Canonical Correlation-based Model Selection for the Multilevel Factors*

In Choi[†] Rui Lin[‡] Yongcheol Shin[§]

This Version: March 2020

Abstract

A great deal of research effort has been devoted to the analysis of the multilevel factor model. To date, however, limited progress has been made on the development of coherent inference for identifying the number of the global factors. We propose a novel approach based on the canonical correlation analysis to identify the number of the global factors. We develop the canonical correlations difference (CCD), which is constructed by the difference between the cross group-averages of the adjacent canonical correlations between factors. We prove that CCD is a consistent selection criterion. Via Monte Carlo simulations, we show that CCD always selects the number of global factors correctly even in small samples. Further, CCD outperforms the existing approaches in 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 framework with an application to the multilevel asset pricing model for the stock return data of 12 industries in the U.S.

JEL: C52, G12.

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

*We are mostly grateful to Jia Chen, Laura Coroneo, Xu Han, Young Hoon Lee, Liangjun Su, Michael Thornton, Takashi Yamagata, Paola Zerilli, Chaowen Zheng, and the seminar participants at University of York as well as conference delegates at the China Meeting of the Econometric Society (CMES, 2019) for helpful comments. Lin is grateful for 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

§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 the 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, have been growing rapidly. Here we have the global factors that influence the individuals in the whole blocks 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.

The different estimation methods have been developed: the Bayesian approach by Kose et al. (2003) and Moench et al. (2013), the classical approach by Wang (2008), Breitung and Eickmeier (2016) and Choi et al. (2018), and the LASSO approach by 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. 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. 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 by Bai and Ng (2002), 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. Via simulations we examine the finite sample performance of the popular selection criteria, IC_{p2} and BIC_3 by Bai and Ng (2002), ED by Onatski (2010) and ER by Ahn and Horenstein (2013), by ignoring the multilevel structure and applying them directly to the whole data matrix. Overall we find that these approaches tend to produce unreliable inference (see Section V in Online Supplement). Thus, aforementioned 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.

Only a small number of studies have attempted to deal with such an important issue of consistently estimating the number of global factors under the multilevel setting. Let r_0 (r_i) be the number of global (local) factors and R be the number of groups. 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 appropriate difference. The above simulation evidence also suggests that Wang's sequential procedure is quite unreliable, see also Han (2019). Chen (2012) and Dias et al. (2013) propose the modified information criteria by penalising the number of local factors more heavily than the number of global factors. As R rises, however, the computation will be almost infeasible because the number of candidate models drastically increases. 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 the numbers of global factors. But, the validity of their approach requires the strong assumption that the idiosyncratic errors are neither serially nor cross-sectionally correlated. Han (2019) proposes a shrinkage estimator that can consistently estimate the factors/loadings, and determine the number of factors, simultaneously. But, his approach requires selecting too many tuning parameters by the additional information criteria. Consequently, an extension to the model with the (mildly) large R would be almost impractical. Furthermore, it is quite challenging to develop a shrinkage estimator fully robust to the correlated local factors.

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, r_0, which can be easily applied to the models with a fixed number of blocks and with $R \to \infty$, and which will be valid in the presence of serially correlated and weakly cross-sectionally correlated idiosyncratic errors as well as of the correlated local factors. 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 maximum number of factors estimated across R blocks. Next, we evaluate the $r_{\max}$ canonical correlations between factors in any two blocks. Using the $R(R-1)$ pairwise canonical correlations, we construct the cross-block average of the canonical correlations. We then develop the proposed 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 $r = 0, 1, ..., r_{\max}$.[1]

We derive asymptotic properties of pairwise canonical correlations and the cross-block average, and show that $CCD(r)$ is a consistent selection criterion for identifying r_0. Next, via extensive Monte Carlo simulations, we investigate the finite sample properties of CCD and CCR together with two existing approaches advanced by Chen (2012) and Andreou et al. (2019). Overall, we find that the performance of CCD is quite satisfactory, and it always selects the correct number of global factors even in small samples. Furthermore, CCD outperforms other competing methods, even in the presence of serially correlated and weakly cross-sectionally correlated idiosyncratic errors as well as the correlated local factors.

Once the number of global factors is consistently estimated by $CCD(r)$, we remove the global factors from the data in each block, and apply the existing criteria to consistently estimate the number of local factors. We have conducted the additional simulations, finding that BIC_3 by Bai and Ng (2002), and ER by Ahn and Horenstein (2013) outperform the other approaches, see Section I in Online Supplement.

Our proposed approach possess a number of advantages. First, it is very simple to apply as it involves the standard PC and CCA methods, unlike other approaches that require us to assess many tuning or control parameters, e.g. Chen (2012) and 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 the $R(R-1)/2$ pairwise canonical correlations, unlike other

[1] We have also developed the $CCR(r)$, using the ratio of the consecutive cross-block averages.

approaches that will be computationally infeasible or impractical, e.g. Andreou et al. (2019) and Han (2019). More importantly, our approach is shown to be robust to the presence of non-zero correlations between the local factors.

We demonstrate the utility of our framework 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. First, we apply CCD and 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, and find one local factor in NoDur, Enrgy, Hlth and Money, and two local factors in Utils. On average, we find that 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.

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 and discusses the underlying properties and assumptions. Section 4 develops the CCD criterion for selecting the number of global factors and derives the asymptotic theory. Section 5 presents Monte Carlo simulation evidence. Section 6 provides an empirical illustration analysing the multilevel factor structure for the stock return data in 12 industries in the U.S. Section 7 offers concluding remarks. All the mathematical proofs are relegated to Appendix. Additional simulation results and theoretical derivations can be found in Online Supplement.

2 Related Literature

For the traditional approximate factor model in the single-level panel data, 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 (PC) 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. Extending Onatski (2006), Onatski (2010) develops the edge distribution (ED) estimator based on the difference between the adjacent eigenvalues arranged in descending order such that $\hat{r}_0 = \max_{1 \le r \le r_{\max}} \{r | \mu_r - \mu_{r+1} \ge \delta\}$, 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 the maximum value of r, $r_{\max}$. Ahn and Horenstein (2013) propose a simpler estimator, called the eigenvalue ratio (ER) given by $\hat{r}_0 = \max_{1\le r\le r_{\max}}\{\mu_r/\mu_{r+1}\}$. Choi and Jeong (2019) have conducted a comprehensive simulation study about traditional 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.

In the presence of the multi-level 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. If we apply the 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 the number of blocks, R is fixed (and small), it is well-established that 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. Then, Han (2019) conjectures that the number of global factors can be consistently estimated asymptotically by the existing selection criteria (see Remark 4).

In Section V in Online Supplement we examine the finite sample performance of the four selection criteria, IC_{p2}, BIC_3 by Bai and Ng (2002), ED by Onatski (2010) and ER by Ahn and Horenstein (2013), by applying them directly to the whole data matrix. We find that the existing approaches tend to produce unreliable inference. If $R = 2$, all of the four criteria tend to 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 the number of global factors. However, for the moderate value of R, e.g. $R = 5$ or 10, they tend to 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 in the presence of cross-sectionally and serially correlated errors. Finally, and more importantly, even for large R, they overestimate the number of global factors significantly in the presence of moderate correlations among the local factors.

Only a few studies have recently 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 an idea similar to the principle of inclusion–exclusion for set cardinality.[2] The simulation evidence provided above also implies that Wang's sequential procedure is quite unreliable. In particular, for large R, it severely overestimates by selecting $r_0 + \sum_{i=1}^{R} r_i/R$ instead of r_0. Furthermore, Han (2019) provides the simulation evidence that this approach can lead to even negative estimates of both the number of global factors and the number of local factors in small samples, for $R = 3$.

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 number of candidate models drastically increases

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

such that the computation will be almost infeasible.[3]

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 the numbers of global factors 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 then extract the $T \times r_{\max}$ matrix of factors, $\widehat{\boldsymbol{K}}_i$, from the data, $\boldsymbol{Y}_i$ for $i = 1, 2$, where $r_{\max} = \min\{\widehat{r_0 + r_1}, \widehat{r_0 + r_2}\}$. They 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, denoted $\hat{\xi}(r)$ (see Theorem 2). By imposing the strong assumption that idiosyncratic errors are neither serially nor cross-sectionally correlated, AGGR are able to derive that $\tilde{\xi}(r)$ follows the standard normal distribution asymptotically under the null hypothesis, $r_0 = r$. Furthermore, this procedure can be used for model selection only if the critical value diverges at a certain rate, γ with $0 < \gamma < 1$. Then 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 suggest to estimate the number of local factors by $\widehat{r_0 + r_i} - \hat{r}_0$ for $i = 1, 2$. Moreover, it would be quite complicated to extend their approach to cover the case with $R > 2$.

