Multifactor asset pricing with a large number of observable risk factors and unobservable common and group-specific factors

Bai, Jushan and Ando, Tomohiro

4 July 2013

This is a revised version of http://dx.doi.org/10.2139/ssrn.2289201

Multifactor asset pricing with a large number of observable risk factors and unobservable common and group-specific factors

December, 2013

Tomohiro Ando [1] and Jushan Bai [2]

Abstract

This paper analyzes multifactor models in the presence of a large number of potential observable risk factors and unobservable common and group-specific pervasive factors. We show how relevant observable factors can be found from a large given set and how to determine the number of common and group-specific unobservable factors. The method allows consistent estimation of the beta coefficients in the presence of correlations between the observable and unobservable factors. The theory and method are applied to the study of asset returns for A-shares/B-shares traded on the Shanghai and Shenzhen stock exchanges, and to the study of risk prices in the cross section of returns.

Key words: factor models, panel data analysis, penalized method, LASSO, SCAD, heterogenous coefficients

JEL Clarification codes: C23, C52, G12

[1]The Graduate School of Business, Keio University, andoh@kbs.keio.ac.jp. Financial support from the Japan Securities Scholarship Foundation is acknowledged.

[2]Department of Economics, Columbia University, jb3064@columbia.edu. Financial support from the National Science Foundation (SES0962410) is acknowledged.

1 Introduction

The arbitrage pricing theory (APT) of Ross (1976), together with multifactor models of asset returns, plays a central role in modern finance theory. Under a multifactor model, the return of each security is expressed as a linear combination of a small number of factor returns and an asset-specific return. In the capital asset-pricing model (CAPM) of Sharpe (1964) and Lintner (1965), for example, the common factor is the market return. There is a growing body of empirical evidence that stock returns are related to factors based on macroeconomic, market- and firm-level characteristics.

Although multifactor models are widely used in practice, there is scope to develop and to implement a new model-building procedure. For example, Goyal et al. (2008) argued that the assumption that all factors influence a large number of assets, so-called pervasive factors, are too strong if an economy is partitioned into several groups. They emphasized that APT allows for the existence of common pervasive factors influencing returns of securities in all groups, and of group-specific pervasive factors affecting returns of securities only in some groups. Connor and Korajczyk (1993) pointed out that industry-specific components may not be pervasive sources of uncertainty for the entire economy. See also Cho et al. (1986), Bekaert et al. (2009). Here, we provide three examples that illustrate the group structure in financial markets.

Example 1: A relevant instance of a group structure in financial markets is the Chinese stock market. The Chinese market is divided into two segments, namely the once-restricted A-shares and the B-shares. The A-shares were initially designated exclusively for domestic investors and are denominated in Chinese renminbi (RMB), whereas the B-shares were initially designated exclusively for foreign investors and are denominated in foreign currency. Although the launch of the qualified foreign institutional investors policy by the Chinese government allowed foreign investors to enter the domestic A-share market, currency barriers may still hinder them from investing in A-shares. The Chinese government also decided to open the B-share market to domestic investors. There is evidence to suggest that the security returns for dual-listed shares on the Chinese A- and B-share markets are priced differently because the two markets are segmented (Ma, 1996; Su, 1999; and Fung et al., 2000).

Example 2: The two main stock exchanges in the United States, the New York Stock Exchange (NYSE) and the National Association of Securities Dealers Automated Quotations (NASDAQ), provide an additional instance of a grouped financial market structure. The NYSE is a specialist-based auction system, whereas the NASDAQ is a computer-based dealer market. Goyal et al. (2008) argued that "While the NYSE and NASDAQ provide the same service, their underlying structures, rules, and governing principles are very different". Empirical evidence indicates that the securities that trade on these two exchanges are different (Naranjo and Protopapadakis, 1997; Fama and French, 2004; Schwert, 2002; Malkiel and Xu, 2003; Baruch and Saar, 2009; Goyal et al., 2008). See also a survey paper by Karolyi and Stulz (2003).

Example 3: Fama and French's (1993) three-factor model uses the portfolio returns formed by sorting stocks on total equity capitalization (size), the ratio of book value to market value of common equity (book-to-market), and the market return. Fama and French (1998) extended the model to a global context and analyzed international stock returns. Griffin (2002) reported that country-specific versions of the three-factor model were more useful in explaining stock returns than a global version of the three-factor model. Fama and French (2012) studied stock returns using size, book-to-market, and momentum factors for four regions (North America, Europe, Japan, and Asia–Pacific), considering both integrated and local models. Lewis (2011) provided a useful review of the current body of research addressing global asset-pricing challenges. This evidence shows that the stock markets in the world can be analyzed by constructing several market groups.

Despite these examples, little work has been done on pinpointing the differences between factor structures across groups (Goyal et al. 2008). As a contribution, in this paper, we develop a new multifactor-modeling procedure to deal with the situation where there are several groups of assets. In particular, we use a factor structure that consists of the common pervasive factors and group-specific pervasive factors. Grouped-factor structure has been considered in a number of economic studies. For example, Moench, Ng, and Potter (2012) proposed multilevel factor models by characterizing within and between block variations as well as idiosyncratic noise in large dynamic panels. Diebold et al. (2008) considered a hierarchical factor model for government bond yield data from several different countries. Kose et al. (2008) used a multilevel factor model to study international business cycle movements; also see Wang (2010) and Moench and Ng (2011).

In addition to unobservable factors, observable factors that are based on some economic theories that often used in a practical situation. Observable risk factors may include macroeconomics variables (such as exchange rates, oil prices, and inflation rates), financial market variables (such as volatility indices, trading volumes, liquidity, and total market values), and firm-level characteristics (such as dividend yields, the cost of capital, cash-flow-to-price ratios, and book-to-market equity ratios, etc). In this paper, we try to select an appropriate set of observable factors among the huge number of possible variables. As the second contribution in this paper, we develop a procedure for efficiently identifying the set of observable risk factors. More specifically, we use the smoothly clipped absolute deviation (SCAD) penalty approach (Fan and Li, 2001). The non-zero coefficients are estimated as if the zero coefficients were known and were imposed (the so called “oracle property’). This is obtained despite the existence of many unobservable factors. The proposed procedure also identifies the number of common pervasive factors and the number of group-specific pervasive factors in each group simultaneously.

This paper includes further theoretical results. In a data-rich environment in which a large number of cross-sectional securities and observable risk factors are available, we investigate the consistency of the estimated regression coefficients on a set of observable risk factors. We show that the proposed estimator is consistent, even in the presence of error correlations and heteroscedasticity in both dimensions. Moreover, the asymptotic normality of the proposed estimator is obtained. Monte Carlo simulations confirm that the proposed multifactor-modeling procedure performs well.

In summary this paper makes the following theoretical contributions. First, we consider a panel data model with heterogeneous slope coefficients in contrast to the homogeneous regression coefficients in Bai (2009). This is a useful extension because the sensitivity of the asset returns to the observable risk factors may vary over the securities. Second, we provide a panel modeling procedure that allows the researcher to identify the number of observable and unobservable factors that are relevant for explaining the returns for different asset groups. We allow a large number of observable risk factors and try to select the set of relevant observable risk factors. To achieve this purpose, we develop the parameter estimation procedure using the SCAD penalty of Fan and Li (2001). The latter procedure was proposed in the context of cross-section regression (non panel data) without a factor structure, and under iid errors. In this paper we establish the variable-selection consistency under much more general setting. Third, we show that the proposed estimator is consistent as N and T getting large. The result is developed under a general situation that allows weak dependence and heteroskedasticity in the error term. Fourth, we propose a new measure for selecting a proper model from among many candidates or, equivalently, determination of the number of common/group-specific pervasive factors, determination of the magnitude of the regularization parameter for implementing the shrinkage approach. We show that the proposed criterion can identify the number of true common/group-specific pervasive factors consistently.

Beyond the theoretical contributions, our paper also makes practical contributions. We apply our proposed modeling procedure to the market structure of A- and B-share markets in China. We address empirical questions such as: How many common and group-specific pervasive factors exist in the stock market in mainland China? What type of observable risk factors explains the market? And, how can the unobservable common factors be understood in terms of observable variables in the economy? For example, by identifying the common and group-specific driving forces underpinning macroeconomic variables, we can obtain a further understanding of the market structure. We find that there are, at most, two common pervasive factors across the groups and four group-specific pervasive factors, three of which belong to the B-share markets. In addition, we find that some variables from overseas economies, such as stock market returns of other countries, exchange rates, and commodity markets are related to the security returns of the A- and B-shares. Moreover, we find that some domestic macroeconomic variables are risk factors.

Briefly, this paper has the following features. First, we introduce a new multifactor model that consists of a large number of observable risk factors as well as unobservable common pervasive factors and group-specific pervasive factors. Asset-specific returns are allowed to be correlated and heteroscedastic in both dimensions (time and cross section). The number of securities can be much larger than the number of time periods. We develop a model estimation procedure for such models. Second, the consistency and the asymptotic normality of the parameter estimates are investigated. Third, we develop a new model evaluation criterion that enables us to determine the relevant observable factors for each asset as well as the number of common pervasive factors and the number of group-specific pervasive factors in each group. Finally, our analysis of the market structure of A- and B-share markets results in a number of interesting empirical findings.

Notation. Let $\|A\| = [tr(A'A)]^{1/2}$ be the usual norm of the matrix A, where "tr" denotes the trace of a square matrix. The equation $a_n = O(b_n)$ states that the deterministic sequence a_n is at most of order b_n, $c_n = O_p(d_n)$ states that the random variable c_n is at most of order d_n in terms of probability, and $c_n = o_p(d_n)$ is of a smaller order in terms of probability. All asymptotic results are obtained under large number of securities N and large lengths of time series T. Restrictions on the relative rates of convergence of N and T are specified in later sections.

2 Model

This paper considers a panel of asset returns with a large number of observable risk factors, a set of common pervasive factors that affect the returns of all securities in all groups, and group-specific pervasive factors that affect the returns of all securities only in a specific group.

2.1 Model setting

Let $t = 1, ..., T$ be an index for time, G be the prespecified number of groups, $N_1,....N_G$ be the number of securities in each group, and $N = \sum_{g=1}^{G} N_g$ be the total number of securities. The asset return of the i-th security, y_{it}, observed at time t, belonging to group $g_i \in \{1, ..., G\}$, is expressed as follows:

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta}_i + \boldsymbol{f}'_{c,t}\boldsymbol{\lambda}_{c,i} + \boldsymbol{f}'_{g_i,t}\boldsymbol{\lambda}_{g_i,i} + \varepsilon_{i,t}, \quad i = 1, \ldots, N, \ t = 1, \ldots, T. \tag{1}$$

In a vector form, the model (1) can be expressed as: $\boldsymbol{y}_i = X_i\boldsymbol{\beta}_i + F_c\boldsymbol{\lambda}_{c,i} + F_{g_i}\boldsymbol{\lambda}_{g_i,i} + \boldsymbol{\varepsilon}_i$, $i = 1, \ldots, N$, where:

$$\boldsymbol{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}, \ X_i = \begin{pmatrix} \boldsymbol{x}'_{i1} \\ \boldsymbol{x}'_{i2} \\ \vdots \\ \boldsymbol{x}'_{iT} \end{pmatrix}, \ F_c = \begin{pmatrix} \boldsymbol{f}'_{c,1} \\ \boldsymbol{f}'_{c,2} \\ \vdots \\ \boldsymbol{f}'_{c,T} \end{pmatrix}, \ F_{g_i} = \begin{pmatrix} \boldsymbol{f}'_{g_i,1} \\ \boldsymbol{f}'_{g_i,2} \\ \vdots \\ \boldsymbol{f}'_{g_i,T} \end{pmatrix}, \ \boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}.$$

Here $\boldsymbol{x}_{it}$ is the $p_i \times 1$ vector of observable risk factors, and the dimension of $\boldsymbol{x}_{it}$ can be very large and may vary over i, $\boldsymbol{f}_{c,t}$ is an $r \times 1$ vector of unobservable common pervasive factors that affect the returns of all securities in all groups, and $\boldsymbol{f}_{g_i,t}$ is an $r_{g_i} \times 1$ vector of unobservable group-specific pervasive factors that affect the returns of securities only in group g_i. The $p_i \times 1$ vectors $\boldsymbol{\beta}_i$ are the unknown regression coefficients, $\boldsymbol{\lambda}_{c,i}$ and $\boldsymbol{\lambda}_{g_i,i}$ are factor loadings, and ε_{it} are the security-specific returns. Some of the observable risk factors $\boldsymbol{x}_{it}$ may be common to all firms ($\boldsymbol{x}_{it}$ does not depend on i), or common to some of the groups, or specific to a particular firm. Again, the dimension of $\boldsymbol{x}_{it}$ may be large.

This paper assumes that the group membership g_i ($i = 1, 2, ..., N$) is known. This assumption is motivated by empirical applications such as Goyal et al. (2008) as well as our own application in this paper. It might be of interest to let g_i be unknown and be estimated. This problem has been considered by Ando and Bai (2003) under the setting that the slope coefficients are homogeneous ($\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i) or there are G set of group-dependent coefficients. Such a model appears to be restrictive for asset pricing models for which the beta coefficients should be asset dependent. It is an interesting future research topic to allow both unknown group membership and asset-dependent coefficients.

In the appendix, we provide the regularity conditions of the model. Here, we briefly describe the assumptions. We assume the existence of r common pervasive factors and r_g group-specific pervasive factors $g = 1, ..., G$. Also, we allow weak serial and cross-sectional correlations on ε_{it}. Heteroscedasticity is also allowed even though ε_{it} is assumed to have a finite eighth moment. This moment condition is a technical assumption that simplifies the theoretical analysis; it is not a necessary condition. For example, for the student-t distribution with 5 degrees of freedom, the simulation shows that the procedure performs very well.

We point out that the observable risk factors can be correlated with the factor loadings, or with the unobserved common/group-specific pervasive factors, or can be correlated with both the factor loadings and the unobserved common/group-specific pervasive factors. Such correlations are allowed in panel models with factor errors (e.g., Bai, 2009). A similar setting was previously considered by Bai (2009), in which the regression coefficients are common (not varying with i), and there are only a small number of explanatory variables, and there are no group-specific factors. We clarify that the number of cross-sectional securities N is not fixed and is assumed to grow. In addition, N can be much greater than the number of time periods, T.

In the absence of the component of observable risk factors and the component of group-specific pervasive factors, model (1) reduces to a pure factor model, which has attracted much research interest in recent years. Factor models are perhaps the most commonly used statistical tool to simplify the analysis of huge panel data sets. Indeed, lately, many efforts in the econometric and statistical literature have been devoted to factor models for analyzing high-dimensional data. There are various types of factor specifications, including a dynamic exact factor model (Geweke, 1977; Sargent and Sims, 1977), a static approximate factor model (Chamberlain and Rothschild, 1983), a generalized dynamic factor model (Forni et al., 2000; Forni and Lippi, 2001; Amengual and Watson, 2007; Hallin and Liska, 2007), and Bayesian factor models (Aguilar and West, 2000; Lopes and West, 2004; Lopes et al., 2008; Ando, 2009; Bhattacharya and Dunson, 2011; Tsay and Ando, 2012).

Remark 1 When a component of $\boldsymbol{x}_{it}$ is set to 1, the model includes alphas (α_i). Alternatively, let $\alpha_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta}_i$, the model allows a time-varying alpha that depends on the observable variables $\boldsymbol{x}_{it}$. In general, $\boldsymbol{x}_{it}$ may consist of two sets of variables, with one set being the predictors of time-varying alphas, with the other being the state variables in the context of intertemporal CAPM (Merton, 1973). Some components of $\boldsymbol{x}_{it}$ can be common (not varying in i), such as market indices, Fama and French (1993)'s three factors (i.e., the excess return on the market, growth factor, size factor). Also, $\boldsymbol{x}_{it}$ may also contain some macroeconomic variables. Chen et al. (1986) studied the role of macroeconomic variables in asset pricing models. They found that some macroeconomic variables, including the spread between long and short interest rates, inflation and industrial production, systemically affect stock market returns. In our empirical applications, we examine whether the observable risk factors as well as unobservable risk factors are priced in the cross section of returns.

Model (1) encompasses a number of often used asset pricing models. If there are no observable risk factors, the model reduces to a pure approximate factor model (Chamberlain and Rothschild, 1983; Connor and Korajczyk, 1986, 1988; Jones, 2001; Bai and Ng, 2002; Bai, 2003), or the grouped-factor model (Krzanowski, 1979; Flury, 1984; Bekaert et al., 2009; Korajczyk and Sadka, 2008; Wang, 2010).

As pointed out by Goyal et al. (2008), in an economy partitioned into several groups, the existence of common pervasive factors and group-specific pervasive factors is not ruled out by APT. Therefore, model (1) can be justified from the APT's perspective. Model (1) is also empirically appealing. Our estimation method will determine the existence of common and group-specific factors.

3 Estimation

There exist several studies in the absence of observable risk-factor components. To estimate a model similar to the model (1) with $\boldsymbol{\beta}_i = 0$ for $i = 1, ..., N$. Bekaert et al. (2009) proposed the two-step inference procedure. Flury (1984) considered the situation in which the S groups have a common subspace for all groups. Schott (1999) considered the estimation procedure for a different setting. Goyal et al. (2008) used Schott's (1999) results and proposed a multigroup factor analysis as an extension of Connor and Korajczyk (1986,1988) in grouped factors. These studies either do not consider observable factors or only a small number of them or there are no unobservable factors. Pesaran (2006) and Song (2013) allow a small number of observable regressors, without group-specific factors. The limitation of Pesaran's estimation procedure is discussed by Westerlund and Urbain (20013).

We consider a situation for which there are a large number of possible observable risk factors, p_i, for security i, whereas the number of truly relevant observable risk factors is not large. In other words, the true underlying structure has a sparse representation and almost all elements of $\boldsymbol{\beta}_i$ are zero, but which coefficients being zero are unknown. To identify the correct sparse representation of the regression coefficients $\boldsymbol{\beta}_i$, we use the lasso-based approach (Tibshirani, 1996) for variable selection. Although the lasso method is widely used, shrinkage introduced by the lasso results in a bias towards zero for large regression coefficients. To diminish this bias, we use the smoothly clipped absolute deviation (SCAD) penalty approach (Fan and Li, 2001). As the SCAD method estimates redundant parameters for the irrelevant observable risk factors as zero (variable selection consistency), the computational cost is much less than the traditional variable selection methods. While the number of observable factors can be large, needless to say, our method works with a small number of observable factors.

3.1 Estimation procedure

To estimate the unknown parameters given the number of common pervasive factors, r, and the number of group-specific pervasive factors $r_1, ..., r_S$, we minimize the least-squares objective function with a penalty term as follows:

$$\begin{aligned}&\ell(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N, F_c, F_1, ..., F_G, \Lambda_c, \Lambda_1, ..., \Lambda_G | r, r_1, ..., r_G, \kappa)\\&= \sum_{i=1}^{N} \|\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i - F_c\boldsymbol{\lambda}_{c,i} - F_{g_i}\boldsymbol{\lambda}_{g_i,i}\|^2 + T\sum_{i=1}^{N} p_{\kappa,\gamma}\left(|\boldsymbol{\beta}_i|\right)\end{aligned} \tag{2}$$

subject to the constraints $F_c'F_c/T = I_r$ and $\Lambda_c'\Lambda_c$ being diagonal for the common pervasive factor and the corresponding $r \times N$ factor-loading matrix $\Lambda_c = (\boldsymbol{\lambda}_{c,1}, ..., \boldsymbol{\lambda}_{c,N})$, and $F_g'F_g/T = I_{r_g}$ $(g = 1, ..., G)$ and $\Lambda_g'\Lambda_g$ $(g = 1, ..., G)$ being diagonal for the group-specific pervasive factor and the corresponding $r_g \times N_g$ factor-loading matrices $\Lambda_g = (\boldsymbol{\lambda}_{g,1}, ..., \boldsymbol{\lambda}_{g,N_g})$. These restrictions are needed to avoid the model-identification problem and are commonly used in the literature (Connor and Korajczyk, 1986; Bai and Ng, 2002; Stock and Watson, 2002). For separating common pervasive and group-specific pervasive factors, we further assume $F_c'F_g = 0$ for $g = 1, ..., G$. As shown by Wang (2010), this orthogonality condition is necessary even for models without regressors. But it can be shown that the estimated beta coefficients are invariant to whether this normalization restriction is made.

Here, $\sum_{i=1}^{N} p_{\kappa,\gamma}\left(|\boldsymbol{\beta}_i|\right)$ is a function of the coefficients indexed by a parameter κ that controls the tradeoff between the fitness and the penalty. To identify a smaller subset of important variables from each X_i, we can search through subsets of potential observable risk factors for an adequate model. However, Breiman (1996) pointed out that this can be unstable and is computationally unfeasible. To avoid such problems, we use penalized regression procedures by shrinking some coefficients so that they are exactly equal to zero. This operation is equivalent to selection of the relevant observable risk factors. Some methods have been introduced for this purpose, including the lasso method (Tibshirani, 1996), the SCAD penalty (Fan and Li, 2001), and the minimax concave penalty (Zhang, 2010). These methods were introduced for non-panel data models. Here we use the penalty method for panel data models and with the presence of factor errors.

The SCAD penalty is defined as $p_{\kappa,\gamma}(|\boldsymbol{\beta}_i|) = \sum_{k=1}^{p_i} p_{\kappa,\gamma}(|\beta_{ik}|)$ with:

$$p_{\kappa,\gamma}(|\beta_{ik}|) = \begin{cases} \kappa|\beta_{ik}| & (|\beta_{ik}| \leq \kappa) \\ \dfrac{\gamma\kappa|\beta_{ik}| - 0.5(\beta_{ik}^2 + \kappa^2)}{\gamma - 1} & (\kappa < |\beta_{ik}| \leq \gamma\kappa) \\ \dfrac{\kappa^2(\gamma^2 - 1)}{2(\gamma - 1)} & (\gamma\kappa < |\beta_{ik}|) \end{cases},$$

for $\kappa > 0$ and $\gamma > 2$. This penalty first applies the same rate of penalization as the regular lasso and then reduces the rate to zero as it moves further away from zero.

Theoretical property of the the SCAD penalty is investigated in Fan and Li (2001) in the context of non-panel data.

If we take an extremely large value of the regularization parameter κ, almost all estimated $\boldsymbol{\beta}_i$ will be estimated as zero even the true values are nonzero. In such a case, we might exclude important observable risk factors. Conversely, too small a regularization parameter might include a number of unrelated observable risk factors because almost all elements of $\boldsymbol{\beta}_i$ will not vanish at zero. Therefore, we need to balance these options and determine a proper size for the regularization parameter κ. We provide the model-selection criterion to select an optimal penalty size in the next section.

Joint minimization of the least-squares objective function with a penalty term can be done using the method by Bai (2009). Under the homogeneous slope coefficients ($\boldsymbol{\beta} = \boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_N$) and the absence of the group-specific pervasive factors, Bai (2009) proposed to estimate the homogeneous slope coefficients jointly with the common pervasive factors and the corresponding factor loadings. His estimator of the homogeneous slope coefficients is $\sqrt{NT}$ consistent even in the presence of serial or cross-sectional correlations and heteroscedasticities of unknown form in the error term.

Given $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\}$, and the effect from the group-specific pervasive factors $F_{g_i}\boldsymbol{\lambda}_{g_i,i}$ $i = 1, ..., N$, we can define the matrix $W_c = (\boldsymbol{w}_{c,1}, \ldots, \boldsymbol{w}_{c,N})$ of dimension $T \times N$ with:

$$\boldsymbol{w}_{c,i} = \boldsymbol{y}_i - X_i\boldsymbol{\beta}_i - F_{g_i}\boldsymbol{\lambda}_{g_i,i}.$$

Then, the original model (1) reduces to $\boldsymbol{w}_{c,i} = F_c\boldsymbol{\lambda}_{c,i} + \boldsymbol{\varepsilon}_i$, which implies that W_c has a pure factor structure.

The least-squares objective function with the penalty is then:

$$\operatorname{tr}\left\{(W_c - F_c\Lambda_c')\,(W_c - F_c\Lambda_c')'\right\} + T\sum_{i=1}^{N} p_{\kappa,\gamma}\left(|\boldsymbol{\beta}_i|\right).$$

From the analysis of pure factor models estimated by the method of least squares (i.e., principal components; see Connor and Korajczyk, 1986; Bai and Ng, 2002; Stock and Watson, 2002; Bai, 2009). By concentrating out $\Lambda_c = W_c'F_c(F_c'F_c)^{-1} = W_c'F_c/T$, the objective function becomes:

$$\operatorname{tr}\{W_c'W_c\} - \operatorname{tr}\{F_c'W_cW_c'F_c\}/T + T\sum_{i=1}^{N} p_{\kappa,\gamma}\left(|\boldsymbol{\beta}_i|\right). \tag{3}$$

Therefore, minimizing the objective function (3) with respect to F_c is equivalent to maximizing $\operatorname{tr}\{F_c'W_cW_c'F_c\}$, subject to the constraint $F_c'F_c/T = I_r$. Noting that the penalty term is not related to F_c, The asymptotic principal-component estimate of F_c subject to the constraint, $\hat{F}_c$, is $\sqrt{T}$ times the eigenvectors corresponding to the r

largest eigenvalues of the $T \times T$ matrix $W_c W_c'$. Given $\hat{F}_c$, the factor-loading matrix can be obtained as $\hat{\Lambda}_c' = \hat{F}_c' W_c / T$. See also Bai and Ng (2002:197–198).

Next, given $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\}$, and the common pervasive factor structure $F_c \Lambda_c$, we define the variable $W_g = (\boldsymbol{w}_{g,1}, \ldots, \boldsymbol{w}_{g,N_g})$ with $\boldsymbol{w}_{g,i} = \boldsymbol{y}_i - X_i \boldsymbol{\beta}_i - F_c \boldsymbol{\lambda}_{c,i}$ as the set of N_g asset return series belonging to the g-th group. Note that only the N_g asset return series will be used for the estimation of the group-specific pervasive factor structures $F_g \boldsymbol{\lambda}_{g,i}$ of the g-th group. Then, based on a similar argument to that made above, the original model (1) reduces to the structure $\boldsymbol{w}_{g,i} = F_g \boldsymbol{\lambda}_{g,i} + \boldsymbol{\varepsilon}_{g,i}$. Again, this implies that the data matrix W_g (dimension of $T \times N_g$) has a pure factor structure and we can estimate F_g and $\boldsymbol{\lambda}_{g,i}$ using the asymptotic principal-component method. Estimates of the group-specific pervasive factor F_g and the corresponding factor loading $\boldsymbol{\lambda}_{g,i}$ can be obtained by minimizing the objective function:

$$\mathrm{tr}\left\{ \left(W_g - F_g \Lambda_g' \right) \left(W_g - F_g \Lambda_g' \right)' \right\},$$

subject to the constraint $F_g' F_g / T = I_{r_g}$, for $g = 1, ..., G$. The asymptotic principal-component estimate subject to the constraint can be obtained in a similar manner as described in the estimation of F_c and Λ_c.

