Article:
Choi, In, Lin, Rui and Shin, Yongcheol (2023) Canonical Correlation-based Model Selection for the Multilevel Factors. Journal of Econometrics. pp. 22-44. ISSN: 0304-4076

https://doi.org/10.1016/j.jeconom.2021.09.008

Canonical Correlation-based Model Selection for the Multilevel Factors *

In Choi† Rui Lin‡ Yongcheol Shin§

This Version: September 2021

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.

JEL: C52, G12.

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

*We are mostly grateful for the insightful comments by the editor, Serena Ng, the associated editor, and two anonymous referees as well as Vladimir Caballero, Jia Chen, Laura Coroneo, Simon Freyaldenhoven, Xu Han, Young Hoon Lee, Liangjun Su, Donggyu Sul, Michael Thornton, Takashi Yamagata, Paola Zerilli and Chaowen Zheng. We are also grateful to the seminar participants at Sogang University and University of York as well as conference delegates at the China Meeting of the Econometric Society, 2019, Korean Economic Review Conference, August 2020, World Congress, Bocconi University, Milan, August 2020 and AME Conference, Mexico, September 2020 for helpful comments. Shin acknowledges partial financial support from the Economic Social Research Council in the UK (Grant number ES/T01573X/1). Lin is grateful to the financial support from China Scholarship Council. The usual disclaimer applies.

†School of Economics, Sogang University. Email: inchoi@gmail.com

‡Department of Economics, University of York. Email: rl1081@york.ac.uk

§The corresponding author, Department of Economics, University of York. Email: yongcheol.shin@york.ac.uk

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 comovement of the variables by a small number of common factors, e.g. Stock and 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 and Eickmeier (2016) and Choi et al. (2018), and the LASSO approach by Ando and Bai (2017) and Han (2019). 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 and 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 comovements, Ando and 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, most 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 (2019) 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 and Ng (2002) and ER by Ahn and 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 (2019). 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 comoves 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 2 provides an overview of the related literature. Section 3 presents the multilevel factor model with the underlying assumptions. Section 4 develops CCD and MCC criteria for selecting the number of global factors and derives the asymptotic theory. Section 5 presents Monte Carlo simulation evidence. Section 6 provides an empirical application. Section 7 offers concluding remarks. The mathematical proofs are relegated to Appendix. Additional simulation results and theoretical derivations are provided in the Online Appendix.

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 and Ng (2002), which take a form: $PC(r) = V(r, \widehat{\boldsymbol{F}}) + rg(N,T)$, where $V(r, \widehat{\boldsymbol{F}})$ is the sum of squared residuals, $\widehat{\boldsymbol{F}}$ 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 and 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 and Jeong (2019) have conducted a comprehensive simulation study on approximate factor models, and documented evidence that BIC_3 by Bai and Ng (2002) and ER by Ahn and Horenstein (2013) outperform other competing estimators. Interestingly, Breitung and 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 and Liska (2007), Alessi et al. (2010) and Kapetanios (2010).

In the presence of the multilevel factors, the existing selection criteria may fail to identify the number of the global factors. Let r_0 (r_i) be the number of global (local) factors and R the number of blocks. If we apply existing approaches to the $T \times M_i$ data matrix, $\boldsymbol{Y}_i$ 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, $\boldsymbol{Y} = [\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_R]$ 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 (2019) conjectures that the number of global factors can be consistently estimated asymptotically by the existing selection criteria (see Remark 4).

In Section I in the Online Appendix we examine the finite sample performance of the four criteria, IC_{p2} and BIC_3 by Bai and Ng (2002), ED by Onatski (2010) and ER by Ahn and 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 (2019) 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 and 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, $\widehat{\boldsymbol{K}}_i$, from the data, $\boldsymbol{Y}_i$ for $i = 1, 2$, where $r_{\min} = \min\{\widehat{r_0 + r_1}, \widehat{r_0 + r_2}\}$. They

[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, $\widehat{r_0 + r_1} + \widehat{r_0 + r_2} - \widehat{r_0 + r_1 + r_2}$.

compute the sum of the r largest canonical correlations between $\widehat{\boldsymbol{K}}_1$ and $\widehat{\boldsymbol{K}}_2$, 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 (2019) 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 matrices. 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.

3 The Model and Assumptions

Consider the multilevel factor model:

$$y_{ijt} = \boldsymbol{\gamma}_{ij}' \boldsymbol{G}_t + \boldsymbol{\lambda}_{ij}' \boldsymbol{F}_{it} + e_{ijt}, i = 1, ..., R, j = 1, ..., M_i, t = 1, ..., T \tag{3.1}$$

where $\boldsymbol{G}_t = [G_{t1}, ..., G_{tr_0}]'$ comprises the $r_0 \times 1$ global factors, $\boldsymbol{F}_{it} = [F_{it1}, ..., F_{itr_i}]'$ is the $r_i \times 1$ vector of local factors in the block $i = 1, \cdots, R$, $\boldsymbol{\gamma}_{ij}$ and $\boldsymbol{\lambda}_{ij}$ are factor loadings and e_{ijt} is the idiosyncratic error. Stacking (3.1) across individuals in block i, we have:

$$\boldsymbol{y}_{it} = \boldsymbol{\Gamma}_i \boldsymbol{G}_t + \boldsymbol{\Lambda}_i \boldsymbol{F}_{it} + \boldsymbol{e}_{it}, \tag{3.2}$$

[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.

[3] From Table 5 in Han (2019), 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.