Han (2019) proposes a shrinkage estimator that can consistently estimate the factors/loadings, and determine the number of factors, simultaneously. In particular, the number of global (local) factors can be estimated by the number of non-zero columns in their respective factor loading matrices. But, his approach requires us to select tuning parameters by imposing different penalty terms for different blocks. Consequently, even when R is (mildly) large, there will be a large number of candidate tuning parameters to be selected coherently. Further, the shrinkage estimation results are not invariant to the order of the blocks. Hence, we need to apply the additional information criteria to determine which block is ordered first. Indeed, he considers the case with a fixed number of groups, but an extension to the case with the 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 he acknowledges that it is challenging to develop a shrinkage estimator fully robust to the local factors correlations.[4]

We propose a novel and simple approach based on the canonical correlation analysis. We first apply the PC estimation to the data in each block and obtain the $r_{\max}$ factors. Next, we evaluate the $r_{\max}$ canonical correlations between factors in any two blocks, and construct their cross-block averages. We then construct $CCD(r)$ as the difference between the consecutive cross-block averages. The number of global factors can be estimated consistently by maximising $CCD(r)$ over $r = 0, 1, ..., r_{\max}$. Our method differentiates from the above approaches in two main aspects. First, our approach can be easily applied to the models with a fixed number of blocks and with $R \to \infty$. More importantly, our approach is shown to be valid in the presence

[3]For example, if we use the modified information criteria with $R = 10$ and $r_{max} = 5$, there are 5×5^{10} candidate models.

[4]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.

of serially correlated and weakly cross-sectionally correlated idiosyncratic errors as well as of 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 block i, $\boldsymbol{\gamma}_{ij}$ and $\boldsymbol{\lambda}_{ij}$ are conformably defined factor loadings and e_{ijt} is the idiosyncratic error. Stacking (3.1) across individuals in block i, we have

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

where

$$\underset{M_i \times 1}{\boldsymbol{y}_{it}} = \begin{bmatrix} y_{i1t} \\ y_{i2t} \\ \vdots \\ y_{iM_it} \end{bmatrix}, \underset{M_i \times 1}{\boldsymbol{e}_{it}} = \begin{bmatrix} e_{i1t} \\ e_{i2t} \\ \vdots \\ e_{iM_it} \end{bmatrix}, \underset{M_i \times r_0}{\boldsymbol{\Gamma}_i} = \begin{bmatrix} \boldsymbol{\gamma}_{i1}' \\ \boldsymbol{\gamma}_{i2}' \\ \vdots \\ \boldsymbol{\gamma}_{iM_i}' \end{bmatrix}, \underset{M_i \times r_i}{\boldsymbol{\Lambda}_i} = \begin{bmatrix} \boldsymbol{\lambda}_{i1}' \\ \boldsymbol{\lambda}_{i2}' \\ \vdots \\ \boldsymbol{\lambda}_{iM_i}' \end{bmatrix}.$$

The model can also be written as

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

where

$$\underset{N \times 1}{\boldsymbol{Y}_t} = \begin{bmatrix} \boldsymbol{y}_{1t} \\ \boldsymbol{y}_{2t} \\ \vdots \\ \boldsymbol{y}_{Rt} \end{bmatrix}, \underset{N \times 1}{\boldsymbol{e}_t} = \begin{bmatrix} \boldsymbol{e}_{1t} \\ \boldsymbol{e}_{2t} \\ \vdots \\ \boldsymbol{e}_{Rt} \end{bmatrix}, \underset{r^+ \times 1}{\boldsymbol{F}_t^+} = \begin{bmatrix} \boldsymbol{G}_t \\ \boldsymbol{F}_{1t} \\ \boldsymbol{F}_{2t} \\ \vdots \\ \boldsymbol{F}_{Rt} \end{bmatrix}, \underset{N \times r^+}{\boldsymbol{\Lambda}^+} = \begin{bmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Lambda}_1 & 0 & \cdots & 0 \\ \boldsymbol{\Gamma}_2 & 0 & \boldsymbol{\Lambda}_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_R & 0 & 0 & \cdots & \boldsymbol{\Lambda}_R \end{bmatrix}$$

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:

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

where

$$\underset{T \times N}{\boldsymbol{Y}} = \begin{bmatrix} \boldsymbol{Y}_1' \\ \boldsymbol{Y}_2' \\ \vdots \\ \boldsymbol{Y}_T' \end{bmatrix}, \underset{T \times r^+}{\boldsymbol{F}^+} = \begin{bmatrix} \boldsymbol{F}_1^{+\prime} \\ \boldsymbol{F}_2^{+\prime} \\ \vdots \\ \boldsymbol{F}_T^{+\prime} \end{bmatrix} \text{ and } \underset{T \times N}{\boldsymbol{e}} = \begin{bmatrix} \boldsymbol{e}_1' \\ \boldsymbol{e}_2' \\ \vdots \\ \boldsymbol{e}_T' \end{bmatrix}.$$

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

$$\underset{T\times 1}{\boldsymbol{Y}_{ij}} = \begin{bmatrix} \boldsymbol{y}_{ij,1} \\ \boldsymbol{y}_{ij,2} \\ \vdots \\ \boldsymbol{y}_{ij,T} \end{bmatrix}, \underset{T\times 1}{\boldsymbol{e}_{ij}} = \begin{bmatrix} \boldsymbol{e}_{ij,1} \\ \boldsymbol{e}_{ij,2} \\ \vdots \\ \boldsymbol{e}_{ij,T} \end{bmatrix}, \underset{T\times r_0}{\boldsymbol{G}} = \begin{bmatrix} \boldsymbol{G}_1' \\ \boldsymbol{G}_2' \\ \vdots \\ \boldsymbol{G}_T' \end{bmatrix}, \underset{T\times r_i}{\boldsymbol{F}_i} = \begin{bmatrix} \boldsymbol{F}_{i1}' \\ \boldsymbol{F}_{i2}' \\ \vdots \\ \boldsymbol{F}_{iT}' \end{bmatrix}$$

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

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

Assumption A.

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* $\boldsymbol{K}_{it} = \begin{bmatrix} \boldsymbol{G}_t \\ \boldsymbol{F}_{it} \end{bmatrix}$.

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,

   (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. In particular, 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 is standard and ensures that the global factors and local factors can be separately identified. Initially, we make Assumption D2 that the local factors are mutually uncorrelated. 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 $R \longrightarrow \infty$ (see Section V in Oniline Supplement).

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 $i = 1, \dots, R$. Select a positive integer, $r_{\max}$ such that $r_{\max} \geq r_0 + r_i$ for all i. 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 $\widehat{\boldsymbol{K}}_h$ are consistent for 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 $\widehat{\boldsymbol{K}}_h$. In order to cover the cases with zero global factor and zero local factor, we set two mock squared canonical correlations at $\ell_{mh,0} = 1$ and $\ell_{mh,r_{\max}+1} = 0$.[5]

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

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

[5] Ahn and Horenstein (2013) set a mock eigenvalue at the beginning only, to cover the possibility of zero factor.

Suppose that the blocks, m and h, share the r_0 global factors. Then, 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 $0 \leq r \leq r_0$, 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

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

Next, we consider an alternative criterion based on the average of the first r largest squared canonical correlations given by

$$\bar{\ell}_{mh}(r) = \frac{1}{r+1} \sum_{k=0}^{r} \ell_{mh,k}.$$

We construct the cross-block average of $\bar{\ell}_{mh}(r)$ by

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

and the canonical correlation ratio (CCR) as

$$CCR(r) = \frac{\bar{\xi}(r)}{\bar{\xi}(r+1)} \text{ for } r = 0, 1, \ldots, r_{\max}$$

We then propose to estimate the number of global factors by

$$\hat{r}_{0,CCR} = \underset{0 \leq r \leq r_{\max}}{argmax}\ CCR(r)$$

We provide Lemmas 2 and 3 about the asymptotic properties of $\xi(r)$, $\bar{\xi}(r)$, CCD and CCR.

Lemma 2. *Under Assumptions A-D, as $M_1, ..., M_R, T \to \infty$,*
(i)

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

(ii)

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

Lemma 2 shows under Assumptions A-D that $\xi(r)$ and $\bar{\xi}(r)$ are equal to 1 for $r = 0, \ldots, r_0$, while $\xi(r)$ is 0 and $\bar{\xi}(r)$ is less than 1 for $r = r_0 + 1, \ldots, r_{\max} + 1$, asymptotically.

Lemma 3. *Suppose that Assumptions A-D hold.*
(i) For $r_0 > 0$, as $M_1, ..., M_R, T \to \infty$, then

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

$$CCR(r) \xrightarrow{p} \begin{cases} 1 & \text{for } r = 0, ..., r_0 - 1 \\ \frac{2+r}{1+r} & \text{for } r = r_0, ..., r_{\max} \end{cases}$$

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

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

$$CCR(r) \xrightarrow{p} \begin{cases} 2 & \text{for } r = 0 \\ \frac{2+r}{1+r} & \text{for } r = 1, ..., r_{\max} \end{cases}$$

The following theorem shows that $\hat{r}_{0,CCD}$ and $\hat{r}_{0,CCR}$ are consistent model selection criteria.