Although the estimates of $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\}$, $\{F_c, \Lambda_c\}$, and $\{F_g, \Lambda_g;\ g = 1, ..., G\}$ depend on each other, the estimators are obtained by using the following iterative algorithm.

Estimation algorithm

Step 1. Fix the regularization parameter, κ, the number of common pervasive factors, r, and the number of group-specific factors $\{r_1, ..., r_G\}$. Initialize the unknown regression coefficients $\{\boldsymbol{\beta}_1^{(0)}, \ldots, \boldsymbol{\beta}_N^{(0)}\}$, the pervasive common factors, and the corresponding factor-loading matrix $\{F_c^{(0)}, \Lambda_c^{(0)}\}$, as well as the group-specific pervasive factors and the corresponding factor-loading matrices $\{F_g^{(0)}, \Lambda_g^{(0)};\ g = 1, ..., G\}$.
Step 2. Given values of $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\}$ and $\{F_g, \Lambda_g;\ g = 1, ..., G\}$, update $\{F_c, \Lambda_c\}$.
Step 3. Given values of $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\}$ and $\{F_c, \Lambda_c\}$, update $\{F_g, \Lambda_g\}$ for $g = 1, ..., G$.
Step 4. Given values of $\{F_c, \Lambda_c\}$ and $\{F_g, \Lambda_g;\ g = 1, ..., G\}$, update $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\}$.
Step 5. Repeat Steps 2 and 4 until convergence.

In Step 1, starting values for $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\}$, $\{F_c, \Lambda_c\}$, and $\{F_g, \Lambda_g;\ g = 1, ..., S\}$ are needed. In the next section, we discuss how to prepare initial parameter values for these parameters.

3.2 Initial parameter values

We can set initial values as follows. First, by ignoring the common pervasive factor structure for all group $\{F_c, \Lambda_c\}$ and the group-specific pervasive factor structure for each group $\{F_g, \Lambda_g;\ g = 1, ..., G\}$, an initial estimate of $\{\boldsymbol{\beta}_1^{(0)}, \ldots, \boldsymbol{\beta}_N^{(0)}\}$ is obtained using the pure SCAD approach. Second, given values of $\{\boldsymbol{\beta}_1^{(0)}, \ldots, \boldsymbol{\beta}_N^{(0)}\}$, an initial estimate of the factor structures $\{F_c, \Lambda_c\}$ is estimated by ignoring the group-specific pervasive factor structure for each group $\{F_g, \Lambda_g;\ g = 1, ..., G\}$. Finally, given values of $\{\boldsymbol{\beta}_1^{(0)}, \ldots, \boldsymbol{\beta}_N^{(0)}\}$ and the common pervasive factor structure for all of group $\{F_c^{(0)}, \Lambda_c^{(0)}\}$, we obtain the starting values of the group-specific pervasive factor structure $\{F_g^{(0)}, \Lambda_g^{(0)}\}$ for $g = 1, ..., G$.

It is known that the least-squares objective function is not globally convex (see also Bai, 2009). In other words, an arbitrary starting value will not necessarily provide the global optimal solution. To maximize the chance of obtaining the global minimum, one may prepare several starting values. After the convergence, one may choose the estimators that give a smaller value of the objective function.

Here is an alternative parameter initialization. First, by ignoring the effect from the observable risk factors, $\{X_i\boldsymbol{\beta}_i;\ i = 1, ..., N\}$, and the group-specific factor structures $\{F_g, \Lambda_g;\ g = 1, ..., G\}$, we obtain an initial estimate of the common pervasive factor structure $\{F_c^{(0)}, \Lambda_c^{(0)}\}$. Then, given $\{F_c^{(0)}, \Lambda_c^{(0)}\}$, by ignoring the effect from the observable risk factors, $\{X_i\boldsymbol{\beta}_i;\ i = 1, ..., N\}$, we obtain a starting value of the group-specific factor structures $\{F_g, \Lambda_g\}$ for $g = 1, ..., G$. Finally, we obtain an initial value of $\{\boldsymbol{\beta}_1^{(0)}, \ldots, \boldsymbol{\beta}_N^{(0)}\}$.

Simulation results in Section 5 indicate that the estimation method above is robust to the starting values. The results reveal that among the 1000 Monte Carlo repetitions, the converged parameter values $\hat{\boldsymbol{\beta}}_i$ with the above two starting values reached the same point more than 95% of times. If the converged values are different, we then select the one that minimizes the objective function.

Remark 2 In principle, when markets are segmented or when markets have different structures and rules, it is reasonable to expect that different factors affect different segments. When one knows a priori that two or more markets are different, one could also conduct separate analyses for each asset group. This would be a good strategy if there exist no common pervasive factors. Separate analysis for each group makes it difficult to tell if these groups share common-pervasive factors, especially unobservable ones. In addition, pooling groups together allows more efficient estimation of unobservable common factors. Therefore, it is desirable to model simultaneously the common pervasive structures, the group-specific pervasive structures, the observable risk factor components by pooling groups.

Remark 3 Instead of the variable selection approach in selecting the observable factors for each asset i, one might use the methodology of Stock and Watson (2005) to extract some principal components from the explanatory variables X_i or other macroeconomic and financial variables. These principal components could be used as regressors in the model and one could evaluate which principal components are important for which asset groups. This principal component method is an alternative way to reduce the dimensionality problem since the dimension of X_i ($p_i \times T$) can be large (p_i is large). This two-step procedure is very useful for forecasting, it is less desirable than the procedure introduced in this paper. The regressors X_i depend on i, they are not common to all individual assets; many observable risk factors in X_i are security-specific, e.g., profitability, firm size, etc. The number of firm-specific risk factors can be large, so that penalty method is a useful approach.

Remark 4 It is straightforward to put an additional penalty term that penalizes the factor loadings on group-specific pervasive factors in (2). However, by the definition, the group-specific pervasive factors affect almost all security returns within each group by the pervasive nature, penalizing these coefficients may not be desirable. Moreover, it is uncommon to face the parameter estimation instability due to the factor loading estimation as the dimension of the group-specific pervasive factors is usually small. Therefore, the penalty term on the factor loadings is not used. In contrast, the number of possible observable risk factors may be potentially very large at the initial modeling stage. For these reasons, we use the shrinkage method only on the observable parameters. Also, the group factor structure has implicitly put many zero restrictions on the loadings (zero loadings for assets outside its own group). Furthermore, the method does apply penalty when estimating the number of factors.

Remark 5 The proposed model can also be estimated by the Bayesian procedure. Instead of using penalization, shrinkage priors on $\boldsymbol{\beta}_i$ can be used, e.g., Hans (2009), Park and Casella (2008), and Polson and Sccot (2012). Also, the priors on the common/group-specific pervasive factors and corresponding factor loadings are considered in the literature, e.g., Tsay and Ando (2012). Because the joint posterior density does not have an analytical expression, one needs to implement the Markov chain Monte Carlo (MCMC) approach. The details are beyond the scope of this paper.

Remark 6 The SCAD penalty shrinks some elements of $\boldsymbol{\beta}_i$ ($i = 1, ..., N$) to exactly zero. This operation is equivalent to selecting relevant observable risk factors. So the set of observable risk factors are automatically determined once the size of regularization parameter κ is fixed. That is, the selection of the set of relevant observable risk factors is equivalent to the determination of the size of regularization parameter κ. In practice, however, we need to determine the size of regularization parameter κ. This problem is considered in Section 5.

Remark 7 As in Assumption D (See Appendix A1), we exclude the situation in which the observable risk factors and underlying unobservable common factors are correlated perfectly. If they are perfectly correlated, then the dimension of unobservable common factors is automatically reduced since they are already included in the regressors. The dimension of the common factors is determined by the information criterion, which will not select a common factor that is already a part of observable factors. Thus the assumption of non-perfect correlation is without loss of generality.

4 Asymptotic theory for statistical test

The previous sections described the model, its assumptions, and the estimation procedure. This section investigates the asymptotic properties of the parameter estimates. We denote the true value of regression coefficients as $\boldsymbol{\beta}_i^0$, the true value of common pervasive factors that affect the returns of all securities as F_c^0, and the true value of group-specific pervasive factors by F_g^0.

We first consider the consistency of the estimators of the regression coefficients $\boldsymbol{\beta}_i$ $i=1,...,N$, the common pervasive factors F_c, and the group-specific pervasive factors $F_g,\ g=1,...,G$. As the dimensions of F_c and $\{F_g,\ g=1,...,S\}$ are increasing, we prove consistency in terms of a matrix norm. Also, we emphasize that the appropriate size of the regularization parameter depends on the length of the time series T, and thus denote κ_T. Then, we have the following theorem.

Theorem 1 *Under Assumptions A-E given in the appendix, and $\kappa_T \to 0$ and $\sqrt{T}\kappa_T \to \infty$ as $T \to \infty$, the estimator $\hat{\boldsymbol{\beta}}_i$ is consistent such that*

$$\hat{\boldsymbol{\beta}}_i \to_p \boldsymbol{\beta}_i^0,$$

In addition, the estimators of the common pervasive factors $\hat{F}_c$ and the group-specific pervasive factors $\{\hat{F}_g,\ g=1,...,G\}$ are consistent in the sense of the following norm:

$$T^{-1/2}\|\hat{F}_c - F_c^0 H_c\| = o_p(1), \quad T^{-1/2}\|\hat{F}_g - F_g^0 H_g\| = o_p(1),$$

where: $H_c^{-1} = V_{c,NT}(F_c^0\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}$, $H_g^{-1} = V_{g,N_gT}(F_g^0\hat{F}_g/T)^{-1}(\Lambda_g'\Lambda_g/N_g)^{-1}$, and $V_{c,NT}$ and V_{g,N_gT} satisfies:

$$\left[\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}(\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_{g_i}\hat{\boldsymbol{\lambda}}_{g_i,i})(\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_{g_i}\hat{\boldsymbol{\lambda}}_{g_i,i})'\right]\hat{F}_c = \hat{F}_c V_{c,NT}$$

and

$$\left[\frac{1}{N_gT}\sum_{i;g_i=g}(\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_c\hat{\boldsymbol{\lambda}}_{c,i})(\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_c\hat{\boldsymbol{\lambda}}_{c,i})'\right]\hat{F}_g = \hat{F}_g V_{g,N_gT}.$$

The proof of Theorem 1 is presented in the Appendix. Given the consistency, we further establish the asymptotic normality of the estimated parameters. Also, we show that our proposed method can identify the set of true explanatory variables. Let $\boldsymbol{\beta}_i^0 = ({\boldsymbol{\beta}_{i10}}', {\boldsymbol{\beta}_{i20}}')'$ be the true parameter value, and $\hat{\boldsymbol{\beta}}_i = (\hat{\boldsymbol{\beta}}_{i1}', \hat{\boldsymbol{\beta}}_{i2}')'$ be the corresponding parameter estimate. Without loss of generality, assume that $\boldsymbol{\beta}_{i20} = \mathbf{0}$. We assume the dimension of $\boldsymbol{\beta}_{i10}$ is small (uniformly bounded over i) but the dimension of $\boldsymbol{\beta}_{i20}$ can be large. We show that the estimator possesses the sparsity property, $\hat{\boldsymbol{\beta}}_{i2} = \mathbf{0}$. We denote $\hat{\boldsymbol{\beta}}_{i1}$ as the parameter estimate of non-zero true coefficients $\boldsymbol{\beta}_{i10}$. We impose the following assumption, which is necessary for the asymptotic normality of $\hat{\boldsymbol{\beta}}_i$. The limiting results are useful for hypothesis testing.

Define the projection matrices

$$
\begin{aligned}
M_{F_c} &= I - F_c(F_c'F_c)^{-1}F_c' \\
M_{F_c,F_g} &= M_{F_c} - M_{F_c}F_g(F_g'M_{F_c}F_g)^{-1}F_g'M_{F_c}
\end{aligned}
$$

The second projection matrix is also equal to $M_G = I - G(G'G)^{-1}G$ with $G = [F_c, F_g]$.

Theorem 2 *Suppose that the i-th security belongs to group g and that Assumptions A-H hold. Furthermore, the regularization parameter satisfies $\kappa_T \to 0$ and $\sqrt{T}\kappa_T \to \infty$ as $T \to \infty$. Then, as $T, N \to \infty$ with $\sqrt{T}/N \to 0$, the following variable-selection consistency holds:*

$$
P(\hat{\boldsymbol{\beta}}_{i2} = \mathbf{0}) \to 1, \quad N, T \to \infty.
$$

Moreover, $\sqrt{T}(\hat{\boldsymbol{\beta}}_{i1} - \boldsymbol{\beta}_{i10})$ is asymptotically normal with mean $\mathbf{0}$ and variance-covariance matrix $R_i(F_c^0, F_g^0)$,

$$
\sqrt{T}(\hat{\boldsymbol{\beta}}_{i1} - \boldsymbol{\beta}_{i10}) \to_d N\left(\mathbf{0}, R_i(F_c^0, F_g^0)\right),
$$

where

$$
R_i(F_c^0, F_g^0) = D_i(F_c^0, F_g^0)^{-1} J_i(F_c^0, F_g^0) D_i(F_c^0, F_g^0)^{-1},
$$

and $D_i(F_c^0, F_g^0)$ and $J_i(F_c^0, F_g^0)$ are the probability limits (in terms of $T \to \infty$) of:

$$
\begin{aligned}
&\frac{1}{T}\left(X_{i,\beta_i^0\neq 0}' M_{F_c^0,F_g^0} X_{i,\beta_i^0\neq 0} + \Sigma_i(\kappa_T)\right), \quad \text{and} \\
&\frac{1}{T}\left(X_{i,\beta_i\neq 0}' M_{F_c^0,F_g^0} E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'] M_{F_c^0,F_g^0} X_{i,\beta_i\neq 0}\right),
\end{aligned}
$$

respectively, with

$$
\Sigma_i(\kappa_T) \;=\; \mathrm{diag}\left\{p_{\kappa_T,\gamma}'(|\beta_{i1}^0|)/|\beta_{i1}^0|, \ldots, p_{\kappa_T,\gamma}'(|\beta_{iq_i}^0|)/|\beta_{iq_i}^0|\right\}.
$$

A proof of Theorem 2 is given in the Appendix. Note that $\sqrt{T}/N \to 0$ is not a strong assumption; the number of securities N can be much larger than the number of time periods T, and the number of time periods T can also be much larger than N. Although restrictions between N and T are needed in terms of simultaneous limit $(N, T \to \infty)$, the theorem holds not only for a particular relationship between N and T, but also for many combinations of N and T. The theorem allows us to perform statistical significance test for coefficients $\boldsymbol{\beta}_i$. We discuss the estimators of $D_i(F_c^0, F_g^0)$ and $J_i(F_c^0, F_g^0)$ in Section 5.

5 Model specification

In practice, however, the number of common pervasive factors, r, and the number of group-specific pervasive factors, $\{r_1, ..., r_G\}$, are unknown. Moreover, we have to select the size of the regularization parameter κ such that the relevant observable risk factors are included, while excluding irrelevant observable risk factors. In this section, we propose a new criterion to select these quantities.

5.1 A new model-selection criterion

Suppose that $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N$ are replicates of the asset returns $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$, given the true value of common pervasive factors F_c and the corresponding factor loadings Λ_c, group-specific pervasive factors F_g, the corresponding factor loadings Λ_g $g = 1, ..., G$, and the observable factors X_i $(i = 1, \ldots, N)$. In other words, we assume that the $\boldsymbol{z}_i$'s are generated from the true underlying structure of the economy. This situation is commonly considered in Bayesian and non-Bayesian model-selection studies; see, for example, Konishi and Kitagawa (1996), Hansen (2005), Ando (2007), Ando and Tsay (2011), and references therein.

To assess the goodness of fit of the estimated model, we use the expected mean squared errors (MSE):

$$\eta(k, k_1, ..., k_G, \kappa) := E_z \left[\frac{1}{NT} \sum_{g=1}^{G} \sum_{i; g_i = g} \left\| \boldsymbol{z}_i - X_i \hat{\boldsymbol{\beta}}_i - \hat{F}_c \hat{\boldsymbol{\lambda}}_{c,i} - \hat{F}_g \hat{\boldsymbol{\lambda}}_{g_i,i} \right\|^2 \right], \tag{4}$$

where k is the number of common pervasive factors, $k_1, ..., k_G$ are the number of group-specific pervasive factors for each group, κ is the regularization parameter, and the expectation is taken with respect to the joint distribution of $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N$. The quantities $k, k_1, ..., k_G, \kappa$ are chosen by minimizing the expected MSE.

A natural estimator of the expected MSE in (4) is the sample-based MSE:

$$\hat{\eta}(k, k_1, ..., k_G, \kappa) := \frac{1}{NT} \sum_{g=1}^{G} \sum_{i; g_i = g} \left\| \boldsymbol{y}_i - X_i \hat{\boldsymbol{\beta}}_i - \hat{F}_c \hat{\boldsymbol{\lambda}}_{c,i} - \hat{F}_g \hat{\boldsymbol{\lambda}}_{g_i,i} \right\|^2.$$

This quantity is formally calculated by replacing the replicates $\boldsymbol{z}_i$ with observed values $\boldsymbol{y}_i$. This sample-based MSE generally has some bias with respect to the expected MSE because, among other reasons, the same data are used to estimate the parameters of the model. We therefore consider a bias-corrected version of the measure.

The bias b of the sample-based MSE with respect to the expected MSE is given by:

$$b(k, k_1, ..., k_G, \kappa) := E_y \left[\eta(k, k_1, ..., k_G, \kappa) - \hat{\eta}(k, k_1, ..., k_G, \kappa) \right], \tag{5}$$

where the expectation is taken with respect to the joint distribution of $\boldsymbol{y}_i (i = 1, \ldots, N)$. We assume that the bias $b(k, k_1, ..., k_G, \kappa)$ can be estimated by some appropriate procedures, yielding $\hat{b}(k, k_1, ..., k_G, \kappa)$. Taking into account the consistency of the proposed model-selection criterion, we suggest minimization of the predictive measure:

$$\begin{aligned} & \hat{\eta}(k, k_1, ..., k_G, \kappa) + \hat{b}(k, k_1, ..., k_G, \kappa) \\ = \; & \frac{1}{NT} \sum_{g=1}^{G} \sum_{i; g_i = g} \left\| \boldsymbol{y}_i - X_i \hat{\boldsymbol{\beta}}_i - \hat{F}_c \hat{\boldsymbol{\lambda}}_{c,i} - \hat{F}_{g_i} \hat{\boldsymbol{\lambda}}_{g_i,i} \right\|^2 + \hat{b}(k, k_1, ..., k_G, \kappa). \end{aligned} \tag{6}$$

The first term on the right-hand side, $\hat{\eta}(k, k_1, ..., k_G, \kappa)$, measures the goodness of fit of the model, whereas the second term, $\hat{b}(k, k_1, ..., k_G, \kappa)$, is a penalty that depends on the complexity of the model. The remaining task is to construct a proper estimator of the penalty term. Another contribution of this paper is the following theorem.

Theorem 3 *Suppose that Assumptions A–E and the condition $\sqrt{T}/N \to 0$ hold. The penalty term is then:*

$$\begin{aligned} \hat{b}(k, k_1, ..., k_G, \kappa) \; = \; & \frac{1}{NT} \sum_{g=1}^{G} \sum_{i; g_i = g} \mathrm{tr} \left[K_i R_i (F_c^0, F_{g_i}^0) \right] \\ & + k \times h(T, N, N_1, ..., N_G) + \sum_{g=1}^{G} k_g \times h_g(T, N, N_1, ..., N_G), \end{aligned}$$

where $K_i = 2X'_{i,\hat{\beta}_i \neq 0} X_{i,\hat{\beta}_i \neq 0}/T$ and $X_{i,\hat{\beta}_i \neq 0}$ is the submatrix of X_i such that the corresponding columns have a nonvanishing component of the parameter estimate, and $R_i(F_c^0, F_{g_i}^0)$ is defined in Theorem 2. The functions $h(T, N, N_1, ..., N_G)$ and $h_g(T, N, N_1, ..., N_G)$ satisfy (a) $h(T, N, N_1, ..., N_G) \to 0$ and (b) $\sqrt{T} h(T, N, N_1, ..., N_G) \to \infty$ as $T, N \to \infty$, and similarly for h_g.

A derivation of the theorem is given in the Appendix.

The first term of $\hat{b}(k, k_1, ..., k_G, \kappa)$ in Theorem 3 controls the size of the regularization parameter. In other words, it is the term for including the relevant observable risk factors only among a large number of observable risk factors. The second term is relevant to the identification of the true number of common pervasive factors. Also, the quantity $k_g \times h_g(T, N, N_1, ..., N_G)$ in the third term is used for selecting the number of group-specific pervasive factors r_g in the group g.

An example of the function $h(T, N, N_1, ..., N_G)$ that satisfies conditions (a) and (b) of the theorem is:

$$h(T, N, N_1, ..., N_G) = \left(\frac{T+N}{TN}\right) \log(TN).$$

Also, the similar function $\left(\frac{T+N_g}{TN_g}\right) \log(TN_g)$ is used for $h_g(T, N, N_1, ..., N_G)$. Substituting these quantities into the predictive measure (6), we have the following model-selection criterion:

$$\begin{aligned} C_p(k, k_1, ..., k_G, \kappa) &= \hat{\eta}(k, k_1, ..., k_G, \kappa) + \frac{1}{NT} \sum_{i=1}^{N} \mathrm{tr}\left[K_i R(\hat{F}_c, \hat{F}_{g_i}, \kappa)\right] \\ &+ k \times \hat{\sigma}^2 \left(\frac{T+N}{TN}\right) \log(TN) + \sum_{g=1}^{G} k_g \times \hat{\sigma}^2 \left(\frac{T+N_g}{TN_g}\right) \log(TN_g), \end{aligned} \tag{7}$$

where $R(\hat{F}_c, \hat{F}_{g_i}, \kappa)$ is a consistent estimate of $R_i(F_c^0, F_{g_i}^0)$, to be discussed in Section 5.2, and $\hat{\sigma}^2$ is a consistent estimate of

$$(NT)^{-1} \sum_{g=1}^{G} \sum_{i; g_i = g} \left\| \boldsymbol{y}_i - X_i \hat{\boldsymbol{\beta}}_i - \hat{F}_c \hat{\boldsymbol{\lambda}}_{c,i} - \hat{F}_g \hat{\boldsymbol{\lambda}}_{g_i,i} \right\|^2.$$

We can choose the number of common pervasive factors, k, the number of group-specific pervasive factors, r_g $(g = 1, ..., G)$, and the size of the regularization parameter, κ, by minimizing the C_p over a specified range of models.

We can regard the proposed model-selection criterion as a generalization of the C_p criterion of Mallows (1973). Like the original C_p criterion, $\hat{\sigma}^2$ provides proper scaling for the penalty term. In applications, it can be replaced by $(NT)^{-1} \sum_{g=1}^{G} \sum_{i; g_i = g} \|\boldsymbol{y}_i - X_i \hat{\boldsymbol{\beta}}_i - \hat{F}_c \hat{\boldsymbol{\lambda}}_{c,i} - \hat{F}_g \hat{\boldsymbol{\lambda}}_{g_i,i}\|^2$, which is obtained under the maximum possible dimension of X_i, the maximum possible number of common pervasive factors $r_{c,max}$, and the maximum possible number of group-specific pervasive factors $r_{g,max}$ $(g = 1, ..., G)$.

The model-specification algorithm for determining the number of common pervasive factors r, the number of group-specific pervasive factors r_g, and the size of regularization parameter κ, is summarized as follows.

Model-specification algorithm

Step 1. Prepare a set of candidate values of the regularization parameter κ, the number of common pervasive factors k, and the number of group-specific pervasive factors $\{k_1, ..., k_G\}$. Then, fix their initial values.

Step 2. Fix the value of regularization parameter κ as one of the candidate values.

Step 3. Given the value of the regularization parameter κ and the number of group-specific pervasive factors $\{k_1, ..., k_G\}$, optimize the number of common pervasive factors k by minimizing the proposed C_p criterion.
Step 4. Given the value of the regularization parameter κ and the number of common pervasive factors k, optimize the number of group-specific pervasive factors $\{k_1, ..., k_G\}$ by minimizing the proposed C_p criterion.
Step 5. Repeat Steps 3 and 4 until convergence.
Step 6. Repeat Steps 2–5 for each of the prepared regularization parameters κ. Then, select the combination of the regularization parameter κ, the number of global factors k, and the number of local factors $\{k_1, ..., k_G\}$, which minimize the proposed C_p criterion.

Remark 8 The matrices $D_i(F_c^0, F_{g_i}^0)$ and $J_i(F_c^0, F_{g_i}^0)$ in the C_p criterion can be obtained by using their empirical versions. If we assume an absence of serial and cross-section correlations in the idiosyncratic errors ($E[\varepsilon_{it}\varepsilon_{is}] = 0$ $(t \neq s)$, $E[\varepsilon_{it}\varepsilon_{jt}] = 0$ $(i \neq j)$), then the calculation of $J_i(F_c^0, F_{g_i}^0)$ can be simplified as follows:

$$J_i(\hat{F}_c, \hat{F}_{g_i}) = \frac{1}{T} X'_{i,\hat{\beta}_i \neq 0} M_{\hat{F}_c, \hat{F}_g} \hat{\Omega}_i M_{\hat{F}_c, \hat{F}_g} X_{i,\hat{\beta}_i \neq 0}, \tag{8}$$

where $\hat{\Omega}_i = \mathrm{diag}\{\hat{\varepsilon}_{i1}^2, ..., \hat{\varepsilon}_{iT}^2\}$ is the diagonal matrix, and $\hat{\varepsilon}_{it} = y_{it} - \hat{\boldsymbol{\beta}}_i \boldsymbol{x}_{it} - \hat{\boldsymbol{f}}'_{c,t} \hat{\boldsymbol{\lambda}}_{c,i} - \hat{\boldsymbol{f}}'_{g_i,t} \hat{\boldsymbol{\lambda}}_{g_i,i}$. If we further assume the absence of heteroskedasticity ($E[\varepsilon_{it}^2] = \sigma^2$), we can estimate $J_i(F_c^0, F_{g_i}^0)$ by:

$$J_i(\hat{F}_c, \hat{F}_{g_i}) = \frac{1}{T} \hat{\sigma}_i^2 X'_{i,\hat{\beta}_i \neq 0} M_{\hat{F}_c, \hat{F}_g} X_{i,\hat{\beta}_i \neq 0},$$

where $\hat{\sigma}_i^2 = (NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \hat{\varepsilon}_{it}^2$ is the variance estimator. Also, thanks to an advantage of the SCAD procedure, the proposed criterion C_p is applicable even when $p_i > T$, where p_i is the number of possible observable risk factors.