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

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

The model can also be written as

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

where

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

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

$$\boldsymbol{Y} = \boldsymbol{F}^+ \boldsymbol{\Lambda}^{+\prime} + \boldsymbol{e}, \tag{3.3}$$

where

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

Alternatively, stacking (3.1) over time, we can rewrite the model as

$$\boldsymbol{Y}_{ij} = \boldsymbol{G}\boldsymbol{\gamma}_{ij} + \boldsymbol{F}_i\boldsymbol{\lambda}_{ij} + \boldsymbol{e}_{ij}, \tag{3.4}$$

where

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

For each block i, we then have

$$\boldsymbol{Y}_i = \boldsymbol{G}\boldsymbol{\Gamma}_i' + \boldsymbol{F}_i\boldsymbol{\Lambda}_i' + \boldsymbol{e}_i \tag{3.5}$$

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

Following Bai and Ng (2002) and Choi et al. (2018), we make the following assumptions.

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

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

2. *Let* $E(\frac{1}{N}\sum_{i=1}^{R}\sum_{j=1}^{M_i} e_{ijs}e_{ijt}) = \omega_N(s,t)$. *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* $E(e_{mjt}e_{hkt}) = \tau_{(mj),(hk),t}$, *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* $E(e_{mjt}e_{hks}) = \tau_{(mj),(hk),(ts)}$ *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

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

Assumption B.

1. $\boldsymbol{G}_t$, $\boldsymbol{F}_{1t},\ldots,\boldsymbol{F}_{Rt}$ *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. $E(\|\boldsymbol{K}_{it}\|^4) < \infty$, *where* $\boldsymbol{K}_{it} = (\boldsymbol{G}_t', \boldsymbol{F}_{it}')'$.

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

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

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

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

Assumption C.

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

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

3. $\boldsymbol{\Sigma}_\Gamma\boldsymbol{\Sigma}_{\boldsymbol{G}}$ *has distinct eigenvalues.*

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

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

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

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

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

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

Assumption D.

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

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

Assumption A is an extended version of Assumption C in Bai and Ng (2002), which implies that the idiosyncratic errors are allowed to be serially and (weakly) cross-sectionally correlated. Assumptions B1–B4 are standard in the literature. Assumption B5 allows weak correlation between global/local factors and idiosyncratic errors. Assumption C is also standard. Assumption C2 allows global factors to have non-trivial contributions to the variance of all the individuals while Assumption C4(c) allows the local factors to have non-trivial contributions to the individual variances within the corresponding block. Assumption D1 ensures that the global factors and local factors can be separately identified. Initially, we make Assumption D2, but we will provide an extension in Subsection 4.1.1 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 $R \to \infty$.

4 Canonical Correlation-based Model Selection

4.1 Estimation of the Number of Global Factors

Using the model (3.5), we describe the estimation algorithms as follows: Let $\boldsymbol{K}_i = [\boldsymbol{G}, \boldsymbol{F}_i]$ 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 (3.5) for any two blocks, m and h, and obtain the estimates of $\boldsymbol{K}_m$ and $\boldsymbol{K}_h$, denoted $\widehat{\boldsymbol{K}}_m$ and $\widehat{\boldsymbol{K}}_h$, where $\widehat{\boldsymbol{K}}_m$ is $\sqrt{T}$ times eigenvectors corresponding to the $r_{\max}$ largest eigenvalues of the $T \times T$ matrix, $\boldsymbol{Y}_m \boldsymbol{Y}_m'$, and similarly for $\widehat{\boldsymbol{K}}_h$. Under Assumptions A–D, $\widehat{\boldsymbol{K}}_m$ and $\widehat{\boldsymbol{K}}_h$ contain the factor spaces spanned by $[\boldsymbol{G}, \boldsymbol{F}_m]$ and $[\boldsymbol{G}, \boldsymbol{F}_h]$, respectively. See Lemma 4 in Appendix.

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

$$(\widehat{\boldsymbol{S}}_{mh} \widehat{\boldsymbol{S}}_{hh}^{-1} \widehat{\boldsymbol{S}}_{hm} - \ell \widehat{\boldsymbol{S}}_{mm}) \boldsymbol{v} = \boldsymbol{0} \tag{4.1}$$

Let $\ell_{mh,r}$ be the r-th largest characteristic root of (4.1), which is the r-th largest sample squared canonical correlation between $\widehat{\boldsymbol{K}}_m$ and $\widehat{\boldsymbol{K}}_h$.

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

$$\ell_{mh,r} \xrightarrow{p} \begin{cases} 1 & \text{for } r = 1, \ldots, r_0 \\ 0 & \text{for } r = r_0 + 1, \ldots, r_{\max} \end{cases} \tag{4.2}$$

Since the blocks, m and h, share the r_0 global factors, the r_0 characteristic roots from (4.1) 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 and 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

Lemma 2. *Under Assumptions A–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 2 shows under Assumptions A–D that $\xi(r)$ is equal to 1 for $r \leq r_0$ while $\xi(r)$ is 0 for $r > r_0$, asymptotically.

Lemma 3. *Suppose that Assumptions A–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. *Suppose that Assumptions A–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.

$r^*_{\max} = \max\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_R}\}$ 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 $i = 1, ..., R$.

4.2 Non-zero correlation between the local factors

Kose et al. (2003), Beck et al. (2016), Choi et al. (2018) and Han (2019) assume that the local factors are all mutually uncorrelated. Wang (2008), Breitung and 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 $\boldsymbol{K}_m$ and $\boldsymbol{K}_h$. 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 1–3 (see the proofs in Section VII in the Online Appendix).

Lemma 1*. *Under Assumptions A–D1, 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 A–D1, 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 A–D1 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 3* 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 $d_{\max}(r) = \max_{r_{0+1} \leq r \leq r_{\max}} (\bar{\rho}_r - \bar{\rho}_{r+1})$. It still allows some pairs to have canonical correlation larger than η, but the average across all pairs cannot exceed η.