Theorem 1. *Suppose Assumptions A-D hold. Then,*

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

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

It is quite intuitive to apply a canonical correlation-based approach to identify the number of global factors. In this regard, our approach shares the similar idea with AGGR in terms of developing the consistent selection criteria by using the fact that the r_0 canonical correlations are equal to one, and 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, AGGR are able to derive that $\tilde{\xi}(r)$ follows the standard normal distribution asymptotically under the null hypothesis, $r_0 = r$.

On the other hand, we share the similar idea with Onatski (2010) and Ahn and Horenstein (2013), and propose to employ the difference between and the ratio of adjacent canonical correlations as the selection criteria. Notice that $CCD(r)$ is free from the value of r_0 whilst $CCR(r)$ is sensitive to the large value of r_0 as $CCR(r)$ gets close to 1. This suggests that CCD tends to outperform CCR.[6] This may be mainly due to the fact that CCD does not require calibrating any threshold as in Onatski (2010) because the r_0 largest canonical correlations are all bounded by one.

4.1.1 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 correlated, but we also notice that such correlations should not be set too high. Otherwise, strong correlations between blocks, m and h, imply that the local factors in block m would influence the individuals in block h, and *vice versa*. In this situation, it is rather difficult to distinguish between the roles played by the global and local factors, which is clearly inconsistent with an original purpose of the multilevel factor model.

In this regard, we now assume:

Assumption D′. *Let* $\bar{\rho}_r = \frac{2}{R(R-1)} \sum_{m=1}^{R-1} \sum_{h=m+1}^{R} \rho_{mh,r}$, *where* $\rho_{mh,r}$ *is the* r*-th population canonical correlation between* $\boldsymbol{K}_m$ *and* $\boldsymbol{K}_h$*. Then, we set:* $\bar{\rho}_{r_0+1} < \frac{1}{2}$.

This assumption allows for non-zero correlations among local factors, but imposes the upper bound on their correlations. Specifically, the average of (r_0+1)-th population canonical correlation between $\boldsymbol{K}_m$ and $\boldsymbol{K}_h$, or equivalently, the average largest population canonical correlation between $\boldsymbol{F}_m$ and $\boldsymbol{F}_h$ cannot exceed $\frac{1}{2}$. Suppose that $r_m \leq r_h$ for $m \neq h$. By construction we have $1 = \rho_{mh,0} = \cdots = \rho_{mh,r_0} > \rho_{mh,r_0+1} \geq \cdots \geq \rho_{mh,r_0+r_m} \geq 0 = \rho_{mh,r_0+r_m+1} = \cdots = \rho_{mh,r_{\max}+1}$. We provide the following Lemmas, which are extensions of Lemmas 1-3, see the proofs in Section II in Online Supplement.

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

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

Lemma 2*. *Under Assumptions A-D(i) and D′, as* $M_1, ..., M_R, T \rightarrow \infty$,

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

[6] In an earlier version, we have also constructed the alternative versions, $CCD^*(r) = \bar{\xi}(r) - \bar{\xi}(r+1)$ and $CCR^*(r) = \xi(r)/\xi(r+1)$. Using Lemma 2, it is easily seen that CCD^* is consistent whilst CCR^* becomes indeterminate as $CCR^*(r) \xrightarrow{p} 0/0$ for $r_0 < r \leq r_{\max}$. Furthermore, the CCD^* is shown to be less accurate and more sensitive to a large $r_{\max}$ than CCD.

Lemma 3^***.** *Suppose that Assumptions A-D(i) and D′ hold.*
(i) For $r_0 > 0$*, as* $M_1, ..., M_R, T \to \infty$*, then*

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

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

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

It is clear from Lemma 3^* that the consistency of CCD requires the condition: $1 - \bar{\rho}_{r_0+1} > \bar{\rho}_r - \bar{\rho}_{r+1}$ for $r_0 < r \leq r_{\max}$, which is satisfied under a sufficient condition, $\bar{\rho}_{r_0+1} < \frac{1}{2}$ in Assumption D′. This implies that if the largest block-average of canonical correlations among the local factors is smaller than $\frac{1}{2}$, then CCD is still a consistent selection criterion:

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

Notice that the consistency of CCR is not guaranteed in the presence of non-zero correlation between the local factors, even if we relax the sufficient condition on the upper bound of $\bar{\rho}_{r_0+1}$. For instance, consider a two-block multilevel factor model with $r_0 = r_i = 1$ and $\bar{\rho}_2 = \rho_{12,2} = \frac{1}{3}$. Then, we obtain: $\bar{\xi}(0) \xrightarrow{p} 1$, $\bar{\xi}(1) \xrightarrow{p} 1$, $\bar{\xi}(2) \xrightarrow{p} 7/9$ and $\bar{\xi}(3) \xrightarrow{p} 7/12$. Hence, we have $CCR(0) = 1$, $CCR(1) = 9/7$, $CCR(2) = 12/9$ so that CCR is maximized incorrectly at $r = 2$.[7]

4.2 Estimation of the Number of Local Factors

Once the number of global factors is consistently estimated by $\hat{r}_0$, then 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). While $\widehat{\boldsymbol{G}}$ from any block-pair would provide a consistent estimator for $\boldsymbol{G}$, 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 $\boldsymbol{Y}_i^G = \boldsymbol{M}^G \boldsymbol{Y}_i$ for $i = 1, ..., R$ where $\boldsymbol{M}^G = \boldsymbol{I}_T - \widehat{\boldsymbol{G}}(\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{G}})\widehat{\boldsymbol{G}}'$. Then, we apply the existing approaches by Bai and Ng (2002) and Ahn and Horenstein (2013) to $\boldsymbol{Y}_i^G$, 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 $\hat{r}_i$.[8] We apply the PC estimation to $\boldsymbol{Y}_i^G$ 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 an OLS regression of y_{ijt} on $\widehat{\boldsymbol{G}}_t$ and $\widehat{\boldsymbol{F}}_{it}$.

[7] See Table A3 in Section 5, where we find that the performance of CCR improves with the sample sizes if the correlation between the local factors is set at 0.2, whilst CCR tends to overestimate the number of global factors even in large smaples if the correlation between the local factors is set at 0.4.

[8] 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.

4.3 Estimation of Global and Local Factors and Loadings

In Section 4.1 and 4.2, we have already obtained $\hat{r}_0$, $\hat{r}_i$ and 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 estimation 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 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 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 eigenvectors corresponding to the $\hat{r}_i$ largest eigenvalues of the $T \times T$ matrix, $\boldsymbol{Y}_i^{\widetilde{\boldsymbol{G}}}\boldsymbol{Y}_i^{\widetilde{\boldsymbol{G}}\prime}$, while local factor loadings by $\widetilde{\boldsymbol{\Lambda}}_i = \widetilde{\boldsymbol{F}}_i'\boldsymbol{Y}_i^{\widetilde{\boldsymbol{G}}}/T$.

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

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

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

$$\gamma_{ijz} \sim iidN(0,1), z = 1, ..., r_0;\ \lambda_{ijz} \sim iidN(0,1), z = 1, ..., r_i$$

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

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

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

$$Var(\boldsymbol{\lambda}_{ij}'\boldsymbol{F}_{it}) = \sum_{z=1}^{r_i} Var(\lambda_{ijz}F_{itz}) = \frac{r_i}{1-\phi_F^2}$$

$$Var(e_{ijt}) = \frac{1+16\beta^2}{1-\phi_e^2}.$$

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 $T \in \{50, 100\}$. The number of replications for each simulation experiment is set at 1,000. We focus on the estimation of r_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.[9] When implementing IC_{chen} and $AGGR$ in the simulation, we assume that the true number of factors $r_0 + r_i$ is known. This assumption prevents us from selecting too many candidate models in IC_{chen}.[10] For $AGGR$, the null hypothesis is sequentially tested from $k = r_0 + r_i$ until rejected.

We only report the results for the cases with $\phi_G = \phi_F = 0.5$ to save space.[11] In the first experiment, we fix the number of factors as $(r_0, r_i) = (2, 2)$ for $i = 1, \ldots, R$ and $r_{\max} = 5$. Panel A of Table A1 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, respectively. For example, $(0|0)$ implies that r_0 is perfectly correctly estimated. CCD performs very well for all sample sizes while CCR tends to be less accurate when R and/or T are small. IC_{Chen} performs reasonably well only for $R = 2$, but it significantly underestimates by detecting only one global factor for $R = 5$ and $R = 10$. $AGGR$ severely overestimates 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 A1 demonstrate that the performance of CCR, IC_{Chen} and $AGGR$ deteriorates substantially as compared to the first experiment. In particular, $AGGR$ produces quite imprecise estimates because their approach is invalid in the case where idiosyncratic errors are serially and/or cross-sectionally correlated (see Assumption A9 and Theorem

[9] See Section III in Online Appendix for the estimation algorithms in details. 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.

[10] For example, in the case where $(r_0, r_i) = (3, 3)$, the candidate models are restricted to $\{(0,6), (1,5), (2,4), (3,3), (4,2), (5,1), (6,0)\}$. If IC_{chen} works properly, it should report $(\hat{r}_0, \hat{r}_i) = (3, 3)$.

[11] We obtain qualitatively similar result for $\phi_G = \phi_F = 0$. These results are available upon request.