Remark 9 The estimated number of common and group-specific pervasive factors allows us to measure the financial integration of markets. Consider a case in which the estimated number of group-specific pervasive factors in each group is zero, while the common pervasive factors exist. In this case, it is natural to regard the corresponding sub-markets are integrated. As a second case, the result may reveal that the number of common pervasive factors is identified as zero, while there exit group-specific pervasive factors in each group. In contrast to the first case, one would naturally think that the market decoupling is observed. In our empirical analysis, both common and group-specific pervasive factors exist. It implies that the Chinese A and B share markets are integrated, while each market has its own characteristics.

Remark 10 Finally, we point out several references that considered the model selection on the standard linear models with factor-augmented regressions. Soofi (1988) studied an information theoretic criterion for factor-augmented regression models. Under the assumption that the response variable follows an exponential family of distributions, Ando and Tsay (2009) proposed information criteria for generalized linear models. Ando and Tsay (2011) developed model-selection criteria for evaluating the quantile regression models with factor-augmented explanatory variables. Ando and Tsay (2014) considered a diffusion-index model-selection problem. These results are not for panel data models. Ando and Tsay (2010) investigated the model-selection problem for large panel data models with the interactive fixed effects of Bai (2009), where the slope coefficients are common to each unit.

5.2 Data-generating processes

The first data-generating model considered is $\boldsymbol{y}_i = X_i\boldsymbol{\beta}_i + F_c\boldsymbol{\lambda}_{c,i} + F_{g_i}\boldsymbol{\lambda}_{g_i,i} + \boldsymbol{\varepsilon}_i$, the r-dimensional common pervasive factor $\boldsymbol{f}_{c,t}$ is a vector of $N(0,1)$ variables, the r_g-dimensional group-specific pervasive factor $\boldsymbol{f}_{g,t}$ $j = 1,..,S$ is also a vector of $N(0,1)$ variables, and each element of the factor-loading matrix Λ_c, $\{\Lambda_1, ..., \Lambda_G\}$ follows $N(0,1)$. The N-dimensional vector $\boldsymbol{\varepsilon}_t$ has a multivariate normal distribution with a mean of $\mathbf{0}$ and covariance matrix I_N. The number of columns of X_i is set to $p_i = 80$, while the true number of regressors is $q_i = 3$ for $i = 1, \ldots, N$. Each of the elements of X_i is generated from the uniform distribution over $[-2, 2]$. The nonzero true parameter values of $\boldsymbol{\beta}_i$ are set to be $(-1, 2, 2)$. These nonzero elements are put into the first three elements of $\boldsymbol{\beta}_i$ and thus the true parameter vector is $\boldsymbol{\beta}_i = (-1, 2, 2, 0, 0, ..., 0)'$ for $i = 1, ..., N$. The true number of common pervasive factors is $r = 5$, and the true numbers of group-specific pervasive factors are $r_1 = 2$ $r_2 = 3$, $r_3 = 4$, $r_4 = 4$, and $r_5 = 5$. We next investigate a case in which the noise term is nonhomoscedastic.

The second data-generating model considered is $\boldsymbol{y}_i = X_i\boldsymbol{\beta}_i + F_c\boldsymbol{\lambda}_{c,i} + F_{g_i}\boldsymbol{\lambda}_{g_i,i} + \boldsymbol{\varepsilon}_i$ and $\varepsilon_{it} = e^1_{it} + \delta_t e^2_{it}$, where $\delta_t = 1$ if t is odd and is zero if t is even, and the N-dimensional vectors $\boldsymbol{e}^1_t = (e^1_{1t}, \ldots, e^1_{Nt})'$ and $\boldsymbol{e}^2_t = (e^2_{1t}, \ldots, e^2_{Nt})'$ follow multivariate normal distributions, with a mean of $\mathbf{0}$ and covariance matrix $S = (s_{ij})$, with $s_{ij} = 0.3^{|i-j|}$, and $\boldsymbol{e}^1_t$ and $\boldsymbol{e}^2_t$ are independent. The noise terms are not serially correlated. The common pervasive factors, the group-specific pervasive factors, the loading matrices, the design matrix X_i, and the true parameter vector $\boldsymbol{\beta}_i$ are generated by the same method as before. The key feature of the model is that the noise terms are not homoscedastic.

As a third example, we investigated the performance of the proposed method when the idiosyncratic errors had some serial and cross-sectional correlations. The model is $\boldsymbol{y}_i = X_i\boldsymbol{\beta}_i + F_c\boldsymbol{\lambda}_{c,i} + F_{g_i}\boldsymbol{\lambda}_{g_i,i} + \boldsymbol{\varepsilon}_i$ with $\varepsilon_{it} = e_{it} + 0.2\varepsilon_{i,t-1}$, where $t = 1, \ldots, T$, the N-dimensional vector $\boldsymbol{e}_t = (e_{1t}, \ldots, e_{Nt})'$ follows multivariate normal distributions with mean $\mathbf{0}$ and covariance matrix $S = (s_{ij})$, where $s_{ij} = 0.3^{|i-j|}$. Other variables are defined as before.

As a fourth example, we generated the data under a situation where the set of true observable risk factors X_i are correlated with a set of r common pervasive factors. Again, the model is $\boldsymbol{y}_i = X_i\boldsymbol{\beta}_i + F_c\boldsymbol{\lambda}_{c,i} + F_{g_i}\boldsymbol{\lambda}_{g_i,i} + \boldsymbol{\varepsilon}_i$ with $\varepsilon_{it} = e_{it} + 0.2e_{i,t-1}$, where $t = 1, \ldots, T$, the noise values $\boldsymbol{e}_t = (e_{1t}, \ldots, e_{Nt})'$ follow multivariate normal distributions with mean $\mathbf{0}$ and covariance matrix $S = (s_{ij})$, where $s_{ij} = 0.3^{|i-j|}$. Also, we generated a set of $r+2$ dimensional random variables $\boldsymbol{z}_t = (z_{1t}, ..., z_{r+2,t})'$, which follow multivariate normal distributions with mean $\mathbf{0}$ and covariance matrix $S = (s_{ij})$, with $s_{ij} = 0.3^{|i-j|}$. Each of the elements of X_i is generated from the uniform distribution over $[-2, 2]$. Then, the first r elements of $\boldsymbol{z}_t$, $(z_{1t}, ..., z_{r,t})'$ are used for the common pervasive factors $\boldsymbol{f}_{c,t}$ and the remaining part of $\boldsymbol{z}_t = (z_{r+1,t}, z_{r+2,t})'$ is added to the first two elements of observable risk factors $\boldsymbol{x}_{it}$, $i = 1, ..., N$. This operation creates a situation in which the common pervasive factors and the observable risk factors have correlation structures. Other variables are defined as before.

As a fifth example, we generated the data under a situation where the set of true observable risk factors X_i are correlated with group-specific pervasive factors. Each of the elements of the observable risk factors X_i is generated from the uniform distribution over $[-2, 2]$. We generated a set of $r_1 + 2$ dimensional random variables $\boldsymbol{z}_t = (z_{1t}, ..., z_{r_1+2,t})'$, which follow multivariate normal distributions with mean $\mathbf{0}$ and covariance matrix $S = (s_{ij})$, with $s_{ij} = 0.3^{|i-j|}$. Then, the first r_1 elements of $\boldsymbol{z}_t$, $(z_{1t}, ..., z_{r_1,t})'$ are used for the group-specific pervasive factors $\boldsymbol{f}_{1,t}$ of group S_1, and then the remaining part of $\boldsymbol{z}_t$, $(z_{r+1,t}, z_{r+2,t})'$ was added to the first elements of observable risk factors $\boldsymbol{x}_{it}$, $i = 1, ..., N$. This operation creates a situation where the group-specific pervasive factors and the true observable risk factors have correlation structures. Other variables are defined as before.

In these simulation settings, we consider two cases: (1) the number of securities in each group is equal, i.e., $N_1 = N_2 = \cdots = N_5$, and (2) the number of securities in each group are different.

5.3 Results

We generated 1,000 replicates using each of the five data-generating models. We then applied the proposed model-selection criterion, C_p, to select simultaneously the number of common pervasive factors, the number of group-specific pervasive factors, and the size of regularization parameter κ. We set the possible numbers of both common and group-specific pervasive factors to range from 0 to 10. Thus, the maximum number of common and group-specific pervasive factors were set to 10, respectively. Possible candidates for the regularization parameter κ are $\kappa = \{10, 1, 0.1, 0.01, 0.001\}$. To speed up the computation of our C_p criterion in (7), we used an estimator of $J_i(\hat{F}_c, \hat{F}_{g_i})$ in (8), which assumes the absence of serial correlations, to calculate the penalty term of our C_p score. However, as our results show that our criterion performs well, even if this assumption, i.e., the absence of serial correlation, does not hold. The model-selection results for the third example indicate that our C_p criterion is robust to the misspecification of the noise characteristics.

Tables 1 $\sim$ 4 report the percentages for correct, under-, and overidentification of the proposed C_p criterion under the five data-generating models. This presentation style is followed by Tsay and Ando (2012). If the proposed method identifies the true number of common pervasive factors (r) 1,000 times out of 1,000 trials, the corresponding three columns with respect to r become 0, 100, and 0 under U, C, and O, respectively. As shown in the tables, the proposed C_p criterion is capable of selecting the true number of common and group-specific pervasive factors. Table 5 shows the identification performance for the true observable risk factors X_i. We measured the identification performance using the true positive rate (TPR) and the true negative rate (TNR):

$$\mathrm{TPR} = \frac{\sum_{i=1}^N \sum_k I\{\hat{\beta}_{ik} \neq 0 \text{ and } \beta_{ik} \neq 0\}}{\sum_{i=1}^N \sum_k I\{\beta_{ik} \neq 0\}} = \frac{\sum_{i=1}^N \sum_k I\{\hat{\beta}_{ik} \neq 0 \text{ and } \beta_{ik} \neq 0\}}{N \times 3}$$
$$\mathrm{TNR} = \frac{\sum_{i=1}^N \sum_k I\{\hat{\beta}_{ik} = 0 \text{ and } \beta_{ik} = 0\}}{\sum_{i=1}^N \sum_k I\{\beta_{ik} = 0\}} = \frac{\sum_{i=1}^N \sum_k I\{\hat{\beta}_{ik} = 0 \text{ and } \beta_{ik} = 0\}}{N \times (80-3)},$$

where $I(\cdot)$ is the indicator function, which takes a value of 1 if it is true and 0 otherwise. As we have the TPR and TNR for each of N securities, we take their averages. As shown in Table 5, the proposed criterion is capable of selecting the true set of observable risk factors. We can also see that the performance improves as T increases.

Finally, we discuss the regression coefficient estimation results. For simplicity, we shall report the results obtained under the fourth data-generating model only, because other data settings have similar results. Simulation results for the parameter estimates of $\hat{\boldsymbol{\beta}}_i$ are reported in Table 1. Because the theoretical properties of the parameter estimates $\hat{\boldsymbol{\beta}}_i$ are common for each i, we report the results for $\hat{\boldsymbol{\beta}}_1$ only. Again, given T and N, similar results are obtained for others $\hat{\boldsymbol{\beta}}_2$,...,$\hat{\boldsymbol{\beta}}_N$. As shown in Table 6, the parameters are well estimated in the simulation studies. Because the length of $\hat{\boldsymbol{\beta}}_1$ is very long (a vector of length 80), we report the estimation results for the true regressors $(\hat{\beta}_{1,1}, \hat{\beta}_{1,2}, \hat{\beta}_{1,3})'$, and those for the first three irrelevant regressors, $(\hat{\beta}_{1,4}, \hat{\beta}_{1,5}, \hat{\beta}_{1,6})$. We point out that the remaining elements of $\hat{\boldsymbol{\beta}}_1$ (i.e., $\hat{\beta}_{1,7}, \hat{\beta}_{1,8}, ..., \hat{\beta}_{1,80}$) are similar to the estimation results of $(\hat{\beta}_{1,4}, \hat{\beta}_{1,5}, \hat{\beta}_{1,6})$ as they are the irrelevant set of predictors. We can see that the time periods T mainly controls the precision of the parameter estimates. This investigation coincides with the asymptotic theory, developed in Section 4.

In summary, our simulation results show that the proposed C_p criterion works well in selecting the number of common pervasive factors, the number of group-specific pervasive factors, and the set of relevant observable risk factors.

6 Analysis of Chinese A- and B-share markets

There are two stock exchange markets in mainland China: the Shanghai and Shenzhen stock exchanges. In these markets, two types of shares are traded, namely A- and B-shares. Although A- and B-shares are listed and traded in the mainland market, the former are denominated in RMB and were originally traded only among Chinese citizens, whereas the latter are denominated in foreign currencies and were originally traded among non-Chinese citizens or Chinese residing overseas. The Chinese government launched the qualified foreign institutional investors (QFII) policy in 2003 and introduced foreign investors into the domestic A-share market. Although Chinese mainlanders have been eligible to trade B-shares with legal foreign currency accounts since March 2001, the mainlanders may prefer to trade only in A-shares owing to the currency barrier. It therefore seems plausible that the underlying asset return structure of A-shares is different from that of B-shares. It is also important to know how these two stock exchange markets respond to the global economy. This paper investigates empirical questions such as the following: How many common and group-specific pervasive factors exist in the stock market in mainland China? What type of observable risk factors explain the market? And, how can the unobservable common factors be understood in terms of observable variables in the economy?

6.1 The Data

We use monthly excess returns of Chinese A- and B-shares from Standard & Poor (S&P)'s Datastream Database. We consider a roughly 11-year sample, covering the March 2002 to December 2012 period, and systematically exclude stocks with missing returns data. We calculate excess returns by subtracting the interest rate on the one-month interbank offered rate from the individual stock returns. Ideally, we would use the one-month Treasury bill rate instead of the interbank offered rate. However, the one-month Treasury bill rate is only available from 2007. Our reported results are robust to the analysis of the returns, which are not subtracted by the interest rate on the one-month interbank offered rate. We partition our original universe of stocks into two groups, the first containing A-shares, and the second containing B-shares. This implies that the number of groups is $S = 2$. The above filtering procedure yields 1,039 A-share firms and 102 B-share firms.

Table 7 provides the descriptive statistics on the asset returns. For each group, we computed these statistics from the time series of each stock and we then report the cross-sectional average. The statistics can therefore be interpreted as those for a representative stock. As shown in Table 7, the size of the panel in this application is sufficiently large. Because the A-share market is the main market on the China mainland, the number of stocks in each group is not quite the same, with the A-share group having a much higher number of stocks. The volatility of stocks on the A- and B-share markets are comparable. It is apparent from the table that B-shares have a higher skewness than their counterparts on the A-share market. We can also see that the kurtosis of the B-share market is higher than that of stocks on the A-share market.

Numerous studies have analyzed the stock market reaction of the developed countries to changes in macroeconomic variables (Fama, 1981, 1990; Mandelker and Tandon, 1985; Chen et al., 1986; Fama and French, 1989; Cheung and Ng, 1998). If the economic outlook reflects the stock market, such information would be helpful for capturing the stock return characteristics. With a view that the effective investment style will change overtime, it has become more common to consider a variety of types of economic information. Therefore, for the observable risk factors, we employed several types of macroeconomic variables, including macroeconomic climate indexes (leading, coincident, and lagging indexes), the money supply, and the inflation rate (the consumer price index). Monetary policy may affect stock prices (Thorbeke, 1997) through at least two channels. Generally, the growth of the money supply is positively related to the inflation rate. An increase in the money supply may lead to an increase in the inflation rate (e.g., Fama, 1981), which may increase the nominal risk-free interest rate (with the real interest rate is fixed), resulting in a negative relationship between the money supply and stock prices. This is because the higher discount rate level lowers the value of the firm through the valuation formula. On the other hand, a corporate earning effect may result in increased future cash flow and stock prices, while the effect of a higher discount rate would be neutralized if cash flows increase with inflation. Also, investors would expect higher dividend payments and hence increase their demand for the stocks. Inflation may also be caused through real factors such as consumption. In Marshall (1992), an expected increase in inflation decreases the expected return to money, and this reduces demand for money and increases the demand for equity, resulting in a positive correlation between inflation and stock prices. On the other hand, there are empirical studies (Fama and Schwert, 1977; Geske and Roll, 1983) that report a negative relationship between inflation and stock prices.

Commodity prices are a major cost factor in various economic activities in China. Therefore, commodity price information is used for the observable risk factors, including the industrial metal price, the aluminum price, the copper price, the crude oil price, the natural gas price, and the nickel price. In addition to these, we use the gold price and the silver price, which affect the price of alternatives to the traditional financial instruments, including stocks and bonds.

Currency movements directly affect the earnings of Chinese firms. There is an exchange rate risk for holding foreign currency. Also, the value of a firm's assets with foreign operations, and its revenue through exports, will be affected by fluctuations in exchange rates. Moreover, the firms that sell goods that compete with imports are subject to the price elasticity of consumer demand and impacted by the cost of imported raw materials. In this paper, we consider the Chinese yuan to the US dollar exchange rate, the Chinese yuan to the Japanese yen exchange rate, the Chinese yuan to the euro exchange rate, the Chinese yuan to the UK pound exchange rate, and the Chinese yuan to the HK dollar exchange rate.

Finally, the international stock market conditions may affect the China mainland stock market. Therefore, we use the S&P 500 index, the MSCI World index, the FTSE 100 index, the MSCI Europe index, the TOPIX index, the Hang Seng index, as well as the MSCI China index. Table 8 provides descriptive statistics of the monthly returns of the above-mentioned observable risk factors. Some observable risk factors are highly skewed. Also, we can see that some variables have heavier tails than normal as their kurtosis levels are above 3.

Figure 1 shows the correlation matrix of the set of 25 observable risk factors. The ordering of the variables in the correlation matrix is identical to the ordering of those in Table 8. The plot indicates that the set of six stock market indexes (S&P 500, MSCI World, FTSE 100, MSCI Europe, TOPIX, Hang Seng, MSCI China) are highly correlated. This implies that the stock markets seem to be connected to each other. We can also see from Figure 1 that some commodity prices (industrial metal, aluminum, copper, crude oil, nickel) have a high level of correlation. Figure 1 also indicates that the Chinese yuan to the Japanese yen exchange rate is negatively correlated with some commodity prices (industrial metal, aluminum, copper) as well as with some stock market indexes (S&P 500, MSCI World, FTSE 100, TOPIX). The MSCI World index is correlated with many other variables, with the exception of the China money supply, the China macroeconomic climate indexes (lagging), the inflation rate (consumer price index), the Chinese yuan to the HK dollar exchange rate, the gold price, and the gas price.

6.2 How many pervasive factors?

We fit the model (1) by minimizing the objective function. Then, we applied the proposed model-selection criterion, C_p, to select simultaneously the number of common pervasive factors, the number of group-specific pervasive factors, and the size of the regularization parameter κ. The possible numbers of both common and group-specific pervasive factors range from 0 to 10. Possible candidates for the regularization parameter κ are $\kappa = \{10, 1, 0.1, 0.01, 0.001\}$. The estimated numbers of common/group-specific pervasive factors are: $\hat{r} = 2$ common pervasive factors, $\hat{r}_1 = 1$ group-specific pervasive factors with respect to A-shares, and $\hat{r}_2 = 3$ group-specific pervasive factors with respect to B-shares. This suggests that there are at least six pervasive factors in the Chinese mainland stock markets.

We next explore the economic meanings of six constructed factors. Here, we use the methods in Bai (2003) and Bai and Ng (2006a). Suppose we observe s_t, the time series of an observable economic variable. We are interested in the relationship between the variable s_t and the unobservable common/group-specific pervasive factors.

Consider the case where we try to explore the meaning of the common pervasive factors $\boldsymbol{f}_{c,t}$. As pointed out by Bai and Ng (2006a), one may regress $\boldsymbol{\varepsilon}_i^c = y_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{f}}_{g_i,t}'\hat{\boldsymbol{\lambda}}_{g_i,i}$ on s_t and then use some measure to assess the explanatory power of s_t. However, Bai and Ng (2006a) mentioned that this is not a satisfactory test because even if s_t is exactly equal to one of the elements of the common pervasive factors $\boldsymbol{f}_{c,t}$, s_t might still be weakly correlated only with $\boldsymbol{\varepsilon}_i^c$ if the variance of the idiosyncratic error is large. In this paper, following Bai and Ng (2006a), we regress s_t on the estimated common pervasive factors $s_t = \hat{\boldsymbol{f}}_{c,t}'\boldsymbol{\gamma}_c + e_{c,t}$, and then conduct the statistical significance test of the least squared estimate $\hat{\boldsymbol{\gamma}}_c$. Then, we use the result of Theorem 1 of Bai and Ng (2006b). Under $\sqrt{T}/N \to 0$, they showed that $\sqrt{T}(\hat{\boldsymbol{\gamma}}_c - \boldsymbol{\gamma}_c)$ asymptotically follow the multivariate normal with zero mean and the covariance matrix Σ_{γ_c} with its consistent estimator, in our setting, is:

$$\widehat{\Sigma}_{\gamma_c} = \left(\frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{f}}_{c,t}\hat{\boldsymbol{f}}_{c,t}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{e}_{c,t}^2\hat{\boldsymbol{f}}_{c,t}\hat{\boldsymbol{f}}_{c,t}'\right)\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{f}}_{c,t}\hat{\boldsymbol{f}}_{c,t}'\right)^{-1},$$

with $\hat{e}_{c,t} = s_t - \hat{\boldsymbol{\gamma}}_c'\hat{\boldsymbol{f}}_{c,t}$. See Bai and Ng (2006b) for more details. We can implement the same idea for exploring the meaning of the group-specific pervasive factors.

To make a link between the estimated common and group-specific pervasive factors, we considered the following six observable economic/market variables: consumer confidence index in China, the Chicago Board Options Exchange (CBOE) volatility index, market excess returns of A-shares, market excess returns of B-shares, and two factors considered by Fama and French (1993), HML and SMB, but computed using the Chinese data. Note that the market excess returns (MSCI China index) were already used as an observable risk factor. The consumer confidence index measures consumer confidence, which is defined as the degree of optimism on the state of the economy. The CBOE volatility index is a key measure of market expectations of near-term volatility conveyed by S&P 500 stock index option prices, and has been considered by many to be the world's premier barometer of investor sentiment and market volatility. Because the market excess returns of A- and B-shares represents the group-specific market factor, it is expected that the estimated group-specific factors relate to these observable risk factors. In particular, the first asymptotic principal component of the A-share group, corresponding to the largest eigenvalue, is related to the market excess returns of A-shares. A similar argument is made in relation to the B-share market. HML factor accounts for the spread in returns between value and growth stocks, and thus shows the value premium. SMB measures the historic excess returns of small caps over big caps. HML and SMB factors are calculated based on the stock returns of Shanghai and Shenzhen stock exchanges. Except for the CBOE volatility index, these 6 variables are from market data in China.

Table 9 summarizes the results. Here, we standardized each of the observable economic/market variables g_t before we regressed them on the estimated factors. In Table 9, for each factor, the first row corresponds to the estimated regression coefficients, whereas the second and third rows correspond to their standard deviations and t-values, respectively. If the absolute value of the t-value is above 2.56, 1.96, or 1.64, the estimated regression coefficient is statistically significant at 1%, 5% and 10% level, respectively. We can see from Table 9 that the first common pervasive factor, the first element of $\boldsymbol{f}_{c,t}$, is relating to the market excess returns of A-shares, the market excess returns of B-shares, and the size factor, SMB. The second common pervasive factor is also relating to SMB. Contrary to findings for the US market, the book-to-market ratio is not included in the common pervasive factors across group. However, as shown in the result for Group B, the third group-specific factor of B-share is relating to HML at the 10% level. This implies that HML factor is effective in the B-share market only. As we expected, the first group-specific factor of the B-share market relates strongly to the market excess returns of B-shares. However, none of the six observable factors relate to the first group-specific factor of the A-shares. It would be an interesting topic to investigate what type of economic/market variables relate to the first group-specific factor of A-shares. This also applies to the second group-specific factor of B-shares. With respect to two of the market risk factors (VIX and CCI, explained below), we could not find the relationships with the estimated factors.

6.3 What types of observable risk factors explain the market?

From Theorem 2, we can implement a statistical significance test for the estimated regression coefficients. Thus, we can check whether the regression coefficients $\hat{\boldsymbol{\beta}}_i$ for each security are statistically significant. Table 10 shows the percentage of statistically significant conclusions of each of the observable risk factors. The percentage for the k-th observable risk factor is calculated as follows:

$$\frac{1}{N_g} \sum_{i;g_i=g} I\{\hat{\beta}_{ik} \text{ is statistically significant}\},$$

where $I(\cdot)$ is the indicator function, which takes a value of 1 if it is true and 0 otherwise, and N_g is the size of group g. The significance level was set as $\alpha = 0.05$. As shown in Table 10, we can make the following investigations. First, among the five Chinese macroeconomic variables, the leading indicator of the macroeconomic climate index, the money supply, and the lagging indicator of the macroeconomic climate index are important in explaining the excess returns of individual stocks both in the A- and B-share markets. On the other hand, the consumer price index does not seem to explain the excess returns of individual stocks during the March 2002 to December 2012 period.

Second, the exchange rate of the Chinese yuan to the UK pound has a large impact on the excess returns of individual stocks in the A-share market. Interestingly, its explanatory power in relation to B-shares, as indicated by the percentage in Table 10, is half that of the A-shares. This implies that the investors in B-shares are greatly concerned with the exchange rate of the Chinese yuan to the UK pound. Although they are smaller than the percentage for the Chinese yuan to the UK pound, the percentages in the table for the exchange rates of the Chinese yuan to the US dollar and the Japanese yen are an important source of A- and B-share market fluctuations. Again, the impact of these exchange rates on the A-share market is half of that on the B-share market.

Third, the commodity prices are important observable risk factors. We can see some contrasts between the A- and B-share markets in this respect. The gold and silver price indexes appear to be more important for A-share investors than they are for B-share investors. On the other hand, the metal, oil, and aluminum price indexes seem to be more important for B-share investors than A-share investors.

Fourth, Table 10 shows that the MSCI China index is important for almost all B-shares, but the index is less important for about half of the A-shares. The MSCI China index consists of securities of B, H, Red Chip and P Chip share classes, but excludes securities of A-share class. Also, the correlation between the market excess returns of A- and B-shares is above 0.8. This may be a reason why half of the A-shares are explained by the MSCI China index, even though the index excludes A-share securities. Also, as shown in Table 9, the first unobserved common factor is mainly the Chinese market excess return, and thus for interpretation, we regard the first unobserved factor as the market excess return.

Fifth, the B-share market participants are more concerned with the FTSE 100 index than are the A-share market participants. The impact of the European, Hong Kong, and Japanese stock markets appears to be less important than that of the China mainland, the US and the UK stock markets for the B-share market participants. Although the European, Hong Kong, and Japanese stock market indexes are not included in almost all $\boldsymbol{\beta}_i$'s, this does not imply that these markets are irrelevant. As shown in Figure 1, the six stock market indexes (S&P 500, MSCI World, FTSE 100, MSCI Europe, TOPIX, Hang Seng, MSCI China) are highly correlated and, thus, some of the indexes are sufficient to explain the variations of individual stock returns of A- and B-shares. Similar arguments apply to the other group of observable risk factors.