Theorem 2. *Suppose that Assumptions A–D1 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* $d_{\max}(r) = \max_{r_{0+1} \leq r \leq r_{\max}}(\bar{\rho}_r - \bar{\rho}_{r+1})$. *Then, we have:*

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

Theorem 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 $\bar{\rho}_{r_0+1} < \frac{1}{2}$.

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 $\delta_{\underline{M}T} = \min(\sqrt{\underline{M}}, \sqrt{T})$ 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

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

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 3.

Theorem 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* $\underline{M} = \min\{M_1, M_2, \ldots, M_R\}$. *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:

$$P(\underline{M},T) = \frac{\ln \underline{M} + \ln T}{\sqrt{\underline{M}T}} \ln\ln(\underline{M}T). \tag{4.3}$$

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 and Ng (2002) and Breitung and Pigorsch (2013). We may consider the popular penalty function in BIC_3 given by

$$BIC_3 = \frac{\underline{M}+T}{\underline{M}T} \ln(\underline{M}T). \tag{4.4}$$

In Section III in the Online Appendix, we provide the simulation results for MCC using BIC_3 in (4.4). Overall, its performance is relatively satisfactory for most cases, but it is outperformed by MCC using $P(\underline{M},T)$ in (4.3). 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 in small samples. Hence, we propose the following data-dependent tuning constant:[6]

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

[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 $r = 0, 1, \ldots, r_{\max}$. 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.

[6]Following Hallin and Liska (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.

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

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

$N = \sum_{i=1}^{R} M_i$, $\widehat{\boldsymbol{K}}_{it}$ 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.

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 $\widehat{\boldsymbol{G}} = \widehat{\boldsymbol{K}}_m \boldsymbol{V}_m^{\hat{r}_0}$, where $\boldsymbol{V}_m^{\hat{r}_0}$ is an $r_{\max} \times \hat{r}_0$ matrix consisting of the characteristic vectors associated with the $\hat{r}_0$ largest characteristic roots of (4.1). $\widehat{\boldsymbol{G}}$ from any block-pair would provide a consistent estimator for $\boldsymbol{G}$, but, in practice, we suggest to use the block-pair that yields the maximum value of $\ell_{mh,1}$.

Next, we concentrate $\widehat{\boldsymbol{G}}$ out in each block by $\boldsymbol{Y}_i^{\widehat{G}} = \boldsymbol{M}^{\widehat{G}}\boldsymbol{Y}_i$ for $i = 1, \ldots, R$ where $\boldsymbol{M}^{\widehat{G}} = \boldsymbol{I}_T - \widehat{\boldsymbol{G}}(\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{G}})\widehat{\boldsymbol{G}}'$. Then, we apply the existing approaches by Bai and Ng (2002) and Ahn and Horenstein (2013) to $\boldsymbol{Y}_i^{\widehat{G}}$, 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 $\boldsymbol{Y}_i^{\widehat{G}}$ and obtain $\widehat{\boldsymbol{F}}_i$ for $i = 1, ..., R$.

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

4.4 Estimation of Global and Local Factors and Loadings

In Sections 4.1–4.3, we have obtained the consistent estimates, $\hat{r}_0$ and $\hat{r}_i$. Given the initial estimates, $\widehat{\boldsymbol{G}}$, $\widehat{\boldsymbol{\Gamma}}_i$, $\widehat{\boldsymbol{F}}_i$ and $\widehat{\boldsymbol{\Lambda}}_i$ for $i = 1, ..., R$, we follow a sequential approach by Choi et al. (2018) and update the factors and loadings as follows:

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

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

[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.

5 Monte Carlo Simulation

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

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

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

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

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

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

$$e_{ijt} = \phi_e e_{ij,t-1} + \varepsilon_{ijt} + \beta\sum_{1\le|h|\le 8}\varepsilon_{i,j-h,t},\ \varepsilon_{ijt} \sim iidN(0,1)$$

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

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

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

Following Choi et al. (2018) and Han (2019), we make the variance contribution of each component equalised. For $r_0 > 0$, we set

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

For $r_0 = 0$, we set

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

We consider 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 $\boldsymbol{Y}_i$ 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 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 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 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

[8] See Section VI in the Online Appendix for the detailed estimation algorithms. Unfortunately, we are unable to implement Han's (2019) algorithm because his code can only be run on Matlab R2013b and R2014a, but not on the later versions.

[9] Ahn and 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 II in the Online Appendix 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 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 6. The other setups are the same as in the second case. From Panel C of Table 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.

Table 1 about here

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 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.

Tables 2 about here

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 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 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.

Tables 3 about here

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

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

where $\boldsymbol{F}_t = [\boldsymbol{F}'_{1t}, \ldots, \boldsymbol{F}'_{Rt}]'$, $\boldsymbol{w}_t = [\boldsymbol{w}'_{1t}, \ldots, \boldsymbol{w}'_{Rt}]'$ and $\boldsymbol{\Phi}_F$ is a diagonal matrix with the common elements, 0.5. We set the common diagonal elements of $\boldsymbol{\Omega}_F$ 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 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 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 3.[12]

Table 4 about here

In Section V in the Online Appendix, 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 and Ng (2002) and ER by Ahn and Horenstein (2013) outperform the other existing approaches.

Finally, in Section VI in the Online Appendix, 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, $\boldsymbol{G}$. 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.

[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 IV in the online Appendix, 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.

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.