2 in $AGGR$). Still, CCD tends to select the correct r_0 in almost all cases. In line with our theoretical prediction, the performance of CCD is mostly invariant to the presence of serially and cross-sectionally correlated idiosyncratic errors.

The third case of the first experiment is a very noisy DGP with $\kappa = 3$ in which the variance share explained by the global factors becomes only 20%.[12] The other setups are the same as in the second case. As reported in Panel C of Table A1, the performance of all the approaches are adversely affected, especially when T is small. CCR is less accurate unless R, M and T become sufficiently large whilst the performances of $AGGR$ are rather unreliable in all cases. The performance of IC_{Chen} improves as M or T rises only for $R = 2$, but it still underestimates for $R = 5$ and $R = 10$ even in large samples. The CCD still outperforms the others and its performance improves sharply as M or T rises for all the values of R.

Table A1 about here

In the second experiment we allow the number of global factors to vary from 0 to 3 by setting $(r_0, r_i) \in \{(0, 2), (1, 1), (3, 3)\}$ and $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$.[13] First, consider the results for the case with $(r_0, r_i) = (0, 2)$, which are reported in Panel A of Table A2. CCD, CCR and $AGGR$ tend to select the zero global factor correctly, but CCD slightly outperforms. IC_{chen} always selects one factor incorrectly. Second, when turning to the case with $(r_0, r_i) = (1, 1)$, we find that both CCD and IC_{chen} estimate $r_0 = 1$ correctly while CCR and $AGGR$ display the less reliable results in small samples, see Panel B of Table A2. Finally, the results for $(r_0, r_i) = (3, 3)$ presented in Panel C show that if $R = 2$, the performance of both CCD and IC_{chen} is reasonably satisfactory and improves sharply with the sample sizes, whereas the performance of CCR and $AGGR$ is still unreliable unless both M and T are substantially large. Next, if $R = 5$ and $R = 10$, the performance of CCD remains quite satisfactory, but IC_{chen} severely underestimates.

Tables A2 about here

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

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

where $\boldsymbol{F} = [\boldsymbol{F}'_{1t}, \ldots, \boldsymbol{F}'_{Rt}]'$, $\boldsymbol{w}_t = [\boldsymbol{w}'_{1t}, \ldots, \boldsymbol{w}'_{Rt}]'$ and Φ_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 and 0.4, respectively. The performance of IC_{Chen} and $AGGR$ is qualitatively similar to Table A1. We find that CCD is still satisfactory, unless M or T is too small. CCR tends to overestimate and its performance does not improve for $\omega_F = 0.4$, even as T and M increase, suggesting that CCR is not always consistent in the presence of non-zero correlations between the local factors as discussed in Section 4.1.1. Overall, CCD outperforms other competing approaches even in the presence of correlated local factors.

[12] This figure matches closely with empirical evidence reported in Table 6 in Section 6.

[13] When $r_0 = 3$, we reset $r_{\max}$ to 8.

Table A3 about here

Finally, we propose a practical guideline for selecting $r_{\max}$. Notice that we are more likely to select the redundant factors, if $r_{\max}$ is set too high.[14] First, we suggest to apply the BIC_3 or ER, to each $\boldsymbol{Y}_i$, with a sufficiently large $r_{\max}$, and obtain $\widehat{r_0 + r_i}$ for all i. Then, we propose to select $\hat{r}_{\max}$ by $\hat{r}_{\max} = \max\{\widehat{r_0 + r_1}, \dots, \widehat{r_0 + r_R}\}$. This procedure selects $\hat{r}_{\max}$ smaller than $r_{\max}$, while ensuring that $Pr(\hat{r}_{\max} \geq r_0 + r_i) \xrightarrow{p} 1$ for all i. In Table A4, we report and compare simulation results for CCD using fixed $r_{\max} = 10$ and $\hat{r}_{\max} = \max\{\widehat{r_0 + r_1}, \dots, \widehat{r_0 + r_R}\}$ selected by BIC_3. We consider four cases with $(\beta, \phi_e) = \{(0, 0), (0.1, 0), (0, 0.5), (0.1, 0.5)\}$. If idiosyncratic errors are serially correlated, the impact of the large $r_{\max}$ on the performance of CCD would be non-negligible such that CCD tends to overestimate r_0, especially if T is small. By contrast, when applying our practical guideline, we find that CCD selects the number of global factors correctly even in small samples.

Table A4 about here

Overall, we demonstrate that CCD tends to select the number of global factors correctly even in small samples while outperforming all other methods even in the presence of serially correlated and weakly cross-sectionally correlated idiosyncratic errors as well as the correlated local factors.

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

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), Fama and French (1993) and Connor and Korajczyk (1988). In practice, however, the studies investigating the role of industry factors explicitly in asset pricing model are relatively scarce. 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

[14] 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.

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.[15] We use the SIC codes to categorise the stocks into twelve industries, listed in the first column of Table A5.[16]. We consider a balanced panel data and include stocks that have the complete return data during the sample period. We also require the stocks to be listed on NYSE and NASDAQ at two years prior to Jan. 2015 so as to mitigate the survivorship bias. We end up with twelve industries ($R = 12$), 2618 firms in total ($N = 2618$) and 105 weeks ($T = 105$). The number of stocks in each industry is reported in the second column of Table A5.

We first report the within correlation and between correlation. The former is evaluated as the average pairwise correlation of individual stock returns within the same industry while the latter is the average correlation between individual returns across two different industries. We visualise them through a heat map in Figure A1, 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 average 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.

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 apply CCD by selecting $\hat{r}_{\max} = 3$, following our practical guideline as described in Section 5. We find that there is only one global factor. Then, we apply BIC_3 to the defactored data in each block by concentrating out the global factor and selecting $\hat{r}_{i,\max} = \hat{r}_{\max} - \hat{r}_0$. 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 4.3, and obtain the full estimation results, that are reported in Table A5.

We evaluate the relative importance ratios[17] of the global factor, the local/industry factors and idiosyncratic errors, that are summarised in columns 4-6 in Table A5. On average, the

[15]The sample period and frequency of the data are chosen, ensuring that we have sufficient observations while the membership of the industries remains stable.

[16]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.

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

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

global factor and local factors can explain 22.6% and 5.8% of the total variation[18] whereas idiosyncratic disturbance components still account for 70.8% of the total variation.

The global factor tends to display the higher relative importance ratios for the cyclical industries such as Durbl (32.8%), Manuf (32.1%), Chems (30%) and Money (27.4%), suggesting that the higher within correlations observed in these industries are likely to reflect the higher loadings to the global factor. On the other hand, the influence of the global factor is below average for the non-cyclical industries such as NoDur (16.5%), Utils (8.3%) and Hlth (10.5%). Interestingly, local factors are more important than the global factor for Enrgy (23.2%), Hlth (9.6%) and Utils (54.2%) industries. The variance share explained by the local factors are also non-negligible for NoDur (9.3%) and Money (10.1%).

Next, we examine the within and between correlations after concentrating out the global and the local factors, respectively. Figure A2 displays the results constructed using the residuals from a regression of the return data on the global factor. In contrast to Figure A1, 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 local factors. We further project out the local factors such that the new residuals would be purely idiosyncratic. Figure A3 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.

Figure A4 displays that the estimated global factor comoves closely with the market factor (mkt) with correlation of 0.95,[19] though the latter is slightly more volatile. This may provide a support for the CAPM, which highlights a predominant role played by the market index in the asset pricing model.

However, it is more complicated 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. In Table A6, we provide the correlation matrix between the local factors and the Fama-French 3 factors, namely mkt, smb and hml. Though

---

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

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

[18] If we only consider the industries that contain local factors, their average relative importance ratio jumps to 11.8%.

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

smb and *hml* tend to share non-zero correlations with the subset of the local factors, it is difficult to identify any systematic relation between the local factors and the 3 factors. Further, notice that the average (absolute) pairwise correlation among the local/industry factors is 0.21, which can provide an empirical support for Assumption D' in Section 4.1.1.

Finally, in Figure A5, 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 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.

Our empirical findings may open up a few important issues for the asset pricing models. First, we find only one global factor which is almost perfectly correlated with the market factor. In this regard, it is an open question whether an inclusion of more common factors would be necessary in pricing individual stocks, which has been a rather standard practice in the literature,[20] e.g. Fama and French (1993), Carhart (1997) and Fama and French (2015). See also Harvey and Liu (2019) for a census of the factor zoo. Second, we find the presence of the local factors in five (NoDrl, Enrgy, Utils, Hlth and Money) out of twelve industries. In portfolio management, the covariance matrix of the return is used to compute the optimal weights for the asset allocation. The large covariance matrix is basically non-invertible in practice, so it must be constructed by the estimates of an asset pricing model, e.g. Bekaert et al. (2009). If the significant local factors are omitted, the covariance matrix would be estimated imprecisely, leading to the misleading asset allocation.

7 Conclusion

We have developed a novel and simple 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), which is constructed by the difference between the block-averages of the adjacent canonical correlations between factors.

