}_i\right)=O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+O_p\left(C_{N_iT}^{-2}\right)$.
4. $(N_iT)^{-1}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i}\left(\widehat{\mathbf{K}}_i-\mathbf{K}_i\widehat{\mathbf{H}}_i\right)=O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+O_p\left(C_{N_iT}^{-2}\right)$.
5. $T^{-1}\mathbf{e}_{ij}'\left(\widehat{\mathbf{K}}_i-\mathbf{K}_i\widehat{\mathbf{H}}_i\right)=O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+O_p\left(C_{N_iT}^{-2}\right)$.
6. $(\sqrt{N_i}T)^{-1}\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\left(\widehat{\mathbf{K}}_i-\mathbf{K}_i\widehat{\mathbf{H}}_i\right)=T^{-1/2}O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+N_i^{-1/2}O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+O_p\left(N_i^{-1/2}\right)$ $+O_p\left(C_{N_iT}^{-2}\right)$.
7. $(N_iT)^{-1}\sum_{j=1}^{N_i}\boldsymbol{\theta}_{ij}'\left(\widehat{\mathbf{K}}_i\widehat{\mathbf{H}}_i^{-1}-\mathbf{K}_i\right)'\mathbf{e}_{ij}=\left(N_iT^{-1}\right)O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+O_p\left(N_i^{-1}\right)+O_p\left(C_{N_iT}^{-2}\right)$.
8. $(N_iT)^{-1}\sum_{j=1}^{N_i}\left(\mathbf{X}_{ij}'\mathbf{K}_i/T\right)\left(\mathbf{K}_i'\mathbf{K}_i/T\right)\left(\widehat{\mathbf{K}}_i\widehat{\mathbf{H}}_i^{-1}-\mathbf{K}_i\right)'\mathbf{e}_{ij}=N_i^{-2}\sum_{j=1}^{N_i}\sum_{K=1}^{N_i}\left(\mathbf{X}_{ij}'\mathbf{K}_i/T\right)\times$ $\left(\mathbf{K}_i'\mathbf{K}_i/T\right)\left(\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i\right)^{-1}\boldsymbol{\theta}_{ij}\left(T^{-1}\sum_{t=1}^{T}e_{ijt}e_{ikt}\right)+\left(N_iT^{-1}\right)O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+N_i^{-1/2}O_p\left(C_{N_iT}^{-2}\right)$.
9. $(N_iT)^{-2}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i}\left(\mathbf{e}_{ik}\mathbf{e}_{ik}'-\boldsymbol{\Omega}_{ik}\right)\widehat{\mathbf{K}}_i\boldsymbol{\Pi}_i\boldsymbol{\theta}_{ij}=O_p\left(T^{-1}N_i^{-1}\right)+(N_iT)^{-1/2}\times$ $\left[O_p\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)+O_p\left(C_{N_iT}^{-1}\right)\right]+N_i^{-1/2}O_p\left(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|^2\right)+N_i^{-1/2}O_p\left(C_{N_iT}^{-2}\right)$ *where* $\boldsymbol{\Omega}_{ik}=E\left(\mathbf{e}_{ik}\mathbf{e}_{ik}'\right)$ *and* $\boldsymbol{\Pi}_i=\left(\mathbf{K}_i'\widehat{\mathbf{K}}_i/T\right)^{-1}\left(\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i\right)^{-1}$.
10. $\widehat{\mathbf{H}}_i\widehat{\mathbf{H}}_i'=\left(\mathbf{K}_i'\mathbf{K}_i/T\right)^{-1}+O_p\left(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|\right)+O_p\left(C_{N_iT}^{-2}\right)$.
11. $\left\|\mathbf{P}_{\widehat{K}_i}-\mathbf{P}_{K_i}\right\|^2=O_p\left(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|\right)+O_p\left(C_{N_iT}^{-2}\right)$.
12. $N_i^{-1}\left\|\widehat{\boldsymbol{\Theta}}_i'-\widehat{\mathbf{H}}_i^{-1}\boldsymbol{\Theta}_i'\right\|^2=O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|^2\right)+O_p\left(C_{N_iT}^{-2}\right)$.
13. $N_i^{-1}\left(\widehat{\boldsymbol{\Theta}}_i'-\widehat{\mathbf{H}}_i^{-1}\boldsymbol{\Theta}_i'\right)\boldsymbol{\Theta}_i=O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+O_p\left(C_{N_iT}^{-2}\right)$.
14. $\widehat{\boldsymbol{\Theta}}_i'\widehat{\boldsymbol{\Theta}}_i/N_i-\widehat{\mathbf{H}}_i^{-1}\left(\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i\right)\widehat{\mathbf{H}}_i'^{-1}=O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+O_p\left(C_{N_iT}^{-2}\right)$.
15. $\left(\widehat{\boldsymbol{\Theta}}_i'\widehat{\boldsymbol{\Theta}}_i/N_i\right)^{-1}-\widehat{\mathbf{H}}_i\left(\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i\right)^{-1}\widehat{\mathbf{H}}_i'=O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+O_p\left(C_{N_iT}^{-2}\right)$.
16. $N_i^{-1}\sum_{j=1}^{N_i}\left\|\widehat{\boldsymbol{\theta}}_{ij}-\widehat{\mathbf{H}}_i^{-1}\boldsymbol{\theta}_{ij}\right\|=O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+O_p\left(C_{N_iT}^{-1}\right)$.
17. $N_i^{-1}\sum_{j=1}^{N_i}\left\|T^{-1/2}\mathbf{X}_{ij}\right\|\left\|\widehat{\boldsymbol{\theta}}_{ij}-\widehat{\mathbf{H}}_i^{-1}\boldsymbol{\theta}_{ij}\right\|=O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+O_p\left(C_{N_iT}^{-1}\right)$.

Lemma C.1.4. *Under Assumptions 3.A–3.E, if $T/N_i^2\rightarrow 0$, then*

$$\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)=\mathbf{D}\left(\widehat{\mathbf{K}}_1,\ldots,\widehat{\mathbf{K}}_R\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\left(\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i}-\frac{1}{N_i}\sum_{k=1}^{N_i}a_{i,kj}\mathbf{X}_{ik}'\mathbf{M}_{\widehat{K}_i}\right)\mathbf{e}_{ij}$$

$$+\sum_{i=1}^{R}\frac{N_i}{\sqrt{NT}}\boldsymbol{\zeta}_i + o_p(1)$$

where $a_{i,jk} = \boldsymbol{\theta}_{ij}'\left(\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i\right)\boldsymbol{\theta}_{ik}$ *and*

$$\boldsymbol{\zeta}_i = -\mathbf{D}\left(\widehat{\mathbf{K}}_1,\ldots,\widehat{\mathbf{K}}_R\right)^{-1}\frac{1}{N_iT}\sum_{j=1}^{R}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i}\left[\frac{1}{N_i}\sum_{k=1}^{N_i}E\left(\mathbf{e}_{ik}\mathbf{e}_{ik}'\right)\right]\widehat{\mathbf{K}}_i\boldsymbol{\Pi}_i\boldsymbol{\theta}_{ij} = O_p(1)$$

with $\boldsymbol{\Pi}_i = \left(\mathbf{K}_i'\widehat{\mathbf{K}}_i/T\right)^{-1}\left(\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i\right)^{-1}$.

Lemma C.1.5. *Under Assumptions 3.A-3.E, we have*

$$\begin{aligned}\frac{1}{\sqrt{NT}}\sum_{i=1}^{R}\sum_{j=1}^{N_i}&\left(\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i}-\frac{1}{N_i}\sum_{k=1}^{N_i}a_{i,jk}\mathbf{X}_{ik}'\mathbf{M}_{\widehat{K}_i}\right)\mathbf{e}_{ij}\\
&=\frac{1}{\sqrt{NT}}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\left(\mathbf{X}_{ij}'\mathbf{M}_{K_i}-\frac{1}{N_i}\sum_{k=1}^{N_i}a_{i,jk}\mathbf{X}_{ik}'\mathbf{M}_{K_i}\right)\mathbf{e}_{ij}\\
&+\sum_{i=1}^{R}\left[-\sqrt{\frac{N_i}{N}}\frac{\sqrt{N_iT}}{N_i}\frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\frac{\left(\mathbf{X}_{ij}-\mathbf{W}_{ij}\right)'\mathbf{K}_i}{T}\boldsymbol{\Pi}_i\boldsymbol{\theta}_{ik}\left(\frac{1}{T}\sum_{t=1}^{T}e_{ijt}e_{ikt}\right)\right.\\
&\left.+O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+\sqrt{T}O_p\left(\frac{1}{C_{N_iT}^2}\right)+\sqrt{\frac{N_i}{N}}\sqrt{T}O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|^2\right)\right]\end{aligned}\quad\text{(C.1.1)}$$

where $\mathbf{W}_{ij} = N_i^{-1}\sum_{k=1}^{N_i}a_{i,jk}\mathbf{X}_{ik}$ *and therefore,* $\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)$ *in Lemma C.1.4 can be expressed as*

$$\begin{aligned}\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)=\mathbf{D}\left(\widehat{\mathbf{K}}_1,\ldots,\widehat{\mathbf{K}}_R\right)^{-1}&\frac{1}{\sqrt{NT}}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\left(\mathbf{X}_{ij}'\mathbf{M}_{K_i}-\frac{1}{N_i}\sum_{k=1}^{N_i}a_{i,kj}\mathbf{X}_{ik}'\mathbf{M}_{K_i}\right)\mathbf{e}_{ij}\\
&+\sum_{i=1}^{R}\sqrt{\frac{N_i}{N}}\sqrt{\frac{N_i}{T}}\boldsymbol{\zeta}_i+\sum_{i=1}^{R}\sqrt{\frac{N_i}{N}}\sqrt{\frac{T}{N_i}}\boldsymbol{\xi}_i+o_p(1)\end{aligned}\quad\text{(C.1.2)}$$

where

$$\boldsymbol{\xi}_i = -\mathbf{D}\left(\widehat{\mathbf{K}}_1,\ldots,\widehat{\mathbf{K}}_R\right)^{-1}\frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\frac{\left(\mathbf{X}_{ij}-\mathbf{W}_{ij}\right)'\mathbf{K}_i}{T}\boldsymbol{\Pi}_i\boldsymbol{\theta}_{ik}\left(\frac{1}{T}\sum_{t=1}^{T}e_{ijt}e_{ikt}\right) = O_p(1).$$

Lemma C.1.6. *Under Assumptions 3.A–3.E, we have*

1. $\mathbf{D}\left(\widehat{\mathbf{K}}_1,\ldots,\widehat{\mathbf{K}}_R\right)^{-1}-\mathbf{D}\left(\mathbf{K}_1,\ldots,\mathbf{K}_R\right)^{-1} = o_p(1)$.
2. $\sqrt{T/N}\left[\mathbf{D}\left(\widehat{\mathbf{K}}_1,\ldots,\widehat{\mathbf{K}}_R\right)^{-1}-\mathbf{D}\left(\mathbf{K}_1,\ldots,\mathbf{K}_R\right)^{-1}\right] = o_p(1)$ *if* $T/N^2\to 0$.
3. $\sqrt{N/T}\left[\mathbf{D}\left(\widehat{\mathbf{K}}_1,\ldots,\widehat{\mathbf{K}}_R\right)^{-1}-\mathbf{D}\left(\mathbf{K}_1,\ldots,\mathbf{K}_R\right)^{-1}\right] = o_p(1)$ *if* $N/T^2\to 0$.

4. *For each i, it holds that* $\sqrt{T/N_i}\left(\boldsymbol{\xi}_i - \mathbb{B}_i\right) = o_p(1)$ *if* $T/N_i^2 \to 0$ *where*

$$\mathbb{B}_i = -\mathbf{D}\left(\mathbf{K}_1, \ldots, \mathbf{K}_R\right)^{-1} \frac{1}{N_i} \sum_{j=1}^{N_i} \sum_{k=1}^{N_i} \frac{\left(\mathbf{X}_{ij} - \mathbf{W}_{ij}\right)' \mathbf{K}_i}{T} \mathbb{P}_i \boldsymbol{\theta}_{ik} \left(\frac{1}{T} \sum_{t=1}^{T} \sigma_{i,(jk),(tt)} \right).$$

5. *For each i, it holds that* $\sqrt{N_i/T}\left(\boldsymbol{\zeta}_i - \mathbb{C}_i\right) = o_p(1)$ *if* N_i/T^2 *where*

$$\mathbb{C}_i = -\mathbf{D}\left(\mathbf{K}_1, \ldots, \mathbf{K}_R\right)^{-1} \frac{1}{N_i T} \sum_{j=1}^{N_i} \mathbf{X}_{ij}' \mathbf{M}_{K_i} \left[\frac{1}{N_i} \sum_{k=1}^{N_i} E\left(\mathbf{e}_{ik} \mathbf{e}_{ik}'\right) \right] \mathbf{K}_i \mathbb{P}_i \boldsymbol{\theta}_{ij}.$$

Lemma C.1.7. *Under Assumptions 3.A-3.E and* $E(e_{mjt} e_{hkt}) = \sigma_{(mh),(jk)}$ *for all m, h, j, k, and t, if* $T/N^2 \to 0$ *and* $N/T^2 \to 0$, *then for each block i we have:*

1. $\sqrt{T/N_i}\left(\widehat{\mathbb{B}}_i - \mathbb{B}_i\right) = o_p(1)$ *where*

$$\widehat{\mathbb{B}}_i = \mathbf{D}\left(\widehat{\mathbf{K}}_1, \ldots, \widehat{\mathbf{K}}_R\right)^{-1} \frac{1}{N_i} \sum_{j=1}^{N_i} \sum_{k=1}^{N_i} \left[\frac{\left(\mathbf{X}_{ij} - \widehat{\mathbf{W}}_{ij}\right)' \widehat{\mathbf{K}}_i}{T} \right] \left(\frac{\widehat{\boldsymbol{\Theta}}_i' \widehat{\boldsymbol{\Theta}}_i}{N_i} \right)^{-1} \widehat{\boldsymbol{\theta}}_{ik} \hat{\sigma}_{i,(jk)}$$

and $\hat{\sigma}_{i,(jk)} = T^{-1} \widehat{\mathbf{e}}_{ij}' \widehat{\mathbf{e}}_{ik}$ *with* $\widehat{\mathbf{e}}_{ij} = \mathbf{Y}_{ij} - \mathbf{X}_{ij} \widehat{\boldsymbol{\beta}} - \widehat{\mathbf{K}}_i \widehat{\boldsymbol{\theta}}_{ij}$ *and* $\widehat{\mathbf{e}}_{ik} = \mathbf{Y}_{ik} - \mathbf{X}_{ik} \widehat{\boldsymbol{\beta}} - \widehat{\mathbf{K}}_i \widehat{\boldsymbol{\theta}}_{ik}$.

2. $\sqrt{N_i/T}\left(\widehat{\mathbb{C}}_i - \mathbb{C}_i\right) = o_p(1)$ *where*

$$\widehat{\mathbb{C}}_i = -\mathbf{D}\left(\widehat{\mathbf{K}}_1, \ldots, \widehat{\mathbf{K}}_R\right)^{-1} \frac{1}{N_i T} \sum_{j=1}^{N_i} \mathbf{X}_{ij}' \mathbf{M}_{\widehat{K}_i} \widehat{\mathbb{S}}_i \widehat{\mathbf{K}}_i \left(\frac{\widehat{\boldsymbol{\Theta}}_i' \widehat{\boldsymbol{\Theta}}_i}{N_i} \right)^{-1} \widehat{\boldsymbol{\theta}}_{ij}$$

with $\widehat{\mathbb{S}}_i$ *being the estimator of* $\mathbb{S}_i$ *whose (t, s)-th element is* $\widehat{\mathbb{S}}_{i,(ts)} = N_i^{-1} \sum_{k=1}^{N_i} \hat{e}_{ikt} \hat{e}_{iks}$.

Lemma C.1.8. *Suppose* $\mathbf{A}$ *and* $\mathbf{A} + \mathbf{E}$ *are* $n \times n$ *symmetric matrices and that* $\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2]$, *where* $\mathbf{Q}_1$ *has size* $n \times r$ *and* $\mathbf{Q}_2$ *has size* $n \times (n - r)$, *is an orthogonal matrix such that* $span\left(\mathbf{Q}_1\right)$ *is an invariant subspace for* $\mathbf{A}$. *that is,* $\mathbf{A} \times span\left(\mathbf{Q}_1\right) \subset span\left(\mathbf{A}\right)$. *Partition the matrices* $\mathbf{Q}'\mathbf{A}\mathbf{Q}$ *and* $\mathbf{Q}'\mathbf{E}\mathbf{Q}$ *as follows:*

$$\mathbf{Q}'\mathbf{A}\mathbf{Q} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix} \text{ and } \mathbf{Q}'\mathbf{E}\mathbf{Q} = \begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12}' \\ \mathbf{E}_{12} & \mathbf{E}_{22} \end{bmatrix}.$$

If $sep\left(\mathbf{D}_1, \mathbf{D}_2\right) = \min_{\lambda \in \lambda(\mathbf{D}_1), \mu \in \lambda(\mathbf{D}_2)} |\lambda - \mu| > 0$, *where* $\lambda\left(\mathbf{M}\right)$ *denotes the set of eigenvalues of the matrix* $\mathbf{M}$, *and* $\|\mathbf{E}\|_2 \le sep\left(\mathbf{D}_1, \mathbf{D}_2\right)/5$ *where* $\|.\|_2$ *is the spectral norm of a matrix, then there exists a matrix* $\mathbf{P} \in \mathbb{R}^{(n-r)\times r}$ *with*

$$\|\mathbf{P}\|_2 \le \frac{4}{sep\left(\mathbf{D}_1, \mathbf{D}_2\right)} \|\mathbf{E}_{12}\|_2$$

such that the columns of $\widehat{\mathbf{Q}}_1 = \left(\mathbf{Q}_1 + \mathbf{Q}_2 \mathbf{P}\right)\left(\mathbf{I} + \mathbf{P}'\mathbf{P}\right)^{-1/2}$ *define an orthonormal basis for a subspace that is invariant for* $\mathbf{A} + \mathbf{E}$.