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 and Korajczyk (1988) and Fama and French (1993). However, the studies investigating the role of industry factors explicitly in asset pricing model are relatively few. Fama and 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 and 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 comovement, 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 comovement 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 and French (1997) and use the SIC codes to categorise the stocks into twelve industries, listed in the first column of Table 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 and 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 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

[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.

the latter is the average correlation between individual returns across two different industries. We visualise them through a heat map in Figure 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 comove within the same industry.

Figure 1 about here

Next, we explore the correlation structure using the multilevel factor model. We standardise the data following Bai and Ng (2002) and Ahn and Horenstein (2013). First, we follow the practical guideline for $r^*_{\max}$ as described in Section 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. (2020).[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 4.4, and report the full estimation results in Table 5.

Table 5 about here

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 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

[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.

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

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

We construct the relative importance ratios for each industry by

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

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

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

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 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, the between correlations decline drastically for all industries, indicating that the market-wide comovement 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 comovements may be captured by the local factors. We further project out the local factors such that the resulting residuals would be purely idiosyncratic. Figure 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.

Figures 2 and 3 about here

Figure 4 displays that the estimated global factor comoves 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 2.

Figure 4 about here

Finally, in Figure 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 comove with the corresponding factors, and *vice versa*. First, almost all individual stock returns are positively loaded on the global factor, suggesting that they comove 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

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

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 comovement.

Figure 5 about here

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 comoves 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 and 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 and 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 and 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 multi-level 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 and Pelger (2020), Bailey et al. (2020) and Uematsu and Yamagata (2020). On the other hand, consistent estimation of both global and local factors and their loadings can be easily achieved in the

the multi-level 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. (2020) and Lettau and Pelger (2020).

References

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

Alessi, L., Barigozzi, M. and Capasso, M. (2010). Improved penalization for determining the number of factors in approximate factor models, *Statistics and Probability Letters* **80**: 1806–1813.

Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis*, Wiley Series in Probability and Statistics, Wiley.

Ando, T. and Bai, J. (2014). Asset pricing with a general multifactor structure, *Journal of Financial Econometrics* **13**(3): 556–604.

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

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

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

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

Bailey, N., Kapetanios, G. and Pesaran, M. H. (2020). Measurement of factor strength: Theory and practice, *Unpublished manuscript, University of Southern California* .

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

Bekaert, G., Hodrick, R. J. and Zhang, X. (2009). International stock return comovements, *The Journal of Finance* **64**(6): 2591–2626.

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

Breitung, J. and 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, *Journal of Finance* **44**(5): 1247–1262.

Castro-González, N., Martínez-Serrano, M. and Robles, J. (2015). Expressions for the moore–penrose inverse of block matrices involving the schur complement, *Linear Algebra and its Applications* **471**: 353–368.

Chen, P. (2012). Common factors and specific factors, *MPRA Paper No. 36082* .

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

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

Chou, P.-H., Ho, P.-H. and 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, *Journal of Finance* **66**: 1047–1108.

Connor, G. and 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.

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

Fama, E. F. and 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. and French, K. R. (1997). Industry costs of equity, *Journal of financial economics* **43**(2): 153–193.

Hallin, M. and Liska, R. (2007). Determining the number of factors in the general dynamic factor model, *Journal of the American Statistical Association* **102**: 603–617.

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

Han, X. (2019). Shrinkage estimation of factor models with global and group-specific factors, *Journal of Business & Economic Statistics* (just-accepted): 1–52.

Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets, *Journal of Business & Economic Statistics* **28**(3): 397–409.

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

Lettau, M. and 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. and Shanken, J. (2010). A skeptical appraisal of asset pricing tests, *Journal of Financial economics* **96**(2): 175–194.

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

Moskowitz, T. J. and 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.

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.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk, *The journal of finance* **19**(3): 425–442.

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

Su, L., Shi, Z. and Phillips, P. (2016). Identifying latent structures in panel data, *Econometrica* **84**: 2215–2264.

Tian, Y. and 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, *Journal of Finance* **41**(2): 347–368.

Uematsu, Y. and Yamagata, T. (2020). Estimation of weak factor models, *Discussion Paper, No. 108, Tohuku University* .

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

Appendices

A Proofs of Lemmas and Theorem

Lemma 4. *Let* $\widetilde{\boldsymbol{K}}_i = \frac{1}{M_i T}\boldsymbol{Y}_i\boldsymbol{Y}_i'\widehat{\boldsymbol{K}}_i$. *Under Assumption A–D, as* $M_i, T \to \infty$, *we have:*

$$\widetilde{\boldsymbol{K}}_i - \boldsymbol{K}_i\boldsymbol{H}_i = O_p\left(\frac{1}{\delta_{M_iT}}\right),\ i = 1, ..., R,$$

where $\boldsymbol{H}_i$ *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 and Ng (2002) are satisfied, the stated result follows directly from Theorem 1 of Bai and Ng (2002). *Q.E.D*

For any two blocks m and h, we apply the PC estimator to (3.5), and obtain consistent estimators of $\boldsymbol{K}_m = [\boldsymbol{G}, \boldsymbol{F}_m]$ and $\boldsymbol{K}_h = [\boldsymbol{G}, \boldsymbol{F}_h]$, denoted $\widehat{\boldsymbol{K}}_m$ and $\widehat{\boldsymbol{K}}_h$. Let $\ell_{mh,r}$ be the r-th largest squared canonical correlation between $\widehat{\boldsymbol{K}}_m$ and $\widehat{\boldsymbol{K}}_h$, which is given by the rth largest characteristic root of