6.4 Price of risk

In the APT framework, the expected returns on assets are approximatively linear in their sensitivities to the factors $E[r] = \nu_0 + \boldsymbol{\lambda}'\boldsymbol{\nu}$, where ν_0 is a constant, $\boldsymbol{\nu}$ is a vector of factor risk premiums, and $\boldsymbol{\lambda}$ is a vector of factor sensitivities. Here, we partition the excess returns into two groups (A-shares and B-shares) and investigate the subset pricing relations based on Fama and MacBeth (1973) type approach. Two stage approach was also used in Goyal et al. (2008), in which the factor structure of excess returns on stocks traded on the NYSE and Nasdaq (two groups) were studied. Through the model construction process, we have already obtained the matrix of factor sensitivities $\hat{\Lambda}_c$ (common pervasive factors), $\hat{\Lambda}_1$ (group-specific pervasive factors with respect to A-shares), and $\hat{\Lambda}_2$ (group-specific pervasive factors with respect to B-shares). We then run the following cross-sectional regression for each group:

$$\hat{\boldsymbol{r}}_g = \nu_{0,g}\mathbf{1} + \hat{\Lambda}_c\boldsymbol{\nu}_{G,g} + \hat{\Lambda}_g\boldsymbol{\nu}_g + \hat{\Lambda}_{FF3,g}\boldsymbol{\nu}_{FF3,g} + \boldsymbol{\xi}_g, \qquad (g = 1, 2),$$

where $\mathbf{1}$ is a vector of ones, $\boldsymbol{\xi}_g$ is a vector of pricing errors, $\hat{\Lambda}_{FF3,1}$ is the matrix of sensitivities to Fama and French's 3 factors for A-shares (i.e., ER-A, HML, SMB in Table 9), $\hat{\Lambda}_{FF3,2}$ is the matrix of sensitivities to Fama and French's 3 factors for B-shares (i.e., ER-B, HML, SMB in Table 9), and $\hat{\boldsymbol{r}}_g$ is a vector of average excess returns, which are observable-risk adjusted, i.e., for the i-th security, $T^{-1}\sum_{t=1}^{T}(y_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}}_i)$ is used. Here $\boldsymbol{x}_{it}$ are listed in Table 8 and do not include HML, SMB , ER-A and ER-B factors. Table 11 reports the results of this cross-sectional regression. The estimates for the risk premium on the common-pervasive factors are statistically significant in each group. Almost all factors seem to be priced. This indicates that our method extracted useful factors that are priced.

One of the main contributions of this paper is to propose a procedure to select the set of relevant observable risk factors. It is also interesting to see whether these selected observable risk factors are priced in the cross-section of asset returns. Similar to the above analysis, we run the following cross-sectional regression for each group:

$$\hat{\boldsymbol{r}}_g = \nu_{0,g}\mathbf{1} + \hat{\Lambda}_{\beta,g}\boldsymbol{\nu}_{\beta,g} + \hat{\Lambda}_{FF3,g}\boldsymbol{\nu}_{FF3,g} + \boldsymbol{\xi}_g, \qquad (g = 1, 2),$$

where $\mathbf{1}$ is a vector of ones, $\hat{\Lambda}_{\beta,1}$ is the matrix of sensitivities to the set of observable risk factors for A-shares in Table 10, $\hat{\Lambda}_{\beta,2}$ is the matrix of sensitivities to the set of observable risk factors for B-shares in Table 10, and $\hat{\boldsymbol{r}}_g$ is a vector of average excess returns, which are unobservable common/group-specific pervasive factor adjusted, i.e., for the i-th security, $T^{-1}\sum_{t=1}^{T}(y_{it} - \hat{\boldsymbol{f}}'_{c,t}\hat{\boldsymbol{\lambda}}_{c,i} - \hat{\boldsymbol{f}}'_{g_i,t}\hat{\boldsymbol{\lambda}}_{g_i,i})$ is used. Table 12 reports the results of this cross-sectional regression. The statistically significant estimates for the risk premium on the observable risk factors varies over the groups. We can see that the estimates on the macroeconomic climate coincident index, Yuan/Yen exchange rate, natural gas, are priced in both groups. Like this, we can investigate the observable risk factors that are priced.

We can see that the number of priced observable risk factors for the A-shares market is much greater than for the B-shares market. Together with the number of priced factors of unobservables, the results imply that the A-shares market exhibits more heterogeneity than the B-shares market in terms of price of risk. Historically, A-shares market investors were mainlanders until 2003. Due to the entry of the qualified foreign institutional investors into the domestic A-share market, the degree of heterogeneity has been increased as the A-share market consists of mainlanders and newly entered foreign investors after 2003. On the other hand, due to the currency barrier of the mainlanders, the investors in B-shares market are still foreign investors. This might be one of the reasons why such differences are observed.

6.5 Robustness check

A unique feature of the Chinese stock market is that many companies issue "twin" A and B shares. Here, the "twin" share has two classes of common shares with identical voting and dividend rights, listed on the same exchanges (Shanghai or Shenzhen stock exchanges), but traded by different participants (see, for instance, Mei et al. (2009)). The dataset contains 50 "twin" A and B shares. To check the robustness of the obtained result, we exclude the 50 "twin" A shares from the dataset, resulting in 989 A-share firms and 102 B-share firms. We then implement the same model construction procedure as in the previous section. The selected numbers of common/group-specific pervasive factors are identical to the case of without excluding the "twin" A shares. Also, similar results are obtained with respect to the observable risk factors. This suggests that the previous results are robust to the presence of twin shares.

The effect from foreign denominated currencies is another market characteristic to be investigated. B-shares are denominated in foreign currencies with Shanghai B-shares traded in U.S. dollars while Shenzhen B-shares in Hong Kong dollars. In the previous section, we analyzed B-shares based on the foreign-currency denominated returns. Here, we take the effect of exchange rates into account. More specifically, we express the B-share returns in Chinese yuan, and then implement the the same model construction procedure as in the previous sections. We use the same dataset without excluding the "twin" A shares. Again, the previously reported results are robust to this change. This is not surprising because the dataset covers a time period in which the value of Chinese yuan was pegged to the U.S. dollar. Although Chinese yuan exchange rate has been allowed to float since 2005, it was in a narrow margin around a fixed base rate determined with reference to a basket of world currencies.

7 Conclusion

We proposed a new econometric modeling procedure for the multifactor asset-pricing model, which has three main features: high-dimensional observable risk factors, unobservable common pervasive factors that influence a large number of assets, and group-specific pervasive factors that influence a subset of assets. Both the number of assets and the number of potential observable risk factors can be larger than the sample size (the number of time periods). We developed a procedure to identify the relevant observable factors from a large number of potentially related factors. We showed that the proposed procedure delivers consistent estimation of the unknown beta coefficients; the estimated beta coefficients are also asymptotically normal. The analysis is nonstandard because of the selection problem in the presence of unobservable factors and a large number of observable factors. We also studied how to determine the number of (unobservable) common pervasive factors and the number of group-specific factors. Monte Carlo simulations demonstrated that the proposed modeling procedure performs well.

We then applied the proposed method to the analysis of the Chinese stock markets and presented a number of empirical findings. Application of the method to the A-share and B-share markets identifies the commonalities and differences in the return structure of the assets across the two markets. The study revealed that the observable risk factors affect the two markets in different ways. The study further demonstrated the existence of two common pervasive factors across the two markets, a single group-specific factor in the A-share market, and three group-specific factors in the B-share market. We also studied the price of risk in the cross section of returns. The findings are robust to the presence of "twin" shares, and robust to the exchange rate fluctuations.

References

[1] Aguilar, O., West, M., 2000. Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. Journal of Business and Economic Statistics 18, 338–357.

[2] Amengual, D., Watson, M. W., 2007. Consistent estimation of the number of dynamic factors in a large N and T panel. Journal of Business and Economic Statistics 25, 91–96.

[3] Ando, T., 2007. Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. Biometrika 94, 443–458.

[4] Ando, T., and J. Bai (2013). Panel data models with grouped factor structure under unknown group membership. Available at SSRN:
http://ssrn.com/abstract=2373629

[5] Ando, T., Tsay, R., 2009. Model selection for generalized linear models with factor-augmented predictors (with discussion). Applied Stochastic Models in Business and Industry 25, 207–235.

[6] Ando, T., Tsay, R., 2010. Information criteria for panel data models with interactive fixed effects. Working paper, Keio University.

[7] Ando, T., Tsay, R., 2011. Quantile regression models with factor-augmented predictors and information criteria. Econometrics Journal 14, 1–24.

[8] Ando, T., Tsay, R., 2014. A predictive approach for selection of diffusion index models. Econometric Reviews 33, 68–99.

[9] Armagan, A., Dunson, D.B., Lee, J., Bajwa, W. 2013. Posterior consistency in linear models under shrinkage priors. Biometrika, forthcoming.

[10] Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–172.

[11] Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77, 1229–1279.

[12] Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

[13] Bai, J., Ng, S., 2006a. Evaluating latent and observed factors in macroeconomics and finance. Journal of Econometrics 131, 507–537.

[14] Bai, J., Ng, S. 2006b. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. Econometrica, 74, 1133–1150.

[15] Bhattacharya, A., Dunson, D., 2011. Sparse Bayesian infinite factor models. Biometrika 98, 291–306.

[16] Baruch, S., Saar, G., 2009. Asset returns and the listing choice of firms. Review of Financial Studies 22, 2239–2274.

[17] Bekaert, G., Hodrick, R. J., Zhang, X., 2009. International stock return comovements. Journal of Finance 64, 2591–2626.

[18] Breiman, L., 1996. Heuristics of instability and stabilization in model selection. Annals of Statistics 24, 2350–2383.

[19] Chen, N. F., Roll, R., Ross, S., 1986. Economic forces and the stock market. Journal of Business 59(3), 383–403.

[20] Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure and mean-variance analysis in large asset markets. Econometrica 51, 1305–1324.

[21] Chen, N.-F., Roll, R., Ross, S.A., 1986. Economic forces and the stock market. Journal of Business, 59, 383–403.

[22] Cheung, Y., Ng, L. K., 1998. International evidence on the stock market and aggregate economic activity. Journal of Empirical Finance 5, 281–296.

[23] Cho, D., Eun, C., L. Senbet., 1986. International Arbitrage Pricing Theory: An Empirical Investigation. Journal of Finance 41, 313–330.

[24] Connor, G., Korajczyk, R., 1986. Performance measurement with the arbitrage pricing theory: a new framework for analysis. Journal of Financial Economics 15, 373–394.

[25] Connor, G., Korajczyk, R. A., 1993. A test for the number of factors in an approximate factor model. Journal of Finance 48, 1263–1291.

[26] Diebold, F., Li, C., Yue, V., 2008. Global yield curve dynamics and interactions: a dynamic Nelson-Siegel approach. Journal of Econometrics 146, 315–363.

[27] Fama, E. F., MacBeth, J.D., 1973. Risk, Return, and Equilibrium: Empirical Tests. Journal of Political Economy, 81, 607–636.

[28] Fama, E. F., 1981. Stock prices, real activity, inflation and money. American Economic Review 71, 545–65.

[29] Fama, E. F., 1990. Stock prices, expected returns, and real activity. Journal of Finance 45, 1089–1108.

[30] Fama, E. F., French, K. R., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 5, 23–49.

[31] Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

[32] Fama, E., French, K., 2004. New lists: fundamentals and survival rates. Journal of Financial Economics 73, 229–269.

[33] Fama, E. F., French, K. R., 2012. Size, value, and momentum in international stock returns. Journal of Financial Economics 105, 457–472.

[34] Fama, E. F., Schwert, G. W., 1977. Asset returns and inflation. Journal of Financial Economics 29, 337–363.

[35] Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96, 1348–1361.

[36] Flury, B., 1984. Common principal components in k groups. Journal of the American Statistical Association 79, 892–898.

[37] Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic factor model: identification and estimation. Review of Economics and Statistics 82, 540–554.

[38] Forni, M., Lippi, M., 2001. The generalized factor model: representation theory. Econometric Theory 17, 1113–1141.

[39] Fung, H. G., Lee, W., Leung, W. K., 2000. Segmentation of the A- and B-share Chinese equity markets. Journal of Financial Research 23, 179–195.

[40] Geske, R., Roll, R., 1983. The fiscal and monetary link between stock returns and inflation, Journal of Finance 38, 1–33.

[41] Geweke, J., 1977. The dynamic factor analysis of economic time series. In: Aigner, D. J., Goldberger, A. S. (eds), Latent Variables in Socio-Economic Models. Amsterdam: North-Holland, pp. 365–383.

[42] Goyal, A., Perignon, C., Villa, C., 2008. How common are common return factors across the NYSE and Nasdaq? Journal of Financial Economics 90, 252–271.

[43] Griffin, J. M., 2002. Are the Fama and French factors global or country specific? Review of Financial Studies 15, 783–803.

[44] Hallin, M., Liska R., 2007. The generalized dynamic factor model: determining the number of factors. Journal of the American Statistical Association 102, 603–617.

[45] Hans, C., 2009. Bayesian lasso regression. Biometrika, 96, 835–845.

[46] Jones, C. S., 2001. Extracting factors from heteroskedastic asset returns. Journal of Financial Economics 62, 293–325.

[47] Karolyi, G. A., Stluz, R.M. 2003. Are financial assets priced locally or globally? Handbook of the Economics of Finance (ed. George, C., Harris, M. and Stulz, R.), Vol. 1, 975–1020.

[48] Korajczyk, R. A., Sadka, R., 2008. Pricing the commonality across alternative measures of liquidity. Journal of Financial Economics 87, 45–72

[49] Kose, A., Otrok, C., Whiteman, C. 2008, Understanding the evolution of world business cycles. International Economic Review 75, 110–130.

[50] Krzanowski, W. J., 1979. Between-groups comparison of principal components. Journal of the American Statistical Association 74, 703–707.

[51] Lewis, K., 2011. Global asset pricing. Annual Review of Financial Economics 3, 435–466.

[52] Lintner, J., 1965. Security prices, risk, and maximal gains from diversification. Journal of Finance 20, 587–615.

[53] Lopes, H. F., West, M., 2004. Bayesian model assessment in factor analysis. Statistica Sinica 14, 41–67.

[54] Lopes, H. F., Salazar, E., Gamerman, D., 2008. Spatial dynamic factor models. Bayesian Analysis 3, 759–792.

[55] Ma, X., 1996. Capital controls, market segmentation and stock prices: evidence from the Chinese stock market. Pacific-Basin Finance Journal 4, 219–239.

[56] Malkiel, B. G., Xu, Y., 2003. Investigating the behavior of idiosyncratic volatility. Journal of Business 76, 613–644.

[57] Mallows, C. L., 1973. Some comments on *Cp*. Technometrics 15, 661–675.

[58] Mandelker, G., Tandon, K., 1985. Common stock returns, real activity, money, and inflation: some international evidence. Journal of International Money and Finance 4, 267–286.

[59] Marshall, D. A., 1992. Inflation and asset returns in a monetary economy. Journal of Finance 47, 1315–1342.

[60] Mei, J., Scheinkman, J. A, and Xiong, W., 2009 Speculative Trading and Stock Prices: Evidence from Chinese A-B Share Premia. Annals of Economics and Finance 10, 225–255.

[61] Merton, R. C., 1973. An intertemporal capital asset pricing model. Econometrica, 41, 867-887.

[62] Moench, E., Ng, S., 2011. A factor analysis of housing market dynamics in the US and the regions. Econometrics Journal 14, 1–24.

[63] Moench, E., Ng, S., Potter, S., 2012. Dynamic hierarchical factor models. Review of Economics and Statistics, forthcoming.

[64] Naranjo, A., Protopapadakis, A., 1997. Financial market integration tests: an investigation using US equity markets. Journal of International Financial Markets, Institutions, and Money 7, 93–135.

[65] Park, T., Casella, G., 2008. The Bayesian lasso. Journal of American Statistical Association, 103, 681–686.

[66] Pesaran, M. H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74, 967–1012.

[67] Polson, N.G., Scott, J.G., 2012. Local shrinkage rules, levy processes and regularized regression. Journal of Royal Statistical Society, B74, 287–311.

[68] Ross, S. A., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341–360.

[69] Sargent, T. J., Sims, C. A., 1977. Business cycle modeling without pretending to have too much a priori economic theory. In: Sims C. et al. (eds), New Methods in Business Cycle Research. Federal Reserve Bank of Minneapolis, Minneapolis.

[70] Schott, J. R., 1999. Partial common principal component subspaces. Biometrika 86, 899–908.

[71] Schwert, G. W., 2002. Stock volatility in the new millennium: how wacky is Nasdaq? Journal of Monetary Economics 49, 1–24.

[72] Sharpe, W., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–442.

[73] Song, M., 2013. Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications, unpublished manuscript, Department of Economics, Columbia University.

[74] Soofi, E. S., 1988. Principal component regression under exchangeability. Communications in Statistics – Theory and Methods A17, 1717–1733.

[75] Stock, J. H., Watson, M. W., 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

[76] Su, D., 1999. Ownership restrictions and stock prices: evidence from Chinese markets. Financial Review 34, 37–55.

[77] Thorbeke, W., 1997. On stock market returns and monetary policy. Journal of Finance 52, 635–653.

[78] Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society B58, 267–288.

[79] Tsay, R., Ando, T., 2012. Bayesian panel data analysis for exploring the impact of recent financial crisis on the US stock market. Computational Statistics and Data Analysis 56, 3345–3365.

[80] Wang, P., 2010. Large dimensional factor models with a multi-level factor structure, unpublished manuscript, Department of Economics, HKUST.

[81] Westerlund, J. and Urbain, P.J. (2013). On The Estimation And Inference In Factor-Augmented Panel Regressions With Correlated Loadings, *Economics Letters*, 247-250, 2013.

[82] Zhang, C. H., 2010. Nearly unbiased variable selection under minimax concave penalty. Annals of Statistics 38, 894–942.

Table 1: Percentage (%) of factor-selection results for correct (C), under- (U), and overidentification (O) over 1,000 replicates. If the proposed method identifies the true number of common pervasive factors 1,000 times out of 1,000 trials, the corresponding three columns with respect to r become 0, 100, and 0 under U, C, and O, respectively. This table contains the model-identification results under the five data structure settings. In each data-generating process, the true number of common pervasive factors is $r = 5$, and the true numbers of group-specific pervasive factors are $r_1 = 2$, $r_2 = 3$, $r_3 = 4$, $r_4 = 4$, and $r_5 = 5$. The data size setting is $T = 200$, $N = 1,000$, $N_1 = 200$, $N_2 = 200$, $N_3 = 200$, $N_4 = 200$, and $N_5 = 200$.

	r			r_1			r_2			r_3			r_4			r_5		
Data structure	U	C	O	U	C	O	U	C	O	U	C	O	U	C	O	U	C	O
1	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0
2	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0
3	0	99	1	0	100	0	0	100	0	0	100	0	0	100	0	0	99	1
4	0	100	0	0	100	0	0	100	0	0	100	0	0	99	1	0	100	0
5	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	99	1

Table 2: Percentage (%) of factor-selection results for U, C, and O over 1,000 replicates. If the proposed method identifies the true number of common pervasive factors 1,000 times out of 1,000 trials, the corresponding three columns with respect to r become 0, 100, and 0 under U, C, and O, respectively. This table contains the model-identification results under the five data structure settings. In each data-generating process, the true number of common pervasive factors is $r = 5$, and the true numbers of group-specific pervasive factors are $r_1 = 2$, $r_2 = 3$, $r_3 = 4$, $r_4 = 4$, and $r_5 = 5$. The panel size setting is $T = 150$, $N = 1,000$, $N_1 = 200$, $N_2 = 200$, $N_3 = 200$, $N_4 = 200$, and $N_5 = 200$.

	r			r_1			r_2			r_3			r_4			r_5		
Data	U	C	O	U	C	O	U	C	O	U	C	O	U	C	O	U	C	O
1	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0
2	0	100	0	0	99	1	0	100	0	0	99	1	0	100	0	0	99	1
3	0	99	1	0	99	1	0	100	0	0	99	1	0	100	0	0	99	1
4	0	99	1	0	99	1	0	99	1	0	99	1	0	99	1	0	99	1
5	1	98	1	0	99	1	0	99	1	0	98	2	0	98	2	0	97	3

Table 3: Percentage (%) of factor-selection results for U, C, and O over 1,000 replicates. If the proposed method identifies the true number of common pervasive factors 1,000 times out of 1,000 trials, the corresponding three columns with respect to r become 0, 100, and 0 under U, C, and O, respectively. This table contains the model-identification results under the five data structure settings. In each data-generating process, the true number of common pervasive factors is $r = 5$, and the true numbers of group-specific pervasive factors are $r_1 = 2$, $r_2 = 3$, $r_3 = 4$, $r_4 = 4$, and $r_5 = 5$. The panel size setting is $T = 130$, $N = 920$, $N_1 = 180$, $N_2 = 150$, $N_3 = 140$, $N_4 = 250$, and $N_5 = 200$.

	r			r_1			r_2			r_3			r_4			r_5		
Data	U	C	O	U	C	O	U	C	O	U	C	O	U	C	O	U	C	O
1	0	99	1	0	100	1	0	99	1	0	100	0	0	100	0	0	99	1
2	0	96	4	0	98	2	0	98	2	0	98	2	6	94	0	3	96	1
3	0	96	4	0	99	1	0	98	2	0	99	1	0	97	3	0	97	3
4	0	96	4	0	98	2	0	99	1	0	98	3	0	96	4	0	97	3
5	2	96	2	0	96	4	0	96	4	0	98	2	0	97	3	0	97	3

Table 4: Percentage (%) of factor-selection results for U, C, and O over 1,000 replicates. If the proposed method identifies the true number of common pervasive factors 1,000 times out of 1,000 trials, the corresponding three columns with respect to r become 0, 100, and 0 under U, C, and O, respectively. This table contains the model-identification results under the five data structure settings. In each data-generating process, the true number of common pervasive factors is $r = 5$, and the true numbers of group-specific pervasive factors are $r_1 = 2$, $r_2 = 3$, $r_3 = 4$, $r_4 = 4$, and $r_5 = 5$. The panel size setting is $T = 150$, $N = 2000$, $N_1 = 400$, $N_2 = 400$, $N_3 = 400$, $N_4 = 400$, and $N_5 = 400$.

	r			r_1			r_2			r_3			r_4			r_5		
Data	U	C	O	U	C	O	U	C	O	U	C	O	U	C	O	U	C	O
1	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0
2	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0
3	0	99	1	0	100	0	0	100	0	0	100	0	0	100	0	0	99	1
4	0	99	1	0	100	0	0	99	1	0	100	0	0	99	1	0	99	1
5	0	99	1	0	100	0	0	100	0	0	99	1	0	100	0	0	99	1

Table 5: The entries are the true positive rate (TPR) and true negative rate (TNR) for the observable risk factors. TPR and TNR are averages taken over i ($i = 1, .., N$) for 1,000 replicates. The panel size settings are: (a) $T = 200$, $N = 1,000$, $N_1 = 200$, $N_2 = 200$, $N_3 = 200$, $N_4 = 200$, and $N_5 = 200$; (b) $T = 150$, $N = 1,000$, $N_1 = 200$, $N_2 = 200$, $N_3 = 200$, $N_4 = 200$, and $N_5 = 200$; (c) $T = 130$, $N = 920$, $N_1 = 180$, $N_2 = 150$, $N_3 = 140$, $N_4 = 250$, and $N_5 = 200$; and (d) $T = 150$, $N = 2000$, $N_1 = 400$, $N_2 = 400$, $N_3 = 400$, $N_4 = 400$, and $N_5 = 400$.

	(a)		(b)		(c)		(d)	
Data	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR
1	0.9999	0.8991	0.9927	0.8475	0.9912	0.8312	0.9920	0.8469
2	0.9999	0.8989	0.9920	0.8450	0.9908	0.8346	0.9917	0.8444
3	0.9999	0.8915	0.9916	0.8528	0.9878	0.8384	0.9901	0.8511
4	0.9999	0.8927	0.9973	0.8440	0.9954	0.8459	0.9853	0.8390
5	0.9999	0.8930	0.9910	0.8461	0.9948	0.8290	0.9908	0.8412

Table 6: Simulation results of the parameter estimates for $\hat{\boldsymbol{\beta}}_1$ based on 1,000 repetitions, under the data-generating process 4. We report the mean and standard deviation (SD) of the regression coefficient estimates. Because the theoretical properties of $\hat{\boldsymbol{\beta}}_i$ are common for each i, we report the results for $\hat{\boldsymbol{\beta}}_1$ only. Also, the true number of regressors are three and thus the first three elements of $\hat{\boldsymbol{\beta}}_i$ take nonzero values. Because the length of $\hat{\boldsymbol{\beta}}_1 = (\hat{\beta}_{1,1}, ..., \hat{\beta}_{1,p})'$ is very long, we report $(\hat{\beta}_{1,1}, \hat{\beta}_{1,2}, \hat{\beta}_{1,3}, \hat{\beta}_{1,4}, \hat{\beta}_{1,5}, \hat{\beta}_{1,6})'$. Similar results are obtained for the remaining elements $\hat{\beta}_{1,7}$,...,$\hat{\beta}_{1,p}$, which are similar to $\hat{\beta}_{1,4}, \hat{\beta}_{1,5}, \hat{\beta}_{1,6}$.

$T = 50,\ N = 2000,\ N_1 = 400,\ N_2 = 400,\ N_3 = 400,\ N_4 = 400,\ N_5 = 400.$

	$\beta_{11} = -1$	$\beta_{12} = 2$	$\beta_{13} = 2$	$\beta_{14} = 0$	$\beta_{15} = 0$	$\beta_{16} = 0$
Estimate	-0.9690	2.0019	2.0121	0.0060	-0.0061	0.0065
SD	0.2990	0.2987	0.3127	0.2301	0.2307	0.2878

$T = 100,\ N = 2000,\ N_1 = 400,\ N_2 = 400,\ N_3 = 400,\ N_4 = 400,\ N_5 = 400.$

	$\beta_{11} = -1$	$\beta_{12} = 2$	$\beta_{13} = 2$	$\beta_{14} = 0$	$\beta_{15} = 0$	$\beta_{16} = 0$
Estimate	-0.9956	1.9955	2.0160	-0.0047	-0.0055	0.0045
SD	0.0955	0.0961	0.1076	0.0364	0.0389	0.0354

$T = 200,\ N = 2000,\ N_1 = 400,\ N_2 = 400,\ N_3 = 400,\ N_4 = 400,\ N_5 = 400.$

	$\beta_{11} = -1$	$\beta_{12} = 2$	$\beta_{13} = 2$	$\beta_{14} = 0$	$\beta_{15} = 0$	$\beta_{16} = 0$
Estimate	-0.9984	1.9959	1.9951	0.0000	0.0012	0.0011
SD	0.0541	0.0553	0.0616	0.0000	0.0118	0.0116

$T = 50,\ N = 1000,\ N_1 = 200,\ N_2 = 200,\ N_3 = 200,\ N_4 = 200,\ N_5 = 200.$

	$\beta_{11} = -1$	$\beta_{12} = 2$	$\beta_{13} = 2$	$\beta_{14} = 0$	$\beta_{15} = 0$	$\beta_{16} = 0$
Estimate	-0.9733	1.9816	2.0132	0.0131	-0.0205	-0.0069
SD	0.3101	0.3196	0.3209	0.2426	0.2465	0.2354

$T = 100,\ N = 1000,\ N_1 = 200,\ N_2 = 200,\ N_3 = 200,\ N_4 = 200,\ N_5 = 200.$

	$\beta_{11} = -1$	$\beta_{12} = 2$	$\beta_{13} = 2$	$\beta_{14} = 0$	$\beta_{15} = 0$	$\beta_{16} = 0$
Estimate	-0.9948	2.0159	2.0112	0.0051	-0.0057	0.0061
SD	0.0991	0.0984	0.1094	0.0366	0.0386	0.0395

Table 7: Descriptive statistics for asset returns of the sample period: March 2002 to December 2012. Each of these statistics is calculated using the time series of individual stocks and then cross-sectionally averaged across stocks.