Lemma C.1.9. *Let $\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{K}}_1, \ldots, \widetilde{\mathbf{K}}_R\right)$ be the estimators from minimising (3.3.8). For each i, let $v_{i,k}$ be the k-th largest eigenvalue of $(\mathbf{K}_i'\mathbf{K}_i/T)(\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i)$ and $\tilde{v}_{i,k}$ be the k-th largest eigenvalue of $\widetilde{\boldsymbol{\Sigma}}_i$ defined in (3.3.11). Under Assumption 3.A-3.D, as $N_1, \ldots, N_R, T \to \infty$, we have*

1. $\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\| = \sum_{i=1}^R \frac{N_i}{N} O_p\left(C_{N_iT}^{-1/2}\right)$.
2. *For each i,* $|\tilde{v}_{i,k} - v_{i,k}| = O_p\left(\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|\right) + C_{N_iT}^{-1}$ *for* $k = 1, \ldots, d_i$.
3. *For each i,* $|\tilde{v}_{i,k}| = O_p\left(\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^2\right) + C_{N_iT}^{-2}$ *for* $k = d_i + 1, \ldots, d_{\max}$.

C.1.2 Proofs of the main results

Proof of Lemma 3.1.

Consider the difference of the objective functions:

$$
\begin{aligned}
&\frac{1}{NT}\mathcal{Q}(\mathbf{b}, \mathcal{K}_1, \ldots, \mathcal{K}_R) - \frac{1}{NT}\mathcal{Q}(\boldsymbol{\beta}, \mathbf{K}_1, \ldots, \mathbf{K}_R) = \frac{1}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}(\boldsymbol{\beta}-\mathbf{b})'\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{X}_{ij}(\boldsymbol{\beta}-\mathbf{b}) \\
&+ \frac{1}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}\boldsymbol{\theta}_{ij}'\mathbf{K}_i'\mathbf{M}_{\mathcal{K}_i}\mathbf{K}_i\boldsymbol{\theta}_{ij} + \frac{1}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{e}_{ij} + \frac{2}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}(\boldsymbol{\beta}-\mathbf{b})'\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{K}_i\boldsymbol{\theta}_{ij} \\
&\quad + \frac{2}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}(\boldsymbol{\beta}-\mathbf{b})'\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{e}_{ij} + \frac{2}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}\boldsymbol{\theta}_{ij}'\mathbf{K}_i'\mathbf{M}_{\mathcal{K}_i}\mathbf{e}_{ij} - \frac{1}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\mathbf{M}_{K_i}\mathbf{e}_{ij}
\end{aligned}
$$

By Lemma C.1.1, the fifth and sixth terms are $o_p(1)$. We also have

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{e}_{ij} - \frac{1}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\mathbf{M}_{K_i}\mathbf{e}_{ij} = \frac{1}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\mathbf{e}_{ij} - \frac{1}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\mathbf{P}_{\mathcal{K}_i}\mathbf{e}_{ij} \\
&-\frac{1}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\mathbf{e}_{ij} + \frac{1}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\mathbf{P}_{K_i}\mathbf{e}_{ij} = \frac{1}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\mathbf{P}_{\mathcal{K}_i}\mathbf{e}_{ij} + \frac{1}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\mathbf{P}_{K_i}\mathbf{e}_{ij} = o_p(1)
\end{aligned}
$$

by Lemma C.1.1. We thus obtain

$$
\begin{aligned}
&\frac{1}{NT}\mathcal{Q}(\mathbf{b}, \mathcal{K}_1, \ldots, \mathcal{K}_R) - \frac{1}{NT}\mathcal{Q}(\boldsymbol{\beta}, \mathbf{K}_1, \ldots, \mathbf{K}_R) = \frac{1}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}(\boldsymbol{\beta}-\mathbf{b})'\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{X}_{ij}(\boldsymbol{\beta}-\mathbf{b}) \\
&\qquad + \frac{1}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}\boldsymbol{\theta}_{ij}'\mathbf{K}_i'\mathbf{M}_{\mathcal{K}_i}\mathbf{K}_i\boldsymbol{\theta}_{ij} + \frac{2}{NT}\sum_{i=1}^R\sum_{j=1}^{N_i}(\boldsymbol{\beta}-\mathbf{b})'\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{K}_i\boldsymbol{\theta}_{ij} + o_p(1)
\end{aligned}
$$

Since the terms on the RHS are all scalars, we can write them as

$$
\begin{aligned}
&\frac{1}{NT}\mathcal{Q}(\mathbf{b}, \mathcal{K}_1, \ldots, \mathcal{K}_R) - \frac{1}{NT}\mathcal{Q}(\boldsymbol{\beta}, \mathbf{K}_1, \ldots, \mathbf{K}_R) \\
&= (\boldsymbol{\beta}-\mathbf{b})'\left(\frac{1}{NT}\sum_{i=1}^R\boldsymbol{\mathcal{A}}_{i,1}\right)(\boldsymbol{\beta}-\mathbf{b}) + \frac{1}{NT}\sum_{i=1}^R\boldsymbol{\eta}_i'\boldsymbol{\mathcal{A}}_{i,2}\boldsymbol{\eta}_i + 2(\boldsymbol{\beta}-\mathbf{b})'\left(\frac{1}{NT}\sum_{i=1}^R\boldsymbol{\mathcal{A}}_{i,3}\right)\boldsymbol{\eta}_i + o_p(1).
\end{aligned}
$$

where $\boldsymbol{\eta}_i = \text{vec}\{\mathbf{M}_{\mathcal{K}_i}\mathbf{K}_i\}$, $\boldsymbol{\mathcal{A}}_{i,1} = \sum_{j=1}^{N_i} \mathbf{X}'_{ij}\mathbf{M}_{\mathcal{K}_i}\mathbf{X}_{ij}$, $\boldsymbol{\mathcal{A}}_{i,2} = (\boldsymbol{\Theta}'_i\boldsymbol{\Theta}_i) \otimes \mathbf{I}_T$, and $\boldsymbol{\mathcal{A}}_{i,3} = \sum_{j=1}^{N_i} \boldsymbol{\theta}'_{ij} \otimes \mathbf{X}'_{ij}\mathbf{M}_{\mathcal{K}_i}$. Completing the squares, we obtain

$$\begin{aligned}\frac{1}{NT}\mathcal{Q}\left(\mathbf{b},\boldsymbol{\mathcal{K}}_1,\ldots,\boldsymbol{\mathcal{K}}_R\right) - \frac{1}{NT}\mathcal{Q}\left(\boldsymbol{\beta},\mathbf{K}_1,\ldots,\mathbf{K}_R\right)& \\ = (\boldsymbol{\beta}-\mathbf{b})'\left[\frac{1}{NT}\sum_{i=1}^{R}\left(\boldsymbol{\mathcal{A}}_{i,1} - \boldsymbol{\mathcal{A}}_{i,3}\boldsymbol{\mathcal{A}}_{i,2}^{-1}\boldsymbol{\mathcal{A}}'_{i,3}\right)\right](\boldsymbol{\beta}-\mathbf{b})& \\ + \frac{1}{NT}\sum_{i=1}^{R}\left[\boldsymbol{\eta}'_i + (\boldsymbol{\beta}-\mathbf{b})'\boldsymbol{\mathcal{A}}_{i,3}\boldsymbol{\mathcal{A}}_{i,2}^{-1}\right]\boldsymbol{\mathcal{A}}_{i,2}\left[\boldsymbol{\eta}_i + \boldsymbol{\mathcal{A}}_{i,2}^{-1}\boldsymbol{\mathcal{A}}'_{i,3}(\boldsymbol{\beta}-\mathbf{b})\right] + o_p(1).&\end{aligned}$$

Let

$$\begin{aligned}\widetilde{\mathcal{Q}}\left(\mathbf{b},\boldsymbol{\mathcal{K}}_1,\ldots,\boldsymbol{\mathcal{K}}_R\right) = (\boldsymbol{\beta}-\mathbf{b})'\left[\sum_{i=1}^{R}\left(\boldsymbol{\mathcal{A}}_{i,1} - \boldsymbol{\mathcal{A}}_{i,3}\boldsymbol{\mathcal{A}}_{i,2}^{-1}\boldsymbol{\mathcal{A}}'_{i,3}\right)\right](\boldsymbol{\beta}-\mathbf{b})& \\ + \sum_{i=1}^{R}\left[\boldsymbol{\eta}'_i + (\boldsymbol{\beta}-\mathbf{b})'\boldsymbol{\mathcal{A}}_{i,3}\boldsymbol{\mathcal{A}}_{i,2}^{-1}\right]\boldsymbol{\mathcal{A}}_{i,2}\left[\boldsymbol{\eta}_i + \boldsymbol{\mathcal{A}}_{i,2}^{-1}\boldsymbol{\mathcal{A}}'_{i,3}(\boldsymbol{\beta}-\mathbf{b})\right].&\end{aligned}$$

Using the fact that

$$\mathbf{D}\left(\mathcal{K}_1,\ldots,\mathcal{K}_R\right) = \frac{1}{NT}\sum_{i=1}^{R}\left(\boldsymbol{\mathcal{A}}_{i,1} - \boldsymbol{\mathcal{A}}_{i,3}\boldsymbol{\mathcal{A}}_{i,2}^{-1}\boldsymbol{\mathcal{A}}'_{i,3}\right)$$

is positive definite by Assumption 3.A and $\boldsymbol{\mathcal{A}}_{i,2}$ is positive definite by Assumption 3.B.2, we have

$$\widetilde{\mathcal{Q}}\left(\mathbf{b},\boldsymbol{\mathcal{K}}_1,\ldots,\boldsymbol{\mathcal{K}}_R\right) \geq 0$$

and it attains its minimum uniquely at $(\boldsymbol{\beta},\mathbf{K}_1\mathbf{H}_1,\ldots,\mathbf{K}_R\mathbf{H}_R)$ where $\mathbf{H}_i$ is any invertible matrix. Since

$$\frac{1}{NT}\mathcal{Q}\left(\mathbf{b},\boldsymbol{\mathcal{K}}_1,\ldots,\boldsymbol{\mathcal{K}}_R\right) - \frac{1}{NT}\mathcal{Q}\left(\boldsymbol{\beta},\mathbf{K}_1,\ldots,\mathbf{K}_R\right) = \frac{1}{NT}\widetilde{\mathcal{Q}}\left(\mathbf{b},\boldsymbol{\mathcal{K}}_1,\ldots,\boldsymbol{\mathcal{K}}_R\right) + o_p(1),$$

and $\left(\widehat{\boldsymbol{\beta}},\widehat{\mathbf{K}}_1,\ldots,\widehat{\mathbf{K}}_R\right)$ is the minimiser of $\frac{1}{NT}\mathcal{Q}\left(\mathbf{b},\boldsymbol{\mathcal{K}}_1,\ldots,\boldsymbol{\mathcal{K}}_R\right)$, it must hold that

$$\frac{1}{NT}\mathcal{Q}\left(\widehat{\boldsymbol{\beta}},\widehat{\mathbf{K}}_1,\ldots,\widehat{\mathbf{K}}_R\right) - \frac{1}{NT}\mathcal{Q}\left(\boldsymbol{\beta},\mathbf{K}_1,\ldots,\mathbf{K}_R\right) = \frac{1}{NT}\widetilde{\mathcal{Q}}\left(\widehat{\boldsymbol{\beta}},\widehat{\mathbf{K}}_1,\ldots,\widehat{\mathbf{K}}_R\right) + o_p(1) \leq 0$$

Let $\widehat{\boldsymbol{\eta}}_i = \mathbf{M}_{\widehat{K}_i}\widehat{\mathbf{K}}_i$. If either $\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\| \geq \mathcal{M} > 0$ or $\sum_{i=1}^{R}\widehat{\boldsymbol{\eta}}'_i\boldsymbol{\mathcal{A}}_{i,2}\widehat{\boldsymbol{\eta}}_i \geq \mathcal{M} > 0$ holds, we will have $(NT)^{-1}\widetilde{\mathcal{Q}}\left(\widehat{\boldsymbol{\beta}},\widehat{\mathbf{K}}_1,\ldots,\widehat{\mathbf{K}}_R\right) = O_p(1)$, which leads to the RHS of the above equation larger than zero, resulting in a contradiction. Therefore, we have

$$\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\| = o_p(1) \text{ and } \sum_{i=1}^{R}\frac{N_i}{N}\operatorname{tr}\left\{\frac{\mathbf{K}'_i\mathbf{M}_{\widehat{K}_i}\mathbf{K}_i}{T}\frac{\boldsymbol{\Theta}'_i\boldsymbol{\Theta}_i}{N_i}\right\} = o_p(1).$$

Since $N_i^{-1}\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i$ is positive definite for each i, it follows that

$$\left\|\mathbf{P}_{\widehat{K}_i}-\mathbf{P}_{K_i}\right\|^2=\operatorname{tr}\left\{\left(\mathbf{P}_{\widehat{K}_i}-\mathbf{P}_{K_i}\right)^2\right\}=2\operatorname{tr}\left\{\mathbf{I}_{r_0+r_i}-\frac{\widehat{\mathbf{K}}_i'\mathbf{P}_{K_i}\widehat{\mathbf{K}}_i}{T}\right\}=o_p(1).$$

Finally, we have

$$\frac{\mathbf{K}_i'\mathbf{M}_{\widehat{K}_i}\mathbf{K}_i}{T}=\frac{\mathbf{K}_i'\mathbf{K}_i}{T}-\frac{\mathbf{K}_i'\widehat{\mathbf{K}}_i}{T}\left(\frac{\mathbf{K}_i'\widehat{\mathbf{K}}_i}{T}\right)^{-1}=o_p(1)\text{ for each } i.$$

Due to Assumption 3.B.1, $T^{-1}\mathbf{K}_i'\mathbf{K}_i$ is invertible, so $T^{-1}\mathbf{K}_i'\widehat{\mathbf{K}}_i$ is also invertible.

Q.E.D

Proof of Lemma 3.2. The proof follows from (C.1.2) in Lamma C.1.5. The first term is $O_p(1)$ implied by Lemma C.1.1. The second and third terms are $O_p(1)$ because N_i/N, N_i/T, and T/N_i are $O(1)$ by our assumption for each i.

Q.E.D

Proof of Theorem 3.1. Combining Lemma C.1.5, Lemma C.1.6, and Assumption 3.F, the asymptotic normality of $\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)$ holds.

Q.E.D

Proof of Theorem 3.2. The asymptotic distribution can be achieved by combining (C.1.2), Lemma C.1.7, and Assumption 3.F. The consistency of $\widehat{\mathbb{D}}_0$ and $\widehat{\mathbb{D}}_{bc}$ follow from Proposition 2 in Bai (2009) with minor changes so the details are omitted.

Q.E.D

Proof of Proposition 3.1.

The result for $k=1,\ldots,d_{\max}-1$ follows from Lemma C.1.9. We note that, using Lemma 1, the sum of the eigenvalues is

$$\sum_{\ell=1}^{C_{N_iT}^2}\tilde{v}_{i,\ell}=O_p(1)+\sum_{k=d_i+1}^{C_{N_iT}^2}\tilde{v}_{i,k}=O_p(1)+\sum_{k=d_i+1}^{C_{N_iT}^2}\sum_{i=1}^{R}\left(\frac{N_i}{N}\right)^2O_p\left(\frac{1}{C_{N_iT}}\right).$$

Recall that $\tilde{v}_{i,0}=\sum_{\ell=1}^{C_{N_iT}}\tilde{v}_{i,\ell}/\log\left(C_{N_iT}^2\right)$. We consider two cases, $d_i>0$ and $d_i=0$. When $d_i>0$, $\tilde{v}_{i,0}/\tilde{v}_{i,1}=o_p(1)$, which does not affect the selection of d_i. When $d_i=0$, we have

$$\frac{\tilde{v}_{i,0}}{\tilde{v}_{i,1}}=\frac{\sum_{k=1}^{C_{N_iT}^2}\sum_{i=1}^{R}(N_i/N)^2O_p\left(C_{N_iT}^{-1}\right)}{\log\left(C_{N_iT}^2\right)\sum_{i=1}^{R}(N_i/N)^2O_p\left(C_{N_iT}^{-1}\right)}=\frac{C_{N_iT}^2\sum_{i=1}^{R}(N_i/N)^2O_p\left(C_{N_iT}^{-1}\right)}{\log\left(C_{N_iT}^2\right)\sum_{i=1}^{R}(N_i/N)^2O_p\left(C_{N_iT}^{-1}\right)}\xrightarrow{p}\infty,$$

which completes the proof.

Q.E.D

Proof of Proposition 3.2 and 3.3.

From Lemma 3.2 and C.1.2, we deduce that $T^{-1/2}\left\|\widehat{\mathbf{K}}_i-\mathbf{K}_i\widehat{\mathbf{H}}_i\right\|=O_p\left(C_{N_iT}^{-1}\right)$ since the slope parameters are $\sqrt{NT}$-consistent. Therefore, Lemma 1 of Lin & Shin (2022) is satisfied and the results follow from their Theorem 1–2 without any changes.

Q.E.D

C.1.3 Proofs of the auxiliary lemmas

Proof of Lemma C.1.1.