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

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

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

where $\boldsymbol{V}_m$ is an $r_{\max} \times r_{\max}$ diagonal matrix consisting of the $r_{\max}$ largest eigenvalues of $\boldsymbol{Y}_m\boldsymbol{Y}_m'$ in descending .order divided by M_mT. This implies that $\widehat{\boldsymbol{K}}_m\boldsymbol{V}_m = \widetilde{\boldsymbol{K}}_m$. Similarly, we obtain $\widehat{\boldsymbol{K}}_h\boldsymbol{V}_h = \widetilde{\boldsymbol{K}}_h$ for block h. Since $r_{\max} < \min\{M_m, T\}$ $(r_{\max} < \min\{M_h, T\})$, the diagonal elements of $\boldsymbol{V}_m$ $(\boldsymbol{V}_h)$ are non-zero. This implies that $\boldsymbol{V}_m$ $(\boldsymbol{V}_h)$ is of full rank, though some diagonal elements may be very small. The canonical correlations between $\widehat{\boldsymbol{K}}_m$ and $\widehat{\boldsymbol{K}}_h$ are equal to those between $\widetilde{\boldsymbol{K}}_m$ and $\widetilde{\boldsymbol{K}}_h$, 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 $\widetilde{\boldsymbol{K}}_m$ and $\widetilde{\boldsymbol{K}}_h$ instead of those between $\widehat{\boldsymbol{K}}_m$ and $\widehat{\boldsymbol{K}}_h$. This enables us to employ Lemma 4 subsequently.

Proof for Lemma 1. For any two blocks m and h, the population covariance between $\boldsymbol{K}_{mt}$ and $\boldsymbol{K}_{ht}$ can be expressed as

$$Var\begin{pmatrix}\boldsymbol{K}_{mt}\\ \boldsymbol{K}_{ht}\end{pmatrix} = \begin{bmatrix}\boldsymbol{\Sigma}_{mm} & \boldsymbol{\Sigma}_{mh}\\ \boldsymbol{\Sigma}_{hm} & \boldsymbol{\Sigma}_{hh}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\Sigma}_G & \mathbf{0} & \boldsymbol{\Sigma}_G & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_{F_m} & \mathbf{0} & \mathbf{0}\\ \boldsymbol{\Sigma}_G & \mathbf{0} & \boldsymbol{\Sigma}_G & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{F_h}\end{bmatrix} \tag{A.1}$$

where $\boldsymbol{\Sigma}_G$, $\boldsymbol{\Sigma}_{F_m}$ and $\boldsymbol{\Sigma}_{F_h}$ are defined in Assumption C. Without loss of generality, we assume $r_m \leq r_h$. Using (A.1), we can rewrite the characteristic equation,

$$\left(\boldsymbol{\Sigma}_{mh}\boldsymbol{\Sigma}_{hh}^{-1}\boldsymbol{\Sigma}_{hm} - \rho\boldsymbol{\Sigma}_{mm}\right)\boldsymbol{v} = 0 \tag{A.2}$$

as

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

where $\rho_{mh,r}$ is the r-th largest squared canonical correlation between $\boldsymbol{K}_m$ and $\boldsymbol{K}_h$. 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 $\boldsymbol{v}_r = [0, \ldots, 0, 1, 0, \ldots, 0]$, which is the unit vector with the rth element being 1 and 0 otherwise.

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

$$Var\left(\begin{bmatrix}\boldsymbol{H}_m'\boldsymbol{K}_{mt}\\ \boldsymbol{H}_h'\boldsymbol{K}_{ht}\end{bmatrix}\right) = \begin{bmatrix}\boldsymbol{H}_m'\boldsymbol{\Sigma}_{mm}\boldsymbol{H}_m & \boldsymbol{H}_m'\boldsymbol{\Sigma}_{mh}\boldsymbol{H}_h\\ \boldsymbol{H}_h'\boldsymbol{\Sigma}_{hm}\boldsymbol{H}_m & \boldsymbol{H}_h'\boldsymbol{\Sigma}_{hh}\boldsymbol{H}_h\end{bmatrix} \tag{A.3}$$

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

$$\left[\boldsymbol{H}_m'\boldsymbol{\Sigma}_{mh}\boldsymbol{H}_h\left(\boldsymbol{H}_h'\boldsymbol{\Sigma}_{hh}\boldsymbol{H}_h\right)^{-}\boldsymbol{H}_h'\boldsymbol{\Sigma}_{hm}\boldsymbol{H}_m - \rho\boldsymbol{H}_m'\boldsymbol{\Sigma}_{mm}\boldsymbol{H}_m\right]\boldsymbol{u} = \mathbf{0} \tag{A.4}$$

where $(\boldsymbol{H}_h'\boldsymbol{\Sigma}_{hh}\boldsymbol{H}_h)^{-}$ is the Moore-Penrose inverse of $\boldsymbol{H}_h'\boldsymbol{\Sigma}_{hh}\boldsymbol{H}_h$. Using the property of Moore-Penrose inverse, we have:[17]

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

which holds if $\boldsymbol{H}_h$ has full row rank. Then, (A.4) becomes:

$$\boldsymbol{H}_m'\left(\boldsymbol{\Sigma}_{mh}\boldsymbol{\Sigma}_{hh}^{-1}\boldsymbol{\Sigma}_{hm} - \rho\boldsymbol{\Sigma}_{mm}\right)\boldsymbol{H}_m\boldsymbol{u} = \mathbf{0}. \tag{A.5}$$

[17] 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 $AA^{-} = I$.

Using (A.1), we rewrite (A.5) as

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

which shows that both (A.2) and (A.4) will produce the same non-zero eigenvalues.