We derive that CCD is a consistent model selection criterion. Via Monte Carlo simulations, we show that CCD selects the number of global factors correctly even in small samples. Further,

[20] In Section IV of Online Appendix, we use *mkt*, *smb* and *hml* as the three global factors, and construct the heat maps for the residuals obtained after concentrating out their combinations. In sum we find that: (i) only *mkt* can explain the market comovement but fails to capture the within industry correlations; (ii) both *smb* and *hml* cannot explain either the market comovement or the within industry comovement.

CCD outperforms other competing approaches even in the presence of serially correlated and weakly cross-sectionally correlated 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 of twelve industries in the U.S. over the period, Jan. 2015 to Dec. 2016. By applying CCD we find that there is only one global factor, which comoves closely with the market factor. Then, by applying BIC_3, we find that the local/industry factors explain non-trivial portions of the stock return variations in 5 out of 12 industries in the U.S.

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.

References

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

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.

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.

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 (Advances in Econometrics, Volume 35) Emerald Group Publishing Limited* **35**: 177–214.

Carhart, M. M. (1997). On persistence in mutual fund performance, *The Journal of finance* **52**(1): 57–82.

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* .
URL: *"https://mpra.ub.uni-muenchen.de/36114/"*

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.

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.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model, *Journal of Financial Economics* **116**(1): 1 – 22.
URL: *http://www.sciencedirect.com/science/article/pii/S0304405X14002323*

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.

Harvey, C. R. and Liu, Y. (2019). A census of the factor zoo, *Available at SSRN 3341728* .

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.

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.

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.

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* $\widetilde{\boldsymbol{K}}_i = \frac{1}{M_i T}\mathbf{Y}_i\mathbf{Y}_i'\widehat{\boldsymbol{K}}_i$. *Under Assumption A-D, as* $M_i, T \longrightarrow \infty$, *we have:*

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

where $\boldsymbol{H}_i$ *is the* $r_{\max} \times (r_0 + r_i)$ *rotation matrix,* $\delta_{M_i T} = \min\left\{\sqrt{M_i}, \sqrt{T}\right\}$ *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 $\mathbf{Y}_m\mathbf{Y}_m'$, we have

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

where $\boldsymbol{V}_m$ is an $r_{\max} \times r_{\max}$ diagonal matrix consisting of the $r_{\max}$ largest eigenvalues of $\mathbf{Y}_m\mathbf{Y}_m'$ in descending order divided by $M_m T$. 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

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

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

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

as

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

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:

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

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

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

where $(\boldsymbol{H}_h'\mathbf{\Sigma}_{hh}\boldsymbol{H}_h)^{-}$ is the Moore-Penrose inverse of $\boldsymbol{H}_h'\mathbf{\Sigma}_{hh}\boldsymbol{H}_h$. Using the property of Moore-Penrose inverse, we have[21]:

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

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

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

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

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

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

[21]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 $AA^- = I$.

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

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

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))

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

where $\boldsymbol{W}$ is an $(r_0+r_m)\times(r_0+r_m)$ orthonormal matrix, $\boldsymbol{D}$ is an $(r_0+r_h)\times(r_0+r_h)$ orthonormal matrix and $\boldsymbol{R}$ is the $(r_0+r_m)\times(r_0+r_m)$ diagonal matrix given by $\boldsymbol{R} = diag(\rho_1, ..., \rho_{r_0}, \rho_{r_0+1}, ..., \rho_{r_0+r_m}) = diag(1, ..., 1, 0, ..., 0)$.[22]

Define the full rank matrices,

$$\boldsymbol{A} = \begin{bmatrix}\boldsymbol{P}_1\boldsymbol{\Delta}_1^{-1}\boldsymbol{W}, \boldsymbol{P}_2\end{bmatrix} \text{ and } \boldsymbol{B} = \begin{bmatrix}\boldsymbol{Q}_1\boldsymbol{\Delta}_2^{-1}\boldsymbol{D}, \boldsymbol{Q}_2\end{bmatrix} \tag{A.8}$$

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

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

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

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

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

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

$$\boldsymbol{A}'\widetilde{\boldsymbol{S}}_{mm}\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I}_{r_0+r_m} + O_p(\delta_{\underline{M}T}^{-1}) & O_p(\delta_{\underline{M}T}^{-1}) \\ O_p(\delta_{\underline{M}T}^{-1}) & O_p(\delta_{\underline{M}T}^{-1}) \end{bmatrix}$$

$$\boldsymbol{B}'\widetilde{\boldsymbol{S}}_{hh}\boldsymbol{B} = \begin{bmatrix} \boldsymbol{I}_{r_0+r_m} + O_p(\delta_{\underline{M}T}^{-1}) & O_p(\delta_{\underline{M}T}^{-1}) & O_p(\delta_{\underline{M}T}^{-1}) \\ O_p(\delta_{\underline{M}T}^{-1}) & \boldsymbol{I}_{r_h-r_m} + O_p(\delta_{\underline{M}T}^{-1}) & O_p(\delta_{\underline{M}T}^{-1}) \\ O_p(\delta_{\underline{M}T}^{-1}) & O_p(\delta_{\underline{M}T}^{-1}) & O_p(\delta_{\underline{M}T}^{-1}) \end{bmatrix}$$

and

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

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

$$rank\left(\begin{bmatrix} \boldsymbol{I}_{r_h-r_m} + O_p(\delta_{\underline{M}T}^{-1}) & O_p(\delta_{\underline{M}T}^{-1}) \\ O_p(\delta_{\underline{M}T}^{-1}) & O_p(\delta_{\underline{M}T}^{-1}) \end{bmatrix}\right) \neq rank\left(\begin{bmatrix} \boldsymbol{I}_{r_h-r_m} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}\right)$$

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.[23] Thus, we

[23] 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)

have:

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

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

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

Using the above results, we obtain:

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

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

$$= \begin{bmatrix} \boldsymbol{R}+O_p(\delta_{\underline{M}T}^{-1}) & O_p(\delta_{\underline{M}T}^{-1}) \\ O_p(\delta_{\underline{M}T}^{-1}) & O_p(\delta_{\underline{M}T}^{-1}) \end{bmatrix}$$

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

$$\left(\begin{bmatrix} \boldsymbol{R} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} - \ell\begin{bmatrix} \boldsymbol{I}_{r_0+r_m} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} + O_p(\delta_{\underline{M}T}^{-1})\right)\boldsymbol{\xi} = \boldsymbol{0}$$

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 First, 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. Next, using the results in (4.2) in Lemma 1, for every block pair m and h, we have

$$\bar{\ell}_{mn}(r) \xrightarrow{p} \begin{cases} 1 & \text{for } r = 0, ..., r_0 \\ \frac{1+r_0}{1+r} & \text{for } r = r_0+1, ..., r_{\min}+1 \end{cases}$$

Using these results in the definition of $\bar{\xi}(r)$, we get the desired result.

Q.E.D

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

Proof for Theorem 1.

(i) For CCD 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 $0 \leq r < r_0$ we have:

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

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

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

(ii) For CCR we need to show that

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

for $r \neq r_0$ and $r \leq r_{\max}$. Consider first the case with $r_0 > 0$. By Lemma 2, it is easily seen that for $r_0 < r \leq r_{\max}$, we have:

$$CCR(r) - CCR(r_0) \xrightarrow{p} \frac{2+r}{1+r} - \frac{2+r_0}{1+r_0} = \frac{1}{1+r} - \frac{1}{1+r_0} < 0$$

whilst for $0 \leq r < r_0$ we have:

$$CCR(r) - CCR(r_0) \xrightarrow{p} 1 - \frac{2+r_0}{1+r_0} = -\frac{1}{1+r} < 0.$$

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

$$CCR(r) - CCR(r_0) \xrightarrow{p} \frac{2+r}{1+r} - 2 = \frac{1}{1+r} - 1 < 0.$$

Q.E.D

B Simulation results

Table A1: Average estimates of the number of global factors for experiments with $r_{max} = 5$, $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (2, 2)$