A-share market		B-share market	
N_1	1039	N_2	102
Mean (%)	-0.084	Mean	-0.096
SD (%)	14.496	SD	14.03
Skewness	0.057	Skewness	0.599
Kurtosis	5.725	Kurtosis	6.950

Table 8: Descriptive statistics of the monthly returns of the observable risk factors. The mean and the standard deviations (SD) are multiplied by 100. The sample periods are from March 2002 to December 2012.

Variables	Mean(%)	SD(%)	Skew.	Kurtosis
China macroeconomic variables				
MACROECONOMIC CLIMATE INDEX (LEADING)	-0.007	0.467	-0.260	3.174
MONEY SUPPLY - M2	1.375	1.076	0.305	3.690
MACROECONOMIC CLIMATE INDEX (COINCIDENT)	0.020	0.598	-0.544	4.362
MACROECONOMIC CLIMATE INDEX (LAGGING)	0.039	0.678	-0.342	3.328
CONSUMER PRICE INDEX	0.013	0.661	-0.419	4.658
Exchange rates				
CHINESE YUAN to US DOLLAR	-0.218	0.424	-1.704	6.538
CHINESE YUAN to YEN (JAPAN)	0.192	2.674	0.306	2.840
CHINESE YUAN to EURO	0.097	3.026	-0.060	4.129
CHINESE YUAN to POUND (UK)	-0.126	2.797	-0.875	5.122
CHINESE YUAN to HK DOLLAR	-0.213	0.449	-1.482	5.970
Commodity price index				
S&P GSCI Industrial Metals Spot	0.787	7.245	-1.040	7.418
S&P GSCI Aluminum Spot	0.284	6.178	-0.381	4.248
S&P GSCI Copper Spot	1.279	8.994	-1.094	8.527
S&P GSCI Crude Oil Spot	1.142	10.657	-0.682	4.398
S&P GSCI Gold Spot	1.383	5.073	-0.160	4.072
S&P GSCI Natural Gas Spot	0.370	14.802	0.260	2.919
S&P GSCI Nickel Spot	0.807	11.185	-0.669	5.165
S&P GSCI Silver Spot	1.545	9.630	-0.503	3.941
Major stock market indexes				
S&P 500 INDEX	0.220	5.472	-1.766	8.622
MSCI WORLD INDEX	0.264	5.692	-1.731	8.287
FTSE 100 INDEX	0.117	5.320	-1.294	5.789
MSCI EUROPE INDEX	0.257	6.925	-1.438	7.181
TOPIX INDEX	-0.211	6.081	-0.745	3.243
HANG SENG INDEX	0.540	6.897	-0.504	4.541
MSCI CHINA INDEX	0.520	8.838	-0.295	4.008

Table 9: The result of regression of the observable economic/market risk factors s_t on the estimated common/group-specific factors. For example, when we regress s_t on the estimated common pervasive factors $\hat{\boldsymbol{f}}_{c,t}$, the regression model is $s_t = \hat{\boldsymbol{f}}_{c,t}^{'}\gamma_c + e_{c,t}$, where $\boldsymbol{\gamma}_c$ are the regression coefficients, and $e_{c,t}$ is the error term. The six observable economic/market risk factors s_t are the consumer confidence index in China (CCI), the CBOE volatility index (VIX), market excess returns of A-shares (ER–A), market excess returns of B-shares (ER–B), the book-to-market ratio (HML), and the market capitalization (SMB). HML and SMB are based on Chinese stock returns. For each factor, the first row corresponds to the estimated regression coefficients $\hat{\boldsymbol{\gamma}}_c$, whereas the second and third rows are the corresponding standard deviations and t-values. If the absolute value of the t-value is above 2.56, 1.96, and 1.64, the estimated regression coefficient is statistically significant at 1%, 5%, and 10% level, respectively.

		CCI	VIX	ER–A	ER–B	HML	SMB
Common factors							
	First	-0.070	-0.017	0.579	0.425	-0.066	0.443
	SD	0.077	0.116	0.084	0.099	0.085	0.083
	t-value	-0.915	-0.151	6.873	4.277	-0.783	5.306
	Second	-0.107	-0.021	0.084	-0.039	0.046	-0.172
	SD	0.079	0.081	0.075	0.091	0.069	0.076
	t-value	-1.355	-0.262	1.109	-0.434	0.670	-2.246
Group A							
	First	0.026	0.005	0.067	0.018	0.056	0.005
	SD	0.073	0.077	0.096	0.081	0.063	0.098
	t-value	0.356	0.068	0.698	0.222	0.892	0.052
Group B							
	First	0.073	-0.058	-0.079	-0.357	-0.094	-0.239
	SD	0.081	0.059	0.101	0.11	0.074	0.095
	t-value	0.903	-0.988	-0.779	-3.235	-1.261	-2.507
	Second	-0.126	-0.097	-0.019	0.065	-0.077	0.057
	SD	0.100	0.085	0.092	0.098	0.079	0.085
	t-value	-1.26	-1.133	-0.204	0.658	-0.978	0.678
	Third	0.043	0.062	-0.095	-0.024	-0.167	-0.029
	SD	0.084	0.060	0.098	0.114	0.085	0.113
	t-value	0.513	1.027	-0.972	-0.214	-1.958	-0.264

Table 10: The percentage of statistically significant observable risk factors across markets. The percentage for the k-th observable risk factor is calculated as $\sum_{i;g_i=g} I\{\hat{\beta}_{ik}$ is statistically significant$\}/N_g$, where $I(\cdot)$ is the indicator function, which takes a value of 1 if it is true and 0 otherwise, and N_g is the size of group g. The significance level was set as $\alpha = 0.05$.

Variables	A-shares	B-shares
China macroeconomic variables		
MACROECONOMIC CLIMATE INDEX (LEADING)	50.33	67.64
MONEY SUPPLY - M2	35.12	24.50
MACROECONOMIC CLIMATE INDEX (COINCIDENT)	4.52	0.98
MACROECONOMIC CLIMATE INDEX (LAGGING)	63.90	60.78
CONSUMER PRICE INDEX	0.00	0.98
Exchange rates		
CHINESE YUAN to US DOLLAR	18.76	32.35
CHINESE YUAN to YEN	16.55	28.43
CHINESE YUAN to EURO	4.13	1.96
CHINESE YUAN to UK POUND	35.70	73.52
CHINESE YUAN to HK DOLLAR	3.75	1.96
Commodity price index		
S&P GSCI Industrial Metals Spot	7.12	22.54
S&P GSCI Aluminum Spot	3.17	34.31
S&P GSCI Copper Spot	17.22	0.00
S&P GSCI Crude Oil Spot	5.10	28.43
S&P GSCI Gold Spot	18.09	9.80
S&P GSCI Natural Gas Spot	0.76	0.00
S&P GSCI Nickel Spot	19.05	24.50
S&P GSCI Silver Spot	14.91	13.72
Major stock market indexes		
S&P 500 INDEX	10.10	14.70
MSCI WORLD INDEX	0.76	0.00
FTSE 100 INDEX	24.63	53.92
MSCI EUROPE INDEX	1.05	0.00
TOPIX INDEX	2.79	7.84
HANG SENG INDEX	11.06	0.98
MSCI CHINA INDEX	47.06	93.13

Table 11: Factor risk premiums for the set of $\hat{r} = 2$ common pervasive factors, $\hat{r}_1 = 1$ group-specific pervasive factors with respect to A-shares, $\hat{r}_2 = 3$ group-specific pervasive factors with respect to B-shares, and risk premiums for the Fama and French three factors, all constructed from the Chinese stock markets. For each group, we run the following cross-sectional regression: $\hat{\boldsymbol{r}}_g = \nu_{0,g}\mathbf{1} + \hat{\Lambda}_c \boldsymbol{\nu}_{c,g} + \hat{\Lambda}_g \boldsymbol{\nu}_g + \hat{\Lambda}_{FF3,g} \boldsymbol{\nu}_{FF3,g} + \boldsymbol{\xi}_g$, $(g = 1, 2)$. Details on this model are described in Section 7.4. The first line associated with each row presents the factor risk price estimates, and the second line reports the p-value (in parenthesis). The R^2_{OLS} denotes the OLS cross-sectional R^2. Note that the A-share group has one group-specific pervasive factor only.

		A-shares	B-shares
Constant	$\nu_{0,g}$	0.0032	0.0081
		(0.0000)	(0.0082)
Common factor	First ($\nu_{c,g1}$)	0.1156	0.1983
		(0.0000)	(0.0057)
	Second ($\nu_{c,g2}$)	0.0889	0.1624
		(0.0000)	(0.0004)
Group-specific factor	First (ν_{g1})	0.0029	0.1402
		(0.7162)	(0.0728)
	Second (ν_{g2})	——	0.1188
			(0.0039)
	Third (ν_{g2})	——	0.0600
			(0.1147)
	ER-A ($\nu_{FF3,11}$)	0.0079	——
		(0.0006)	
	ER-B ($\nu_{FF3,21}$)	——	0.0299
			(0.0566)
	HML ($\nu_{FF3,g2}$)	-0.0033	-0.0103
		(0.0000)	(0.0037)
	SMB ($\nu_{FF3,g3}$)	0.0018	0.0075
		(0.1171)	(0.2038)
R^2_{OLS}		0.2282	0.4565

Table 12: Risk premiums for the set of observable risk factors. For each group, we run the following cross-sectional regression: $\hat{\boldsymbol{r}}_g = \nu_{0,g}\mathbf{1} + \hat{\Lambda}_{\beta,g}\boldsymbol{\nu}_{\beta,g} + \hat{\Lambda}_{FF3,g}\boldsymbol{\nu}_{FF3,g} + \boldsymbol{\xi}_g$, $(g = 1, 2)$. Details on this model are described in Section 7.4. The first/third columns associated with each row present the factor risk price estimates, and the second/fourth columns report the p-value (in parenthesis). The R^2_{OLS} denotes the OLS cross-sectional R^2.

	A-shares	B-shares
Constant	0.0024 (0.0001)	-0.0012 (0.6638)
China macroeconomic variables		
MACROECONOMIC CLIMATE INDEX (LEADING)	0.0000 (0.5911)	0.0000 (0.8308)
MONEY SUPPLY - M2	0.0000 (0.9778)	-0.0006 (0.3343)
MACROECONOMIC CLIMATE INDEX (COINCIDENT)	-0.0003 (0.0140)	-0.0014 (0.0169)
MACROECONOMIC CLIMATE INDEX (LAGGING)	-0.0004 (0.0020)	-0.0010 (0.0937)
CONSUMER PRICE INDEX	0.0004 (0.0012)	0.0000 (0.9977)
Exchange rates		
CHINESE YUAN to US DOLLAR	-0.0001 (0.0079)	0.0000 (0.7837)
CHINESE YUAN to YEN	-0.0011 (0.0065)	-0.0039 (0.0389)
CHINESE YUAN to EURO	-0.0004 (0.3994)	-0.0014 (0.6107)
CHINESE YUAN to UK POUND	-0.0001 (0.7636)	-0.0014 (0.5310)
CHINESE YUAN to HK DOLLAR	-0.0003 (0.0000)	0.0000 (0.8974)
Commodity price index		
S&P GSCI Industrial Metals Spot	-0.0007 (0.6018)	-0.0070 (0.2837)
S&P GSCI Aluminum Spot	-0.0012 (0.2669)	-0.0078 (0.1439)
S&P GSCI Copper Spot	0.0002 (0.8681)	-0.0046 (0.5211)
S&P GSCI Crude Oil Spot	0.0012 (0.5040)	0.0036 (0.7235)
S&P GSCI Gold Spot	0.0003 (0.6289)	0.0023 (0.5299)
S&P GSCI Natural Gas Spot	-0.0057 (0.0287)	-0.0331 (0.0047)
S&P GSCI Nickel Spot	-0.0039 (0.0540)	-0.0184 (0.0600)
S&P GSCI Silver Spot	0.0041 (0.0061)	0.0058 (0.4283)
Major stock market indexes		
S&P 500 INDEX	-0.0034 (0.0006)	-0.0051 (0.3041)
MSCI WORLD INDEX	-0.0030 (0.0037)	-0.0022 (0.6540)
FTSE 100 INDEX	-0.0037 (0.0000)	0.0016 (0.7128)
MSCI EUROPE INDEX	-0.0034 (0.0064)	0.0000 (0.9908)
TOPIX INDEX	0.0006 (0.5056)	0.0086 (0.0677)
HANG SENG INDEX	-0.0016 (0.1777)	0.0064 (0.2372)
MSCI CHINA INDEX	-0.0003 (0.8229)	0.0087 (0.2669)
R^2_{OLS}	0.2576	0.7647

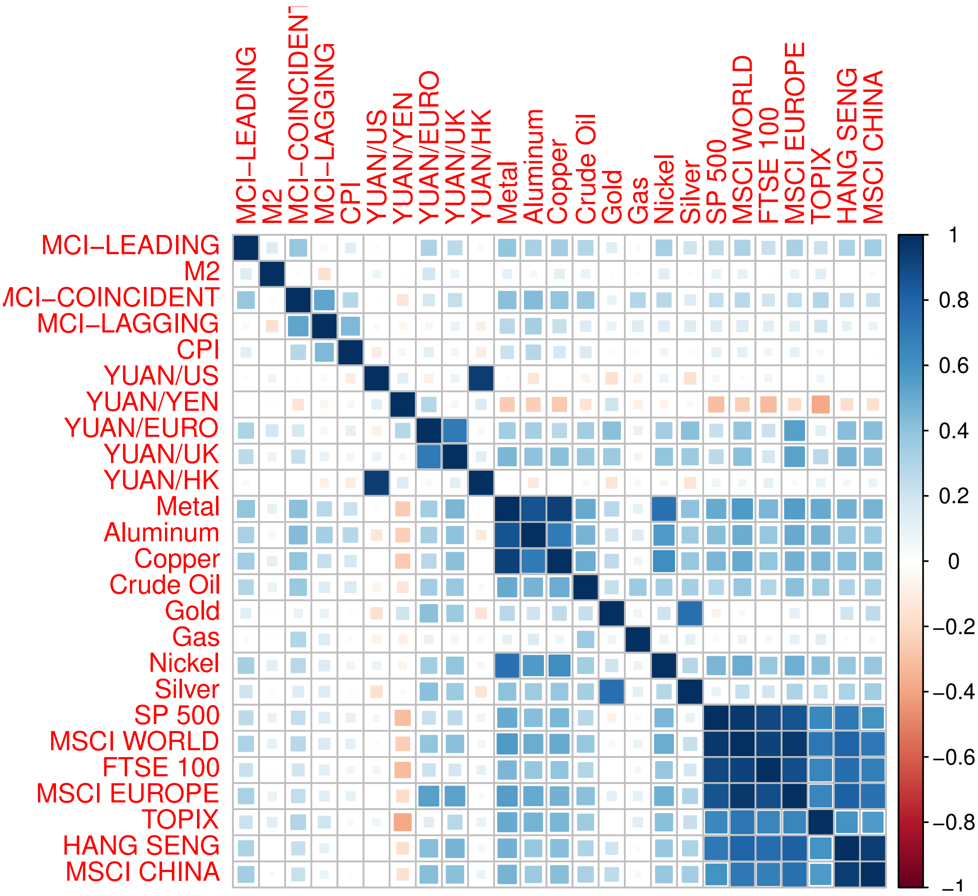

Figure 1: Correlation matrix. The ordering of the variables in the correlation matrix is identical to the ordering of those in Table 8. MCI: macroeconomic climate index, M2: money supply, and CPI: consumer price index.

Supplementary document, for online posting only

Appendix

A1. Assumptions

We first state the assumptions needed for the asymptotic analysis. Following each assumption, its meaning is briefly explained.

Assumption A: Common/group-specific pervasive factors

The common and group-specific pervasive factors satisfy $E\|\boldsymbol{f}_{c,t}\|^4 < \infty$ and $E\|\boldsymbol{f}_{g,t}\|^4 < \infty$ $(g = 1, ..., G)$, respectively. Also

$$T^{-1}\sum_{t=1}^{T}\boldsymbol{f}_{c,t}\boldsymbol{f}_{c,t}{}' \to \Sigma_{F_c} \quad \text{and} \quad T^{-1}\sum_{t=1}^{T}\boldsymbol{f}_{g,t}\boldsymbol{f}_{g,t}{}' \to \Sigma_{F_g}$$

as $T \to \infty$, where Σ_{F_c} is an $r \times r$ positive definite matrix, and Σ_{F_g} is an $r_g \times r_g$ positive definite matrix. The vector of common/group-specific factor $\boldsymbol{f} = (\boldsymbol{f}'_{c,t}, \boldsymbol{f}'_{1,t}, ..., \boldsymbol{f}'_{G,t})'$ has a positive definite covariance matrix. Also, we assume orthogonality between the common and group-specific factors $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f}_{c,t}\boldsymbol{f}_{g,t}{}' = 0$.

The full rank assumption of Σ_{F_c} and Σ_{F_g} is necessary for the number of common factors to be r and the number of group-specific factors to be r_g $(g = 1, 2, ..., G)$. The Assumption B below is for the same reason. The last part of the assumption A assumes that the common factors $\boldsymbol{f}_{c,t}$ and the group-specific factors $\boldsymbol{f}_{g,t}$ are orthogonal. Wang (2010) demonstrates that this assumption is needed to separately identify the common (global) and the group-specific factors. However, it can be shown that the estimated slope coefficients $\boldsymbol{\beta}_i$ does not depend on this assumption.

Assumption B: Factor loadings

The factor-loading matrix for the common pervasive factors $\Lambda_c = [\boldsymbol{\lambda}_{c,1}, \ldots, \boldsymbol{\lambda}_{c,N}]'$ satisfies $E\|\boldsymbol{\lambda}_{c,i}^4\| < \infty$ and:

$$\|N^{-1}\Lambda_c'\Lambda_c - \Sigma_{\Lambda_c}\| \to \mathbf{0} \quad \text{as} \quad N \to \infty,$$

where Σ_{Λ_c} is an $r \times r$ positive definite matrix. Let Λ_g denote the $N_g \times r$ factor loading matrix for the group factor $\boldsymbol{f}_{g,t}$ (for assets belong to group g). For example, if the first N_1 assets belong to group 1, then $\Lambda_1 = [\boldsymbol{\lambda}_{1,1}, \boldsymbol{\lambda}_{1,2}, \ldots, \boldsymbol{\lambda}_{1,N_1}]'$. We assume $E\|\boldsymbol{\lambda}_{g,i}^4\| < \infty$ and

$$\|N_g^{-1}\Lambda_g'\Lambda_g - \Sigma_{\Lambda_g}\| \to \mathbf{0} \quad \text{as} \quad N_g \to \infty,$$

where Σ_{Λ_g} is an $r_g \times r_g$ positive definite matrix, $g = 1, ..., G$. In addition, $[\Lambda_{cg}, \Lambda_g]$ is of full column rank, where Λ_{cg} consists of rows of Λ_c corresponding to group g.

Assumption C: Security-specific returns

There exists a positive constant $C < \infty$ such that for all N and T,
(C1): $E[\varepsilon_{it}] = 0$, $E[|\varepsilon_{it}|^8] < C$ for all i and t;
(C2): $E[\varepsilon_{it}\varepsilon_{js}] = \tau_{ij,ts}$ with $|\tau_{ij,ts}| \le |\tau_{ij}|$ for some τ_{ij} for all (t,s), and $N^{-1}\sum_{i,j=1}^{N}|\tau_{ij}| < C$; and $|\tau_{ij,ts}| \le |\eta_{ts}|$ for some η_{ts} for all (i,j), and $T^{-1}\sum_{t,s=1}^{N}|\eta_{ts}| < C$. In addition, $(TN)^{-1}\sum_{i,j,t,s=1}|\tau_{ij,ts}| < C$.
(C3): For every (s,t), $E[|N^{-1/2}\sum_{i=1}^{N}(\varepsilon_{is}\varepsilon_{it} - E[\varepsilon_{is}\varepsilon_{it}])|^4] < C$.
(C4): $T^{-2}N^{-1}\sum_{t,s,u,v}\sum_{i,j}|\mathrm{cov}(\varepsilon_{is}\varepsilon_{it}, \varepsilon_{ju}\varepsilon_{jv})| < C$ and $T^{-1}N^{-2}\sum_{t,s}\sum_{i,j,k,l}|\mathrm{cov}(\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{ks}\varepsilon_{ls})| < C$.
(C5): ε_{it} is independent of $\boldsymbol{x}_{js}$, $\boldsymbol{\lambda}_{c,i}$, $\boldsymbol{\lambda}_{g,i}$, $\boldsymbol{f}_{c,s}$ and $\boldsymbol{f}_{g,s}$ for all i, j, t, s, g.

Assumption C is used in Bai (2003, 2009) and others. These assumptions permit cross-sectional and serial correlations and heteroskedasticities in the idiosyncratic errors. It can shown that if the ε_{it} are independent and have bounded eighth moment, then this assumption holds.

Assumption D: Observable risk factors

We assume $E\|\boldsymbol{x}_{it}\|^4 < C$. Let the i-th unit belong to the g-th group (i.e., $g_i = g$), let $\boldsymbol{\beta}_{i,\beta_i\neq 0}$ be nonzero elements of the parameter vector of $\boldsymbol{\beta}_i$, and let $X_{i,\beta_i\neq 0}$ be the submatrix of X_i, corresponding to the columns of nonzero elements of the parameter vector $\boldsymbol{\beta}_i$. We use q_i to denote the number of nonzero elements of $\boldsymbol{\beta}_i$. We assume the $q_i \times q_i$ matrix

$$\frac{1}{T}\Big[X'_{i,\beta_i\neq 0} M_{F_c^0,F_g^0} X_{i,\beta_i\neq 0}\Big] \tag{9}$$

is positive definite, where $M_{F_c,F_g} = M_{F_c} - M_{F_c}F_g({F_g}'M_{F_c}F_g)^{-1}{F_g}'M_{F_c} = M_{F_c} - M_{F_c}P_{F_g}M_{F_c}$, $M_{F_c} = I - F_c(F_c'F_c)^{-1}F_c'$, $P_{F_g} = F_g(F_g'F_g)^{-1}F_g'$, and $M_{F_c^0,F_g^0}$ is equal to M_{F_c,F_g} when evaluated at the true common and group-specific factors (F_c^0, F_g^0).

Also, we define

$$C_i = (C_{ci}, C_{gi}), \quad B_i = \begin{pmatrix} B_{ci} & B_{cgi} \\ B'_{cgi} & B_{gi} \end{pmatrix},$$

with $A_i = \frac{1}{T}X_i'M_{F_c,F_g}X_i$, $B_{ci} = (\boldsymbol{\lambda}_{c,i}\boldsymbol{\lambda}'_{c,i}) \otimes I_T$, $B_{gi} = (\boldsymbol{\lambda}_{g_i,i}\boldsymbol{\lambda}'_{g_i,i}) \otimes I_T$, $B_{cgi} = (\boldsymbol{\lambda}_{c,i}\boldsymbol{\lambda}'_{g_i,i}) \otimes I_T$, $C_{ci} = \frac{1}{\sqrt{T}}\boldsymbol{\lambda}'_{c,i} \otimes (X_i'M_{F_c,F_g})$, $C_{gi} = \frac{1}{\sqrt{T}}\boldsymbol{\lambda}'_{g_i,i} \otimes (X_i'M_{F_c,F_g})$. Let $\mathcal{A}$ be the collection of (F_c, F_g) such that $\mathcal{A} = \{(F_c, F_g) : F_c'F_c/T = I,\ F_g'F_g/T = I\}$. We assume

$$\inf_{F_c,F_g\in\mathcal{A}}\Big[\frac{1}{N}\sum_{i;g_i=g} E_i(F_c, F_g)\Big] \quad \text{is positive definite,} \tag{10}$$

where $E_i(F_c, F_g) = B_i - C_i' A_i^- C_i$ and A_i^- is the generalized inverse of A_i.

A few comments are in order for this assumption. First, we assume the matrix in (9) is positive definite. This is a necessary assumption for consistent estimation of the regression coefficients $\boldsymbol{\beta}_i$ even if the factors (F_c^0, F_g^0) are observable. This is the usual rank condition for identification. We do not require the said matrix to be positive definite for all $(F_c, F_g) \in \mathcal{A}$ (it would not be satisfied). Second, in (10) we require the matrix to be positive definite over $\mathcal{A}$. This is used to prove the consistency for the estimates of (F_c^0, F_g^0), which is unknown; $\mathcal{A}$ is the parameter space of the factors. We do not require A_i to be positive definite over $\mathcal{A}$, thus a generalized inverse is used. Note that if $A_i = 0$, it implies that $C_i = 0$, thus $C_i A_i^- C_i$ is well defined, and in this case, $E_i = B_i$. For each i, the matrix E_i is semipositive definite. The summation of E_i over a large number of observations (Assumption E below) should be positive definite. Song (2013) assumes that the matrix in (9) is positive definite for all (F_c, F_g) in our notation (not just for (F_c^0, F_g^0)), which is more difficult to satisfy. Also, he does not consider the regularization problem, nor the co-existence of common and group factors. Our assumption requires a new proof of consistency.

Assumption E: Number of securities

This economy is divided into a finite number G of groups, with gth group containing N_g securities such that $0 < \underline{a} < N_g/N < \bar{a} < 1$, which implies that the number of securities in g-th groups will increase as the entire number of securities N grows.

Assumption F: Central limit theory

Let $X_{i,\beta_i \neq 0}$ be the submatrix of X_i corresponding to columns of nonzero elements of the true parameter vector $\boldsymbol{\beta}_i^0$. We assume the central limit theory

$$\frac{1}{\sqrt{T}} X_{i,\beta_i \neq 0}' M_{F_c^0, F_g^0} \boldsymbol{\varepsilon}_i \rightarrow_d N(\mathbf{0}, J_i(F_c^0, F_g^0)),$$

where $J_i(F_c^0, F_g^0)$ is the probability limit of (as T going to infinity)

$$\frac{1}{T} X_{i,\beta_i \neq 0}' M_{F_c^0, F_g^0} E[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i'] M_{F_c^0, F_g^0} X_{i,\beta_i \neq 0}.$$

This assumption is required for the asymptotic normality of the estimated $\boldsymbol{\beta}_i$.