We now consider the following spectral decompositions:

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

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

$$\begin{aligned} \boldsymbol{H}_m' \boldsymbol{\Sigma}_{mm} \boldsymbol{H}_m &= \begin{bmatrix} \boldsymbol{P}_1 & \boldsymbol{P}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Delta}_1^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{P}_1 & \boldsymbol{P}_2 \end{bmatrix}' \\ \boldsymbol{H}_h' \boldsymbol{\Sigma}_{hh} \boldsymbol{H}_h &= \begin{bmatrix} \boldsymbol{Q}_1 & \boldsymbol{Q}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Delta}_2^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}_1 & \boldsymbol{Q}_2 \end{bmatrix}' \end{aligned} \tag{A.6}$$

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

$$\boldsymbol{\Delta}_1^{-1} \boldsymbol{P}_1' \left(\boldsymbol{H}_m' \boldsymbol{\Sigma}_{mh} \boldsymbol{H}_h\right) \boldsymbol{Q}_1 \boldsymbol{\Delta}_2^{-1} = \boldsymbol{W} \begin{bmatrix} \boldsymbol{R}^{\frac{1}{2}} & \mathbf{0} \end{bmatrix} \boldsymbol{D}' \tag{A.7}$$

where $\boldsymbol{W}$ is an $(r_0 + r_m) \times (r_0 + r_m)$ orthonormal matrix, $\boldsymbol{D}$ an $(r_0 + r_h) \times (r_0 + r_h)$ orthonormal matrix and $\boldsymbol{R}$ the $(r_0 + r_m) \times (r_0 + r_m)$ diagonal matrix given by $\boldsymbol{R} = diag(\rho_1, ..., \rho_{r_0}, \rho_{r_0+1}, ..., \rho_{r_0+r_m}) = diag(1, ..., 1, 0, ..., 0)$.[18]

Define the full rank matrices,

$$\boldsymbol{A} = \left[\boldsymbol{P}_1 \boldsymbol{\Delta}_1^{-1} \boldsymbol{W}, \boldsymbol{P}_2\right] \text{ and } \boldsymbol{B} = \left[\boldsymbol{Q}_1 \boldsymbol{\Delta}_2^{-1} \boldsymbol{D}, \boldsymbol{Q}_2\right] \tag{A.8}$$

Combining (A.6), (A.7) and (A.8), it is straightforward to show that

$$Var\left(\begin{bmatrix} \boldsymbol{A}' \boldsymbol{H}_m' \boldsymbol{K}_{mt} \\ \boldsymbol{B}' \boldsymbol{H}_h' \boldsymbol{K}_{ht} \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{I}_{r_0+r_m} & \mathbf{0} & \boldsymbol{R}^{\frac{1}{2}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{R}^{\frac{1}{2}} & \mathbf{0} & \boldsymbol{I}_{r_0+r_m} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I}_{r_h-r_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{A.9}$$

[18] Notice that $\boldsymbol{R}$ contains the same non-zero roots as in (A.4), see Rao (1981).

From (A.9), we obtain the characteristic equation between $\boldsymbol{A}'\boldsymbol{H}_m'\boldsymbol{K}_{mt}$ and $\boldsymbol{B}'\boldsymbol{H}_h'\boldsymbol{K}_{ht}$ by

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

which can be simplified as

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

Hence,

$$\left(\begin{bmatrix}\boldsymbol{R} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}-\rho\begin{bmatrix}\boldsymbol{I}_{r_0+r_m} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\right)\boldsymbol{u}=\boldsymbol{0} \tag{A.10}$$

Obviously, (A.10) has the same characteristic roots from (A.4) and the same non-zero characteristic roots from (A.2), consequently.

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

$$Var\begin{pmatrix}\widetilde{\boldsymbol{K}}_m\\ \widetilde{\boldsymbol{K}}_h\end{pmatrix}=\frac{1}{T}\begin{bmatrix}\widetilde{\boldsymbol{K}}_m'\widetilde{\boldsymbol{K}}_m & \widetilde{\boldsymbol{K}}_m'\widetilde{\boldsymbol{K}}_h\\ \widetilde{\boldsymbol{K}}_h'\widetilde{\boldsymbol{K}}_m & \widetilde{\boldsymbol{K}}_h'\widetilde{\boldsymbol{K}}_h\end{bmatrix}=\begin{bmatrix}\widetilde{\boldsymbol{S}}_{mm} & \widetilde{\boldsymbol{S}}_{mh}\\ \widetilde{\boldsymbol{S}}_{hm} & \widetilde{\boldsymbol{S}}_{hh}\end{bmatrix}$$

Consider the full rank transformation $\widetilde{\boldsymbol{K}}_m\boldsymbol{A}$ and $\widetilde{\boldsymbol{K}}_h\boldsymbol{B}$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are defined in (A.8). The canonical correlations between them are equivalent to those between $\widetilde{\boldsymbol{K}}_m$ and $\widetilde{\boldsymbol{K}}_h$. By Lemma 4, we obtain: $\boldsymbol{A}'\widetilde{\boldsymbol{S}}_{mm}\boldsymbol{A}\xrightarrow{p}\boldsymbol{A}'\boldsymbol{H}_m'\boldsymbol{\Sigma}_{mm}\boldsymbol{H}_m\boldsymbol{A}$, $\boldsymbol{B}'\widetilde{\boldsymbol{S}}_{hh}\boldsymbol{B}\xrightarrow{p}\boldsymbol{B}'\boldsymbol{H}_h'\boldsymbol{\Sigma}_{hh}\boldsymbol{H}_h\boldsymbol{B}$ and $\boldsymbol{A}'\widetilde{\boldsymbol{S}}_{mh}\boldsymbol{B}\xrightarrow{p}\boldsymbol{A}'\boldsymbol{H}_m'\boldsymbol{\Sigma}_{mh}\boldsymbol{H}_h\boldsymbol{B}$. Let $\underline{M}=\min\{M_m,M_h\}$ and $\delta_{\underline{M}T}=\min\{\sqrt{\underline{M}},\sqrt{T}\}$. Applying (A.9) and Lemma 4, we can rewrite these transformed variance/covariance matrices as