		CCD	CCR	IC_{chen}	$AGGR$	CCD	CCR	IC_{chen}	CCD	CCR	IC_{chen}
Panel A: $(\beta, \phi_e, \kappa) = (0, 0, 1)$											
M	T	$R = 2$				$R = 5$			$R = 10$		
20	50	1.99(0.7|1.6)	2.27(24.4|0.2)	2.02(2.9|1.2)	2.83(65.3|2.3)	2(0|0.2)	2.03(2.6|0)	1(0|100)	2(0|0.2)	2(0|0)	1(0|100)
50	50	2.01(0.5|0)	2.21(19.1|0)	2(0.2|0)	1.98(0|2.2)	2(0|0)	2.01(0.8|0)	1.24(0|75.8)	2(0|0)	2(0|0)	1(0|100)
100	50	2(0.1|0)	2.21(19.3|0)	2(0|0)	2(0|0.4)	2(0|0)	2.01(0.5|0)	2(0|0.2)	2(0|0)	2(0|0)	1.02(0|97.6)
200	50	2(0|0)	2.19(17.6|0)	2(0|0)	2(0|0)	2(0|0)	2(0.3|0)	2(0|0)	2(0|0)	2(0|0)	2(0|0)
20	100	2(0|0.1)	2.01(1.4|0)	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)	2.01(1|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)	2.01(0.6|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(0.2|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)	2(0|0)	1.92(0|7.9)	2.39(45.9|7.1)	2(0|0)	2(0|0)	1(0|100)	2(0|0)	2(0|0)	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	2.05(5.9|1.5)	2.69(50.4|0.3)	1.76(0.2|24)	2.61(65.2|16.5)	2(0|0.2)	2.47(40.5|0)	1(0|100)	2(0|0)	2.34(33|0)	1(0|100)
50	50	2.04(3.8|0.1)	2.85(57|0)	2(0.3|0.5)	1.62(0|35)	2(0|0)	2.63(51.5|0)	1.11(0|89.3)	2(0|0)	2.5(45.8|0)	1(0|100)
100	50	2.06(4.9|0)	3.09(67.1|0)	2(0|0)	1.88(0|11)	2(0|0)	3.15(78.7|0)	1.94(0|6.1)	2(0|0)	3.32(90.7|0)	1.01(0|99.4)
200	50	2.18(16.4|0)	3.41(81|0)	2(0|0)	1.95(0|4.7)	2(0.1|0)	3.83(98|0)	2(0|0)	2(0|0)	3.92(100|0)	1.93(0|6.7)
20	100	2(0|0.2)	2.07(7.3|0)	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)	2.07(7|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)	2.09(9|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)	2.2(18.8|0)	2(0|0)	1.98(0|2.3)	2(0|0)	2.01(1.1|0)	2(0|0)	2(0|0)	2(0|0)	1.18(0|81.9)
20	200	2(0|0)	2(0|0)	1.54(0|46)	2.01(43.9|32.3)	2(0|0)	2(0|0)	1(0|100)	2(0|0)	2(0|0)	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(0.3|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	1.53(12.9|43.6)	2.48(50.4|20.5)	1.65(2.7|37.9)	3.12(96.2|1.5)	1.24(0.3|46.8)	2.8(64.6|7.9)	1(0|100)	1.34(0|37.6)	2.98(78.4|3.4)	1(0|100)
50	50	2.05(15.7|12.8)	3.04(65.6|2.6)	1.94(7.3|13.4)	0.66(0|95.5)	1.95(0.1|4.7)	3.27(81.7|0.1)	1(0|100)	1.97(0|3.5)	3.4(92.1|0.1)	1(0|100)
100	50	2.22(18.7|1)	3.37(75.7|0.2)	2.16(16.1|0.3)	0.95(0|84)	2(0.5|0.8)	3.68(93.9|0)	1.07(0|93.2)	2(0|0.3)	3.86(99.3|0)	1(0|100)
200	50	2.32(25.5|0)	3.68(86.4|0)	2.12(12.4|0)	1.21(0|69)	2.03(2.6|0)	4.17(99.9|0)	1.85(0|14.7)	2(0.1|0)	4.13(100|0)	1.01(0|99.4)
20	100	1.38(0.2|38.9)	1.84(10.8|17.7)	1.29(0.1|71.6)	2.89(92|3.6)	1.44(0|30.8)	1.92(0.2|5.4)	1(0|100)	1.53(0|24.8)	1.98(0|1.4)	1(0|100)
50	100	1.95(0.1|4.5)	2.13(12.3|0.7)	1.69(0|31.4)	2.21(60.2|25.6)	1.99(0|1.3)	2(0.2|0.2)	1(0|100)	2(0|0.3)	2(0|0)	1(0|100)
100	100	2(0|0.1)	2.14(13.3|0)	2(0|0.4)	1.12(0|75.4)	2(0|0)	2(0.2|0)	1(0|99.9)	2(0|0)	2(0|0)	1(0|100)
200	100	2(0.1|0)	2.24(22|0)	2(0|0)	1.55(0|42.8)	2(0|0)	2.01(0.9|0)	1.74(0|26.5)	2(0|0)	2(0|0)	1(0|100)
20	200	1.57(0|24.7)	1.88(0.2|7.7)	1.14(0|85.6)	2.78(87.2|6.7)	1.72(0|14.1)	1.99(0|0.4)	1(0|100)	1.86(0|7.3)	2(0|0)	1(0|100)
50	200	1.99(0|0.8)	2(0|0.1)	1.57(0|43.5)	1.79(44.2|40.4)	2(0|0.1)	2(0|0)	1(0|100)	2(0|0)	2(0|0)	1(0|100)
100	200	2(0|0)	2(0.2|0)	2(0|0.3)	1.42(21.7|55.5)	2(0|0)	2(0|0)	1(0|99.6)	2(0|0)	2(0|0)	1(0|100)
200	200	2(0|0)	2(0.2|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)

Table A2: Average estimates of the number of global factors for experiments with $r_{max} = 5$, $(\phi_G, \phi_F) = (0.5, 0.5)$, $(\beta, \phi_e, \kappa) = (0.1, 0.5, 1)$

		CCD	CCR	IC_{chen}	$AGGR$	CCD	CCR	IC_{chen}	CCD	CCR	IC_{chen}
Panel A: $(r_0, r_i) = (0, 2)$											
M	T		$R = 2$				$R = 5$			$R = 10$	
20	50	0(0|0)	0.04(3.8|0)	1(100|0)	0.33(32.1|0)	0(0|0)	0(0|0)	1(100|0)	0(0|0)	0(0|0)	1(100|0)
50	50	0(0.1|0)	0.03(2.7|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)
100	50	0(0|0)	0.01(1|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(0|0)	0.01(1|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(0|0)	1.09(9.1|0)	1(0|0)	0.93(4.7|11.3)	1(0|0)	1(0.1|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
50	50	1(0.4|0)	1.06(6.2|0)	1(0|0)	0.92(0|8.1)	1(0|0)	1(0.1|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
100	50	1(0|0)	1.01(1.2|0)	1(0|0)	0.96(0|3.9)	1(0|0)	1(0.1|0)	1(0|0)	1(0|0)	1(0|0)	1(0|0)
200	50	1(0|0)	1(0.3|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(0.2|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	2.68(11.4|36.6)	3.67(55.4|10.3)	2.92(10.5|18.6)	4.9(88.1|7)	2.8(0.3|19.3)	4.17(80.7|0.3)	1(0|100)	2.9(0|9.8)	4.41(96.3|0.1)	1(0|100)
50	50	2.94(4|10.7)	3.66(51.5|2.1)	3.04(6.1|2.1)	2.1(0.6|74.1)	2.95(0|4.9)	3.85(68.6|0.1)	1.17(0|99.7)	2.98(0|2)	4.1(85.5|0)	1(0|100)
100	50	2.98(1|2.8)	3.56(46.8|1.2)	3.03(3.3|0)	2.58(0.7|40.7)	3(0|0.3)	3.46(45.7|0.2)	2.7(0|28.2)	3(0|0.2)	3.48(46.9|0.1)	1.01(0|100)
200	50	2.97(0.7|3.4)	3.48(43.3|1.3)	3.02(1.6|0)	2.82(0.4|18.7)	2.99(0|1.1)	3.39(39.1|0.4)	3(0|0.1)	3(0|0)	3.35(34.7|0)	2.64(0|32.3)
20	100	2.34(0.2|49.2)	2.86(10.8|19.9)	2.54(0.5|44.3)	4.25(76.9|13.9)	2.64(0|33)	2.97(0.4|3.3)	1(0|100)	2.77(0|22.4)	2.99(0|0.7)	1(0|100)
50	100	2.97(0|3.2)	3.09(10|0.9)	2.96(0|3.9)	3.72(55.6|20.7)	3(0|0)	3(0|0.1)	1(0|100)	3(0|0)	3(0|0)	1(0|100)
100	100	3(0|0)	3.07(6.8|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)	3.07(6.6|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	2.15(0|57.6)	2.62(0.5|29)	2.29(0|64.2)	3.88(69|20.3)	2.55(0|39.9)	2.94(0|6.3)	1(0|100)	2.69(0|30.6)	3(0|0.4)	1(0|100)
50	200	3(0|0.3)	3(0.4|0.1)	2.94(0|6.5)	3.29(40.1|27.6)	3(0|0)	3(0|0)	1(0|100)	3(0|0)	3(0|0)	1(0|100)
100	200	3(0|0)	3(0.3|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(0.1|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)

Table A3: Average estimates of the number of global factors for experiments with correlated local factors, $r_{max} = 5$, $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (2, 2)$, $(\beta, \phi_e, \kappa) = (0, 0, 1)$