Assumption G: Let $\Omega_{k\ell} = E[\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_\ell']$. Then, we assume that

$$B_{NT} = \frac{1}{N^2 T} \sum_{k \neq i}^{N} \sum_{\ell \neq i}^{N} X_k' M_{F_c^0, F_g^0} \Omega_{k\ell} M_{F_c^0, F_g^0} X_\ell = o_p(1).$$

This assumption holds trivially if $\boldsymbol{\varepsilon}_{it}$ are i.i.d over i and t because $\Omega_{kk} = \sigma_k^2 I$ and $\Omega_{k\ell} = 0$ $(k \neq \ell)$ and B_{NT} reduces to $B_{NT} = \frac{1}{N^2T}\sum_{k\neq i}^N \sigma_k^2 X_k' M_{F_c^0,F_g^0} X_k = O_p(1/N) = o_p(1)$. Cross-sectional independence (without i.i.d) leads to $\Omega_{k\ell} = 0$ $(k \neq \ell)$, and $B_{NT} = \frac{1}{N^2T}\sum_{k\neq i}^N X_k' M_{F_c^0,F_g^0}\Omega_{kk} M_{F_c^0,F_g^0} X_k$, which can also be shown to be $O_p(1/N)$, thus $o_p(1)$. Assumption G still allows weak cross-sectional dependence and serial correlations.

A2. Proof of Theorem 1

We first consider an alternative expression for the objective function in (3). Concentrating out Λ_c, and substituting $\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i - F_{g_i}\boldsymbol{\lambda}_i$ for $\boldsymbol{w}_{c,i}$, (3) is equal to (ignore the penalty term)

$$\begin{aligned}
&\sum_{g=1}^G \sum_{i;g_i=g} (\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i - F_{g_i}\boldsymbol{\lambda}_{g_i,i})' M_{F_c}(\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i - F_{g_i}\boldsymbol{\lambda}_{g_i,i}) \\
&= \sum_{g=1}^G \operatorname{tr}\left\{(W_{\beta g} - F_g\Lambda_g')' M_{F_c}(W_{\beta g} - F_g\Lambda_g')\right\},
\end{aligned}$$

where $M_{F_c} = I - F_c(F_c'F_c)^{-1}F_c' = I - F_cF_c'/T$. Here, we denote $W_{\beta g}$ as the $T \times N_g$ matrix such that each column consists of $\boldsymbol{w}_i = \boldsymbol{y}_i - X_i\boldsymbol{\beta}_i$ for all i in group g. By further concentrating out Λ_g with $\Lambda_g = W_{\beta g} M_{F_c} F_g(F_g' M_{F_c} F_g)^{-1}$, the above objective function becomes:

$$\begin{aligned}
&\sum_{g=1}^G \operatorname{tr}\left\{W_{\beta g}'(I - M_{F_c}F_g(F_g'M_{F_c}F_g)^{-1}F_g')' M_{F_c}(I - F_g(F_g'M_{F_c}F_g)^{-1}F_g'M_{F_c})W_{\beta g}\right\}, \\
&= \sum_{g=1}^G \operatorname{tr}\left\{W_{\beta g}' M_{F_c} W_{\beta g}\right\} - \operatorname{tr}\left\{W_{\beta g}' M_{F_c} F_g(F_g'M_{F_c}F_g)^{-1}F_g' M_{F_c} W_{\beta g}\right\}, \\
&= \sum_{g=1}^G \operatorname{tr}\left\{W_{\beta g}' M_{F_c,F_g} W_{\beta g}\right\},
\end{aligned}$$

where $M_{F_c,F_g} = M_{F_c} - M_{F_c}F_g(F_g'M_{F_c}F_g)^{-1}F_g'$.

In summary, the true parameters $\{\boldsymbol{\beta}_1^0, ..., \boldsymbol{\beta}_N^0\}$, F_c^0, and $\{F_1^0, ..., F_G^0\}$ are obtained by minimizing the following concentrated (and also centered objective function):

$$\begin{aligned}
&S_{NT}(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_N, F_c, F_1, ..., F_G) \\
= \ &\sum_{g=1}^G \operatorname{tr}\left\{W_{\beta g}' M_{F_c,F_g} W_{\beta g}\right\} + T\sum_{i=1}^N p_{\kappa,\gamma}\left(|\boldsymbol{\beta}_i|\right) - \sum_{g=1}^G \sum_{i;g_i=g} \boldsymbol{\varepsilon}_i' M_{F_c^0,F_g^0}\boldsymbol{\varepsilon}_i \qquad (11)
\end{aligned}$$

The last term is for the purpose of centering, it does not involve unknown parameters. This alternative expression of the objective function is useful for proofing the consistency of the estimated parameters.

The proof extends the that of Bai (2009) to heterogeneous regression coefficients ($\boldsymbol{\beta}_i$ is not restricted to be common) and to be the presence of group-specific factors).

Without loss of generality, we assume that $\boldsymbol{\beta}_i^0 = \mathbf{0}$, $i = 1, ..., N$ (for notational simplicity). Noting that the true data generating process is $\boldsymbol{y}_i = F_c^0 \boldsymbol{\lambda}_{c,i} + F_{g_i}^0 \boldsymbol{\lambda}_{g_i,i} + \boldsymbol{\varepsilon}_i$ ($X_i \boldsymbol{\beta}_i^0 = \mathbf{0}$), we have

$$
\begin{aligned}
& \frac{1}{NT} S_{NT}(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_N, F_c, F_1, ..., F_G) \\
= \;& \frac{1}{NT} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\beta}_i' X_i' M_{F_c,F_g} X_i \boldsymbol{\beta}_i + \frac{2}{NT} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\beta}_i' X_i' M_{F_c,F_g} F_c^0 \boldsymbol{\lambda}_{c,i} \\
& + \frac{2}{NT} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\beta}_i' X_i' M_{F_c,F_g} F_g^0 \boldsymbol{\lambda}_{g,i} + \frac{1}{NT} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\lambda}_{c,i}' F_c^{0\prime} M_{F_c,F_g} F_c^0 \boldsymbol{\lambda}_{c,i} \\
& + \frac{2}{NT} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\lambda}_{c,i}' F_c^{0\prime} M_{F_c,F_g} F_{g_i}^0 \boldsymbol{\lambda}_{g,i} + \frac{1}{NT} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\lambda}_{g_i,i}' F_{g_i}^{0\prime} M_{F_c,F_g} F_{g_i}^0 \boldsymbol{\lambda}_{g,i} \\
& + \frac{2}{NT} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\beta}_i' X_i' M_{F_c,F_g} \boldsymbol{\varepsilon}_i + \frac{2}{NT} \sum_{g=1}^{G} \sum_{i;g_i=g} (F_c^0 \boldsymbol{\lambda}_{c,i})' M_{F_c,F_g} \boldsymbol{\varepsilon}_i \\
& + \frac{1}{NT} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\varepsilon}_i' (M_{F_c,F_g} - M_{F_c^0,F_g^0}) \boldsymbol{\varepsilon}_i + \frac{2}{NT} \sum_{g=1}^{G} \sum_{i;g_i=g} (F_g^0 \boldsymbol{\lambda}_{g,i})' M_{F_c,F_g} \boldsymbol{\varepsilon}_i \\
& + \frac{1}{N} \sum_{i=1}^{N} p_{\kappa,\gamma} \left(|\boldsymbol{\beta}_i| \right),
\end{aligned}
$$

where Λ_{cg} ($N_g \times r$) consists of the factor loadings associated with the common factor (f_c) with respect to the g-th group. We can show that terms involving $\boldsymbol{\varepsilon}_i$ are $o_p(1)$, that is

$$
\begin{aligned}
& \frac{1}{NT} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\beta}_i' X_i' M_{F_c,F_g} \boldsymbol{\varepsilon}_i = o_p(1), \\
& \frac{1}{NT} \sum_{g=1}^{G} \sum_{i;g_i=g} (F_c^0 \boldsymbol{\lambda}_{c,i})' M_{F_c,F_g} \boldsymbol{\varepsilon}_i = o_p(1), \\
& \frac{1}{NT} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\varepsilon}_i' (M_{F_c,F_g} - M_{F_c^0,F_g^0}) \boldsymbol{\varepsilon}_i = o_p(1), \\
& \frac{1}{NT} \sum_{g=1}^{G} \sum_{i;g_i=g} (F_g^0 \boldsymbol{\lambda}_{g,i})' M_{F_c,F_g} \boldsymbol{\varepsilon}_i = o_p(1),
\end{aligned}
$$

where $o_p(1)$ holds uniformly over $\|\beta_i\| \leq C$, and uniformly over F_c and F_g such that $F_c' F_c / T = I_r$ and $F_g' F_g / T = I_{r_g}$. This follows from Lemma A1 of Bai (2009). Thus the first six terms in the S_{NT} dominates the next four terms. Let

$$
\frac{1}{NT} \tilde{S}_{NT}(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_N, F_c, F_1, ..., F_G)
$$

denote the first six terms in $\frac{1}{NT} S_{NT}$. Note that the term $\frac{1}{N} \sum_{i=1}^{N} p_{\kappa,\gamma} \left(|\boldsymbol{\beta}_i| \right)$ is $o_p(1)$ from the assumption on the regularization parameter. We can rewrite

$$
\frac{1}{NT} S_{NT}(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_N, F_c, F_1, ..., F_G) = \frac{1}{NT} \tilde{S}_{NT}(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_N, F_c, F_1, ..., F_G) + o_p(1) \quad (12)
$$

From

$$F_g^{0\prime} M_{F_c^0, F_g^0} = \mathbf{0}$$
$$F_c^{0\prime} M_{F_c^0, F_g^0} = \mathbf{0},$$

we have $\tilde{S}_{NT}(\boldsymbol{\beta}_1^0, ..., \boldsymbol{\beta}_N^0, F_c^0 H_c, F_1^0 H_1, ..., F_G^0 H_G) = 0$ for any $r \times r$ invertible matrix H_c and the $r_g \times r_g$ invertible matrices H_g, $g = 1, , , ., G$.

Introduce

$$A_i = \frac{1}{T} X_i' M_{F_c, F_g} X_i, \quad B_{ci} = (\boldsymbol{\lambda}_{c,i} \boldsymbol{\lambda}_{c,i}') \otimes I_T, \quad B_{gi} = (\boldsymbol{\lambda}_{g_i,i} \boldsymbol{\lambda}_{g_i,i}') \otimes I_T,$$
$$B_{cgi} = (\boldsymbol{\lambda}_{c,i} \boldsymbol{\lambda}_{g_i,i}') \otimes I_T, \quad C_{ci}' = \frac{1}{\sqrt{T}} \boldsymbol{\lambda}_{c,i}' \otimes (X_i' M_{F_c,F_g}), \quad C_{gi}' = \frac{1}{\sqrt{T}} \boldsymbol{\lambda}_{g_i,i}' \otimes (X_i' M_{F_c,F_g}),$$
$$\boldsymbol{\eta}_c = \frac{1}{\sqrt{T}} \mathrm{vec}(M_{F_c,F_g} F_c^0), \quad \boldsymbol{\eta}_g = \frac{1}{\sqrt{T}} \mathrm{vec}(M_{F_c,F_g} F_g^0).$$

Then

$$\begin{aligned}
& \frac{1}{NT} \tilde{S}_{NT}(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_N, F_c, F_1, ..., F_G) \\
= & \frac{1}{N} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\beta}_i' A_i \boldsymbol{\beta}_i + \frac{2}{N} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\beta}_i' C_{ci} \boldsymbol{\eta}_c + \frac{2}{N} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\beta}_i' C_{gi} \boldsymbol{\eta}_g \\
& + \frac{1}{N} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\eta}_c' B_{ci} \boldsymbol{\eta}_c + \frac{2}{N} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\eta}_c' B_{cgi} \boldsymbol{\eta}_g + \frac{1}{N} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\eta}_g' B_{gi} \boldsymbol{\eta}_g
\end{aligned}$$

Completing the square,

$$\begin{aligned}
& \frac{1}{NT} \tilde{S}_{NT}(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_N, F_c, F_1, ..., F_G) \\
= & \frac{1}{N} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\beta}_i' A_i \boldsymbol{\beta}_i + \frac{2}{N} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\beta}_i' C_i \boldsymbol{\eta}_{c,g} + \frac{1}{N} \sum_{g=1}^{G} \sum_{i;g_i=g} \boldsymbol{\eta}_{c,g}' B_i \boldsymbol{\eta}_{c,g} \\
= & \sum_{g=1}^{G} \boldsymbol{\eta}_{c,g}' \left(\frac{1}{N} \sum_{i;g_i=g} E_i \right) \boldsymbol{\eta}_{c,g} + \frac{1}{N} \sum_{g=1}^{G} \sum_{i;g_i=g} \left(\boldsymbol{\beta}_i + A_i^- C_i \boldsymbol{\eta}_{c,g} \right)' A_i \left(\boldsymbol{\beta}_i + A_i^- C_i \boldsymbol{\eta}_{c,g} \right),
\end{aligned}$$

where

$$C_i = (C_{ci}, C_{gi}), \quad B_i = \begin{pmatrix} B_{ci} & B_{cgi} \\ B_{cgi}' & B_{gi} \end{pmatrix}, \quad \boldsymbol{\eta}_{c,g} = (\boldsymbol{\eta}_c', \boldsymbol{\eta}_g')',$$

and $E_i = B_i - C_i' A_i^- C_i$ with A_i^- being a generalized inverse of A_i. Note that even if $A_i = 0$, C_i must be zero because $X_i' M_{F_c,F_g} = 0$. As a result, the term $C_i' A_i^- C_i$ becomes $C_i' A_i^- C_i = 0$. Thus, $C_i' A_i^- C_i$ is well defined even if $A_i = 0$.

Notice that $\tilde{S}_{NT}$ is quadratic in $\boldsymbol{\eta}_{c,g}$ and in $\boldsymbol{\beta}_i + A_i^- C_i \boldsymbol{\eta}_{c,g}$. By Assumption D, $\frac{1}{N} \sum_{i;g_i=g} E_i$ is positive definite and A_i is semipositive definite, it follows that

$$\tilde{S}_{NT}(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_N, F_c, F_1, ..., F_G) \geq 0$$

for all $(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_N, F_c, F_1, ..., F_G)$.

Note that the centered objective function satisfies

$$\frac{1}{NT}S_{NT}(\boldsymbol{\beta}_1^0, ..., \boldsymbol{\beta}_N^0, F_c^0, F_1^0, ..., F_G^0) = \frac{1}{N}\sum_{i=1}^{N} p_{\kappa,\gamma}\left(|\boldsymbol{\beta}_i^0|\right) = o_p(1)$$

here we have used $\frac{1}{NT}\tilde{S}_{NT}(\boldsymbol{\beta}_1^0, ..., \boldsymbol{\beta}_N^0, F_c^0, F_1^0, ..., F_G^0) = 0$, as noted earlier. Note that

$$S_{NT}(\hat{\boldsymbol{\beta}}_1, ..., \hat{\boldsymbol{\beta}}_N, \hat{F}_c, \hat{F}_1, ..., \hat{F}_G) \leq \frac{1}{NT}S_{NT}(\boldsymbol{\beta}_1^0, ..., \boldsymbol{\beta}_N^0, F_c^0, F_1^0, ..., F_G^0) = o_p(1)$$

by definition of $\{\hat{\boldsymbol{\beta}}_1, ..., \hat{\boldsymbol{\beta}}_N, \hat{F}_c, \hat{F}_1, ..., \hat{F}_G\}$. Therefore, we have

$$\begin{aligned} o_p(1) &\geq \frac{1}{NT}S_{NT}(\hat{\boldsymbol{\beta}}_1, ..., \hat{\boldsymbol{\beta}}_N, \hat{F}_c, \hat{F}_1, ..., \hat{F}_G) \\ &= \frac{1}{NT}\tilde{S}_{NT}(\hat{\boldsymbol{\beta}}_1, ..., \hat{\boldsymbol{\beta}}_N, \hat{F}_c, \hat{F}_1, ..., \hat{F}_G) + o_p(1). \end{aligned}$$

where the equality follows from (12). Combined with $\tilde{S}_{NT}(\hat{\boldsymbol{\beta}}_1, ..., \hat{\boldsymbol{\beta}}_N, \hat{F}_c, \hat{F}_1, ..., \hat{F}_G) \geq 0$, it must be true that

$$\frac{1}{NT}\tilde{S}_{NT}(\hat{\boldsymbol{\beta}}_1, ..., \hat{\boldsymbol{\beta}}_N, \hat{F}_c, \hat{F}_1, ..., \hat{F}_G) = o_p(1).$$

This implies that

$$\sum_{g=1}^{G} \hat{\boldsymbol{\eta}}_{c,g}\left(\frac{1}{N}\sum_{i;g_i=g}\hat{E}_i\right)\hat{\boldsymbol{\eta}}_{c,g} = o_p(1), \tag{13}$$

$$\frac{1}{N}\sum_{g=1}^{G}\sum_{i;g_i=g}\left(\hat{\boldsymbol{\beta}}_i + \hat{A}_i^{-}\hat{C}_i\hat{\boldsymbol{\eta}}_{c,g}\right)'\hat{A}_i\left(\hat{\boldsymbol{\beta}}_i + \hat{A}_i^{-}\hat{C}_i\hat{\boldsymbol{\eta}}_{c,g}\right) = o_p(1). \tag{14}$$

where $\hat{\boldsymbol{\eta}}_{c,g}$, $\hat{A}_i$, $\hat{B}_i$, $\hat{C}_i$, and E_i correspond to $\boldsymbol{\eta}_{c,g}$, A_i, B_i, C_i and E_i, all evaluated at the estimates $\{\hat{\boldsymbol{\beta}}_1, ..., \hat{\boldsymbol{\beta}}_N, \hat{F}_c, \hat{F}_1, ..., \hat{F}_G\}$.

From Assumption D, the matrix $N^{-1}\sum_{i;g_i=g}\hat{E}_i$ [note that $\hat{E}_i = E_i(\hat{F}_c, \hat{F}_g)$] is positive definite, and thus the first claim (13) implies that $\|\hat{\boldsymbol{\eta}}_{c,g}\|^2 = o_p(1)$ for $g = 1, ..., G$. That is, we have proved that

$$\frac{1}{T}\|M_{\hat{F}_c,\hat{F}_g}(F_c^0, F_g^0)\|^2 = o_p(1)$$

This result implies that

$$\|M_{\hat{F}_c,\hat{F}_g} - M_{F_c^0,F_g^0}\| = \|P_{\hat{F}_c,\hat{F}_g} - P_{F_c^0,F_g^0}\| = o_p(1) \tag{15}$$

See Bai (2009, page 1265). That is, the space spanned by (F_c^0, F_g^0) and the space spanned by the estimated factors $(\hat{F}_c, \hat{F}_g)$ are asymptotically the same. Because the common factors F_c and the group specific factors are orthogonal (Assumption A), the preceding result implies that

$$\|M_{F_c^0} - M_{\hat{F}_c}\| = o_p(1), \quad \|M_{F_g^0} - M_{\hat{F}_g}\| = o_p(1) \tag{16}$$

We next prove the consistency of $\hat{\boldsymbol{\beta}}_i$. From $\|\hat{\boldsymbol{\eta}}_{c,g}\|^2 = o_p(1)$ for $g = 1, ..., G$, equation (14) implies that

$$\frac{1}{N}\sum_{g=1}^{G}\sum_{i;g_i=g}\hat{\boldsymbol{\beta}}_i'\hat{A}_i\hat{\boldsymbol{\beta}}_i = o_p(1).$$

This implies an average consistency of $\hat{\beta}_i$, but it does not imply individual consistency for each i. We shall use (15) to prove individual consistency. First, note that $\hat{\boldsymbol{\beta}}_i$ satisfies

$$\hat{\boldsymbol{\beta}}_i = \text{argmin}_{\beta_i}\Big[\frac{1}{T}(\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i)'M_{\hat{F}_c,\hat{F}_{g_i}}(\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i) + p_{\kappa,\gamma}(|\boldsymbol{\beta}_i|)\Big] \tag{17}$$

we have used $M_{\hat{F}_c,\hat{F}_g}\hat{F}_c = 0$ and $M_{\hat{F}_c,\hat{F}_g}\hat{F}_{g_i} = 0$. Using (15), we can easily show that

$$\Big|\frac{1}{T}(\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i)'M_{F_c^0,F_{g_i}^0}(\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i) - \frac{1}{T}(\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i)'M_{\hat{F}_c,\hat{F}_{g_i}}(\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i)\Big| = o_p(1). \tag{18}$$

Let $\tilde{\boldsymbol{\beta}}_i$ be the infeasible estimator defined as

$$\begin{aligned}\tilde{\boldsymbol{\beta}}_i &= \text{argmin}_{\beta_i}\Big[\frac{1}{T}(\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i)'M_{F_c^0,F_{g_i}^0}(\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i) + p_{\kappa,\gamma}(|\boldsymbol{\beta}_i|)\Big] \\ &= \text{argmin}_{\beta_i}\Big[\frac{1}{T}(\boldsymbol{y}_i^* - X_i\boldsymbol{\beta}_i)'M_{F_c^0,F_{g_i}^0}(\boldsymbol{y}_i^* - X_i\boldsymbol{\beta}_i) + p_{\kappa,\gamma}(|\boldsymbol{\beta}_i|)\Big],\end{aligned}$$

where $\boldsymbol{y}_i^* = X_i\boldsymbol{\beta}_i^0 + \boldsymbol{\varepsilon}_i$. In view of (18), $\tilde{\boldsymbol{\beta}}_i$ and $\hat{\boldsymbol{\beta}}_i$ are asymptotically equivalent,

$$\|\tilde{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_i\| = o_p(1)$$

It remains to show that $\tilde{\boldsymbol{\beta}}_i$ is consistent. Let

$$R_i(\boldsymbol{\beta}_i) = \frac{1}{T}(\boldsymbol{y}_i^* - X_i\boldsymbol{\beta}_i)'M_{F_c^0,F_g^0}(\boldsymbol{y}_i^* - X_i\boldsymbol{\beta}_i) + p_{\kappa,\gamma}(|\boldsymbol{\beta}_i|)$$

and let $\alpha_{iT} = T^{-1/2} + d_{iT}$ with $d_{iT} = \max\{p'_{\kappa_T,\gamma}(|\beta_{ik}^0|); \beta_{ik}^0 \neq 0\}$. Under $\max\{p''_{\kappa_T,\gamma}(|\beta_{ik}^0|); \beta_{ik}^0 \neq 0\} \to 0$ (which holds for the SCAD considered in this paper), we now show that there exists a local minimizer $\tilde{\boldsymbol{\beta}}_i$ of $R_i(\boldsymbol{\beta}_i)$ such that $\|\tilde{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0\| = O_p(T^{-1/2} + d_{iT})$.

As in Fan and Li (2001), it is enough to show that for any given $e > 0$, there exists a large constant C such that

$$P\left[\min_{\|\boldsymbol{u}\|=C} R_i(\boldsymbol{\beta}_i^0 + \alpha_{iT}\boldsymbol{u}) > R_i(\boldsymbol{\beta}_i^0)\right] \geq 1 - e, \tag{19}$$

because with probability at least $1-e$ that there exists a local minimum in the ball $\{\boldsymbol{\beta}_i^0 + \alpha_{iT}\boldsymbol{u}; \|\boldsymbol{u}\| \le C\}$, which implies that there exists a local minimizer such that $\|\tilde{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0\| = O_p(\alpha_{iT})$. From $p_{\kappa_T,\gamma}(0) = 0$, we have

$$\begin{aligned}
& R_i(\boldsymbol{\beta}_i^0 + \alpha_{iT}\boldsymbol{u}) - R_i(\boldsymbol{\beta}_i^0) \\
\ge\ & \frac{1}{T}(\boldsymbol{y}_i^* - X_i(\boldsymbol{\beta}_i^0 + \alpha_{iT}\boldsymbol{u}))' M_{F_c^0,F_g^0}(\boldsymbol{y}_i^* - X_i(\boldsymbol{\beta}_i^0 + \alpha_{iT}\boldsymbol{u})) + \sum_{k=1}^{q_i} p_{\kappa,\gamma}(|\beta_{ik}^0 + \alpha_{iT}u_k|) \\
& -\frac{1}{T}(\boldsymbol{y}_i^* - X_i\boldsymbol{\beta}_i^0)' M_{F_c^0,F_g^0}(\boldsymbol{y}_i^* - X_i\boldsymbol{\beta}_i^0) - \sum_{k=1}^{q_i} p_{\kappa,\gamma}(|\beta_{ik}^0|)
\end{aligned}$$

where q_i is the number of components of $\boldsymbol{\beta}_i^0$. By using the Taylor expansion, we have

$$\begin{aligned}
& R_i(\boldsymbol{\beta}_i^0 + \alpha_{iT}\boldsymbol{u}) - R_i(\boldsymbol{\beta}_i^0) \\
\ge\ & -2\alpha_{iT}\frac{1}{T}(\boldsymbol{y}_i^* - X_i\boldsymbol{\beta}_i^0)' M_{F_c^0,F_g^0} X_i\boldsymbol{u} + \boldsymbol{u}'\frac{1}{T}\tilde{X}_i' M_{F_c^0,F_g^0}\tilde{X}_i\boldsymbol{u}\alpha_{iT}^2\{1 + o_p(1)\} \\
& + \sum_{k=1}^{q_i}\left[\alpha_{iT} p'_{\kappa,\gamma}(|\beta_{ik}^0|)\mathrm{sgn}(\beta_{ik}^0)u_k + \alpha_{iT}^2 p''_{\kappa,\gamma}(|\beta_{ik}^0|)u_k^2\{1 + o_p(1)\}\right],
\end{aligned}$$

where $\tilde{X}_i$ is $T \times q_i$ matrix that consist of true regressors (with non-zero coefficients), and $\tilde{X}_i' M_{F_c^0,F_g^0}\tilde{X}_i/T$ is a positive definite matrix from Assumption D. From $\frac{1}{T}(\boldsymbol{y}_i^* - X_i\boldsymbol{\beta}_i^0)' M_{F_c^0,F_g^0} X_i = \frac{1}{T}\boldsymbol{\varepsilon}' M_{F_c^0,F_g^0} X_i = O_p(T^{-1/2})$, the first term is on the order $O_p(\alpha_T T^{-1/2})\|\boldsymbol{u}\|$; the second term is on the order $O_p(\alpha_T^2)\|\boldsymbol{u}\|^2$. By choosing a sufficiently large constant C, the second term dominates the first term uniformly in $\|\boldsymbol{u}\| = C$. The third term is bounded by $\sqrt{q_i}\alpha_{iT} d_{iT}\|\boldsymbol{u}\| + \alpha_{iT}^2 \max\{p''_{\kappa_T,\gamma}(|\beta_{ik}^0|); \beta_{ik}^0 \neq 0\}\|\boldsymbol{u}\|^2$, which is also dominated by the second term. In summary, by choosing a sufficiently large constant C, (19) holds. Thus, there exists a local minimizer $\tilde{\boldsymbol{\beta}}_i$ of $R_i(\boldsymbol{\beta}_i)$ such that $\|\tilde{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0\| = O_p(T^{-1/2} + \alpha_{iT}) = o_p(1)$. It is clear that with a proper κ_T, the asymptotical equivalence between $\hat{\boldsymbol{\beta}}_i$ and $\tilde{\boldsymbol{\beta}}_i$ implies the consistency of $\hat{\boldsymbol{\beta}}_i$, i.e., for any $\delta > 0$, we have

$$P\left(\|\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0\| > \delta\right) \to 0, \quad T \to \infty. \tag{20}$$

Using the consistency of $\hat{\boldsymbol{\beta}}_i$ ($\hat{\boldsymbol{\beta}}_i = \boldsymbol{\beta}_i^0 + o_p(1)$), together with (16), we can further show that

$$\frac{1}{\sqrt{T}}\|F_c^0 H_c - \hat{F}_c\| = o_p(1), \quad \text{and} \quad \frac{1}{\sqrt{T}}\|F_g^0 H_g - \hat{F}_g\| = o_p(1), \quad g = 1, ..., G,$$

for some rotation matrices H_c and H_g. The above is slightly stronger than (16). The details are omitted. In fact, later we shall prove an even stronger result, which will include the above as a special case.