We show the first equation of the above lemma. Using $\mathbf{M}_{\mathcal{K}_i}=\mathbf{I}_T-T^{-1}\mathcal{K}_i\mathcal{K}_i'$, the first equation becomes

$$\sup_{\boldsymbol{\mathcal{K}}_1,\dots,\boldsymbol{\mathcal{K}}_R\in\mathbb{K}^R}\left\|\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{e}_{ij}\right\|\le\left\|\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\mathbf{e}_{ij}\right\|+\sup_{\boldsymbol{\mathcal{K}}_1,\dots,\boldsymbol{\mathcal{K}}_R\in\mathbb{K}^R}\left\|\frac{1}{N}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\frac{\mathbf{X}_{ij}'\mathcal{K}_i}{T}\frac{\mathcal{K}_i'\mathbf{e}_{ij}}{T}\right\|$$

We consider the first term of the RHS of the above equation

$$\left\|\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\mathbf{e}_{ij}\right\|^2=\left\|\sum_{i=1}^{R}\frac{N_i}{N}\frac{1}{N_iT}\sum_{j=1}^{N_i}\sum_{t=1}^{T}\mathbf{X}_{ijt}e_{ijt}\right\|^2$$
$$\le\sum_{i=1}^{R}\frac{N_i}{N}\left\|\frac{1}{N_iT}\sum_{j=1}^{N_i}\sum_{t=1}^{T}\mathbf{X}_{ijt}e_{ijt}\right\|^2=\sum_{i=1}^{R}\frac{N_i}{N}\left[\frac{1}{N_i^2T^2}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{X}_{ijt}'\mathbf{X}_{iks}e_{ijt}e_{iks}\right]$$

where the inequality follows from the Cauchy-Schwarz inequality. Taking the expectation, we obtain

$$E\left\{\sum_{i=1}^{R}\frac{N_i}{N}\left[\frac{1}{N_i^2T^2}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{X}_{ijt}'\mathbf{X}_{iks}e_{ijt}e_{iks}\right]\right\}$$
$$\le\sum_{i=1}^{R}\frac{N_i}{N}\left[\frac{1}{N_i^2T^2}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\sum_{t=1}^{T}\sum_{s=1}^{T}\sqrt{E\left(\|\mathbf{X}_{ijt}\|^4\right)}\sqrt{E\left(\|\mathbf{X}_{iks}\|^4\right)}\sigma_{i,(jk),(ts)}\right]=\sum_{i=1}^{R}\frac{N_i}{N}O_p\left(\frac{1}{N_iT}\right)$$

where the inequality is due to Cauchy-Schwarz inequality and the equality is a result of Assumptions 3.A, 3.C.2, 3.D, and 3.E. Therefore, we obtain

$$\left\|\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\mathbf{e}_{ij}\right\|=\sum_{i=1}^{R}\sqrt{\frac{N_i}{N}}O_p\left(\frac{1}{\sqrt{N_iT}}\right)=o_p(1).$$

Next, we consider the second term using Cauchy-Schwarz inequality:

$$\left\|\frac{1}{N}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\frac{\mathbf{X}_{ij}'\mathcal{K}_i}{T}\frac{\mathcal{K}_i'\mathbf{e}_{ij}}{T}\right\|\le\sum_{i=1}^{R}\frac{N_i}{N}\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\left\|\frac{1}{T}\mathbf{X}_{ij}'\mathcal{K}_i\right\|^2\right)^{1/2}\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\left\|\frac{1}{T}\mathcal{K}_i'\mathbf{e}_{ij}\right\|^2\right)^{1/2}$$

Obviously, $N_i^{-1}\sum_{j=1}^{N_i}\left\|T^{-1}\mathbf{X}_{ij}'\mathcal{K}_i\right\|^2=O_p(1)$ and therefore, it suffices to show that the last term is small. We then have

$$\frac{1}{N_i}\sum_{j=1}^{N_i}\left\|\frac{1}{T}\mathcal{K}_i'\mathbf{e}_{ij}\right\|^2 = \frac{1}{N_i}\sum_{j=1}^{N_i}\left\|\frac{1}{T}\sum_{t=1}^{T}\mathcal{K}_{it}e_{ijt}\right\|^2 = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathcal{K}_{it}'\mathcal{K}_{is}\frac{1}{N_i}\sum_{j=1}^{N_i}e_{ijt}e_{ijs}$$
$$= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathcal{K}_{it}'\mathcal{K}_{is}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(e_{ijt}e_{ijs}-\sigma_{i,(jj),(ts)}\right) + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathcal{K}_{it}'\mathcal{K}_{is}\frac{1}{N_i}\sum_{j=1}^{N_i}\sigma_{i,(jj),(ts)}$$

Applying Cauchy-Schwarz inequality to the first term, we obtain

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathcal{K}_{it}'\mathcal{K}_{is}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(e_{ijt}e_{ijs}-\sigma_{i,(jj),(ts)}\right)$$
$$\leq \left(\frac{1}{T}\sum_{t=1}^{T}\|\mathcal{K}_{it}\|^2\frac{1}{T}\sum_{s=1}^{T}\|\mathcal{K}_{is}\|^2\right)^{1/2}\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|\frac{1}{N_i}\sum_{j=1}^{N_i}\left(e_{ijt}e_{ijs}-\sigma_{i,(jj),(ts)}\right)\right|^2\right)^{1/2}$$
$$= \left(\frac{1}{T}\sum_{t=1}^{T}\|\mathcal{K}_{it}\|^2\frac{1}{T}\sum_{s=1}^{T}\|\mathcal{K}_{is}\|^2\right)^{1/2}\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|\frac{1}{\sqrt{N_i}}\sum_{j=1}^{N_i}\left(e_{ijt}e_{ijs}-\sigma_{i,(jj),(ts)}\right)\right|^2\right)^{1/2}\frac{1}{\sqrt{N_i}}$$

The first part of the above expression is simply $O_p(1)$ and the second part is also $O_p(1)$ by Assumption 3.C.3.

Moreover, since $\left|N_i^{-1}\sum_{j=1}^{N_i}\sigma_{i,(jj),(ts)}\right| \leq \tau_{i,(ts)}$ by Assumption 3.C.2, we have

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathcal{K}_{it}'\mathcal{K}_{is}\frac{1}{N_i}\sum_{j=1}^{N_i}\sigma_{i,(jj),(ts)} \leq \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathcal{K}_{it}'\mathcal{K}_{is}\tau_{i,(ts)}$$
$$\leq \left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\|\mathcal{K}_{it}'\mathcal{K}_{is}\|^2\right)^{1/2}\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{i,(ts)}^2\right)^{1/2} = O_p(1)\times O\left(\frac{1}{\sqrt{T}}\right),$$

where the second inequality is a result of Cauchy-Schwarz inequality. Consequently, we have

$$\sup_{\boldsymbol{\mathcal{K}}_1,\ldots,\boldsymbol{\mathcal{K}}_R\in\mathbb{K}^R}\left\|\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\frac{\mathbf{X}_{ij}'\mathcal{K}_i}{T}\frac{\mathcal{K}_i'\mathbf{e}_{ij}}{T}\right\| = o_p(1).$$

Combining these results, it can be concluded that

$$\sup_{\boldsymbol{\mathcal{K}}_1,\ldots,\boldsymbol{\mathcal{K}}_R\in\mathbb{K}^R}\left\|\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{e}_{ij}\right\| = \sum_{i=1}^{R}\frac{N_i}{N}O_p\left(\frac{1}{C_{N_iT}^{1/2}}\right) = o_p(1).$$

The rest parts of the lemma can be proved in a similar fashion and hence are not shown.

Q.E.D

Proof of Lemma C.1.2.

By plugging $\mathbf{Y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{K}_i\boldsymbol{\theta}_{ij} + \mathbf{e}_{ij}$ in (3.2.7), we obtain

$$\widehat{\mathbf{K}}_i\widehat{\mathbf{V}}_i - \mathbf{K}_i\left(\frac{\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i}{N_i}\right)\left(\frac{\mathbf{K}_i'\widehat{\mathbf{K}}_i}{T}\right) = \frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{X}_{ij}\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right)\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right)'\mathbf{X}_{ij}'\widehat{\mathbf{K}}_i$$
$$+\frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{X}_{ij}\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right)\boldsymbol{\theta}_{ij}'\mathbf{K}_i'\widehat{\mathbf{K}}_i + \frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{K}_i\boldsymbol{\theta}_{ij}\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right)'\mathbf{X}_{ij}'\widehat{\mathbf{K}}_i + \frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{X}_{ij}\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right)\mathbf{e}_{ij}'\widehat{\mathbf{K}}_i$$
$$+\frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{e}_{ij}\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right)'\mathbf{X}_{ij}'\widehat{\mathbf{K}}_i + \frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{K}_i\boldsymbol{\theta}_{ij}\mathbf{e}_{ij}'\widehat{\mathbf{K}}_i + \frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{e}_{ij}\boldsymbol{\theta}_{ij}'\mathbf{K}_i'\widehat{\mathbf{K}}_i + \frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{e}_{ij}\mathbf{e}_{ij}'\widehat{\mathbf{K}}_i$$
$$= \mathcal{J}_{i,1} + \cdots + \mathcal{J}_{i,8}$$

It then follows that

$$\widehat{\mathbf{K}}_i\widehat{\mathbf{V}}_i\left(\frac{\mathbf{K}_i'\widehat{\mathbf{K}}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i}{N_i}\right)^{-1} - \mathbf{K}_i = \left(\mathcal{J}_{i,1} + \cdots + \mathcal{J}_{i,8}\right)\left(\frac{\mathbf{K}_i'\widehat{\mathbf{K}}_i}{T}\right)^{-1}\left(\frac{\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i}{N_i}\right)^{-1}$$

and

$$\left(\frac{\mathbf{K}_i'\widehat{\mathbf{K}}_i}{T}\right)\widehat{\mathbf{V}}_i - \left(\frac{\mathbf{K}_i'\mathbf{K}_i}{T}\right)\left(\frac{\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i}{N_i}\right)\left(\frac{\mathbf{K}_i'\widehat{\mathbf{K}}_i}{T}\right) = \frac{1}{T}\mathbf{K}_i'\left(\mathcal{J}_{i,1} + \cdots + \mathcal{J}_{i,8}\right).$$

The first term $\mathcal{J}_{i,1}$ is bounded by

$$\frac{1}{\sqrt{T}}\left\|\mathcal{J}_{i,1}\right\| \leq \frac{1}{N_i}\sum_{j=1}^{N_i}\left(\frac{1}{T}\left\|\mathbf{X}_{ij}\right\|^2\right)\left\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right\|^2\left\|\frac{1}{\sqrt{T}}\widehat{\mathbf{K}}_i\right\| = O_p\left(\left\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right\|^2\right) = o_p\left(\left\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right\|\right)$$

The same argument applies for the terms from $\mathcal{J}_{i,2}$ to $\mathcal{J}_{i,5}$ and each of these has stochastic order $O_p\left(\left\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right\|\right)$. The last three terms are the same as those in Theorem 1 of Bai & Ng (2002) and each of these is $O_p(C_{N_iT}^{-1})$.

Therefore, we deduce that

$$\left(\frac{\mathbf{K}_i'\widehat{\mathbf{K}}_i}{T}\right)\widehat{\mathbf{V}}_i - \left(\frac{\mathbf{K}_i'\mathbf{K}_i}{T}\right)\left(\frac{\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i}{N_i}\right)\left(\frac{\mathbf{K}_i'\widehat{\mathbf{K}}_i}{T}\right) = o_p(1).$$

This implies that $\left(\mathbf{K}_i'\widehat{\mathbf{K}}_i/T\right)$ are the non-normalised eigenvectors of $\left(\mathbf{K}_i'\mathbf{K}_i/T\right)\left(\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i\right)$ in the limit. Therefore, we conclude that $\widehat{\mathbf{V}}_i \xrightarrow{p} \mathbf{V}_i$. Furthermore, combining the stochastic orders for $\mathscr{J}_{i,1}, \ldots, \mathscr{J}_{i,8}$, we also have

$$\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{K}}_i - \mathbf{K}_i\widehat{\mathbf{H}}_i\right\| = O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right) + O_p\left(\frac{1}{C_{N_iT}}\right).$$

Q.E.D

Proof of Lemma C.1.3.

Since the equations in this lemma are with respect to each block i and do not involve summation across blocks, the proof follows directly from Bai (2009). Lemmas C.1.3.1–C.1.3.9 correspond to their Lemmas A.3–A.5. Lemma C.1.3.10–C.1.3.11 correspond to their Lemmas A.6–A.7. The rest of the statements correspond to their Lemma A.10. We omit the details thereby.

Q.E.D

Proof of Lemma C.1.4.

From (3.2.6), we obtain

$$\left(\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\boldsymbol{M}_{\widehat{K}_i}\mathbf{X}_{ij}\right)\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)=\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\boldsymbol{M}_{\widehat{K}_i}\mathbf{K}_i\boldsymbol{\theta}_{ij}+\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\boldsymbol{M}_{\widehat{K}_i}\mathbf{e}_{ij}. \tag{C.1.3}$$

Using $\mathbf{M}_{\widehat{K}_i}\mathbf{K}_i=\mathbf{M}_{\widehat{K}_i}\left(\mathbf{K}_i-\widehat{\mathbf{K}}_i\widehat{\mathbf{H}}_i^{-1}\right)$ and Lemma C.1.2, the first term of the RHS of the above equation can be expressed as

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\boldsymbol{M}_{\widehat{K}_i}\mathbf{K}_i\boldsymbol{\theta}_{ij}&=\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\boldsymbol{M}_{\widehat{K}_i}\left(\mathcal{J}_{i,1}+\cdots+\mathcal{J}_{i,8}\right)\boldsymbol{\Pi}_i\boldsymbol{\theta}_{ij}\\&=\sum_{i=1}^{R}\frac{N_i}{N}\frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\boldsymbol{M}_{\widehat{K}_i}\left(\mathcal{J}_{i,1}+\cdots+\mathcal{J}_{i,8}\right)\boldsymbol{\Pi}_i\boldsymbol{\theta}_{ij}=\sum_{i=1}^{R}\frac{N_i}{N}\left(\mathscr{J}_{i,1}+\cdots+\mathscr{J}_{i,8}\right)\end{aligned}$$

Each of $\mathscr{J}_{i,1},\ldots,\mathscr{J}_{i,8}$ is the same as the corresponding terms in Proposition A.2 of Bai (2009), so they have the same stochastic orders except that N is replaced by N_i. Given Lemma C.1.3.1–C.1.3.9, following the same argument of Proposition A.2 in Bai (2009), we have

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\boldsymbol{M}_{\widehat{K}_i}\mathbf{K}_i\boldsymbol{\theta}_{ij}=\sum_{i=1}^{R}\frac{N_i}{N}\Bigg(&\mathscr{J}_{i,2}+\mathscr{J}_{i,7}-\frac{1}{N_iT^2}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i}\left[\frac{1}{N_i}\sum_{k=1}^{N_i}E\left(\mathbf{e}_{ik}\mathbf{e}_{ik}'\right)\right]\widehat{\mathbf{K}}_i\boldsymbol{\Pi}_i\boldsymbol{\theta}_{ij}\\&+o_p\left(\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\|\right)+O_p\left(\frac{1}{T\sqrt{N_i}}\right)+\frac{1}{\sqrt{N_i}}O_p\left(\frac{1}{C_{N_iT}^2}\right)\Bigg).\end{aligned}$$

Plugging the above equation in (C.1.3) and using the fact that

$$\sum_{i=1}^{R}\frac{N_i}{N}\frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i}\mathbf{X}_{ij}-\sum_{i=1}^{R}\frac{N_i}{N}\mathscr{J}_{i,2}=\mathbf{D}\left(\widehat{\mathbf{K}}_1,\ldots,\widehat{\mathbf{K}}_R\right),$$

we obtain

$$\begin{aligned}\left[\mathbf{D}\left(\widehat{\mathbf{K}}_1,\ldots,\widehat{\mathbf{K}}_R\right)+o_p(1)\right]\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)=&\frac{1}{\sqrt{NT}}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\left(\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i}-\frac{1}{N_i}\sum_{k=1}^{N_i}a_{i,jk}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i}\right)\mathbf{e}_{ij}\\&-\sum_{i=1}^{R}\frac{N_i}{\sqrt{NT}}\zeta_i+\sum_{i=1}^{R}\frac{N_i\sqrt{T}}{\sqrt{N}}\left[O_p\left(\frac{\sqrt{N_i}}{\sqrt{NT}}\right)+\frac{\sqrt{N_iT}}{\sqrt{N}}O_p\left(\frac{1}{C_{N_iT}^2}\right)\right].\end{aligned}$$

Since $N_i/N=O(1)$ for each i by Assumption 3.E, if $T/N_i^2\to 0$ for each i, the last term is simply

$o_p(1)$. It is shown by Lemma A.6 of Bai (2009) that $\boldsymbol{\zeta}_i = O_p(1)$. Pre-multiplying $\mathbf{D}\left(\widehat{\mathbf{K}}_1, \ldots, \widehat{\mathbf{K}}_R\right)^{-1}$ on both sides of the above equation yields the expression in Lemma C.1.4.

Q.E.D

Proof of Lemma C.1.5.