			$\omega_F = 0.2$				$\omega_F = 0.4$			
R	M	T	CCD	CCR	IC_{chen}	$AGGR$	CCD	CCR	IC_{chen}	$AGGR$
2	20	50	2.01(3.9|2.5)	2.38(35.7|0.3)	1.99(2.2|2.8)	2.92(70.4|3.2)	2.25(29.1|3.6)	2.86(77.9|0.3)	2.02(5.6|4.1)	3.28(84.1|2)
2	50	50	2.02(1.8|0)	2.4(37.8|0)	2(0.3|0)	1.98(0.1|2.6)	2.23(23.4|0)	2.89(83.3|0)	2.01(0.9|0)	1.97(0|2.5)
2	100	50	2.01(1.4|0)	2.41(39.2|0)	2(0|0)	2(0|0.3)	2.18(18.1|0)	2.92(83.1|0)	2(0.3|0)	2(0|0.4)
2	200	50	2(0.3|0)	2.4(38|0)	2(0|0)	2(0|0.1)	2.19(18.7|0)	2.94(86.4|0)	2(0|0)	2(0|0.1)
2	20	100	2(0.5|0.2)	2.21(20.6|0)	1.94(0.3|6.4)	2.74(64.9|3.4)	2.22(22.5|0.7)	2.87(86.3|0)	1.93(0.2|7.5)	3.07(80.1|2.1)
2	50	100	2(0|0)	2.2(19.7|0)	2(0|0)	2.45(44.8|0.3)	2.09(9.3|0)	2.89(88.8|0)	2(0|0.1)	2.74(72.6|0.1)
2	100	100	2(0|0)	2.21(21.2|0)	2(0|0)	2(0|0.1)	2.07(6.5|0)	2.88(87.3|0)	2(0|0)	2(0|0)
2	200	100	2(0|0)	2.2(19.5|0)	2(0|0)	2(0|0)	2.06(6|0)	2.91(90.1|0)	2(0|0)	2(0|0.1)
2	20	200	2(0|0)	2.09(9.4|0)	1.92(0.1|8.6)	2.57(60.7|7.1)	2.1(9.7|0.2)	2.92(92.2|0)	1.9(0|9.6)	2.99(81.5|3.3)
2	50	200	2(0|0)	2.1(10|0)	2(0|0)	2.39(39.8|0.5)	2.03(2.7|0)	2.94(94.4|0)	2(0|0)	2.73(73.5|0.5)
2	100	200	2(0|0)	2.08(8.1|0)	2(0|0)	2.29(28.8|0)	2.01(1|0)	2.97(96.8|0)	2(0|0)	2.69(69.2|0)
2	200	200	2(0|0)	2.09(9.4|0)	2(0|0)	2(0|0)	2.01(0.6|0)	2.97(96.8|0)	2(0|0)	2(0|0)
5	20	50	2(0|0.1)	2.27(27.4|0)	1(0|100)		2.19(19.7|0.8)	2.94(93.8|0)	1(0|100)	
5	50	50	2(0|0)	2.29(29.1|0)	1.24(0|76.2)		2.1(10|0)	2.96(95.6|0)	1.11(0|89.4)	
5	100	50	2(0|0)	2.29(28.8|0)	2(0|0.2)		2.07(7.3|0)	2.97(96.7|0)	1.99(0|1.3)	
5	200	50	2(0|0)	2.29(29.1|0)	2(0|0)		2.06(5.9|0)	2.97(96.7|0)	2(0|0)	
5	20	100	2(0|0)	2.04(4.3|0)	1(0|100)		2.08(8.2|0.1)	2.96(96|0)	1(0|100)	
5	50	100	2(0|0)	2.06(6.3|0)	1(0|100)		2.03(3|0)	2.98(98.4|0)	1(0|100)	
5	100	100	2(0|0)	2.07(6.9|0)	1.71(0|29.2)		2.01(0.9|0)	2.98(98.2|0)	1.48(0|52.1)	
5	200	100	2(0|0)	2.07(6.8|0)	2(0|0)		2.01(1.2|0)	2.99(98.6|0)	2(0|0)	
5	20	200	2(0|0)	2.01(0.8|0)	1(0|100)		2.03(2.5|0)	2.99(99.1|0)	1(0|100)	
5	50	200	2(0|0)	2.01(0.8|0)	1(0|100)		2(0|0)	3(99.9|0)	1(0|100)	
5	100	200	2(0|0)	2.01(0.7|0)	1(0|100)		2(0|0)	3(100|0)	1(0|100)	
5	200	200	2(0|0)	2.01(0.6|0)	2(0|0.1)		2(0|0)	3(99.8|0)	1.97(0|3.1)	
10	20	50	2(0|0)	2.24(24|0)	1(0|100)		2.17(17.5|0.5)	2.96(96.3|0)	1(0|100)	
10	50	50	2(0|0)	2.24(23.7|0)	1(0|100)		2.06(6.3|0)	2.99(98.8|0)	1(0|100)	
10	100	50	2(0|0)	2.21(20.5|0)	1.01(0|98.8)		2.05(4.6|0)	2.99(98.8|0)	1(0|99.6)	
10	200	50	2(0|0)	2.24(24.3|0)	2(0|0.1)		2.03(2.9|0)	2.98(98.1|0)	1.98(0|2)	
10	20	100	2(0|0)	2.04(3.7|0)	1(0|100)		2.06(5.5|0)	2.98(98.1|0)	1(0|100)	
10	50	100	2(0|0)	2.02(2.1|0)	1(0|100)		2.01(1.4|0)	2.99(99.3|0)	1(0|100)	
10	100	100	2(0|0)	2.03(2.9|0)	1(0|100)		2(0.4|0)	2.99(99.2|0)	1(0|100)	
10	200	100	2(0|0)	2.03(2.9|0)	1.2(0|80.5)		2.01(0.5|0)	2.99(99.3|0)	1.05(0|95.3)	
10	20	200	2(0|0)	2(0.1|0)	1(0|100)		2.02(2.1|0)	3(99.6|0)	1(0|100)	
10	50	200	2(0|0)	2(0.2|0)	1(0|100)		2(0|0)	3(100|0)	1(0|100)	
10	100	200	2(0|0)	2(0|0)	1(0|100)		2(0|0)	3(99.7|0)	1(0|100)	
10	200	200	2(0|0)	2(0|0)	1(0|100)		2(0|0)	3(100|0)	1(0|100)	

Table A4: Average *CCD* estimates of the number of global factors for experiments with excessively large $r_{\max} = 10$ and the practical guide $\hat{r}_{\max} = \max\{\widehat{r_0 + r_1}, \ldots, \widehat{r_0 + r_R}\}$, $(\phi_G, \phi_F) = (0.5, 0.5)$, $(r_0, r_i) = (2, 2)$, $\kappa = 1$

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

C Empirical Results

Table A5: The main empirical results for industries

	M_i	$\hat{r}_i$	IRG	IRF	IRE
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. IRG, IRF and IRE stand for the relative importance ratios for the global, local and idiosyncratic components, respectively.

Table A6: Correlations among the local factors and Fama-French 3 Factors

	NoDur	Enrgy	Utils 1	Utils 2	Hlth	Money	mkt	smb	hml
NoDur	1								
Enrgy	-0.14	1							
Utils 1	0.58*	−0.21*	1						
Utils 2	-0.08	0.55*	0	1					
Hlth	-0.18	−0.24*	-0.08	-0.05	1				
Money	0.32*	0.14	0.23*	0.4*	-0.01	1			
mkt	-0.15	0.02	-0.1	-0.04	−0.21*	0.03	1		
smb	0.26*	0.03	0.21*	0.12	−0.46*	0.22*	0.37*	1	
hml	0.22*	−0.45*	0.13	-0.06	0.62*	0.52*	0.02	-0.04	1

* indicates that the correlation is significant at the 5% significance level.

Figure A1: Average pairwise correlations of returns

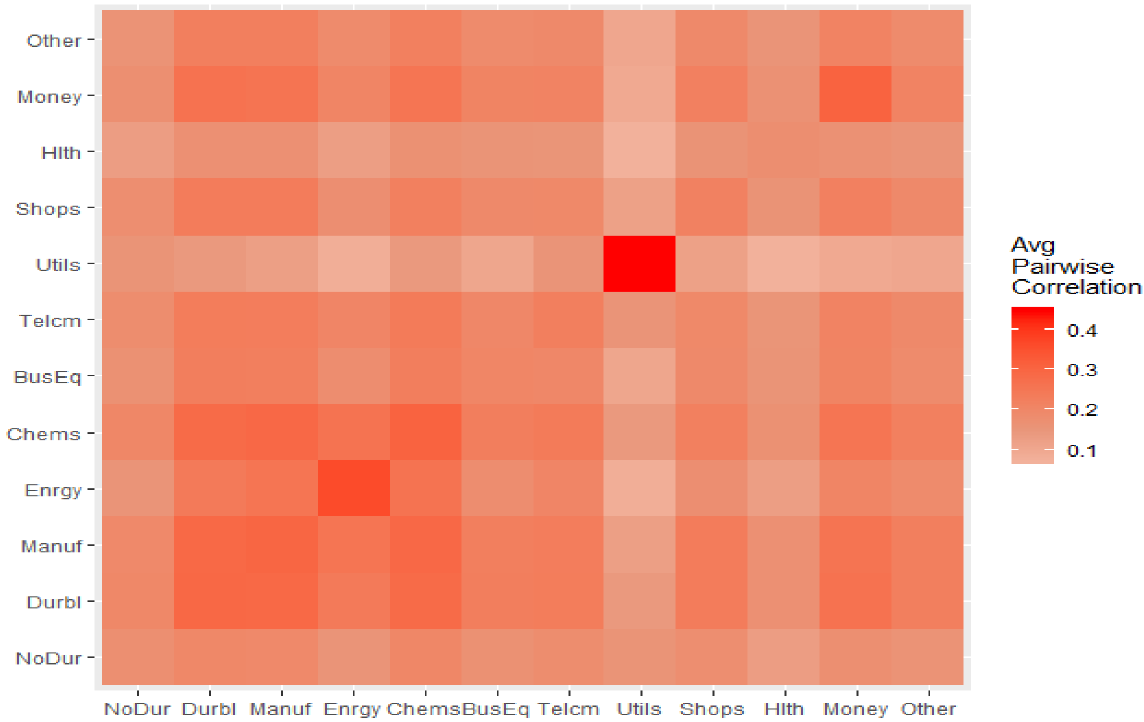

Figure A2: Average pairwise correlations of residuals after concentrating out $\widehat{\boldsymbol{G}}$