A3. Lemma A1

Under assumptions in Theorem 1, we have

$$\frac{1}{\sqrt{N}}\left\|\hat{\Lambda}_c' - H_c^{-1}\Lambda_c{}'\right\| = o_p(1),$$
$$\frac{1}{\sqrt{N_g}}\left\|\hat{\Lambda}_g' - H_g^{-1}\Lambda_g{}'\right\| = o_p(1), \quad g = 1, ..., G.$$

For notational simplicity, in the following proof, we shall use I_r in place of H_c and H_g to avoid carrying these cumbersome notations in many places, especially note that $\hat{\Lambda}_c'\hat{F}_c$ and $\hat{\Lambda}_g'\hat{F}_g$ do not depend on these matrices. Thus we shall write the above in simplified notation:

$$\frac{1}{\sqrt{N}}\left\|\hat{\Lambda}_c' - \Lambda_c{}'\right\| = o_p(1), \quad \frac{1}{\sqrt{N_g}}\left\|\hat{\Lambda}_g' - \Lambda_g{}'\right\| = o_p(1), \quad g = 1, ..., G.$$

Proof of Lemma A1 We use the following facts. $T^{-1}\|X_i\|^2 = T^{-1}\sum_{t=1}^{T}\|\boldsymbol{x}_{it}\|^2 = O_p(1)$, or $T^{-1/2}\|X_i\| = O_p(1)$. Averaging over i, $(TN)^{-1}\sum_{i=1}^{N}\|X_i\|^2 = O_p(1)$. Similarly, $T^{-1/2}\|F_c\| = O_p(1)$, $T^{-1}\|X_i'F_c\| = O_p(1)$, and so forth.

Using the similar argument of Lemma A.10 of Bai (2009), we can express the factor loading estimate for the common factor as $\hat{\Lambda}_c = T^{-1}\hat{F}_c'\hat{W}_c$ where $\hat{W}_c$ is $T \times N$ matrix so that i-th column consists of $\hat{\boldsymbol{w}}_{c,i} = \boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_{g_i}\hat{\boldsymbol{\lambda}}_{g_i,i}$. With respect to the security returns that belong to g-th group, the corresponding factor loading matrix on the common factor, denoted as $\hat{\Lambda}_c^g$, can be expressed as $\hat{\Lambda}_c^g = T^{-1}\hat{F}_c'\hat{W}_c^g$ where $\hat{W}_c^g$ is $T \times N_g$ matrix so that i-th column consists of $\hat{\boldsymbol{w}}_{c,i}$ with each i belong to the g-th group. Thus, we have

$$\hat{W}_c^g = Y^g - X^g\hat{B}^g - \hat{F}_g\hat{\Lambda}_g' = F_c^0\Lambda_c^{g\prime} + E_g + X^g(B^g - \hat{B}^g) + (F_g^0\Lambda_g' - \hat{F}_g\hat{\Lambda}_g'),$$

where Y^g is $T \times N_g$ matrix so that each column consists of the security return vector $\boldsymbol{y}_i$ that belong to the g-th group, and X^g is $T \times N_j \times p_i$ (three-dimensional matrix), so that X^gB^g and $X^g\hat{B}^g$ are $T \times N_j$ matrices so that each column consists of $X_i\boldsymbol{\beta}_i^0$ and $X_i\hat{\boldsymbol{\beta}}_i$, respectively.

From $F_c^0 = (F_c^0 - \hat{F}_c) + \hat{F}_c$ and $\hat{F}_c'\hat{F}_c/T = I$, we have

$$\begin{aligned}
\hat{\Lambda}_c^{g\prime} - \Lambda_c^{g\prime} &= T^{-1}\hat{F}_c'\hat{W}_c^g - \Lambda_c^{g\prime} \\
&= T^{-1}\hat{F}_c'(Y^g - X^g\hat{B}^g - \hat{F}_g\hat{\Lambda}_g') - \Lambda_c^{g\prime} \\
&= T^{-1}\hat{F}_c'\{F_c^0\Lambda_c^{g\prime} + E_g + X^g(B^g - \hat{B}^g) + (F_g^0\Lambda_g' - \hat{F}_g\hat{\Lambda}_g')\} - \Lambda_c^{g\prime} \\
&= T^{-1}\hat{F}_c'(F_c^0 - \hat{F}_c)\Lambda_c^{g\prime} + T^{-1}\hat{F}_c'E_g + T^{-1}\hat{F}_c'X^g(B^g - \hat{B}^g) \\
&\quad + T^{-1}\hat{F}_c'[(F_g^0 - \hat{F}_g)\Lambda_g' + \hat{F}_g(\Lambda_g' - \hat{\Lambda}_g')].
\end{aligned}$$

Using the same argument, we also have

$$\begin{aligned}\hat{\Lambda}_g' - \Lambda_g{}' &= T^{-1}\hat{F}_g'(F_g^0 - \hat{F}_g)\Lambda_g{}' + T^{-1}\hat{F}_g'E_g + T^{-1}\hat{F}_g'X^g(B^g - \hat{B}^g) \\ &\quad + T^{-1}\hat{F}_g'[(F_c^0 - \hat{F}_c)\Lambda_c^{g\prime} + \hat{F}_c(\Lambda_c^{g\prime} - \hat{\Lambda}_c^{g\prime})].\end{aligned}$$

Putting the expression of $\hat{\Lambda}_c^{g\prime} - \Lambda_c^{g\prime}$ into $\hat{\Lambda}_g' - \Lambda_g'$, we have

$$\begin{aligned}\hat{\Lambda}_g' - \Lambda_g{}' &= T^{-1}\hat{F}_g'(F_g^0 - \hat{F}_g)\Lambda_g{}' + T^{-1}\hat{F}_g'E_g + T^{-1}\hat{F}_g'X^g(B^g - \hat{B}^g) + T^{-1}\hat{F}_g'(F_c - \hat{F}_c)\Lambda_c^{g\prime} \\ &\quad + T^{-2}\hat{F}_g'\hat{F}_c\hat{F}_g'(F_g^0 - \hat{F}_g)\Lambda_g{}' + T^{-2}\hat{F}_g'\hat{F}_c\hat{F}_c'E_g + T^{-2}\hat{F}_g'\hat{F}_c\hat{F}_c'X^g(B^g - \hat{B}^g) \\ &\quad + T^{-2}\hat{F}_g'\hat{F}_c\hat{F}_c'(F_g^0 - \hat{F}_g)\Lambda_g' + T^{-2}\hat{F}_g'\hat{F}_c\hat{F}_c'\hat{F}_g(\Lambda_g' - \hat{\Lambda}_g'),\end{aligned}$$

which leads to

$$\begin{aligned}\hat{\Lambda}_g' - \Lambda_g{}' &= (I + T^{-1}\hat{F}_g'(\hat{F}_c\hat{F}_c'/T)\hat{F}_g)^{-1}\Big[T^{-1}\hat{F}_g'(F_g^0 - \hat{F}_g)\Lambda_g{}' + T^{-1}\hat{F}_g'E_g + T^{-1}\hat{F}_g'X^g(B^g - \hat{B}^g) \\ &\quad + T^{-1}\hat{F}_g'(F_c^0 - \hat{F}_c)\Lambda_c^{g\prime} + T^{-2}\hat{F}_g'\hat{F}_c\hat{F}_g'(F_g^0 - \hat{F}_g)\Lambda_g{}' + T^{-2}\hat{F}_g'\hat{F}_c\hat{F}_c'E_g \\ &\quad + T^{-2}\hat{F}_g'\hat{F}_c\hat{F}_c'X^g(B^g - \hat{B}^g) + T^{-2}\hat{F}_g'\hat{F}_c\hat{F}_c'(F_g^0 - \hat{F}_g)\Lambda_g'\Big].\end{aligned}$$

Because $T^{-1}\|F_g'E_g\|^2 = o_p(1)$, $N^{-1}\|B^g - \hat{B}^g\|^2 = o_p(1)$, $T^{-1}\|F_c^0 - \hat{F}_c\|^2 = o_p(1)$, $T^{-1}\|F_g^0 - \hat{F}_g\|^2 = o_p(1)$, $N_g^{-1}\|\Lambda_c^g\|^2 = O_p(1)$, $N_g^{-1}\|\Lambda_g\|^2 = O_p(1)$, and $\|I + T^{-1}\hat{F}_g'(\hat{F}_c\hat{F}_c'/T)\hat{F}_g\| = O_p(1)$, we obtain

$$\frac{1}{\sqrt{N_g}}\left\|\hat{\Lambda}_g' - \Lambda_g{}'\right\| = o_p(1), \quad g = 1, ..., G.$$

Also, putting the expression of $\hat{\Lambda}_g' - \Lambda_g'$ into $\hat{\Lambda}_c^{g\prime} - \Lambda_c^{g\prime}$, we obtain the similar expression of $\hat{\Lambda}_c^{g\prime} - \Lambda_c^{g\prime}$. We then obtain $\frac{1}{\sqrt{N}}\left\|\hat{\Lambda}_c' - \Lambda_c{}'\right\| = o_p(1)$. This completes the proof of Lemma A1.

A3. Lemma A2

Under the assumptions of Theorem 1, we have

$$T^{-1/2}\|\hat{F}_c - F_c^0 H_c\| = \frac{1}{N}\sum_{i=1}^{N} O_p\left(\|\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i\|\right) + O_p\left(\frac{1}{\min\{N^{1/2}, T^{1/2}\}}\right),$$

and

$$T^{-1/2}\|\hat{F}_g - F_g^0 H_g\| = \frac{1}{N}\sum_{i=1}^{N} O_p\left(\|\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i\|\right) + O_p\left(\frac{1}{\min\{N_j^{1/2}, T^{1/2}\}}\right),$$

for $g = 1, ..., G$, Here H_c and H_g are defined as $H_c^{-1} = V_{c,NT}(F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}$, and $H_g^{-1} = V_{g,N_gT}(F_g^{0\prime}\hat{F}_g/T)^{-1}(\Lambda_g'\Lambda_g/N_g)^{-1}$.

Proof of Lemma A2 We first obtained the convergence rate of $\hat{F}_c$. We recall

$$\left[\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}(\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_{g_i,i}\hat{\boldsymbol{\lambda}}_{g_i,i})(\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_{g_i,i}\hat{\boldsymbol{\lambda}}_{g_i,i})'\right]\hat{F}_c = \hat{F}_c V_{c,NT}$$

From $\hat{\boldsymbol{\lambda}}_{g_i,i} = \hat{\boldsymbol{\lambda}}_{g_i,i} - H_{g_i}^{-1}\boldsymbol{\lambda}_{g_i,i} + H_{g_i}^{-1}\boldsymbol{\lambda}_{g_i,i}$, we can show that terms involving $\hat{\boldsymbol{\lambda}}_{g_i,i} - H_{g_i}^{-1}\boldsymbol{\lambda}_{g_i,i}$ are negligible, thus we consider the equation, up to an negligible term,

$$\Bigg[\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\Big\{\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_{g_i,i}(H_{g_i}^{-1}\boldsymbol{\lambda}_{g_i,i}) - \hat{F}_{g_i,i}(\hat{\boldsymbol{\lambda}}_{g_i,i} - H_{g_i}^{-1}\boldsymbol{\lambda}_{g_i,i})\Big\}$$
$$\times\Big\{\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_{g_i,i}(H_{g_i}^{-1}\boldsymbol{\lambda}_{g_i,i}) - \hat{F}_{g_i,i}(\hat{\boldsymbol{\lambda}}_{g_i,i} - H_{g_i}^{-1}\boldsymbol{\lambda}_{g_i,i})\Big\}'\Bigg]\hat{F}_c = \hat{F}_c V_{c,NT}.$$

Hereafter, we consider

$$\left[\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}(\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_{g_i,i}H_{g_i}^{-1}\boldsymbol{\lambda}_{g_i,i})(\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_{g_i,i}H_{g_i}^{-1}\boldsymbol{\lambda}_{g_i,i})'\right]\hat{F}_c = \hat{F}_c V_{c,NT}.$$

Using $\boldsymbol{y}_i = X_i\boldsymbol{\beta}_i^0 + F_c^0\boldsymbol{\lambda}_{c,i} + F_{g_i,i}^0\boldsymbol{\lambda}_{g_i,i} + \boldsymbol{\varepsilon}_i$, we have

$$\begin{aligned}
&\hat{F}_c V_{c,NT}\\
=\ & \frac{1}{NT}\sum_{i=1}^{N}X_i(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)'X_i'\hat{F}_c + \frac{1}{NT}\sum_{i=1}^{N}X_i(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{c,i}F_c^{0\prime}\hat{F}_c\\
&+\frac{1}{NT}\sum_{i=1}^{N}X_i(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)\boldsymbol{\varepsilon}_i'\hat{F}_c + \frac{1}{NT}\sum_{i=1}^{N}F_c^0\boldsymbol{\lambda}_{c,i}(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)'X_i'\hat{F}_c\\
&+\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)'X_i'\hat{F}_c + \frac{1}{NT}\sum_{i=1}^{N}F_c^0\boldsymbol{\lambda}_{c,i}\boldsymbol{\varepsilon}_i'\hat{F}_c + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{\lambda}_{c,i}'F_c^{0\prime}\hat{F}_c\\
&+\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\hat{F}_c + \frac{1}{NT}\sum_{i=1}^{N}F_c^0\boldsymbol{\lambda}_{c,i}\boldsymbol{\lambda}_{c,i}'F_c^{0\prime}\hat{F}_c\\
&+\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{g_i,i}'(F_{g_i}^0 - \hat{F}_{g_i}H_{g_i}^{-1})'\hat{F}_c\\
&+\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}(F_{g_i}^0 - \hat{F}_{g_i}H_{g_i}^{-1})\boldsymbol{\lambda}_{g_i,i}(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)'X_i'\hat{F}_c\\
&+\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}(F_{g_i}^0 - \hat{F}_{g_i}H_{g_i}^{-1})\boldsymbol{\lambda}_{g_i,i}\boldsymbol{\varepsilon}_i'\hat{F}_c + \frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\boldsymbol{\varepsilon}_i\boldsymbol{\lambda}_{g_i,i}'(F_{g_i}^0 - \hat{F}_{g_i}H_{g_i}^{-1})'\hat{F}_c\\
&+\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}(F_{g_i}^0 - \hat{F}_{g_i}H_{g_i}^{-1})\boldsymbol{\lambda}_{g_i,i}\boldsymbol{\lambda}_{g_i,i}'(F_{g_i}^0 - \hat{F}_{g_i}H_{g_i}^{-1})'\hat{F}_c\\
&+\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}(F_{g_i}^0 - \hat{F}_{g_i}H_{g_i}^{-1})\boldsymbol{\lambda}_{g_i,i}\boldsymbol{\lambda}_{c,i}'F_c^{0\prime}\hat{F}_c\\
&+\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}F_c^0\boldsymbol{\lambda}_{c,i}\boldsymbol{\lambda}_{g_i,i}'(F_{g_i}^0 - \hat{F}_{g_i}H_{g_i}^{-1})'\hat{F}_c\\
=\ & I_1 + \cdots + I_{16}.
\end{aligned}$$

Multiplying $(F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}$ on each side of the prior formula, we have

$$\|\hat{F}_c V_{c,NT}(F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1} - F_c^0\|$$

$$
\begin{aligned}
&= \sum_{k=1,k\neq 9}^{16} I_k (F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1} \\
&\leq \sum_{k=1,k\neq 9}^{16} \|I_k\| \times \|(F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\|.
\end{aligned}
$$

We now consider each term on the right.

$$
\|I_1\| \leq \frac{\sqrt{T}}{N}\sum_{i=1}^{N}\left(\frac{\|X_i\|^2}{T}\right)\|\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i\|^2\left(\frac{\|\hat{F}_c\|}{\sqrt{T}}\right) = \frac{\sqrt{T}}{N}\sum_{i=1}^{N} o_p\left(\|\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i\|\right).
$$

Using the same argument, the next four terms satisfy

$$
\|I_k\| \leq \frac{\sqrt{T}}{N}\sum_{i=1}^{N} O_p\left(\|\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i\|\right),
$$

$k = 2, 3, 4, 5$. The next three terms have the same expressions as those in Bai (2009) and thus $T^{-1/2}\|I_k\| \leq O_p(1/\min\{\sqrt{N}, \sqrt{T}\})$, $k = 6, 7, 8$.

Next, we need to evaluate the terms that contain $(F_g^0 - \hat{F}_g H_g^{-1})$. From the definition of $\hat{F}_g$, we have

$$
\hat{F}_g V_{g,N_gT} = \left[\frac{1}{N_gT}\sum_{i;g_i=g}(\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_c\hat{\boldsymbol{\lambda}}_{c,i})(\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_c\hat{\boldsymbol{\lambda}}_{c,i})'\right]\hat{F}_g
$$

where N_g is the number of assets in the g-th group. Because of the result of Lemma A1, we can show that terms involving $\hat{F}_c(\hat{\boldsymbol{\lambda}}_{c,i} - H_c^{-1}\boldsymbol{\lambda}_{c,i})$ are negligible. We consider the following equation, up to an negligible term,

$$
\hat{F}_g V_{g,N_gT} = \left[\frac{1}{N_gT}\sum_{i;g_i=g}(\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_c H_c^{-1}\boldsymbol{\lambda}_{c,i})(\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_c H_c^{-1}\boldsymbol{\lambda}_{c,i})'\right]\hat{F}_g.
$$

From $\boldsymbol{y}_i = X_i\boldsymbol{\beta}_i^0 + F_c^0\boldsymbol{\lambda}_{c,i} + F_{g_i,i}^0\boldsymbol{\lambda}_{g_i,i} + \boldsymbol{\varepsilon}_i$, we have

$$
\begin{aligned}
&\hat{F}_g V_{g,N_gT} \\
= &\frac{1}{N_gT}\sum_{i;g_i=g} X_i(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)'X_i'\hat{F}_g + \frac{1}{N_gT}\sum_{i;g_i=g} X_i(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{g_i,i}F_g^{0\prime}\hat{F}_g \\
&+\frac{1}{N_gT}\sum_{i;g_i=g} X_i(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)\boldsymbol{\varepsilon}_i'\hat{F}_g + \frac{1}{N_gT}\sum_{i;g_i=g} F_g^0\boldsymbol{\lambda}_{g_i,i}(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)'X_i'\hat{F}_g \\
&+\frac{1}{N_gT}\sum_{i;g_i=g} \boldsymbol{\varepsilon}_i(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)'X_i'\hat{F}_g + \frac{1}{N_gT}\sum_{i;g_i=g} F_g^0\boldsymbol{\lambda}_{g_i,i}\boldsymbol{\varepsilon}_i'\hat{F}_g + \frac{1}{N_gT}\sum_{i;g_i=g}\boldsymbol{\varepsilon}_i\boldsymbol{\lambda}_{g_i,i}'F_g^{0\prime}\hat{F}_g \\
&+\frac{1}{N_gT}\sum_{i;g_i=g}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\hat{F}_g + \frac{1}{N_gT}\sum_{i;g_i=g} F_g^0\boldsymbol{\lambda}_{g_i,i}\boldsymbol{\lambda}_{g_i,i}'F_g^{0\prime}\hat{F}_g \\
&+\frac{1}{N_gT}\sum_{i;g_i=g} X_i(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{c,i}'(F_c^0 - \hat{F}_c H_c^{-1})'\hat{F}_g
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{N_gT}\sum_{i;g_i=g}(F_c^0-\hat{F}_cH_c^{-1})\boldsymbol{\lambda}_{c,i}(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)'X_i'\hat{F}_g\\
&+\frac{1}{N_gT}\sum_{i;g_i=g}(F_c^0-\hat{F}_cH_c^{-1})\boldsymbol{\lambda}_{c,i}\boldsymbol{\varepsilon}_i'\hat{F}_g+\frac{1}{N_gT}\sum_{i;g_i=g}\boldsymbol{\varepsilon}_i\boldsymbol{\lambda}_{c,i}'(F_c^0-\hat{F}_cH_c^{-1})'\hat{F}_g\\
&+\frac{1}{N_gT}\sum_{i;g_i=g}(F_c^0-\hat{F}_cH_c^{-1})\boldsymbol{\lambda}_{c,i}\boldsymbol{\lambda}_{c,i}'(F_c^0-\hat{F}_cH_c^{-1})'\hat{F}_g\\
&+\frac{1}{N_gT}\sum_{i;g_i=g}(F_c^0-\hat{F}_cH_c^{-1})\boldsymbol{\lambda}_{c,i}\boldsymbol{\lambda}_{g_i,i}'F_g^{0\prime}\hat{F}_g\\
&+\frac{1}{N_gT}\sum_{i;g_i=g}F_g^0\boldsymbol{\lambda}_{g_i,i}\boldsymbol{\lambda}_{c,i}'(F_c^0-\hat{F}_cH_c^{-1})'\hat{F}_g\\
=\;& I_1^g+\cdots+I_{16}^g.
\end{aligned}
$$

Multiplying $(F_g^{0\prime}\hat{F}_g/T)^{-1}(\Lambda_g'\Lambda_g/N_g)^{-1}$ on each side of the prior formula, we have

$$
\hat{F}_gH_g^{-1}-F_g^0=\sum_{k=1,k\neq 9}^{16}I_k^g(F_g^{0\prime}\hat{F}_g/T)^{-1}(\Lambda_g'\Lambda_g/N_g)^{-1},
$$

where $H_g^{-1}=V_{g,N_gT}(F_g^{0\prime}\hat{F}_g/T)^{-1}(\Lambda_g'\Lambda_g/N_g)^{-1}$.

Putting this expression into I_{10}, we have

$$
\begin{aligned}
I_{10} &= \frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{g_i,i}'\left[\sum_{k=1,k\neq 9}^{16}I_k^g(F_g^{0\prime}\hat{F}_g/T)^{-1}(\Lambda_g'\Lambda_g/N_g)^{-1}\right]'\hat{F}_c\\
&= \frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\sum_{k=1,k\neq 9}^{16}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{g_i,i}'(I_k^gG_g)'\hat{F}_c,
\end{aligned}
$$

where we denote $(F_g^{0\prime}\hat{F}_g/T)^{-1}(\Lambda_g'\Lambda_g/N_g)^{-1}$ as G_g. We then evaluate each of the terms in I_{10}.