Consider

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\left(\mathbf{M}_{K_i}-\mathbf{M}_{\widehat{K}_i}\right)\mathbf{e}_{ij} = \sum_{i=1}^{R}\sqrt{\frac{N_i}{N}}\frac{1}{\sqrt{N_iT}}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\left(\mathbf{M}_{K_i}-\mathbf{M}_{\widehat{K}_i}\right)\mathbf{e}_{ij}.$$

For each i, it is shown by Lemma A.8 of Bai (2009) that

$$\frac{1}{\sqrt{N_iT}}\sum_{j=1}^{N_i}\mathbf{X}_{ij}'\left(\mathbf{M}_{K_i}-\mathbf{M}_{\widehat{K}_i}\right)\mathbf{e}_{ij} = -\frac{\sqrt{N_iT}}{N_i}\frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\frac{\mathbf{X}_{ij}'\mathbf{K}_i}{T}\boldsymbol{\Pi}_i\boldsymbol{\theta}_{ik}\left(\frac{1}{T}\sum_{t=1}^{T}e_{ijt}e_{ikt}\right)$$
$$+\sqrt{T}O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|^2\right)+O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+\sqrt{T}O_p\left(\frac{1}{C_{N_iT}^2}\right).$$

Obviously, replacing $\mathbf{X}_{ij}$ with $\mathbf{W}_{ij}$ in the RHS of the above equation does not change the stochastic orders, so we also have

$$\frac{1}{\sqrt{N_iT}}\sum_{j=1}^{N_i}\mathbf{W}_{ij}'\left(\mathbf{M}_{K_i}-\mathbf{M}_{\widehat{K}_i}\right)\mathbf{e}_{ij} = -\frac{\sqrt{N_iT}}{N_i}\frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{k=1}^{N_i}\frac{\mathbf{W}_{ij}'\mathbf{K}_i}{T}\boldsymbol{\Pi}_i\boldsymbol{\theta}_{ik}\left(\frac{1}{T}\sum_{t=1}^{T}e_{ijt}e_{ikt}\right)$$
$$+\sqrt{T}O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|^2\right)+O_p\left(\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|\right)+\sqrt{T}O_p\left(\frac{1}{C_{N_iT}^2}\right).$$

Combining these results gives (C.1.1), which can be plugged into the expression given in Lemma C.1.4, then (C.1.2) of Lemma C.1.5 follows.

Q.E.D

Proof of Lemma C.1.6.

The proof of statements 1–3 is identical to Lemma A.9(i)–(iii) in Bai (2009). Since statements 4 and 5 are with respect to each block i and do not involve summation across blocks, their proof follow from Lemma A.9(iv)–(v) without any changes.

Q.E.D

Proof of Lemma C.1.7. Again, since these results do not involve summation across i, they can be shown in the same fashion as in Lemmas A.11 and A.12 of Bai (2009) using Lemmas C.1.3.10–C.1.3.17 above.

Q.E.D

Lemma C.1.8 is a useful result from Theorem 8.1.10 of Golub & Van Loan (2013) so we omit the proof. In what follows, we use a similar strategy as in Lam et al. (2011) to show the consistency of $\hat{d}_i$ for each i.

Proof of Lemma C.1.9.

The asymptotic results of Lemma C.1.1, C.1.2, and 3.1 are unchanged by replacing $\mathbf{M}_{\widehat{K}_i}$'s with $\mathbf{M}_{\widetilde{K}_i}$'s for all i, except that now $\left(\mathbf{K}_i'\widetilde{\mathbf{K}}_i/T\right)$ is non-invertible. Consequently, from the establishment of Lemma C.1.1, the first part of this lemma follows.

For convenience, we use $\mathbf{A}^{(a:b)}$ to denote the matrix containing from a to b-th columns of a matrix $\mathbf{A}$ and $\mathbf{A}^{(a)}$ stands for the a-th column vector of $\mathbf{A}$. From Lemma C.1.2, we have

$$\frac{1}{\sqrt{T}}\left\|\widetilde{\mathbf{K}}_i^{(1:d_i)} - \mathbf{K}_i\widetilde{\mathbf{H}}_i^{(1:d_i)}\right\| = O_p\left(\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|\right) + O_p\left(\frac{1}{C_{N_iT}}\right)$$

where $\widetilde{H}_i = \left(\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i\right)\left(\mathbf{K}_i'\widetilde{\mathbf{K}}_i/T\right)\widetilde{\mathbf{V}}_i^{-1}$.

Let $\boldsymbol{\Sigma}_i = T^{-1}\mathbf{K}_i\left(\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i\right)\mathbf{K}_i'$. By the development of Lemma C.1.2, it is straightforward that

$$\left\|\widetilde{\boldsymbol{\Sigma}}_i - \boldsymbol{\Sigma}_i\right\| = O_p\left(\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|\right) + O_p\left(\frac{1}{C_{N_iT}}\right)$$

and

$$\left(\frac{\mathbf{K}_i'\widetilde{\mathbf{K}}_i^{(1:d_i)}}{T}\right)\widetilde{\mathbf{V}}_i^{(1:d_i)} - \left(\frac{\mathbf{K}_i'\mathbf{K}_i}{T}\right)\left(\frac{\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i}{N_i}\right)\left(\frac{\mathbf{K}_i'\widetilde{\mathbf{K}}_i^{(1:d_i)}}{T}\right) = O_p\left(\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|\right) + O_p\left(\frac{1}{C_{N_iT}}\right).$$

By continuity of the eigenvalues, the first equation suggests that the non-zero eigenvalues of $\widetilde{\boldsymbol{\Sigma}}_i$ converges to the non-zero eigenvalues of $\boldsymbol{\Sigma}_i$. Notice that $\left(\mathbf{K}_i'\mathbf{K}_i/T\right)\left(\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i\right)$ has the same eigenvalues as $\boldsymbol{\Sigma}_i$, so the second equation implies the convergence rate of the first d_i largest eigenvalues of $\widetilde{\boldsymbol{\Sigma}}_i$. This completes the proof of the second statement.

For each i, let $\mathbf{K}_i^{\perp}$ be a $T \times (T - d_i)$ matrix such that

$$\frac{1}{T}\left[\mathbf{K}_i^{\perp}, \mathbf{K}_i\mathbf{H}_i\right]'\left[\mathbf{K}_i^{\perp}, \mathbf{K}_i\mathbf{H}_i\right] = \mathbf{I}_T$$

where $\mathbf{H}_i = \left(\mathbf{K}_i'\mathbf{K}_i/T\right)^{-1/2}$. The existence of such $\mathbf{K}_i^{\perp}$ can be easily verified since $\mathbf{K}_i\mathbf{H}_i$ is full rank. Moreover, because $\mathbf{H}_i$ is non-zero, the above equation implies that $\mathbf{K}_i^{\perp\prime}\mathbf{K}_i = \mathbf{0}$. We apply Lemma C.1.8 by replacing $\mathbf{Q}_1$, $\mathbf{Q}_2$, $\mathbf{A}$, and $\mathbf{E}$ with $T^{-1/2}\mathbf{K}_i^{\perp}$, $T^{-1/2}\mathbf{K}_i\mathbf{H}_i$, $\boldsymbol{\Sigma}_i$, and $\widetilde{\boldsymbol{\Sigma}}_i - \boldsymbol{\Sigma}_i$. Then, it is straightforward that $\mathbf{Q}'\mathbf{A}\mathbf{Q}$ in Lemma C.1.8 becomes

$$\begin{aligned}\frac{1}{\sqrt{T}}\left[\mathbf{K}_i^{\perp}, \mathbf{K}_i\mathbf{H}_i\right]'\boldsymbol{\Sigma}_i\frac{1}{\sqrt{T}}\left[\mathbf{K}_i^{\perp}, \mathbf{K}_i\mathbf{H}_i\right] &= \frac{1}{\sqrt{T}}\left[\mathbf{K}_i^{\perp}, \mathbf{K}_i\mathbf{H}_i\right]'\frac{1}{\sqrt{T}}\mathbf{K}_i\left(\frac{\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i}{N_i}\right)\frac{1}{\sqrt{T}}\mathbf{K}_i'\frac{1}{\sqrt{T}}\left[\mathbf{K}_i^{\perp}, \mathbf{K}_i\mathbf{H}_i\right] \\ &= \begin{bmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_i'\left(\frac{\mathbf{K}_i'\mathbf{K}_i}{T}\right)\left(\frac{\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i}{N_i}\right)\left(\frac{\mathbf{K}_i'\mathbf{K}_i}{T}\right)\mathbf{H}_i\end{bmatrix}\end{aligned}$$

so $\mathbf{D}_1$ in Lemma C.1.8 becomes $\mathbf{0}$ and, $\mathbf{D}_2$ is now the lower block-diagonal element of the above matrix, which has the same eigenvalues as $\boldsymbol{\Sigma}_i$. Additionally, we also have an orthonormal basis for a subspace that is invariant of $\widetilde{\boldsymbol{\Sigma}}_i$ such that

$$\widetilde{\mathbf{K}}_i^{\perp} = \frac{1}{\sqrt{T}}\left(\mathbf{K}_i^{\perp} + \mathbf{K}_i\mathbf{H}_i\mathbf{P}_i\right)\left(\mathbf{I}_{T-d_i} + \mathbf{P}_i'\mathbf{P}_i\right)^{-1/2}.$$

By Lemma C.1.8, we have

$$\|\mathbf{P}_i\| \leq d_i \times \|\mathbf{P}_i\|_2 \leq \frac{4d_i}{sep\left(\mathbf{0}, \boldsymbol{\Sigma}_i\right)}\left\|\widetilde{\boldsymbol{\Sigma}}_i - \boldsymbol{\Sigma}_i\right\|_2 \leq O_p(1) \times \left\|\widetilde{\boldsymbol{\Sigma}}_i - \boldsymbol{\Sigma}_i\right\| = O_p\left(\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|\right) + O_p\left(\frac{1}{C_{N_iT}}\right)$$

where the first and last inequality follows from the property of the spectral norm. Moreover, we have

$$\begin{aligned}
\left\|\widetilde{\mathbf{K}}_i^{\perp} - \frac{1}{\sqrt{T}}\mathbf{K}_i^{\perp}\right\| &= \frac{1}{\sqrt{T}}\left\|\left[\left(\mathbf{K}_i^{\perp} + \mathbf{K}_i\mathbf{H}_i\mathbf{P}_i\right) - \mathbf{K}_i^{\perp}\left(\mathbf{I}_{T-d_i} + \mathbf{P}_i'\mathbf{P}_i\right)^{1/2}\right]\left(\mathbf{I}_{T-d_i} + \mathbf{P}_i'\mathbf{P}_i\right)^{-1/2}\right\| \\
&\leq \frac{1}{\sqrt{T}}\left\|\mathbf{K}_i^{\perp}\left[\mathbf{I}_{T-d_i} - \left(\mathbf{I}_{T-d_i} + \mathbf{P}_i'\mathbf{P}_i\right)^{1/2}\right]\left(\mathbf{I}_{T-d_i} + \mathbf{P}_i'\mathbf{P}_i\right)^{-1/2}\right\| + \left\|\mathbf{P}_i\left(\mathbf{I}_{T-d_i} + \mathbf{P}_i'\mathbf{P}_i\right)^{-1/2}\right\| \\
&\leq O_p(1)\left\|\mathbf{I}_{T-d_i} - \left(\mathbf{I}_{T-d_i} + \mathbf{P}_i'\mathbf{P}_i\right)^{-1/2}\right\| + O_p(1)\|\mathbf{P}_i\| = O_p\left(\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|\right) + O_p\left(\frac{1}{C_{N_iT}}\right)
\end{aligned}$$

where the inequalities follow from the Cauchy-Schwarz inequality. Finally, we can obtain $\tilde{v}_{i,d_i+\ell}$ by

$$\begin{aligned}
\tilde{v}_{i,d_i+\ell} &= \widetilde{\mathbf{K}}_i^{\perp(\ell)}\widetilde{\boldsymbol{\Sigma}}_i'\widetilde{\mathbf{K}}_i^{\perp(\ell)} \\
&= \left(\widetilde{\mathbf{K}}_i^{\perp(\ell)} - \frac{1}{\sqrt{T}}\mathbf{K}_i^{\perp} + \frac{1}{\sqrt{T}}\mathbf{K}_i^{\perp}\right)'\left(\widetilde{\boldsymbol{\Sigma}}_i' - \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_i\right)\left(\widetilde{\mathbf{K}}_i^{\perp(\ell)} - \frac{1}{\sqrt{T}}\mathbf{K}_i^{\perp} + \frac{1}{\sqrt{T}}\mathbf{K}_i^{\perp}\right) \\
&\leq \left\|\widetilde{\mathbf{K}}_i^{\perp(\ell)} - \frac{1}{\sqrt{T}}\mathbf{K}_i^{\perp}\right\|^2\left\|\widetilde{\boldsymbol{\Sigma}}_i' - \boldsymbol{\Sigma}_i\right\| + 2\left\|\widetilde{\mathbf{K}}_i^{\perp(\ell)} - \frac{1}{\sqrt{T}}\mathbf{K}_i^{\perp}\right\|\left\|\widetilde{\boldsymbol{\Sigma}}_i' - \boldsymbol{\Sigma}_i\right\|\left\|\frac{1}{\sqrt{T}}\mathbf{K}_i^{\perp}\right\| \\
&\qquad + \left\|\widetilde{\mathbf{K}}_i^{\perp(\ell)} - \frac{1}{\sqrt{T}}\mathbf{K}_i^{\perp}\right\|^2\|\boldsymbol{\Sigma}_i\| = O_p\left(\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^2\right) + O_p\left(\frac{1}{C_{N_iT}^2}\right)
\end{aligned}$$

where the inequality follows from Cauchy-Schwarz inequality and the fact that the other summands are simply zeros by construction. This completes the proof of the third statement.

Q.E.D

C.2 Extension to the Heterogeneous Coefficient Model

In this section, we extend the model (3.2.1) to a heterogeneous coefficient model. The model and estimators are presented in Section C.2.1. Sections C.2.2–C.2.4 provide the mains asymptotic results for the estimators. Section C.2.5 outlines the auxiliary lemmas that are useful for the proofs of the main results in Section C.2.6. The proofs of the auxiliary lemmas can be found in Section C.2.7.

C.2.1 The model

The heterogeneous coefficient model is specified as follows:

$$\begin{aligned} y_{ijt} &= \mathbf{X}_{ijt}'\boldsymbol{\beta}_{ij} + u_{ijt},\ i = 1, ..., R,\ j = 1, ..., N_i,\ t = 1, ..., T \\ u_{ijt} &= \boldsymbol{\gamma}_{ij}'\mathbf{G}_t + \boldsymbol{\lambda}_{ij}'\mathbf{F}_{it} + e_{ijt} \end{aligned} \tag{C.2.4}$$

where $\boldsymbol{\beta}_{ij} = \left[\beta_{ij}^1, \ldots, \beta_{ij}^p\right]'$ is now individually specific while the other model setups remain the same as in (3.2.1). Let $\{\mathbf{b}_{ij}\}_{i=1,\ldots,R}^{j=1,\ldots,N_i}$ be the set of slope coefficients for all the individuals. The above heterogeneous coefficient model extends the ones considered by Pesaran (2006), Song (2013), and Li et al. (2020) among others.

We define the objective function as:

$$\mathcal{S}\left(\{\mathbf{b}_{ij}\}_{i=1,\ldots,R}^{j=1,\ldots,N_i}, \boldsymbol{\mathcal{K}}_1, \ldots, \boldsymbol{\mathcal{K}}_R\right) = \sum_{i=1}^{R}\sum_{j=1}^{N_i} \left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\mathbf{b}_{ij}\right)' \mathbf{M}_{\mathcal{K}_i} \left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\mathbf{b}_{ij}\right) \tag{C.2.5}$$

where $\boldsymbol{\mathcal{K}}_i$ is a $T \times d_i$ matrix for all i. The estimators are given by

$$\left(\left\{\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}\right\}_{i=1,\ldots,R}^{j=1,\ldots,N_i}, \widehat{\mathbf{K}}_1^{\dagger}, \ldots, \widehat{\mathbf{K}}_R^{\dagger}\right) = \underset{\left(\{\boldsymbol{\beta}_{ij}\}_{i=1,\ldots,R}^{j=1,\ldots,N_i}, \boldsymbol{\mathcal{K}}_1,\ldots,\boldsymbol{\mathcal{K}}_R\right)\in\mathcal{D}^{\dagger}}{\arg\min} \mathcal{S}\left(\{\mathbf{b}_{ij}\}_{i=1,\ldots,R}^{j=1,\ldots,N_i}, \boldsymbol{\mathcal{K}}_1, \ldots, \boldsymbol{\mathcal{K}}_R\right)$$

where $\mathcal{D}^{\dagger} = \mathbb{R}^p \times \cdots \times \mathbb{R}^p \times \mathbb{K}^R$ and $\mathbb{K} = \left\{\mathbf{K} | T^{-1}\mathbf{K}'\mathbf{K} = \mathbf{I}_{d_i}\right\}$. The above objective function implies that the solutions are the following

$$\widehat{\boldsymbol{\beta}}_{ij}^{\dagger} = \left(\mathbf{X}_{ij}'\boldsymbol{M}_{\widehat{K}_i^{\dagger}}\mathbf{X}_{ij}\right)^{-1}\mathbf{X}_{ij}'\boldsymbol{M}_{\widehat{K}_i^{\dagger}}\mathbf{Y}_{ij} \text{ for each } i \text{ and } j \tag{C.2.6}$$

$$\widehat{\mathbf{K}}_i^{\dagger}\widehat{\mathbf{V}}_i^{\dagger} = \left[\frac{1}{N_i T}\sum_{j=1}^{N_i}\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}\right)\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}\right)'\right]\widehat{\mathbf{K}}_i^{\dagger} \text{ for each } i \tag{C.2.7}$$

where $\widehat{\mathbf{V}}_i^{\dagger}$ is a diagonal matrix consisting of the $d_i = r_0 + r_i$ largest eigenvalues of the term in the square bracket of the above term. Similar to the homogeneous model in (3.2.1), the estimates $\left\{\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}\right\}_{i=1,\ldots,R}^{j=1,\ldots,N_i}$ and $\left(\widehat{\mathbf{K}}_1^{\dagger}, \ldots, \widehat{\mathbf{K}}_R^{\dagger}\right)$ are obtained iteratively.

C.2.2 Consistency and asymptotic distribution

To ensure consistency of the slope coefficients and factor estimates, we make the following assumption for identification similar to Assumption 3.A:

Assumption C.A.