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

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

and

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

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

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

Similarly, $\left(\boldsymbol{B}'\widetilde{\boldsymbol{S}}_{hh}\boldsymbol{B}\right)^{-}$ does not converge to $\begin{bmatrix}\boldsymbol{I}_{r_0+r_m} & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_{r_h-r_m} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}$. See Theorem 1 in Karabiyik et al. (2017). But, the Moore-Penrose inverse follows the Banachiewicz-Schur form.[19] Thus,

$$\begin{bmatrix}\boldsymbol{I}_{r_h-r_m}+O_p(\delta_{\underline{M}T}^{-2}) & O_p(\delta_{\underline{M}T}^{-2})\\ O_p(\delta_{\underline{M}T}^{-2}) & O_p(\delta_{\underline{M}T}^{-2})\end{bmatrix}^{-}=\begin{bmatrix}\boldsymbol{I}_{r_h-r_m}+O_p(\delta_{\underline{M}T}^{-2}) & -O_p(1)\\ -O_p(1) & O_p(\delta_{\underline{M}T}^{2})\end{bmatrix}=O_p(\delta_{\underline{M}T}^{2}) \tag{A.11}$$

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

$$\left(\boldsymbol{B}'\widetilde{\boldsymbol{S}}_{hh}\boldsymbol{B}\right)^{-}=\begin{bmatrix}\boldsymbol{I}_{r_0+r_m}+O_p(\delta_{\underline{M}T}^{-2}) & -O_p(1)\\ -O_p(1) & O_p(\delta_{\underline{M}T}^{2})\end{bmatrix}. \tag{A.12}$$

Using the above results, we obtain:

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

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

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

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

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

can be rewritten as

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

which is analogous to (A.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 for Lemma 2 Using Lemma 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*

[19] 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 and Takane (2009) and Castro-González et al. (2015)

Proof for Lemma 3. By applying Lemma 1 to the definition of $CCD(r)$, it is straightforward to derive the main results in Lemmas 3. *Q.E.D*

Proof for Theorem 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 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

B Simulation results

Table 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 $r^*_{\max}$.

Table 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 $r^*_{\max} = \max\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_R}\}$

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.

,

Table 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.

Table 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 $\boldsymbol{F}_t = \boldsymbol{\Phi}_F \boldsymbol{F}_{t-1} + \boldsymbol{w}_t$ with $\boldsymbol{w}_t \sim iidN(0, \boldsymbol{\Omega}_F)$, where $\boldsymbol{F}_t = [\boldsymbol{F}'_{1t}, \ldots, \boldsymbol{F}'_{Rt}]'$, $\boldsymbol{w}_t = [\boldsymbol{w}'_{1t}, \ldots, \boldsymbol{w}'_{Rt}]'$ and $\boldsymbol{\Phi}_F$ is a diagonal matrix with the common elements, 0.5. We set the common diagonal elements of $\boldsymbol{\Omega}_F$ 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.

C Empirical Results

Table 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.

Figure 1: Average pairwise correlations of returns

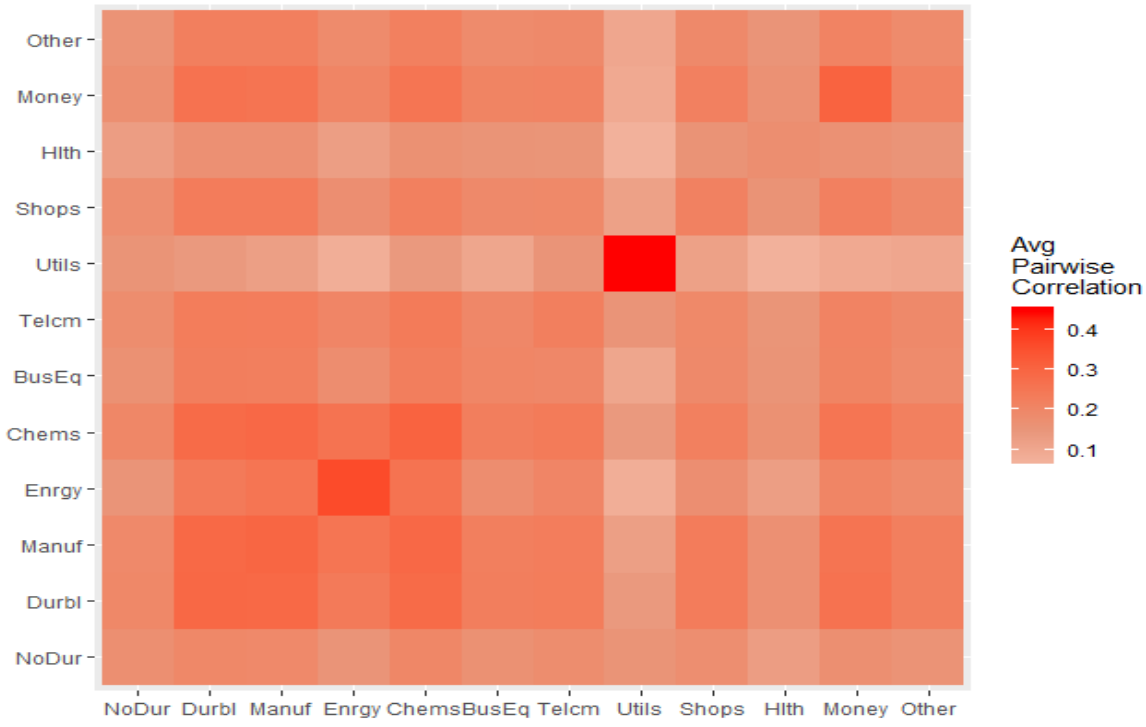

Figure 2: Average pairwise correlations of residuals after concentrating out $\widehat{\boldsymbol{G}}$