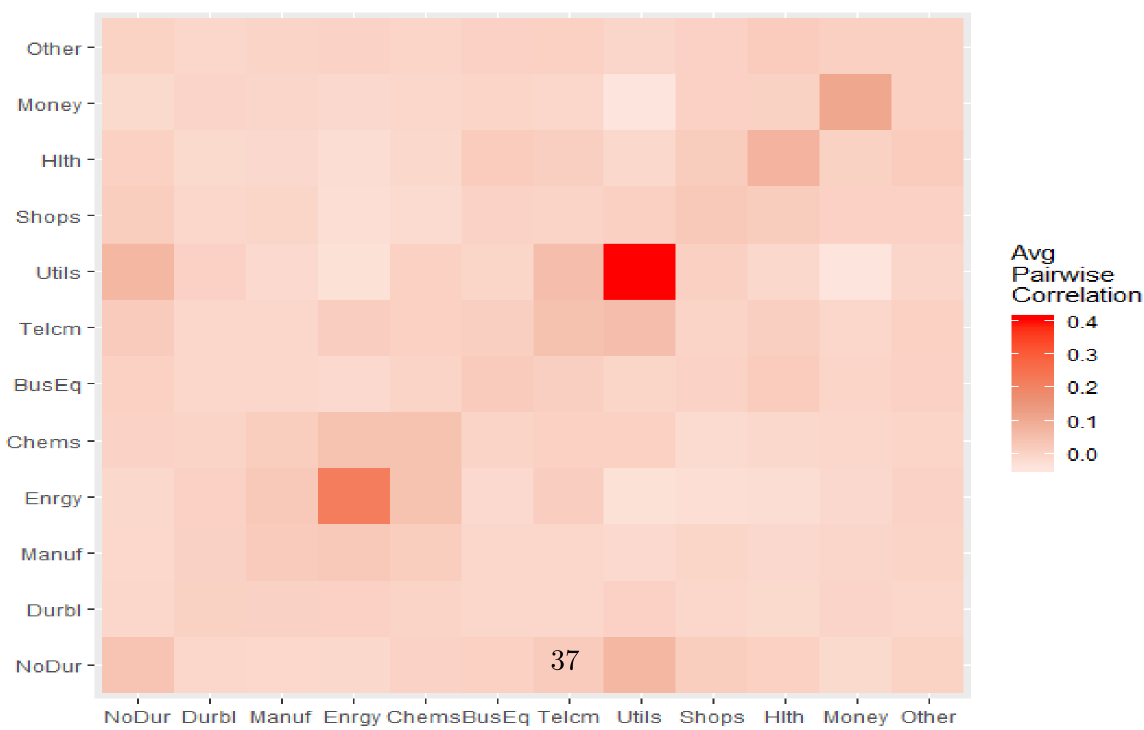

Figure A3: Average pairwise correlations of residuals after concentrating out $\widehat{\boldsymbol{G}}$ and $\widehat{\boldsymbol{F}}_i$'s

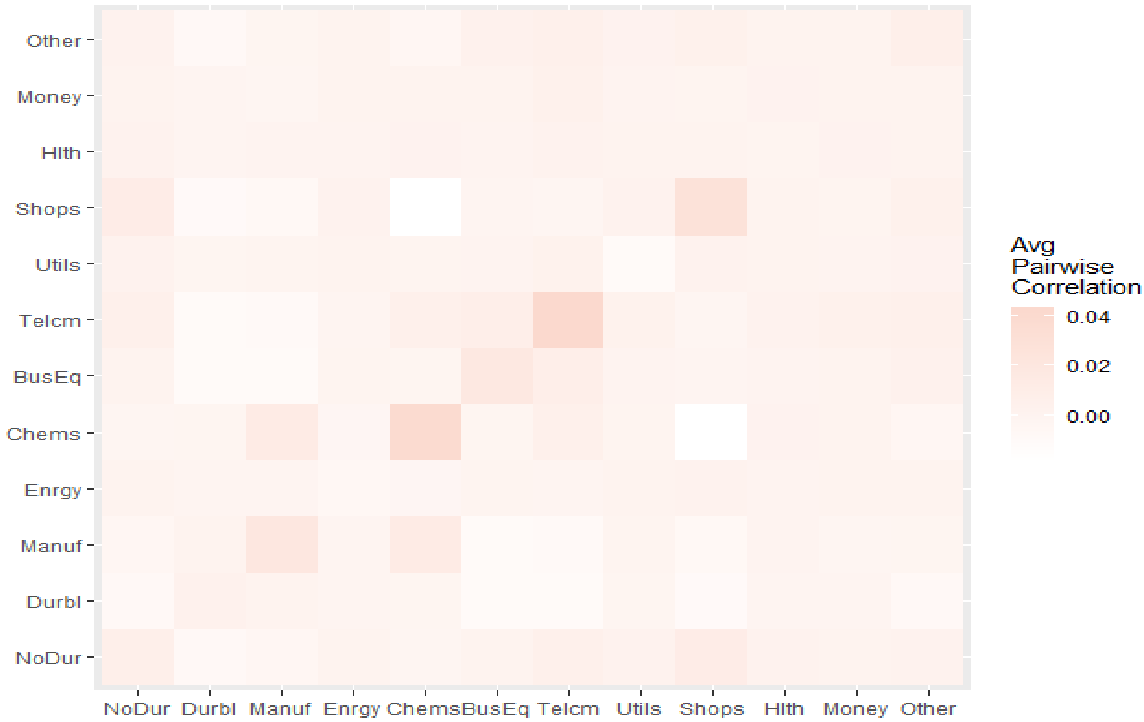

Figure A4: The global factor and market factor

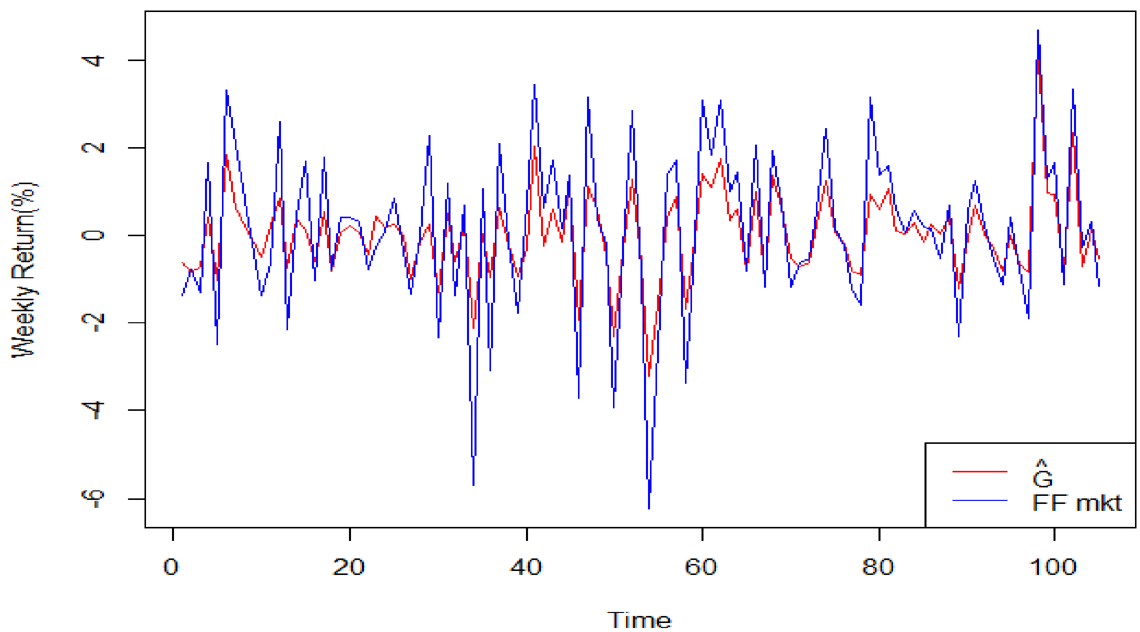

Figure A5: Density plots of the global and local factor loadings