1. *For all i, j, and t , we have $E\left(\left\|\mathbf{X}_{ijt}\right\|^4\right) \leq \mathcal{M}$. Furthermore, for all i and j, we have*

$$\inf_{\mathcal{K}_i \in \mathbb{K}} \frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{X}_{ij} > 0 \text{ and } \frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{X}_{ij} \xrightarrow{p} \mathbf{\Omega}_{ij} > 0.$$

2. *For all i, we have:*

$$\inf_{\mathcal{K}_i \in \mathbb{K}} \mathbf{D}_i\left(\mathcal{K}_i\right) > 0$$

where

$$\mathbf{D}_i\left(\mathcal{K}_i\right) = \frac{1}{N_i}\sum_{j=1}^{N_i}\left(\mathcal{A}_{ij,1} - \mathcal{A}_{ij,3}\mathcal{A}_{ij,2}^{-1}\mathcal{A}_{ij,3}'\right)$$

with $\mathcal{A}_{ij,1} = T^{-1}\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{X}_{ij}$, $\mathcal{A}_{ij,2} = \left(\boldsymbol{\theta}_{ij}\boldsymbol{\theta}_{ij}'\right) \otimes T^{-1}\mathbf{I}_T$, and $\mathcal{A}_{ij,3} = \boldsymbol{\theta}_{ij}' \otimes \left(T^{-1}\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\right)$.

Assumption C.A.1 resembles Assumption 5 of Pesaran (2006), which guarantees the identification of each $\boldsymbol{\beta}_{ij}$. Assumption C.A.1 rules out time-invariant regressors and the case that $\mathbf{X}_{ij}$ is perfectly collinear with the true factors $\mathbf{K}_i$. Assumption C.A.2 extends Assumption 3.A and ensures the consistency of $\widehat{\boldsymbol{\beta}}_{ij}$'s and $\widehat{\mathbf{K}}_i$'s. Similar assumptions can be found in Song (2013) and Ando & Bai (2014). It is shown by Lemma C.2.1 that the consistent estimation of $\boldsymbol{\beta}_{ij}$'s and $\mathbf{K}_i$'s for all i and j does not require separate identification of $\mathbf{G}$ and $\mathbf{F}_i$'s. Therefore, minimising (C.2.5) is equivalent to implementing *IPC* in Song (2013) block by block. To derive the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}$, we make the following assumption:

Assumption C.B. *For each i and j, we have*

1.
$$\frac{1}{N_i^2 T}\sum_{k=1}^{N_i}\sum_{\ell=1}^{N_i}\mathbf{X}_{ik}'\mathbf{M}_{K_i}E\left(\mathbf{e}_{ik}\mathbf{e}_{i\ell}'\right)\mathbf{M}_{K_i}\mathbf{X}_{i\ell} = o_p(1).$$

2.
$$\frac{1}{\sqrt{T}}\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{e}_{ij} \xrightarrow{d} N\left(\mathbf{0}, \mathbb{D}_{ij}\right)$$

where

$$\mathbb{D}_{ij} = \text{plim}\,\frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{K_i}E\left(\mathbf{e}_{ij}\mathbf{e}_{ij}'\right)\mathbf{M}_{K_i}\mathbf{X}_{ij}$$

is a non-random positive definite matrix.

Assumption C.B.1 follows from Assumption G of Ando & Bai (2014) which restricts the cross-section correlation. It is less restrictive than the cross-section independence assumed in Song (2013) as it still allows weak cross-section and serial correlation among the error terms. Assumption C.B.2 is the central limit theorem for deriving the asymptotic distribution of $\widehat{\boldsymbol{\beta}}^{\dagger}_{ij}$'s.

The asymptotic normality is summarised in the following theorem:

Theorem C.2.1. *Under Assumptions 3.B–3.E and C.A–C.B, as $N_1, \dots, N_R, T \to \infty$, and $T/N_i^2 \to 0$ for all i, it follows that*

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}^{\dagger}_{ij} - \boldsymbol{\beta}_{ij}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Omega}_{ij}^{-1}\mathbb{D}_{ij}\boldsymbol{\Omega}_{ij}^{-1}\right)$$

where

$$\boldsymbol{\Omega}_{ij} = \operatorname{plim}\frac{\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{X}_{ij}}{T} \text{ and } \mathbb{D}_{ij} = \operatorname{plim}\frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{K_i}E\left(\mathbf{e}_{ij}\mathbf{e}_{ij}'\right)\mathbf{M}_{K_i}\mathbf{X}_{ij}.$$

C.2.3 Determining the numbers of factors

So far, we have assumed that the number of factors d_i's are know. Lemma C.2.1 together with (C.2.7) offers a consistent model selection alternative to the information criterion proposed by Ando & Bai (2014). Following Section 3.3.2, we set a finite integer that is common to all blocks, $d_{\max}$, such that $d_{\max} \le \max_i\{d_i\}$. The initial estimates $\widetilde{\boldsymbol{\beta}}^{\dagger}_{ij}$'s can be obtained consistently via minimising the objective function defined as follows:

$$\mathscr{S}\left(\{\mathbf{b}_{ij}\}_{i=1,\dots,R}^{j=1,\dots,N_i}, \mathscr{K}_1, \dots, \mathscr{K}_R\right) = \sum_{i=1}^{R}\sum_{j=1}^{N_i}\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\mathbf{b}_{ij}\right)'\mathbf{M}_{\mathscr{K}_i}\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\mathbf{b}_{ij}\right) \tag{C.2.8}$$

where $\mathscr{K}_i$ is a $T \times d_{\max}$ matrix for all i. The estimators are given by

$$\left(\left\{\widetilde{\boldsymbol{\beta}}^{\dagger}_{ij}\right\}_{i=1,\dots,R}^{j=1,\dots,N_i}, \widetilde{\mathbf{K}}^{\dagger}_1, \dots, \widetilde{\mathbf{K}}^{\dagger}_R\right) = \underset{\left(\{\boldsymbol{\beta}_{ij}\}_{i=1,\dots,R}^{j=1,\dots,N_i}, \mathscr{K}_1,\dots,\mathscr{K}_R\right)\in\mathscr{D}^{\dagger}}{\arg\min} \mathscr{S}\left(\{\mathbf{b}_{ij}\}_{i=1,\dots,R}^{j=1,\dots,N_i}, \mathscr{K}_1, \dots, \mathscr{K}_R\right)$$

where $\mathscr{D}^{\dagger} = \mathbb{R}^p \times \cdots \times \mathbb{R}^p \times \widetilde{\mathbb{K}}^R$ and $\widetilde{\mathbb{K}} = \left\{\mathbf{K} | T^{-1}\mathbf{K}'\mathbf{K} = \mathbf{I}_{d_{\max}}\right\}$. The above objective function implies that the solutions are the following

$$\widetilde{\boldsymbol{\beta}}^{\dagger}_{ij} = \left(\mathbf{X}_{ij}'\boldsymbol{M}_{\widetilde{K}_i^{\dagger}}\mathbf{X}_{ij}\right)^{-1}\mathbf{X}_{ij}'\boldsymbol{M}_{\widetilde{K}_i^{\dagger}}\mathbf{Y}_{ij} \text{ for each } i \text{ and } j$$

$$\widetilde{\mathbf{K}}^{\dagger}_i\widetilde{\mathbf{V}}^{\dagger}_i = \left[\frac{1}{N_iT}\sum_{j=1}^{N_i}\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\widetilde{\boldsymbol{\beta}}^{\dagger}_{ij}\right)\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\widetilde{\boldsymbol{\beta}}^{\dagger}_{ij}\right)'\right]\widetilde{\mathbf{K}}^{\dagger}_i \text{ for each } i$$

where $\widetilde{\mathbf{V}}^{\dagger}_i$ is a diagonal matrix consisting of the $d_i = r_0 + r_i$ largest eigenvalues of the term in the square bracket of the above term. Define the covariance matrix for each block i:

$$\widetilde{\boldsymbol{\Sigma}}^{\dagger}_i = \frac{1}{N_iT}\sum_{j=1}^{N_i}\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\widetilde{\boldsymbol{\beta}}^{\dagger}_{ij}\right)\left(\mathbf{Y}_{ij} - \mathbf{X}_{ij}\widetilde{\boldsymbol{\beta}}^{\dagger}_{ij}\right)'. \tag{C.2.9}$$

Let $\tilde{v}_{i,k}^{\dagger}$ be the k-th largest eigenvalue of $\widetilde{\boldsymbol{\Sigma}}_i^{\dagger}$. It can be shown that for $k = 1, \ldots, d_i$, $\tilde{v}_{i,k}^{\dagger}$ converges to its population counterpart $v_{i,k}$, which is larger than zero. On the contrary, $\tilde{v}_{i,k}^{\dagger}$ is asymptotically zero for $k = d_i + 1, \ldots, d_{\max}$. d_i can be estimated as:

$$\hat{d}_i^{\dagger} = \underset{k=0,\ldots,d_{\max}-1}{\arg\max} \frac{\tilde{v}_{i,k}^{\dagger}}{\tilde{v}_{i,k+1}^{\dagger}} \text{ for } i = 1, \ldots, R$$

where $\tilde{v}_{i,0}^{\dagger} = \sum_{\ell=1}^{C_{N_iT}^2} \tilde{v}_{i,\ell}^{\dagger} / \log\left(C_{N_iT}^2\right)$ is a mock eigenvalue allowing the case where $d_i = 0$. It can be shown that, for each i, it is straightforward that for $k = 1, \ldots, d_i - 1, d_i + 1, \ldots, d_{\max} - 1$, $\tilde{v}_{i,k}^{\dagger} / \tilde{v}_{i,k+1}^{\dagger} = O_p(1)$, whilst for $k = d_i$ we have $\tilde{v}_{i,k}^{\dagger} / \tilde{v}_{i,k+1}^{\dagger} \to \infty$. Therefore, the ratio of the eigenvalues falls sharply at $k = d_i$. The mock eigenvalue $\tilde{v}_{i,0}^{\dagger}$ is smaller in magnitude than the largest d_i eigenvalues but larger than the rest, which allows the case $d_i = 0$. The consistency of $\hat{d}_i^{\dagger}$'s is summarised in the following proposition:

Proposition C.2.1. *Under Assumptions 3.B–3.E and C.A–C.B, we have*

$$\lim_{N_i,T\to\infty} \Pr\left(\hat{d}_i^{\dagger} = r_0 + r_i\right) = 1.$$

C.2.4 Disentangling the global and local factors

Given the consistent estimates $\hat{d}_i^{\dagger}$'s and $\widehat{\mathbf{K}}_i^{\dagger}$'s, we construct the following $T(R-1)R/2 \times \hat{d}^*$ system-wide matrix where $\hat{d}^{\dagger *} = \sum_{l=1}^{R} \hat{d}_l^{\dagger}$:

$$\widehat{\boldsymbol{\Phi}}^{\dagger} = \begin{bmatrix} \widehat{\mathbf{K}}_1^{\dagger} & -\widehat{\mathbf{K}}_2^{\dagger} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \widehat{\mathbf{K}}_1^{\dagger} & \mathbf{0} & -\widehat{\mathbf{K}}_3^{\dagger} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ & & & & \vdots & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \widehat{\mathbf{K}}_{R-1}^{\dagger} & -\widehat{\mathbf{K}}_R^{\dagger} \end{bmatrix} \tag{C.2.10}$$

and perform an SVD to $\widehat{\boldsymbol{\Phi}}^{\dagger}$ as $\widehat{\boldsymbol{\Phi}}^{\dagger} = \widehat{\mathbf{P}}^{\dagger}\widehat{\boldsymbol{\Delta}}^{\dagger}\widehat{\mathbf{Q}}^{\dagger\prime}$. Let $\hat{\delta}_1^{\dagger}, \ldots, \hat{\delta}_{\hat{d}^*}^{\dagger}$ be the diagonal elements of $\widehat{\boldsymbol{\Delta}}^{\dagger}$ in *ascending order*. Then, the number of global factors r_0 can be obtained by

$$\hat{r}_0^{\dagger} = \underset{k=0,\ldots,\hat{d}_{\min}^{\dagger}}{\arg\max} \frac{\hat{\delta}_{k+1}^{\dagger 2}}{\hat{\delta}_k^{\dagger 2}}, \; \hat{d}_{\min}^{\dagger} = \min\left\{\hat{d}_1^{\dagger}, \ldots, \hat{d}_R^{\dagger}\right\}.$$

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

$$\hat{\delta}_0^{\dagger 2} = \frac{1}{C_{\underline{N}T}\hat{d}^{\dagger *}} \sum_{k=1}^{\hat{d}^*} \hat{\delta}_k^{\dagger 2}.$$

The number of local factors can be simply obtained by $\hat{r}_i^{\dagger} = \hat{d}_i^{\dagger} - \hat{r}_0^{\dagger}$ for each block i. The consistency of $\hat{r}_0^{\dagger}$ and $\hat{r}_i^{\dagger}$'s are summarised in the following proposition:

Proposition C.2.2. *Under Assumptions 3.B–3.E, 3.G, and C.A–C.B, we have:*

$$\lim_{N_1,\ldots,N_R,T\to\infty} \Pr\left(\hat{r}_0^\dagger = r_0\right) = 1 \text{ and } \lim_{N_1,\ldots,N_R,T\to\infty} \Pr\left(\hat{r}_i^\dagger = r_i\right) = 1$$

Let $\widehat{\mathbf{Q}}^{\dagger\hat{r}_0^\dagger} = \left[\widehat{\mathbf{Q}}_1^{\dagger\hat{r}_0^\dagger\prime}, \ldots, \widehat{\mathbf{Q}}_R^{\dagger\hat{r}_0^\dagger\prime}\right]'$ as the first $\hat{r}_0^\dagger$ columns of $\widehat{\mathbf{Q}}^\dagger$ from the above SVD, and construct the $T \times R\hat{r}_0^\dagger$ matrix, $\widehat{\mathbf{\Psi}}^\dagger = \left[\widehat{\mathbf{K}}_1^\dagger \widehat{\mathbf{Q}}_1^{\dagger\hat{r}_0^\dagger}, \ldots, \widehat{\mathbf{K}}_R^\dagger \widehat{\mathbf{Q}}_R^{\dagger\hat{r}_0^\dagger}\right]$. We perform the eigen decomposition,

$$T^{-1}\widehat{\mathbf{\Psi}}^\dagger \widehat{\mathbf{\Psi}}^{\dagger\prime} = \widehat{\mathbf{L}}^\dagger \widehat{\mathbf{\Xi}}^\dagger \widehat{\mathbf{L}}^{\dagger\prime}$$

where $\widehat{\mathbf{L}}^\dagger$ is a $T \times R\hat{r}_0^\dagger$ orthonormal matrix and $\widehat{\mathbf{\Xi}}^\dagger$ is a $T \times T$ diagonal matrix consisting of the eigenvalues in *descending order*. The consistent estimator of the global factors, denoted $\widehat{\mathbf{G}}^\dagger$, by the $\hat{r}_0$ vectors of $\widehat{\mathbf{L}}^\dagger$ corresponding to the r_0 largest eigenvalues multiplied by $\sqrt{T}$. Let $\widehat{\mathbf{Y}}_{ij}^\dagger = \mathbf{Y}_{ij} - \mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}_{ij}^\dagger$ and $\widehat{\mathbf{Y}}_i^\dagger = \left[\widehat{\mathbf{Y}}_{i1}^\dagger, \ldots, \widehat{\mathbf{Y}}_{iN_i}^\dagger\right]$. The global factor loadings can be estimated by $\widehat{\mathbf{\Gamma}}_i^\dagger = T^{-1}\widehat{\mathbf{Y}}_i^{\dagger\prime}\widehat{\mathbf{G}}^\dagger$. Finally, let $\widehat{\mathbf{Y}}_{ij}^{\dagger G} = \mathbf{Y}_{ij} - \mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}_{ij}^\dagger - \widehat{\mathbf{G}}^\dagger \widehat{\boldsymbol{\gamma}}_{ij}^\dagger$, the local factors and loadings can be obtained by applying PC to each block $\widehat{\mathbf{Y}}_i^{\dagger G} = \left[\widehat{\mathbf{Y}}_{i1}^{\dagger G}, \ldots, \widehat{\mathbf{Y}}_{iN_i}^{\dagger G}\right]$. $\widehat{\mathbf{F}}_i^\dagger$ is $\sqrt{T}$ multiply the $\hat{r}_i^\dagger$ eigenvectors of $\widehat{\mathbf{Y}}_i^{\dagger G}\widehat{\mathbf{Y}}_i^{\dagger G\prime}$ corresponding to the r_i largest eigenvalues. The local factor loadings can be obtained by $\widehat{\mathbf{\Lambda}}_i^\dagger = T^{-1}\widehat{\mathbf{Y}}_i^{\dagger G\prime}\widehat{\mathbf{F}}_i^\dagger$. The consistency of the *GCC* estimates are summarised by the following proposition:

Proposition C.2.3.

Under Assumptions 3.B–3.E, 3.G, and C.A, as $N_1, N_2, \ldots, N_R, T \to \infty$, we have:

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

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

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

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

where $\mathbb{H} = T^{-1/2}\mathbf{G}'\mathbf{J}^{r_0}\mathbf{U}$ is an $r_0 \times r_0$ rotation matrix, $\mathbf{J}^{r_0} = \mathbf{L}^{r_0}(\mathbf{\Xi}^{r_0})^{-1}$, $\mathbf{\Xi}^{r_0}$ is an $r_0 \times r_0$ diagonal matrix consisting of the r_0 non-zero eigenvalues of $T^{-1}\mathbf{G}\mathbf{G}'$ in descending order, $\mathbf{L}^{r_0}$ is a $T \times r_0$ matrix of the corresponding eigenvectors, and $\mathbf{U}$ is an $r_0 \times r_0$ orthogonal matrix. $\widehat{\mathscr{H}}_i^\dagger = (\mathbf{\Lambda}_i'\mathbf{\Lambda}_i/N_i)\left(\widehat{\mathbf{F}}_i^{\dagger\prime}\mathbf{F}/T\right)\widehat{\mathbf{\Upsilon}}_i^{\dagger-1}$ is an $r_i \times r_i$ rotation matrix, $\widehat{\mathbf{\Upsilon}}_i^\dagger$ is an $r_i \times r_i$ diagonal matrix consisting of the r_i largest eigenvalues of $(N_iT)^{-1}\widehat{\mathbf{Y}}_i^\dagger\widehat{\mathbf{Y}}_i^{\dagger\prime}$ in descending order, $\widehat{\mathbf{Y}}_i^\dagger = \mathbf{Y}_i - \widehat{\mathbf{G}}^\dagger\widehat{\mathbf{\Gamma}}_i^{\dagger\prime}$. Moreover, $C_{\underline{N},T} = \min\{\sqrt{\underline{N}}, \sqrt{T}\}$ with $\underline{N} = \min\{N_1, N_2, \ldots, N_R\}$.