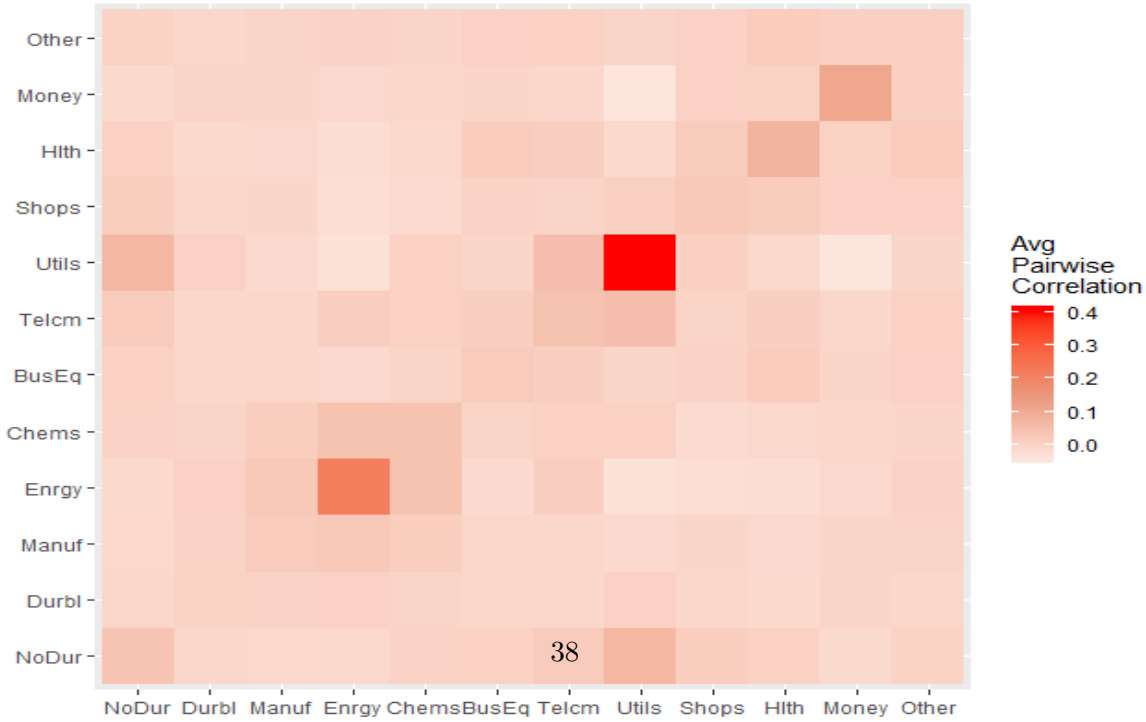

Figure 3: Average pairwise correlations of residuals after concentrating out $\widehat{\boldsymbol{G}}$ and $\widehat{\boldsymbol{F}}_i$'s

Figure 4: The global factor and market factor

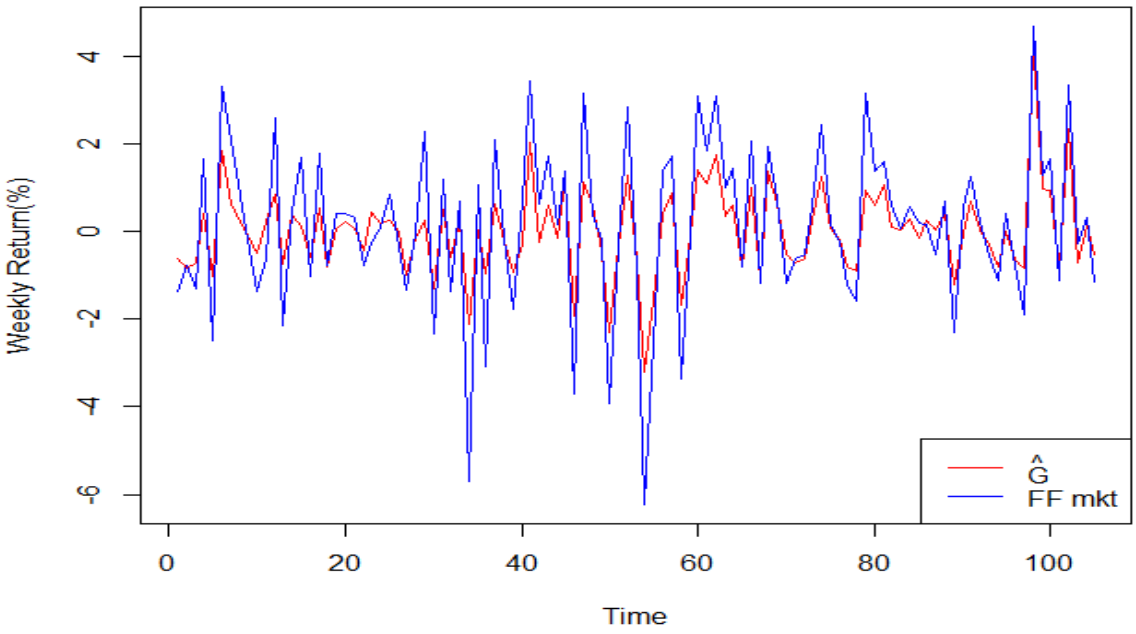

Figure 5: Density plots of the global and local factor loadings