C.2.5 Auxiliary lemmas

To facilitate the proofs of the main result in Sections C.2.2–C.2.4, we outline some useful auxiliary lemmas in this section. The proofs of these lemmas can be found in Section C.2.7.

Lemma C.2.1. *Let* $\left(\left\{\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}\right\}_{i=1,\ldots,R}^{j=1,\ldots,N_i}, \widehat{\mathbf{K}}_1^{\dagger}, \ldots, \widehat{\mathbf{K}}_R^{\dagger}\right)$ *be the estimators from minimising (C.2.5). Under Assumption 3.B-3.E and C.A, as* $N_1, \ldots, N_R, T \to \infty$, *the following statements holds:*

1. $\widehat{\boldsymbol{\beta}}_{ij}^{\dagger} - \boldsymbol{\beta}_{ij} \xrightarrow{p} \mathbf{0}$ *for all* i *and* j.

2. $\left\|\mathbf{P}_{\widehat{K}_i^{\dagger}} - \mathbf{P}_{K_i}\right\| \xrightarrow{p} 0$ *for all* i.

Lemma C.2.2. *Let* $\widehat{\mathbf{H}}_i^{\dagger} = (\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i)\left(\mathbf{K}_i^{\dagger\prime}\widehat{\mathbf{K}}_i/T\right)\widehat{\mathbf{V}}_i^{\dagger-1}$. *Under Assumptions 3.B–3.E and C.A–C.B,* $\widehat{\mathbf{H}}_i^{\dagger}$ *is an* $(r_0 + r_i) \times (r_0 + r_i)$ *invertible matrix,* $\widehat{\mathbf{V}}_i^{\dagger} \xrightarrow{p} \mathbf{V}_i$ *where* $\mathbf{V}_i$ *is a* $(r_0 + r_i) \times (r_0 + r_i)$ *diagonal matrix consisting of the eigenvalues of* $\boldsymbol{\Sigma}_{\mathbf{K}_i}\boldsymbol{\Sigma}_{\boldsymbol{\Theta}_i}$ *which are defined in Assumption 3.B, and*

$$\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{K}}_i^{\dagger} - \mathbf{K}_i\widehat{\mathbf{H}}_i^{\dagger}\right\| = O_p\left(B_{i,N_iT}\right) + O_p\left(\frac{1}{C_{N_iT}}\right).$$

where $B_{i,N_iT} = \max_j \left\{\left\|\widehat{\boldsymbol{\beta}}_{ij}^{\dagger} - \boldsymbol{\beta}_{ij}\right\|\right\}$ *for all* i.

Lemma C.2.3. *Under Assumptions 3.B–3.E and C.A–C.B, we have*

1. $T^{-1}\mathbf{K}_i'\left(\widehat{\mathbf{K}}_i - \mathbf{K}_i\widehat{\mathbf{H}}_i\right) = O_p\left(B_{i,N_iT}\right) + O_p\left(C_{N_iT}^{-2}\right)$.

2. $T^{-1}\widehat{\mathbf{K}}_i'\left(\widehat{\mathbf{K}}_i - \mathbf{K}_i\widehat{\mathbf{H}}_i\right) = O_p\left(B_{i,N_iT}\right) + O_p\left(C_{N_iT}^{-2}\right)$.

3. $T^{-1}\mathbf{X}_{ij}'\left(\widehat{\mathbf{K}}_i^{\dagger} - \mathbf{K}_i\widehat{\mathbf{H}}_i^{\dagger}\right) = O_p\left(B_{i,N_iT}\right) + O_p\left(C_{N_iT}^{-2}\right)$.

4. $T^{-1}\mathbf{e}_{ij}'\left(\widehat{\mathbf{K}}_i^{\dagger} - \mathbf{K}_i\widehat{\mathbf{H}}_i^{\dagger}\right) = O_p\left(B_{i,N_iT}\right) + O_p\left(C_{N_iT}^{-2}\right)$.

5. $\widehat{\mathbf{H}}_i^{\dagger}\widehat{\mathbf{H}}_i^{\dagger\prime} = \left(\mathbf{K}_i'\mathbf{K}_i/T\right)^{-1} + O_p\left(B_{i,N_iT}\right) + O_p\left(C_{N_iT}^{-2}\right)$.

Lemma C.2.4. *Let* $\left(\left\{\widetilde{\boldsymbol{\beta}}_{ij}^{\dagger}\right\}_{i=1,\ldots,R}^{j=1,\ldots,N_i}, \widetilde{\mathbf{K}}_1^{\dagger}, \ldots, \widetilde{\mathbf{K}}_R^{\dagger}\right)$ *be the estimators from minimising (C.2.8). For each* i, *let* $v_{i,k}$ *be the* k*-th largest eigenvalue of* $\left(\mathbf{K}_i'\mathbf{K}_i/T\right)\left(\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i\right)$ *and* $\tilde{v}_{i,k}^{\dagger}$ *be the* k*-th largest eigenvalue of* $\widetilde{\boldsymbol{\Sigma}}_i^{\dagger}$ *defined in (C.2.9). Under Assumption 3.B–3.E and C.A–C.B, as* $N_1, \ldots, N_R, T \to \infty$, *we have*

1. $\left\|\widetilde{\boldsymbol{\beta}}_{ij}^{\dagger} - \boldsymbol{\beta}_{ij}\right\| = O_p\left(C_{N_iT}^{-1/2}\right)$.

2. *For each* i, $|\tilde{v}_{i,k} - v_{i,k}| = O_p\left(B_{i,N_iT}\right) + C_{N_iT}^{-1}$ *for* $k = 1, \ldots, d_i$.

3. *For each* i, $|\tilde{v}_{i,k}| = O_p\left(B_{i,N_iT}^2\right) + C_{N_iT}^{-2}$ *for* $k = d_i + 1, \ldots, d_{\max}$.

C.2.6 Proofs of the main results

Proof of Theorem C.2.1.

Using (C.2.7) and $\mathbf{Y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta}_{ij} + \mathbf{K}_i\boldsymbol{\theta}_{ij} + \mathbf{e}_{ij}$, we obtain:

$$\left(\frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^\dagger}\mathbf{X}_{ij}\right)\left(\widehat{\boldsymbol{\beta}}_{ij}^\dagger - \boldsymbol{\beta}_{ij}\right) = \frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^\dagger}\mathbf{K}_i\boldsymbol{\theta}_{ij} + \frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^\dagger}\mathbf{e}_{ij}.$$

Plugging in $\mathbf{K}_i = \mathbf{K}_i - \widehat{\mathbf{K}}_i^\dagger\widehat{\mathbf{H}}_i^\dagger + \widehat{\mathbf{K}}_i^\dagger\widehat{\mathbf{H}}_i^\dagger$ and (C.2.15), the above equation becomes

$$\left(\frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^\dagger}\mathbf{X}_{ij}\right)\left(\widehat{\boldsymbol{\beta}}_{ij}^\dagger - \boldsymbol{\beta}_{ij}\right) = \frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^\dagger}\mathbf{e}_{ij} - \frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^\dagger}\left(\mathcal{J}_{i,1}^\dagger + \cdots + \mathcal{J}_{i,8}^\dagger\right)\boldsymbol{\Pi}_i^\dagger\boldsymbol{\theta}_{ij}. \quad \text{(C.2.11)}$$

We consider the first term of (C.2.11):

$$\frac{1}{\sqrt{T}}\mathbf{X}_{ij}'\left(\mathbf{M}_{\widehat{K}_i^\dagger} - \mathbf{M}_{K_i}\right)\mathbf{e}_{ij} = \frac{1}{T}\mathbf{X}_{ij}'\mathbf{K}_i\left(\frac{\mathbf{K}_i'\mathbf{K}_i}{T}\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{K}_i'\mathbf{e}_{ij} - \frac{1}{\sqrt{T}}\mathbf{X}_{ij}'\frac{1}{T}\widehat{\mathbf{K}}_i^\dagger\widehat{\mathbf{K}}_i^{\dagger\prime}\mathbf{e}_{ij}$$

Using $\widehat{\mathbf{K}}_i^\dagger = \widehat{\mathbf{K}}_i^\dagger - \mathbf{K}_i\widehat{\mathbf{H}}_i^\dagger + \mathbf{K}_i\widehat{\mathbf{H}}_i^\dagger$, the above equation becomes

$$\begin{aligned}\frac{1}{\sqrt{T}}\mathbf{X}_{ij}'\left(\mathbf{M}_{\widehat{K}_i^\dagger} - \mathbf{M}_{K_i}\right)\mathbf{e}_{ij} &= \frac{1}{T}\mathbf{X}_{ij}'\mathbf{K}_i\left(\frac{\mathbf{K}_i'\mathbf{K}_i}{T}\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{K}_i'\mathbf{e}_{ij} - \frac{1}{\sqrt{T}}\mathbf{X}_{ij}'\frac{1}{T}\mathbf{K}_i\widehat{\mathbf{H}}_i^\dagger\widehat{\mathbf{H}}_i^{\dagger\prime}\mathbf{K}_i'\mathbf{e}_{ij} \\ &- \frac{1}{\sqrt{T}}\mathbf{X}_{ij}'\frac{1}{T}\left(\widehat{\mathbf{K}}_i^\dagger - \mathbf{K}_i\widehat{\mathbf{H}}_i^\dagger\right)\left(\widehat{\mathbf{K}}_i^\dagger - \mathbf{K}_i\widehat{\mathbf{H}}_i^\dagger\right)'\mathbf{e}_{ij} - \frac{1}{\sqrt{T}}\mathbf{X}_{ij}'\frac{1}{T}\left(\widehat{\mathbf{K}}_i^\dagger - \mathbf{K}_i\widehat{\mathbf{H}}_i^\dagger\right)\widehat{\mathbf{H}}_i^{\dagger\prime}\mathbf{K}_i'\mathbf{e}_{ij} \\ &- \frac{1}{T}\mathbf{X}_{ij}'\mathbf{K}_i\widehat{\mathbf{H}}_i^\dagger\frac{1}{\sqrt{T}}\left(\widehat{\mathbf{K}}_i^\dagger - \mathbf{K}_i\widehat{\mathbf{H}}_i^\dagger\right)'\mathbf{e}_{ij}\end{aligned}$$

Using Lemma C.2.3, we obtain

$$\frac{1}{\sqrt{T}}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^\dagger}\mathbf{e}_{ij} = \frac{1}{\sqrt{T}}\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{e}_{ij} + O_p\left(\sqrt{T}B_{i,N_iT}\right) + O_p\left(\frac{\sqrt{T}}{C_{N_iT}^2}\right).$$

Using the same arguments, we also have

$$\frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^\dagger}\mathbf{X}_{ij} = \frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{X}_{ij} + O_p\left(B_{i,N_iT}\right) + O_p\left(\frac{1}{C_{N_iT}^2}\right).$$

We now analysis each term in the second part (C.2.11). For the first term, we have

$$\begin{aligned}\left\|\frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^\dagger}\mathcal{J}_{i,1}^\dagger\boldsymbol{\Pi}_i^\dagger\boldsymbol{\theta}_{ij}\right\| &= \left\|\frac{1}{N_i}\sum_{k=1}^{N_i}\frac{\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^\dagger}\mathbf{X}_{ik}}{T}\left(\boldsymbol{\beta}_{ik} - \widehat{\boldsymbol{\beta}}_{ik}^\dagger\right)\left(\boldsymbol{\beta}_{ik} - \widehat{\boldsymbol{\beta}}_{ik}^\dagger\right)'\frac{\mathbf{X}_{ik}'\widehat{\mathbf{K}}_i^\dagger}{T}\boldsymbol{\Pi}_i^\dagger\boldsymbol{\theta}_{ij}\right\| \\ &\leq \frac{1}{N_i}\sum_{k=1}^{N_i}\left\|\frac{\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^\dagger}\mathbf{X}_{ik}}{T}\right\|\left\|\boldsymbol{\beta}_{ik} - \widehat{\boldsymbol{\beta}}_{ik}^\dagger\right\|^2\left\|\frac{\mathbf{X}_{ik}'\widehat{\mathbf{K}}_i^\dagger}{T}\right\|\left\|\boldsymbol{\Pi}_i^\dagger\right\|\left\|\boldsymbol{\theta}_{ij}\right\| = O_p\left(B_{i,N_iT}^2\right) = o_p(B_{i,N_iT}).\end{aligned}$$

The second term has the same stochastic order as the LHS of (C.2.11), so we keep this term and deal with it later. The third term is bounded by

$$\left\| \frac{1}{T}\mathbf{X}'_{ij}\mathbf{M}_{\widehat{K}_i^\dagger}\mathcal{J}_{i,3}^\dagger\mathbf{\Pi}_i^\dagger\boldsymbol{\theta}_{ij} \right\| = \left\| \frac{1}{T}\mathbf{X}'_{ij}\mathbf{M}_{\widehat{K}_i^\dagger}\left(\mathbf{K}_i - \widehat{\mathbf{K}}_i^\dagger\widehat{\mathbf{H}}_i^{\dagger-1}\right)\frac{1}{N_i}\sum_{k=1}^{N_i}\boldsymbol{\theta}_{ik}\left(\boldsymbol{\beta}_{ik} - \widehat{\boldsymbol{\beta}}_{ik}^\dagger\right)'\left(\frac{\mathbf{X}'_{ij}\widehat{\mathbf{K}}_i}{T}\right)\mathbf{\Pi}_i\boldsymbol{\theta}_{ij} \right\|$$
$$\leq \left\| \frac{1}{T}\mathbf{X}'_{ij}\mathbf{M}_{\widehat{K}_i^\dagger}\left(\mathbf{K}_i - \widehat{\mathbf{K}}_i^\dagger\widehat{\mathbf{H}}_i^{\dagger-1}\right) \right\| \left\| \frac{1}{N_i}\sum_{k=1}^{N_i}\boldsymbol{\theta}_{ik}\left(\boldsymbol{\beta}_{ik} - \widehat{\boldsymbol{\beta}}_{ik}^\dagger\right)'\left(\frac{\mathbf{X}'_{ij}\widehat{\mathbf{K}}_i}{T}\right)\mathbf{\Pi}_i\boldsymbol{\theta}_{ij} \right\|$$

Using $\mathbf{M}_{\widehat{K}_i^\dagger} = \mathbf{I}_T - T^{-1}\widehat{\mathbf{K}}_i^\dagger\widehat{\mathbf{K}}_i'^\dagger$, Lemma 2, and Lemma 3, the first part of the above expression is bounded by $O_p(B_{i,N_iT}) + O_p\left(C_{N_iT}^{-2}\right)$. Therefore, it follows that

$$\left\| \frac{1}{T}\mathbf{X}'_{ij}\mathbf{M}_{\widehat{K}_i^\dagger}\mathcal{J}_{i,3}^\dagger\mathbf{\Pi}_i^\dagger\boldsymbol{\theta}_{ij} \right\| = \left[O_p(B_{i,N_iT}) + O_p\left(C_{N_iT}^{-2}\right)\right]O_p(B_{i,N_iT}) = o_p(B_{i,N_iT}).$$

Using similar arguments, it can be shown that

$$\left\| \frac{1}{T}\mathbf{X}'_{ij}\mathbf{M}_{\widehat{K}_i^\dagger}\mathcal{J}_{i,6}^\dagger\mathbf{\Pi}_i^\dagger\boldsymbol{\theta}_{ij} \right\| = o_p(B_{i,N_iT}).$$

Next, consider the fourth term

$$\left\| \frac{1}{T}\mathbf{X}'_{ij}\mathbf{M}_{\widehat{K}_i^\dagger}\mathcal{J}_{i,4}^\dagger\mathbf{\Pi}_i^\dagger\boldsymbol{\theta}_{ij} \right\| = \left\| \frac{1}{N_i}\sum_{k=1}^{N_i}\left(\frac{\mathbf{X}'_{ij}\mathbf{M}_{\widehat{K}_i^\dagger}\mathbf{X}_{ik}}{T}\right)\left(\boldsymbol{\beta}_{ik} - \widehat{\boldsymbol{\beta}}_{ik}^\dagger\right)'\left(\frac{\mathbf{e}'_{ik}\widehat{\mathbf{K}}_i}{T}\right)\mathbf{\Pi}_i\boldsymbol{\theta}_{ij} \right\| = o_p(B_{i,N_iT})$$

where the last equality follows from the fact that $\left\|T^{-1}\mathbf{e}'_{ik}\widehat{\mathbf{K}}_i\right\| = o_p(1)$ established in the proof of Lemma C.1.1. Similarly, one can show that

$$\left\| \frac{1}{T}\mathbf{X}'_{ij}\mathbf{M}_{\widehat{K}_i^\dagger}\mathcal{J}_{i,5}^\dagger\mathbf{\Pi}_i^\dagger\boldsymbol{\theta}_{ij} \right\| = o_p(B_{i,N_iT}).$$

Consider

$$\frac{1}{T}\mathbf{X}'_{ij}\mathbf{M}_{\widehat{K}_i^\dagger}\mathcal{J}_{i,7}^\dagger\mathbf{\Pi}_i^\dagger\boldsymbol{\theta}_{ij} = \frac{1}{T}\mathbf{X}'_{ij}\mathbf{M}_{\widehat{K}_i^\dagger}\left(\frac{1}{N_iT}\sum_{k=1}^{N_i}\mathbf{e}_{ik}\boldsymbol{\theta}'_{ik}\mathbf{K}'_i\widehat{\mathbf{K}}_i^\dagger\right)\mathbf{\Pi}_i^\dagger\boldsymbol{\theta}_{ij} = \frac{1}{N_iT}\sum_{k=1}^{N_i}\mathbf{X}'_{ij}\mathbf{M}_{\widehat{K}_i}\mathbf{e}_{ik}a_{i,kj}$$
$$= \frac{1}{N_iT}\sum_{k=1}^{N_i}\mathbf{X}'_{ij}\mathbf{e}_{ik}a_{i,kj} + \left(\frac{\mathbf{X}'_{ij}\widehat{\mathbf{K}}_i^\dagger}{T}\right)\frac{1}{N_iT}\sum_{k=1}^{N_i}\widehat{\mathbf{K}}_i^{\dagger\prime}\mathbf{e}_{ik}a_{i,kj}$$

where $a_{i,kj} = \boldsymbol{\theta}'_{ik}\left(\mathbf{\Theta}'_i\mathbf{\Theta}_i/N_i\right)^{-1}\boldsymbol{\theta}_{ij}$. It is straightforward that the first term of the above equation is $O_p\left(N_i^{-1/2}T^{-1/2}\right)$ and we focus on the second term.

$$\left\| \frac{1}{N_iT}\sum_{k=1}^{N_i}\widehat{\mathbf{K}}_i^{\dagger\prime}\mathbf{e}_{ik}a_{i,kj} \right\|^2 = \left\| \frac{1}{N_i}\sum_{k=1}^{N_i}a_{i,kj}\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{K}}_{it}^\dagger\mathbf{e}_{ikt} \right\|^2 \leq \frac{1}{N_i^2}\sum_{k=1}^{N_i}\left\| a_{i,kj}\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{K}}_{it}^\dagger\mathbf{e}_{ikt} \right\|^2$$
$$\leq O(1)\frac{1}{N_i^2}\sum_{k=1}^{N_i}\left\| \frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{K}}_{it}^\dagger\mathbf{e}_{ikt} \right\|^2 = \frac{1}{N_i}O_p\left(\frac{1}{C_{N_iT}}\right)$$

where the last equality follows from $N_i^{-1}\sum_{k=1}^{N_i}\left\|T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{K}}_{it}^{\dagger}e_{ikt}\right\|^2 = O_p\left(C_{N_iT}^{-1}\right)$ established in the proof of Lemma C.1.1. Consequently, we obtain

$$\left\|\frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^{\dagger}}\mathcal{J}_{i,7}^{\dagger}\mathbf{\Pi}_i^{\dagger}\boldsymbol{\theta}_{ij}\right\| = \frac{1}{\sqrt{N_i}}O_p\left(C_{N_iT}^{-1/2}\right).$$

Finally, we consider

$$\begin{aligned}\left\|\frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^{\dagger}}\mathcal{J}_{i,8}^{\dagger}\mathbf{\Pi}_i^{\dagger}\boldsymbol{\theta}_{ij}\right\| &= \left\|\frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^{\dagger}}\left(\frac{1}{N_iT}\sum_{k=1}^{N_i}\mathbf{e}_{ik}\mathbf{e}_{ik}'\widehat{\mathbf{K}}_i^{\dagger}\right)\mathbf{\Pi}_i^{\dagger}\boldsymbol{\theta}_{ij}\right\| \\ &\leq \left\|\frac{1}{\sqrt{T}}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^{\dagger}}\right\|\left\|\frac{1}{N_iT}\sum_{k=1}^{N_i}\mathbf{e}_{ik}\mathbf{e}_{ik}'\right\|\left\|\frac{1}{\sqrt{T}}\widehat{\mathbf{K}}_i^{\dagger}\right\|\left\|\mathbf{\Pi}_i^{\dagger}\right\|\left\|\boldsymbol{\theta}_{ij}\right\|\end{aligned}$$

Notice that

$$\left\|\frac{1}{N_iT}\sum_{k=1}^{N_i}\mathbf{e}_{ik}\mathbf{e}_{ik}'\right\| = O_p\left(\frac{1}{C_{N_iT}^2}\right)$$

is implied by Assumption 3.C.4. Therefore, we obtain

$$\left\|\frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{\widehat{K}_i^{\dagger}}\mathcal{J}_{i,8}^{\dagger}\mathbf{\Pi}_i^{\dagger}\boldsymbol{\theta}_{ij}\right\| = O_p\left(\frac{1}{C_{N_iT}^2}\right).$$

Combining these results, we obtain

$$\begin{aligned}\left(\frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{X}_{ij}\right)\left(\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}-\boldsymbol{\beta}_{ij}\right) &= \frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{e}_{ij} - \frac{1}{N_i}\sum_{k=1}^{N_i}\left(\frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{X}_{ik}\right)\left(\boldsymbol{\beta}_{ik}-\widehat{\boldsymbol{\beta}}_{ik}^{\dagger}\right)a_{i,jk} \\ &\quad + O_p(B_{i,N_iT}) + O_p\left(\frac{1}{C_{N_iT}^2}\right) + o_p(B_{i,N_iT}) + O_p\left(\frac{1}{\sqrt{N_iC_{N_iT}}}\right)\end{aligned}$$

which leads to

$$\begin{aligned}\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}-\boldsymbol{\beta}_{ij}\right) &= \left(\frac{\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{X}_{ij}}{T}\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{e}_{ij} \\ &\quad - \left(\frac{\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{X}_{ij}}{T}\right)^{-1}\frac{1}{N_i}\sum_{k=1}^{N_i}\left(\frac{1}{T}\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{X}_{ik}\right)\sqrt{T}\left(\boldsymbol{\beta}_{ik}-\widehat{\boldsymbol{\beta}}_{ik}^{\dagger}\right)a_{i,jk} + o_p(1) \qquad \text{(C.2.12)}\end{aligned}$$

if $T/N_i^2 \to 0$.

To derive the individual asymptotic distribution of $\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}-\boldsymbol{\beta}_{ij}\right)$, we define the following matrices:

$$\mathbb{A}_i = \begin{bmatrix}\left(\frac{\mathbf{X}_{i1}'\mathbf{M}_{K_i}\mathbf{X}_{i1}}{T}\right)^{-1} & & & \\ & \left(\frac{\mathbf{X}_{i2}'\mathbf{M}_{K_i}\mathbf{X}_{i2}}{T}\right)^{-1} & & \\ & & \ddots & \\ & & & \left(\frac{\mathbf{X}_{iN_i}'\mathbf{M}_{K_i}\mathbf{X}_{iN_i}}{T}\right)^{-1}\end{bmatrix},$$

$$\mathbb{X}_i = \begin{bmatrix} \frac{\mathbf{X}_{i1}'\mathbf{M}_{K_i}\mathbf{X}_{i1}}{T} & \frac{\mathbf{X}_{i1}'\mathbf{M}_{K_i}\mathbf{X}_{i2}}{T} & \cdots & \frac{\mathbf{X}_{i1}'\mathbf{M}_{K_i}\mathbf{X}_{iN_i}}{T} \\ \frac{\mathbf{X}_{i2}'\mathbf{M}_{K_i}\mathbf{X}_{i1}}{T} & \frac{\mathbf{X}_{i2}'\mathbf{M}_{K_i}\mathbf{X}_{i2}}{T} & \cdots & \frac{\mathbf{X}_{i2}'\mathbf{M}_{K_i}\mathbf{X}_{iN_i}}{T} \\ & & \vdots & \\ \frac{\mathbf{X}_{iN_i}'\mathbf{M}_{K_i}\mathbf{X}_{i1}}{T} & \frac{\mathbf{X}_{iN_i}'\mathbf{M}_{K_i}\mathbf{X}_{i2}}{T} & \cdots & \frac{\mathbf{X}_{iN_i}'\mathbf{M}_{K_i}\mathbf{X}_{iN_i}}{T} \end{bmatrix},$$

$$\boldsymbol{\varsigma}_i = \begin{bmatrix} \frac{1}{\sqrt{T}}\mathbf{X}_{i1}'\mathbf{M}_{K_i}\mathbf{e}_{i1} \\ \vdots \\ \frac{1}{\sqrt{T}}\mathbf{X}_{iN_i}'\mathbf{M}_{K_i}\mathbf{e}_{iN_i} \end{bmatrix}, \text{ and } \boldsymbol{\alpha} = \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{i1}^{\dagger} - \boldsymbol{\beta}_{i1} \\ \vdots \\ \widehat{\boldsymbol{\beta}}_{iN_i}^{\dagger} - \boldsymbol{\beta}_{iN_i} \end{bmatrix}.$$

Then, stacking (C.2.12) across j, we obtain

$$\sqrt{T}\boldsymbol{\alpha} = \mathbb{A}_i\boldsymbol{\varsigma}_i + \frac{1}{N_i}\mathbb{A}_i\mathbb{X}_i\sqrt{T}\boldsymbol{\alpha} + o_p(1),$$

which leads to

$$\sqrt{T}\boldsymbol{\alpha} = \left(\mathbf{I}_{N_i} - \frac{1}{N_i}\mathbb{A}_i\mathbb{X}_i\right)^{-1}\mathbb{A}_i\boldsymbol{\varsigma}_i + o_p(1).$$

Using Taylor expansion, we have

$$\left(\mathbf{I}_{N_i} - \frac{1}{N_i}\mathbb{A}_i\mathbb{X}_i\right)^{-1} = \mathbf{I}_{N_i} + \frac{1}{N_i}\mathbb{A}_i\mathbb{X}_i + o_p(1).$$

Therefore, it follows that

$$\sqrt{T}\boldsymbol{\alpha} = \mathbb{A}_i\boldsymbol{\varsigma}_i + \mathbb{A}_i\frac{1}{N_i}\mathbb{X}_i\mathbb{A}_i\boldsymbol{\varsigma}_i + o_p(1)$$

whose j-th element is

$$\begin{aligned}\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{ij}^{\dagger} - \boldsymbol{\beta}_{ij}\right) &= \left(\frac{\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{X}_{ij}}{T}\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{e}_{ij} \\ &+ \left(\frac{\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{X}_{ij}}{T}\right)^{-1}\frac{1}{N_i}\sum_{k=1}^{N_i}\left(\frac{\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{X}_{ik}}{T}\right)\left(\frac{\mathbf{X}_{ij}'\mathbf{M}_{K_i}\mathbf{X}_{ij}}{T}\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{X}_{ik}'\mathbf{M}_{K_i}\mathbf{e}_{ik} + o_p(1).\end{aligned}$$

The second term of the above equation is $o_p(1)$ due to Assumption C.B.1. Finally, we obtain the stated result

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{ij}^{\dagger} - \boldsymbol{\beta}_{ij}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Omega}_{ij}^{-1}\mathbb{D}_{ij}\boldsymbol{\Omega}_{ij}^{-1}\right)$$

by Assumption C.B.2.

Q.E.D

Proof of Propositions C.2.1–C.2.3.

Using Lemma C.2.4, we can show that Propositions C.2.1–C.2.3 following the same strategy as in the proof of Propositions 3.1–3.3. The details are omitted.

Q.E.D

C.2.7 Proofs of the auxiliary lemmas

Proof of Lemma C.2.1.

Consider the difference of the objective functions:

$$\frac{1}{NT}\mathcal{S}\left(\{\mathbf{b}_{ij}\}_{i=1,\ldots,R}^{j=1,\ldots,N_i},\boldsymbol{\mathcal{K}}_1,\ldots,\boldsymbol{\mathcal{K}}_R\right)-\frac{1}{NT}\mathcal{S}\left(\{\boldsymbol{\beta}_{ij}\}_{i=1,\ldots,R}^{j=1,\ldots,N_i},\mathbf{K}_1,\ldots,\mathbf{K}_R\right)$$
$$=\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}(\boldsymbol{\beta}_{ij}-\mathbf{b}_{ij})'\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{X}_{ij}(\boldsymbol{\beta}_{ij}-\mathbf{b}_{ij})+\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\boldsymbol{\theta}_{ij}'\mathbf{K}_i'\mathbf{M}_{\mathcal{K}_i}\mathbf{K}_i\boldsymbol{\theta}_{ij}+\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{e}_{ij}$$
$$+\frac{2}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}(\boldsymbol{\beta}_{ij}-\mathbf{b}_{ij})'\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{K}_i\boldsymbol{\theta}_{ij}+\frac{2}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}(\boldsymbol{\beta}_{ij}-\mathbf{b}_{ij})'\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{e}_{ij}$$
$$+\frac{2}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\boldsymbol{\theta}_{ij}'\mathbf{K}_i'\mathbf{M}_{\mathcal{K}_i}\mathbf{e}_{ij}-\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\mathbf{M}_{K_i}\mathbf{e}_{ij}$$

By Lemma C.1.1, the fifth and sixth terms are $o_p(1)$. Using the same arguments in the proof of Proposition 3.1, we also have

$$\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{e}_{ij}-\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\mathbf{e}_{ij}'\mathbf{M}_{K_i}\mathbf{e}_{ij}=o_p(1).$$

We thus obtain

$$\frac{1}{NT}\mathcal{S}\left(\{\mathbf{b}_{ij}\}_{i=1,\ldots,R}^{j=1,\ldots,N_i},\boldsymbol{\mathcal{K}}_1,\ldots,\boldsymbol{\mathcal{K}}_R\right)-\frac{1}{NT}\mathcal{S}\left(\{\boldsymbol{\beta}_{ij}\}_{i=1,\ldots,R}^{j=1,\ldots,N_i},\mathbf{K}_1,\ldots,\mathbf{K}_R\right)$$
$$=\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}(\boldsymbol{\beta}_{ij}-\mathbf{b}_{ij})'\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{X}_{ij}(\boldsymbol{\beta}_{ij}-\mathbf{b}_{ij})$$
$$+\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\boldsymbol{\theta}_{ij}'\mathbf{K}_i'\mathbf{M}_{\mathcal{K}_i}\mathbf{K}_i\boldsymbol{\theta}_{ij}+\frac{2}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}(\boldsymbol{\beta}_{ij}-\mathbf{b}_{ij})'\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{K}_i\boldsymbol{\theta}_{ij}+o_p(1)$$

The above equation can be rewritten as

$$\frac{1}{NT}\mathcal{S}\left(\{\mathbf{b}_{ij}\}_{i=1,\ldots,R}^{j=1,\ldots,N_i},\boldsymbol{\mathcal{K}}_1,\ldots,\boldsymbol{\mathcal{K}}_R\right)-\frac{1}{NT}\mathcal{S}\left(\{\boldsymbol{\beta}_{ij}\}_{i=1,\ldots,R}^{j=1,\ldots,N_i},\mathbf{K}_1,\ldots,\mathbf{K}_R\right)$$
$$=\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}(\boldsymbol{\beta}_{ij}-\mathbf{b}_{ij})'\boldsymbol{\mathcal{A}}_{ij,1}(\boldsymbol{\beta}_{ij}-\mathbf{b}_{ij})+\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\boldsymbol{\eta}_i'\boldsymbol{\mathcal{A}}_{ij,2}\boldsymbol{\eta}_i$$
$$+\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}2(\boldsymbol{\beta}_{ij}-\mathbf{b}_{ij})'\boldsymbol{\mathcal{A}}_{ij,3}\boldsymbol{\eta}_i+o_p(1).$$

where $\boldsymbol{\eta}_i=\operatorname{vec}\{\mathbf{M}_{\mathcal{K}_i}\mathbf{K}_i\}$, $\boldsymbol{\mathcal{A}}_{ij,1}=\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}\mathbf{X}_{ij}$, $\boldsymbol{\mathcal{A}}_{ij,2}=\left(\boldsymbol{\theta}_{ij}\boldsymbol{\theta}_{ij}'\right)\otimes\mathbf{I}_T$, and $\boldsymbol{\mathcal{A}}_{ij,3}=\boldsymbol{\theta}_{ij}'\otimes\mathbf{X}_{ij}'\mathbf{M}_{\mathcal{K}_i}$.

Completing the squares, we obtain

$$\frac{1}{NT}\mathcal{S}\left(\{\mathbf{b}_{ij}\}_{i=1,\ldots,R}^{j=1,\ldots,N_i},\mathcal{K}_1,\ldots,\mathcal{K}_R\right)-\frac{1}{NT}\mathcal{S}\left(\{\boldsymbol{\beta}_{ij}\}_{i=1,\ldots,R}^{j=1,\ldots,N_i},\mathbf{K}_1,\ldots,\mathbf{K}_R\right)$$
$$=\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\boldsymbol{\eta}_i'\left(\boldsymbol{\mathcal{A}}_{ij,1}-\boldsymbol{\mathcal{A}}_{ij,3}\boldsymbol{\mathcal{A}}_{ij,2}^{-1}\boldsymbol{\mathcal{A}}_{ij,3}'\right)\boldsymbol{\eta}_i$$
$$+\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\left[\boldsymbol{\eta}_i'+(\boldsymbol{\beta}_{ij}-\mathbf{b}_{ij})'\boldsymbol{\mathcal{A}}_{ij,3}\boldsymbol{\mathcal{A}}_{ij,2}^{-1}\right]\boldsymbol{\mathcal{A}}_{ij,2}\left[\boldsymbol{\eta}_i+\boldsymbol{\mathcal{A}}_{ij,2}^{-1}\boldsymbol{\mathcal{A}}_{ij,3}'(\boldsymbol{\beta}_{ij}-\mathbf{b}_{ij})\right]+o_p(1).$$

Using the positive definiteness by Assumption 3.B.2 and C.A.2, the first two terms of the above equation are non-negative and attain their minimum uniquely at

$$\left(\{\mathbf{b}_{ij}\}_{i=1,\ldots,R}^{j=1,\ldots,N_i},\mathcal{K}_1,\ldots,\mathcal{K}_R\right)=\left(\{\boldsymbol{\beta}_{ij}\}_{i=1,\ldots,R}^{j=1,\ldots,N_i},\mathbf{K}_1\mathbf{H}_1,\ldots,\mathbf{K}_R\mathbf{H}_R\right)$$

where $\mathbf{H}_i$ is any invertible matrix for all i. Moreover,

$$\frac{1}{NT}\mathcal{S}\left(\left\{\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}\right\}_{i=1,\ldots,R}^{j=1,\ldots,N_i},\widehat{\mathbf{K}}_1^{\dagger},\ldots,\widehat{\mathbf{K}}_R^{\dagger}\right)-\frac{1}{NT}\mathcal{S}\left(\{\boldsymbol{\beta}_{ij}\}_{i=1,\ldots,R}^{j=1,\ldots,N_i},\mathbf{K}_1,\ldots,\mathbf{K}_R\right)\leq 0$$

since $\left(\left\{\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}\right\}_{i=1,\ldots,R}^{j=1,\ldots,N_i},\widehat{\mathbf{K}}_1^{\dagger},\ldots,\widehat{\mathbf{K}}_R^{\dagger}\right)$ is the minimiser of $\mathcal{S}$. This implies that

$$\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\widehat{\boldsymbol{\eta}}_i^{\dagger\prime}\left(\boldsymbol{\mathcal{A}}_{ij,1}-\boldsymbol{\mathcal{A}}_{ij,3}\boldsymbol{\mathcal{A}}_{ij,2}^{-1}\boldsymbol{\mathcal{A}}_{ij,3}'\right)\widehat{\boldsymbol{\eta}}_i^{\dagger}=o_p(1) \tag{C.2.13}$$

and

$$\frac{1}{NT}\sum_{i=1}^{R}\sum_{j=1}^{N_i}\left[\widehat{\boldsymbol{\eta}}_i^{\dagger\prime}+\left(\boldsymbol{\beta}_{ij}-\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}\right)'\boldsymbol{\mathcal{A}}_{ij,3}\boldsymbol{\mathcal{A}}_{ij,2}^{-1}\right]\boldsymbol{\mathcal{A}}_{ij,2}\left[\widehat{\boldsymbol{\eta}}_i^{\dagger}+\boldsymbol{\mathcal{A}}_{ij,2}^{-1}\boldsymbol{\mathcal{A}}_{ij,3}'\left(\boldsymbol{\beta}_{ij}-\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}\right)\right]=o_p(1). \tag{C.2.14}$$

Combining (C.2.13) and Assumption C.A.2, it follows that $T^{-1}\left\|\mathbf{M}_{\widehat{K}_i^{\dagger}}\mathbf{K}_i\right\|=o_p(1)$ which implies $\left\|\mathbf{P}_{\widehat{K}_i^{\dagger}}-\mathbf{P}_{K_i}\right\|=o_p(1)$ for all i.

(C.2.14) suggests an average consistency of $\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}$ rather than the consistency of $\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}$ for all i and j individually. To show individual consistency, we consider an infeasible estimator:

$$\widehat{\mathbf{b}}_{ij}=\arg\min_{\mathbf{b}_{ij}}\left(\mathbf{Y}_{ij}-\mathbf{X}_{ij}\mathbf{b}_{ij}\right)'\mathbf{M}_{K_i}\left(\mathbf{Y}_{ij}-\mathbf{X}_{ij}\mathbf{b}_{ij}\right).$$

Obviously, for all i and j, we have $\widehat{\mathbf{b}}_{ij}-\mathbf{b}_{ij}=o_p(1)$. Notice that each $\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}$ satisfies

$$\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}=\arg\min_{\mathbf{b}_{ij}}\left(\mathbf{Y}_{ij}-\mathbf{X}_{ij}\mathbf{b}_{ij}\right)'\mathbf{M}_{\widehat{K}_i^{\dagger}}\left(\mathbf{Y}_{ij}-\mathbf{X}_{ij}\mathbf{b}_{ij}\right)$$

Therefore, using $T^{-1}\left\|\mathbf{M}_{\widehat{K}_i^{\dagger}}\mathbf{K}_i\right\|=o_p(1)$, it is straightforward that $\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}-\widehat{\mathbf{b}}_{ij}=o_p(1)$. Finally, we achieve the desired result that $\widehat{\boldsymbol{\beta}}_{ij}^{\dagger}-\mathbf{b}_{ij}=o_p(1)$ for all i and j.

Q.E.D

Proof of Lemma C.2.2.

Using $\mathbf{Y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta}_{ij} + \mathbf{K}_i\boldsymbol{\theta}_{ij} + \mathbf{e}_{ij}$, (C.2.7) can be expanded as

$$\widehat{\mathbf{K}}_i^{\dagger}\widehat{\mathbf{V}}_i^{\dagger} - \mathbf{K}_i\left(\frac{\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i}{N_i}\right)\left(\frac{\mathbf{K}_i'\widehat{\mathbf{K}}_i^{\dagger}}{T}\right) = \frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{X}_{ij}\left(\boldsymbol{\beta}_{ij} - \widehat{\boldsymbol{\beta}}_{ij}^{\dagger}\right)\left(\boldsymbol{\beta}_{ij} - \widehat{\boldsymbol{\beta}}_{ij}^{\dagger}\right)'\mathbf{X}_{ij}'\widehat{\mathbf{K}}_i^{\dagger}$$
$$+\frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{X}_{ij}\left(\boldsymbol{\beta}_{ij} - \widehat{\boldsymbol{\beta}}_{ij}^{\dagger}\right)\boldsymbol{\theta}_{ij}'\mathbf{K}_i'\widehat{\mathbf{K}}_i^{\dagger} + \frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{K}_i\boldsymbol{\theta}_{ij}\left(\boldsymbol{\beta}_{ij} - \widehat{\boldsymbol{\beta}}_{ij}^{\dagger}\right)'\mathbf{X}_{ij}'\widehat{\mathbf{K}}_i^{\dagger}$$
$$+\frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{X}_{ij}\left(\boldsymbol{\beta}_{ij} - \widehat{\boldsymbol{\beta}}_{ij}^{\dagger}\right)\mathbf{e}_{ij}'\widehat{\mathbf{K}}_i^{\dagger} + \frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{e}_{ij}\left(\boldsymbol{\beta}_{ij} - \widehat{\boldsymbol{\beta}}_{ij}^{\dagger}\right)'\mathbf{X}_{ij}'\widehat{\mathbf{K}}_i^{\dagger} + \frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{K}_i\boldsymbol{\theta}_{ij}\mathbf{e}_{ij}'\widehat{\mathbf{K}}_i^{\dagger}$$
$$+\frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{e}_{ij}\boldsymbol{\theta}_{ij}'\mathbf{K}_i'\widehat{\mathbf{K}}_i^{\dagger} + \frac{1}{N_iT}\sum_{j=1}^{N_i}\mathbf{e}_{ij}\mathbf{e}_{ij}'\widehat{\mathbf{K}}_i^{\dagger} = \mathcal{J}_{i,1}^{\dagger} + \cdots + \mathcal{J}_{i,8}^{\dagger}.$$

It then follows that

$$\widehat{\mathbf{K}}_i^{\dagger}\widehat{\mathbf{H}}_i^{\dagger-1} - \mathbf{K}_i = \left(\mathcal{J}_{i,1}^{\dagger} + \cdots + \mathcal{J}_{i,8}^{\dagger}\right)\boldsymbol{\Pi}_i^{\dagger} \tag{C.2.15}$$

where $\boldsymbol{\Pi}_i^{\dagger} = \left(\mathbf{K}_i^{\dagger\prime}\widehat{\mathbf{K}}_i/T\right)^{-1}(\boldsymbol{\Theta}_i'\boldsymbol{\Theta}_i/N_i)^{-1}$. The rest of the proof follows from Lemma C.1.2 with slight modification by replacing $O_p\left(\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|\right)$ terms with $O_p\left(B_{i,N_iT}\right)$. The details are omitted.

Q.E.D

Proof of Lemma C.2.3.

Using the result from Lemma C.2.2, these results can be shown in a similar way as the ones of Lemma C.1.3 with slight modification by replacing $O_p\left(\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|\right)$ terms with $O_p\left(B_{i,N_iT}\right)$. The details are omitted.

Q.E.D

Bibliography

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

Ahn, S. C. & Zhang, X. (2023), 'Model selection for general multi-level group factor models with global, regional and local factors', *Regional and Local Factors (April 23, 2023)* .

Alessi, L., Barigozzi, M. & Capasso, M. (2010), 'Improved penalization for determining the number of factors in approximate factor models', *Statistics & Probability Letters* **80**(23-24), 1806–1813.

Anderson, T. (2003), *An Introduction to Multivariate Statistical Analysis*, Wiley Series in Probability and Statistics, Wiley.

Ando, T. & Bai, J. (2014), 'Asset pricing with a general multifactor structure', *Journal of Financial Econometrics* **13**(3), 556–604.

Ando, T. & 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. & Rubin, M. (2019), 'Inference in group factor models with an application to mixed-frequency data', *Econometrica* **87**(4), 1267–1305.

Andreou, E., Gagliardini, P., Ghysels, E. & Rubin, M. (2023), 'Spanning latent and observable factors', *Available at SSRN 4349003* .

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

Bada, O. & Liebl, D. (2014), 'The r-package phtt: Panel data analysis with heterogeneous time trends', *arXiv preprint arXiv:1407.6484* .

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

Bai, J. (2009), 'Panel data models with interactive fixed effects', *Econometrica* **77**(4), 1229–1279.

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

Bailey, N., Kapetanios, G. & Pesaran, M. H. (2021), 'Measurement of factor strength: Theory and practice', *Journal of Applied Econometrics* **36**(5), 587–613.

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

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

Bekaert, G., Hodrick, R. J. & Zhang, X. (2009), 'International stock return comovements', *The Journal of Finance* **64**(6), 2591–2626.

Breitung, J. & 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.

Breitung, J. & Pigorsch, U. (2013), 'A canonical correlation approach for selecting the number of dynamic factors', *Oxford Bulletin of Economics and Statistics* **75**(1), 23–36.

Brown, S. J. (1989), 'The number of factors in security returns', *The Journal of Finance* **44**(5), 1247–1262.

Castro-González, N., Martínez-Serrano, M. & Robles, J. (2015), 'Expressions for the moore–penrose inverse of block matrices involving the schur complement', *Linear Algebra and its Applications* **471**, 353–368.

Chamberlain, G. & Rothschild, M. (1983), 'Arbitrage, factor structure, and mean-variance analysis on large asset markets', *Econometrica* **51**(1), 1305–1324.

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. & McQuade, T. J. (2021), The 2000s housing cycle with 2020 hindsight: A neo-kindlebergerian view, Technical report, National Bureau of Economic Research.

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

Chou, P.-H., Ho, P.-H. & Ko, K.-C. (2012), 'Do industries matter in explaining stock returns and asset-pricing anomalies?', *Journal of Banking & Finance* **36**(2), 355–370.

Cochrane, J. (2011), 'The number of factors in security returns', *The Journal of Finance* **66**, 1047–1108.

Conley, T. G., Gonçalves, S., Kim, M. S. & Perron, B. (2023), 'Bootstrap inference under cross-sectional dependence', *Quantitative Economics* **14**(2), 511–569.

Connor, G. & Korajczyk, R. A. (1988), 'Risk and return in an equilibrium apt: Application of a new test methodology', *Journal of Financial Economics* **21**(2), 255–289.

Coroneo, L., Jackson, L. E. & Owyang, M. T. (2020), 'International stock comovements with endogenous clusters', *Journal of Economic Dynamics and Control* **116**, 103904.

Damette, O. & Seghir, M. (2013), 'Energy as a driver of growth in oil exporting countries?', *Energy Economics* **37**, 193–199.

Del Negro, M. & 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. & 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.

Fama, E. F. & French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3–56.

Fama, E. F. & French, K. R. (1997), 'Industry costs of equity', *Journal of Financial Economics* **43**(2), 153–193.

Feng, G., Gao, J., Liu, F. & Peng, B. (2023), 'Estimation and inference for three-dimensional panel data models', *Available at SSRN 4547778* .

Feng, G., Giglio, S. & Xiu, D. (2020), 'Taming the factor zoo: A test of new factors', *The Journal of Finance* **75**(3), 1327–1370.

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

Golub, G. H. & Van Loan, C. F. (2013), *Matrix computations*, JHU press.

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

Hallin, M. & Liška, R. (2007), 'Determining the number of factors in the general dynamic factor model', *Journal of the American Statistical Association* **102**(478), 603–617.

Hallin, M. & 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.

Harvey, C. R., Liu, Y. & Zhu, H. (2016), '... and the cross-section of expected returns', *The Review of Financial Studies* **29**(1), 5–68.

Holly, S., Pesaran, M. H. & 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. & Grimes, A. (2008), 'Is there long-run convergence among regional house prices in the uk?', *Urban Studies* **45**(8), 1531–1544.

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

Jalil, A. (2014), 'Energy–growth conundrum in energy exporting and importing countries: Evidence from heterogeneous panel methods robust to cross-sectional dependence', *Energy Economics* **44**, 314–324.

Jin, S., Lu, X. & Su, L. (2023), 'Three-dimensional factor models with global and local factors', *Working Paper* .
URL: *"https://teacher.econ.cuhk.edu.hk/ xunlu/3D_factor_20230328.pdf"*

Kapetanios, G., Serlenga, L. & Shin, Y. (2021), 'Estimation and inference for multi-dimensional heterogeneous panel datasets with hierarchical multi-factor error structure', *Journal of Econometrics* **220**(2), 504–531.

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

Karabiyik, H., Reese, S. & Westerlund, J. (2017), 'On the role of the rank condition in cce estimation of factor-augmented panel regressions', *Journal of Econometrics* **197**(1), 60–64.

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

Lam, C. & Yao, Q. (2012), 'Factor modeling for high-dimensional time series: inference for the number of factors', *The Annals of Statistics* pp. 694–726.

Lam, C., Yao, Q. & Bathia, N. (2011), 'Estimation of latent factors for high-dimensional time series', *Biometrika* **98**(4), 901–918.

Leamer, E. E. (2007), Housing is the business cycle, Technical report, National Bureau of Economic Research.

Lettau, M. & Pelger, M. (2020), 'Factors that fit the time series and cross-section of stock returns', *The Review of Financial Studies* **33**(5), 2274–2325.

Lewellen, J., Nagel, S. & Shanken, J. (2010), 'A skeptical appraisal of asset pricing tests', *Journal of Financial Economics* **96**(2), 175–194.

Li, K., Cui, G. & Lu, L. (2020), 'Efficient estimation of heterogeneous coefficients in panel data models with common shocks', *Journal of Econometrics* **216**(2), 327–353.

Lin, R. & Shin, Y. (2022), 'Generalised canonical correlation estimation of the multilevel factor model', *Available at SSRN 4295429* .

Mao, Y., Peng, B., Silvapulle, M. J., Silvapulle, P. & Yang, Y. (2021), 'Decomposition of bilateral trade flows using a three-dimensional panel data model', *Available at SSRN 3825385* .

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

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

Moon, H. R. & Weidner, M. (2015), 'Linear regression for panel with unknown number of factors as interactive fixed effects', *Econometrica* **83**(4), 1543–1579.

Moon, H. R. & Weidner, M. (2017), 'Dynamic linear panel regression models with interactive fixed effects', *Econometric Theory* **33**(1), 158–195.

Moskowitz, T. J. & Grinblatt, M. (1999), 'Do industries explain momentum?', *The Journal of Finance* **54**(4), 1249–1290.

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

Ozturk, I. (2010), 'A literature survey on energy–growth nexus', *Energy Policy* **38**(1), 340–349.

Pesaran, M. H. (2006), 'Estimation and inference in large heterogeneous panels with a multifactor error structure', *Econometrica* **74**(4), 967–1012.

Rao, C. (1981), 'A lemma on g-inverse of a matrix and computation of correlation coefficients in the singular case', *Communications in Statistics-Theory and Methods* **10**(1), 1–10.

Rodríguez-Caballero, C. V. (2022), 'Energy consumption and gdp: A panel data analysis with multi-level cross-sectional dependence', *Econometrics and Statistics* **23**, 128–146.

Rodríguez-Caballero, C. V. & 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.

Sharpe, W. F. (1964), 'Capital asset prices: A theory of market equilibrium under conditions of risk', *The Journal of Finance* **19**(3), 425–442.

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

Stewart, G. W. & Sun, J.-g. (1990), *Matrix Perturbation Theory*, Academic Press, Boston.

Stock, J. H. & Watson, M. W. (2002), 'Forecasting using principal components from a large number of predictors', *Journal of the American Statistical Association* **97**(460), 1167–1179.

Su, L., Shi, Z. & Phillips, P. C. (2016), 'Identifying latent structures in panel data', *Econometrica* **84**(6), 2215–2264.

Tian, Y. & Takane, Y. (2009), 'More on generalized inverses of partitioned matrices with banachiewicz–schur forms', *Linear Algebra and its Applications* **430**(5-6), 1641–1655.

Trzcinka, C. (1986), 'On the number of factors in the arbitrage pricing model', *The Journal of Finance* **41**(2), 347–368.

Uematsu, Y. & 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. & Tao, D. (2019), 'A survey on canonical correlation analysis', *IEEE Transactions on Knowledge and Data Engineering* **33**(6), 2349–2368.

Yu, Y., Wang, T. & Samworth, R. J. (2015), 'A useful variant of the davis–kahan theorem for statisticians', *Biometrika* **102**(2), 315–323.