$$
\begin{aligned}
&T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{g_i,i}'(I_1^gG_g)'\hat{F}_c\Big\|\\
=\;& T^{-1/2}\left\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{g_i,i}'\left[\frac{1}{N_gT}\sum_{\ell;g_\ell=g}X_\ell(\boldsymbol{\beta}_\ell-\hat{\boldsymbol{\beta}}_\ell)(\boldsymbol{\beta}_\ell-\hat{\boldsymbol{\beta}}_\ell)'X_\ell'\hat{F}_gG_g\right]'\hat{F}_c\right\|\\
\leq\;& \sum_{g=1}^{G}\sum_{i;g_i=g}\frac{1}{N_g}\times o_p(\|\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i\|).
\end{aligned}
$$

$$
\begin{aligned}
&T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{g_i,i}'(I_2^gG_g)'\hat{F}_c\Big\|\\
=\;& T^{-1/2}\left\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{g_i,i}'\left[\frac{1}{N_gT}\sum_{\ell;g_\ell=g}X_\ell(\boldsymbol{\beta}_\ell-\hat{\boldsymbol{\beta}}_\ell)\boldsymbol{\lambda}_{g,\ell}F_g^{0\prime}\hat{F}_gG_g\right]'\hat{F}_c\right\|\\
\leq\;& \sum_{g=1}^{G}\sum_{i;g_i=g}\frac{1}{N_g}\times o_p(\|\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i\|).
\end{aligned}
$$

$$
\begin{aligned}
& T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}(I_3^gG_g)'\hat{F}_c\Big\| \\
= \ & T^{-1/2}\left\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}\left[\frac{1}{N_gT}\sum_{\ell;g_\ell=g}X_\ell(\boldsymbol{\beta}_\ell-\hat{\boldsymbol{\beta}}_\ell)\boldsymbol{\varepsilon}'_\ell\hat{F}_gG_g\right]'\hat{F}_c\right\| \\
\leq \ & \sum_{g=1}^{G}\sum_{i;g_i=g}\frac{1}{N_g}\times o_p(\|\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i\|).
\end{aligned}
$$

$$
\begin{aligned}
& T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}(I_4^gG_g)'\hat{F}_c\Big\| \\
= \ & T^{-1/2}\left\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}\left[\frac{1}{N_gT}\sum_{\ell;g_\ell=g}F_g^0\boldsymbol{\lambda}_{g,\ell}(\boldsymbol{\beta}_\ell-\hat{\boldsymbol{\beta}}_\ell)'X'_\ell\hat{F}_gG_g\right]'\hat{F}_c\right\| \\
\leq \ & \sum_{g=1}^{G}\sum_{i;g_i=g}\frac{1}{N_g}\times o_p(\|\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i\|).
\end{aligned}
$$

$$
\begin{aligned}
& T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}(I_5^gG_g)'\hat{F}_c\Big\| \\
= \ & T^{-1/2}\left\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}\left[\frac{1}{N_gT}\sum_{\ell;g_\ell=g}\boldsymbol{\varepsilon}_\ell(\boldsymbol{\beta}_\ell-\hat{\boldsymbol{\beta}}_\ell)'X'_\ell\hat{F}_gG_g\right]'\hat{F}_c\right\| \\
\leq \ & \sum_{g=1}^{G}\sum_{i;g_i=g}\frac{1}{N_g}\times o_p(\|\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i\|).
\end{aligned}
$$

$$
\begin{aligned}
& T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}(I_6^gG_g)'\hat{F}_c\Big\| \\
= \ & T^{-1/2}\left\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}\left[\frac{1}{N_gT}\sum_{\ell;g_\ell=g}F_g^0\boldsymbol{\lambda}_{g,\ell}\boldsymbol{\varepsilon}'_\ell\hat{F}_gG_g\right]'\hat{F}_c\right\| \\
\leq \ & \sum_{g=1}^{G}\sum_{i;g_i=g}\frac{1}{N_g}\times O_p(\|\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i\|).
\end{aligned}
$$

$$
\begin{aligned}
& T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}(I_7^gG_g)'\hat{F}_c\Big\| \\
= \ & T^{-1/2}\left\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}\left[\frac{1}{N_gT}\sum_{\ell;g_\ell=g}\boldsymbol{\varepsilon}_\ell\boldsymbol{\lambda}'_{g,\ell}F_g^{0\prime}\hat{F}_gG_g\right]'\hat{F}_c\right\| \\
\leq \ & \sum_{g=1}^{G}\sum_{i;g_i=g}\frac{1}{N_g}\times O_p(\|\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i\|).
\end{aligned}
$$

$$
\begin{aligned}
& T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}(I_8^gG_g)'\hat{F}_c\Big\| \\
= \;& T^{-1/2}\Bigg\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}\left[\frac{1}{N_gT}\sum_{\ell;g_\ell=g}\boldsymbol{\varepsilon}_\ell\boldsymbol{\varepsilon}'_\ell\hat{F}_gG_g\right]'\hat{F}_c\Bigg\| \\
\leq \;& \sum_{g=1}^{G}\sum_{i;g_i=g}\frac{1}{N_g}\times O_p(\|\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i\|).
\end{aligned}
$$

$$
\begin{aligned}
& T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}(I_{10}^gG_g)'\hat{F}_c\Big\| \\
= \;& T^{-1/2}\Bigg\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}\left[\frac{1}{N_gT}\sum_{\ell;g_\ell=g}X_\ell(\boldsymbol{\beta}_\ell-\hat{\boldsymbol{\beta}}_\ell)\boldsymbol{\lambda}'_{c,\ell}(F_c^0-\hat{F}_cH_c^{-1})'\hat{F}_gG_g\right]'\hat{F}_c\Bigg\| \\
\leq \;& o_p(1)\times\frac{\|F_c^0-\hat{F}_cH_c^{-1}\|}{\sqrt{T}}
\end{aligned}
$$

$$
\begin{aligned}
& T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}(I_{11}^gG_g)'\hat{F}_c\Big\| \\
= \;& T^{-1/2}\Bigg\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}\left[\frac{1}{N_gT}\sum_{\ell;g_\ell=g}(F_c^0-\hat{F}_cH_c^{-1})\boldsymbol{\lambda}_{c,\ell}(\boldsymbol{\beta}_\ell-\hat{\boldsymbol{\beta}}_\ell)'X'_\ell\hat{F}_gG_g\right]'\hat{F}_c\Bigg\| \\
\leq \;& o_p(1)\times\frac{\|F_c^0-\hat{F}_cH_c^{-1}\|}{\sqrt{T}}.
\end{aligned}
$$

$$
\begin{aligned}
& T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}(I_{12}^gG_g)'\hat{F}_c\Big\| \\
= \;& T^{-1/2}\Bigg\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}\left[\frac{1}{N_gT}\sum_{\ell;g_\ell=g}(F_c^0-\hat{F}_cH_c^{-1})\boldsymbol{\lambda}_{c,\ell}\boldsymbol{\varepsilon}'_\ell\hat{F}_gG_g\right]'\hat{F}_c\Bigg\| \\
\leq \;& o_p(1)\times\frac{\|F_c^0-\hat{F}_cH_c^{-1}\|}{\sqrt{T}}.
\end{aligned}
$$

$$
\begin{aligned}
& T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}(I_{13}^gG_g)'\hat{F}_c\Big\| \\
= \;& T^{-1/2}\Bigg\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}\left[\frac{1}{N_gT}\sum_{\ell\in S_j}\boldsymbol{\varepsilon}_\ell\boldsymbol{\lambda}'_{G,\ell}(F_c^0-\hat{F}_cH_c^{-1})'\hat{F}_gG_g\right]'\hat{F}_c\Bigg\| \\
\leq \;& o_p(1)\times\frac{\|F_c^0-\hat{F}_cH_c^{-1}\|}{\sqrt{T}}.
\end{aligned}
$$

$$
T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}'_{g_i,i}(I_{14}^gG_g)'\hat{F}_c\Big\|
$$

$$
\begin{aligned}
&= T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{g_i,i}'\Big[\frac{1}{N_gT}\sum_{\ell;g_\ell=g}(F_c^0-\hat{F}_cH_c^{-1})\boldsymbol{\lambda}_{c,\ell}\boldsymbol{\lambda}_{c,\ell}'(F_c^0-\hat{F}_cH_c^{-1})'\hat{F}_gG_g\Big]'\hat{F}_c\Big\| \\
&\leq o_p(1)\times\frac{\|F_c^0-\hat{F}_cH_c^{-1}\|^2}{T}.
\end{aligned}
$$

$$
\begin{aligned}
& T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{g_i,i}'(I_{15}^gG_g)'\hat{F}_c\Big\| \\
&= T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{g_i,i}'\left[\frac{1}{N_gT}\sum_{\ell;g_\ell=g}(F_c^0-\hat{F}_cH_c^{-1})\boldsymbol{\lambda}_{c,\ell}\boldsymbol{\lambda}_{g,\ell}'F_g^{0\prime}\hat{F}_gG_g\right]'\hat{F}_c\Big\| \\
&\leq o_p(1)\times\frac{\|F_c^0-\hat{F}_cH_c^{-1}\|}{\sqrt{T}}.
\end{aligned}
$$

The final term is

$$
\begin{aligned}
& T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{g_i,i}'(I_{16}^gG_g)'\hat{F}_c\Big\| \\
&= T^{-1/2}\Big\|\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}X_i(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{g_i,i}'\left[\frac{1}{N_gT}\sum_{\ell;g_\ell=g}F_g^0\boldsymbol{\lambda}_{j,\ell}\boldsymbol{\lambda}_{c,\ell}'(F_c^0-\hat{F}_cH_c^{-1})'\hat{F}_gG_g\right]'\hat{F}_c\Big\| \\
&\leq o_p(1)\times\frac{\|F_c^0-\hat{F}_cH_c^{-1}\|}{\sqrt{T}}.
\end{aligned}
$$

Summarizing these evaluations, we finally have

$$
T^{-1/2}\|I_{10}\|\leq\sum_{g=1}^{G}\sum_{i;g_i=g}\frac{1}{N_g}\times O_p(\|\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i\|)+o_p(1)\times\frac{\|F_c^0-\hat{F}_cH_c^{-1}\|}{\sqrt{T}}.
$$

The terms $I_{11}\sim I_{16}$ can be evaluated in a similar manner and

$$
T^{-1/2}\|I_k\|\leq\sum_{g=1}^{G}\sum_{i;g_i=g}\frac{1}{N_g}\times O_p(\|\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i\|)+o_p(1)\times\frac{\|F_c^0-\hat{F}_cH_c^{-1}\|}{\sqrt{T}},
$$

for $k=11,...,16$. Finally, we have

$$
\begin{aligned}
& T^{-1/2}\|\hat{F}_cV_{c,NT}(F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}-F_c^0\| \\
&= \frac{1}{N}\sum_{i=1}^{N}O_p\left(\|\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i\|\right)+O_p\left(\frac{1}{\min\{N^{1/2},T^{1/2}\}}\right)+o_p(1)\times O_p\left(\left\|\frac{F_c^0-\hat{F}_cH_c^{-1}}{\sqrt{T}}\right\|\right)
\end{aligned}
$$

This implies the claim

$$
T^{-1/2}\|\hat{F}_c-F_c^0H_c\|=\frac{1}{N}\sum_{i=1}^{N}O_p\left(\|\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i\|\right)+O_p\left(\frac{1}{\min\{N^{1/2},T^{1/2}\}}\right),
$$

The proof for the second part of the lemma concerning $\hat{F}_g$ is similar. This completes the proof.

A4. Lemma A3

Define $E[\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k'] = \Omega_k$. Under assumptions in Theorem 1, we have

$$
\begin{aligned}
& \frac{1}{NT^2} \sum_{k=1}^{N} X_i' M_{\hat{F}_c,\hat{F}_g} (\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k' - \Omega_k) \hat{F}_c (F_c^{0\prime} \hat{F}_c/T)^{-1} (\Lambda_c' \Lambda_c/N)^{-1} \boldsymbol{\lambda}_{c,i} \\
= \; & O_p\left(\frac{1}{T\sqrt{N}}\right) + \frac{1}{\sqrt{NT}} \left[\frac{1}{N} \sum_{k=1}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right) + O_p\left(\frac{1}{\min\{N^{1/2}, T^{1/2}\}}\right)\right] \\
& + \frac{1}{\sqrt{N}} \times \left[\frac{1}{N} \sum_{i=1}^{N} O_p\left(\|\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i\|\right) + O_p\left(\frac{1}{\min\{N, T\}}\right)\right].
\end{aligned}
$$

Proof: Using $M_{F_c,F_g} = M_{F_c} - M_{F_c} F_g (F_g' M_{F_c} F_g)^{-1} F_g' M_{F_c}$, we rewrite the term as

$$
\begin{aligned}
& \frac{1}{NT^2} \sum_{k=1}^{N} X_i' M_{\hat{F}_c,\hat{F}_g} (\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k' - \Omega_k) \hat{F}_c (F_c^{0\prime} \hat{F}_c/T)^{-1} (\Lambda_c' \Lambda_c/N)^{-1} \boldsymbol{\lambda}_{c,i} \\
= \; & \frac{1}{NT^2} \sum_{k=1}^{N} X_i' M_{\hat{F}_c} (\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k' - \Omega_k) \hat{F}_c (F_c^{0\prime} \hat{F}_c/T)^{-1} (\Lambda_c' \Lambda_c/N)^{-1} \boldsymbol{\lambda}_{c,i} \\
- \; & \frac{1}{NT^2} \sum_{k=1}^{N} X_i' K_{\hat{F}_c,\hat{F}_g} (\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k' - \Omega_k) \hat{F}_c (F_c^{0\prime} \hat{F}_c/T)^{-1} (\Lambda_c' \Lambda_c/N)^{-1} \boldsymbol{\lambda}_{c,i} \\
= \; & I + II,
\end{aligned}
$$

where $K_{F_c,F_g} = M_{F_c} F_g (F_g' M_{F_c} F_g)^{-1} F_g' M_{F_c}$. The first term is written as

$$
\begin{aligned}
I \; = \; & \frac{1}{NT^2} \sum_{k=1}^{N} X_i' (\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k' - \Omega_k) \hat{F}_c (F_c^{0\prime} \hat{F}_c/T)^{-1} (\Lambda_c' \Lambda_c/N)^{-1} \boldsymbol{\lambda}_{c,i} \\
- \; & \frac{1}{NT^2} \sum_{k=1}^{N} X_i' (\hat{F}_c \hat{F}_c'/T) (\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k' - \Omega_k) \hat{F}_c (F_c^{0\prime} \hat{F}_c/T)^{-1} (\Lambda_c' \Lambda_c/N)^{-1} \boldsymbol{\lambda}_{c,i} \\
= \; & I' + I''.
\end{aligned}
$$

Adding and subtracting terms yields

$$
\begin{aligned}
I' \; = \; & \frac{1}{NT^2} \sum_{k=1}^{N} X_i' (\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k' - \Omega_k) F_c^0 H_c (F_c^{0\prime} \hat{F}_c/T)^{-1} (\Lambda_c' \Lambda_c/N)^{-1} \boldsymbol{\lambda}_{c,i} \\
+ \; & \frac{1}{NT^2} \sum_{k=1}^{N} X_i' (\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k' - \Omega_k) (\hat{F}_c - F_c^0 H_c) (F_c^{0\prime} \hat{F}_c/T)^{-1} (\Lambda_c' \Lambda_c/N)^{-1} \boldsymbol{\lambda}_{c,i}
\end{aligned}
$$

The first term on the right is equal to

$$
\begin{aligned}
& \frac{1}{NT^2} \sum_{k=1}^{N} \left\{\sum_{t=1}^{T} \sum_{s=1}^{T} \boldsymbol{x}_{it} (\varepsilon_{kt} \varepsilon_{ks} - E[\varepsilon_{kt} \varepsilon_{ks}]) \boldsymbol{f}_{c,s}^{0\ \prime}\right\} H_c (F_c^{0\prime} \hat{F}_c/T)^{-1} (\Lambda_c' \Lambda_c/N)^{-1} \boldsymbol{\lambda}_{c,i} \\
= \; & \frac{1}{\sqrt{NT}} \times \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \left\{\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \boldsymbol{x}_{it} (\varepsilon_{kt} \varepsilon_{ks} - E[\varepsilon_{kt} \varepsilon_{ks}]) \boldsymbol{f}_{c,s}^{0\ \prime}\right\} H_c (F_c^{0\prime} \hat{F}_c/T)^{-1} (\Lambda_c' \Lambda_c/N)^{-1} \boldsymbol{\lambda}_{c,i} \\
= \; & O_p\left(\frac{1}{\sqrt{NT}}\right),
\end{aligned}
$$

which follows by Lemma A.2(ii) in Bai (2009). Denote

$$a_s = \frac{1}{\sqrt{NT}} \sum_{k=1}^{N} \sum_{t=1}^{T} \boldsymbol{x}_{it} (\varepsilon_{kt}\varepsilon_{ks} - E[\varepsilon_{kt}\varepsilon_{ks}]) = O_p(1),$$

the second term of I' is

$$\begin{aligned} & \frac{1}{\sqrt{NT}} \left\{ \frac{1}{T} \sum_{s=1}^{T} a_s (\hat{\boldsymbol{f}}_{c,s} - \boldsymbol{f}_{c,s}^0 H_c)' \right\} (F_c^{0\prime} \hat{F}_c / T)^{-1} (\Lambda_c' \Lambda_c / N)^{-1} \boldsymbol{\lambda}_{c,i} \\ = & \frac{1}{\sqrt{NT}} \times \left[\frac{1}{N} \sum_{i=1}^{N} O_p\left(\|\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i\| \right) + O_p\left(\frac{1}{\min\{N^{1/2}, T^{1/2}\}} \right) \right]. \end{aligned}$$

We next consider I''.

$$\begin{aligned} I'' & \leq \frac{\|X_i' \hat{F}_c\|}{T} \|(F_c^{0\prime} \hat{F}_c / T)^{-1} (\Lambda_c' \Lambda_c / N)^{-1} \boldsymbol{\lambda}_{c,i}\| \left\| \frac{1}{NT^2} \sum_{k=1}^{N} \hat{F}_c' (\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k' - \Omega_k) \hat{F}_c \right\| \\ & = O_p(1) \times \left\| \frac{1}{NT^2} \sum_{k=1}^{N} \hat{F}_c' (\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k' - \Omega_k) \hat{F}_c \right\|. \end{aligned}$$

Using the result of Lemma A.5 in Bai (2009), this term becomes

$$\begin{aligned} I'' & \leq O_p\left(\frac{1}{T\sqrt{N}} \right) + \frac{1}{\sqrt{NT}} \left[\frac{1}{N} \sum_{k=1}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\| \right) + O_p\left(\frac{1}{\min\{N^{1/2}, T^{1/2}\}} \right) \right] \\ & \quad + \frac{1}{\sqrt{N}} \times \left[\frac{1}{N} \sum_{i=1}^{N} O_p\left(\|\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i\| \right) + O_p\left(\frac{1}{\min\{N, T\}} \right) \right]. \end{aligned}$$

In a similar manner, the term II is written as

$$\begin{aligned} II & \leq \frac{\|X_i' K_{F_c, F_g}\|}{T} \|(F_c^{0\prime} \hat{F}_c / T)^{-1} (\Lambda_c' \Lambda_c / N)^{-1} \boldsymbol{\lambda}_{c,i}\| \left\| \frac{1}{NT^2} \sum_{k=1}^{N} \hat{F}_c' (\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k' - \Omega_k) \hat{F}_c \right\| \\ & = O_p(1) \times \left\| \frac{1}{NT^2} \sum_{k=1}^{N} \hat{F}_c' (\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k' - \Omega_k) \hat{F}_c \right\| \\ & \leq O_p\left(\frac{1}{T\sqrt{N}} \right) + \frac{1}{\sqrt{NT}} \left[\frac{1}{N} \sum_{k=1}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\| \right) + O_p\left(\frac{1}{\min\{N^{1/2}, T^{1/2}\}} \right) \right] \\ & \quad + \frac{1}{\sqrt{N}} \times \left[\frac{1}{N} \sum_{i=1}^{N} O_p\left(\|\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i\| \right) + O_p\left(\frac{1}{\min\{N, T\}} \right) \right]. \end{aligned}$$

Combining these results, we obtain the result.

A4. Lemma A4

Under the assumptions of Theorem 1, we have

$$\begin{aligned} & \frac{1}{NT} \sum_{g=1}^{G} \sum_{k; g_k = g} \boldsymbol{\lambda}_{g,k} \boldsymbol{\varepsilon}_k' \hat{F}_c \\ & = \frac{1}{N^{3/2}} \sum_{k=1}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\| \right) + \frac{1}{N^{1/2}} O_p\left(\frac{1}{\min\{N^{1/2}, T^{1/2}\}} \right). \end{aligned}$$

Proof: Because $\hat{F}_c = F_c^0 H_c + (\hat{F}_c - F_c^0 H_c)$, we have

$$\begin{aligned}
&\frac{1}{NT}\sum_{g=1}^{G}\sum_{k;g_k=g} \boldsymbol{\lambda}_{g,k}\boldsymbol{\varepsilon}_k'\hat{F}_c \\
&= \frac{1}{NT}\sum_{k=1}^{N} \boldsymbol{\lambda}_{c,k}\boldsymbol{\varepsilon}_k' F_c^0 H_c + \frac{1}{NT}\sum_{k=1}^{N} \boldsymbol{\lambda}_{c,k}\boldsymbol{\varepsilon}_k'(\hat{F}_c - F_c^0 H_c).
\end{aligned}$$

The first term is $\frac{1}{NT}\sum_{k=1}^{N} \boldsymbol{\lambda}_{c,k}\boldsymbol{\varepsilon}_k' F_c^0 H_c = O_p(1/\sqrt{NT})$. The second term can be evaluated using the result of Lemma A2. Thus, we obtain the claim.

A5. Proof of Theorem 2

First, for notational simplicity, we denote the non-zero element of the true parameter $\boldsymbol{\beta}_{i10}$ as $\boldsymbol{\beta}_i$, and the corresponding sub-matrix $X_{i,\beta_i \neq 0}$ of X_i as X_i. Suppose that i belong to the gth group. An alternative expression for the solution of the regression coefficients of $\boldsymbol{\beta}_i$ is

$$\hat{\boldsymbol{\beta}}_i(M_{\hat{F}_c,\hat{F}_g}) = \left(X_i' M_{\hat{F}_c,\hat{F}_g} X_i + \Sigma_i(\kappa)\right)^{-1} X_i' M_{\hat{F}_c,\hat{F}_g}\boldsymbol{y}_i,$$

where $M_{\hat{F}_c,\hat{F}_g}$ is defined in Section 3, $\Sigma(\kappa) = \mathrm{diag}\{p'_{\kappa,\gamma}(|\hat{\beta}_{i1}|)/|\hat{\beta}_{i1}|, ..., p'_{\kappa,\gamma}(|\hat{\beta}_{iq_i}|)/|\hat{\beta}_{iq_i}|\}$ is defined in Theorem 2.

Noting that $\boldsymbol{y}_i = X_i\boldsymbol{\beta}_i + F_c^0\boldsymbol{\lambda}_{c,i} + F_{g_i}^0\boldsymbol{\lambda}_{g_i,i} + \boldsymbol{\varepsilon}_i$, we have

$$\begin{aligned}
&\frac{1}{T}\left(X_i' M_{\hat{F}_c,\hat{F}_g} X_i + \Sigma_i(\kappa)\right)(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) + \frac{1}{T}\Sigma_i(\kappa)\boldsymbol{\beta}_i \\
&= \frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g} F_c^0\boldsymbol{\lambda}_{c,i} + \frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g} F_g^0\boldsymbol{\lambda}_{g_i,i} + \frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g}\boldsymbol{\varepsilon}_i
\end{aligned}$$

Using the result in the proof of Theorem 1,

$$F_c^0 = \hat{F}_c H_c^{-1} - \left[\sum_{k=1,k\neq 9}^{16} I_k\right](F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}$$

we have

$$\frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g} F_c^0\boldsymbol{\lambda}_{c,i} = -\frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g}\left[\sum_{k=1,k\neq 9}^{16} I_k\right](F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i},$$

where we used $M_{\hat{F}_c,\hat{F}_g}\hat{F}_c H_c^{-1} = 0$. We next examine each of the components in the right hand side of the equation.

$$\begin{aligned}
&\frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g} I_1 (F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= \frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g}\left[\frac{1}{NT}\sum_{k=1}^{N} X_k(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k)(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k)' X_k'\hat{F}\right](F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= o_p(1)\times(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) + \frac{1}{N}\sum_{k\neq i}^{N} o_p(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)
\end{aligned}$$

where we used $\|X_i M_{\hat{F}_c,\hat{F}_g}\|/T^{1/2} = O_p(1)$, $\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| = o_p(1)$, $\|(F_c^{0\prime}\hat{F}_c/T)^{-1}\| = O_p(1)$, $\|(\Lambda_c'\Lambda_c/N)^{-1}\| = O_p(1)$, Note that the second term is $\frac{1}{N}\sum_{k\neq i}^{N} o_p(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) = o_p(T^{-1/2})$, which will be shown later. Next, we have

$$
\begin{aligned}
&\frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g} I_2 (F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= \frac{1}{NT}X_i' M_{\hat{F}_c,\hat{F}_g}\left[\sum_{k=1}^{N} X_k(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k)\boldsymbol{\lambda}_{c,k}'(\Lambda_c'\Lambda_c/N)^{-1}\right]\boldsymbol{\lambda}_{c,i} \\
&= \frac{1}{NT}\sum_{k=1}^{N} X_i' M_{\hat{F}_c,\hat{F}_g} X_k \boldsymbol{\lambda}_{c,k}'(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i}(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k) \\
&= \frac{1}{NT}X_i' M_{\hat{F}_c,\hat{F}_g} X_i \boldsymbol{\lambda}_{c,i}'(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i}(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i) \\
&\quad + \frac{1}{NT}\sum_{k\neq i}^{N} X_i' M_{\hat{F}_c,\hat{F}_g} X_k \boldsymbol{\lambda}_{c,k}'(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i}(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k).
\end{aligned}
$$

The second term in the last line is $\frac{1}{NT}\sum_{k\neq i}^{N} X_i' M_{\hat{F}_c,\hat{F}_g} X_k \boldsymbol{\lambda}_{c,k}'(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i}(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k) = o_p(T^{-1/2})$, which will be shown later. The third term can be evaluated as

$$
\begin{aligned}
&\frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g} I_3 (F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
= \;& \frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g}\left[\frac{1}{NT}\sum_{k=1}^{N} X_k(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k)\boldsymbol{\varepsilon}_k'\hat{F}_c\right](F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
= \;& \frac{1}{N}\sum_{k=1}^{N}\left(\frac{X_i' M_{\hat{F}_c,\hat{F}_g} X_k}{T}\right)(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k)\left(\frac{\boldsymbol{\varepsilon}_k'\hat{F}_c}{T}\right)(F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
= \;& \left[\frac{1}{N}\sum_{k\neq i}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right) + O_p\left(\frac{1}{\min\{N_1^{1/2}, T^{1/2}\}}\right)\right] \\
& \times \left[\frac{1}{N}\sum_{k\neq i}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right)\right] + o_p(1)\times O_p(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i),
\end{aligned}
$$

where we used the following relation $\boldsymbol{\varepsilon}_k'\hat{F}_c/T = \boldsymbol{\varepsilon}_k' F_c^0 H_c/T + \boldsymbol{\varepsilon}_k'(\hat{F}_c - F_c^0 H_c)/T = O_p(1/\sqrt{T}) + N^{-1}\sum_{i=1}^{N} O_p\left(\|\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i\|\right) + O_p(1/\min\{N^{1/2}, T^{1/2}\})$. Using $M_{\hat{F}_c,\hat{F}_g}F_c^0 = M_{\hat{F}_c,\hat{F}_g}(F_c^0 - \hat{F}_c H_c^{-1})$, the next term is

$$
\begin{aligned}
&\frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g} I_4 (F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= \frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g}\left[\frac{1}{NT}\sum_{i=1}^{N} F_c^0\boldsymbol{\lambda}_{c,i}(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)' X_i'\hat{F}_c\right](F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= \frac{1}{\sqrt{T}}\left(\frac{X_i' M_{\hat{F}_c,\hat{F}_g}}{\sqrt{T}}\right)\left(\frac{F_c^0 - \hat{F}_c H_c^{-1}}{\sqrt{T}}\right)\left[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\lambda}_{c,i}(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)'\left(\frac{X_i'\hat{F}_c}{\sqrt{T}}\right)\right] \\
&\quad \times (F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= o_p(1)\times O_p(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) + o_p(1)\times\frac{1}{N}\sum_{k\neq i}^{N} O_p(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) + o_p(T^{-1/2}),
\end{aligned}
$$

where $T^{-1/2}\|F_c^0 - \hat{F}_c H_c^{-1}\| = o_p(1)$ was used. Also

$$\begin{aligned}
&\frac{1}{T} X_i' M_{\hat{F}_c,\hat{F}_g} I_5 (F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= \frac{1}{T} X_i' M_{\hat{F}_c,\hat{F}_g}\left[\frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{\varepsilon}_k(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k)' X_k'\hat{F}_c\right](F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= \frac{1}{\sqrt{T}}\left(\frac{X_i' M_{\hat{F}_c,\hat{F}_g}}{\sqrt{T}}\right)\left[\frac{1}{N}\sum_{k=1}^{N}\left(\frac{\boldsymbol{\varepsilon}_k}{\sqrt{T}}\right)(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k)'\left(\frac{X_i'\hat{F}_c}{\sqrt{T}}\right)\right](F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= o_p(1)\times O_p(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) + o_p(1)\times\sum_{k\neq i}^{N} O_p(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) + o_p(T^{-1/2}).
\end{aligned}$$

Next,

$$\begin{aligned}
&\frac{1}{T} X_i' M_{\hat{F}_c,\hat{F}_g} I_6 (F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= \frac{1}{T} X_i' M_{\hat{F}_c,\hat{F}_g}\left[\frac{1}{NT}\sum_{k=1}^{N} F_c^0\boldsymbol{\lambda}_{c,k}\boldsymbol{\varepsilon}_k'\hat{F}_c\right](F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= \frac{1}{T} X_i' M_{\hat{F}_c,\hat{F}_g}\left[\frac{1}{NT}\sum_{k=1}^{N}(F_c^0 - \hat{F}_c H_c^{-1})\boldsymbol{\lambda}_{c,k}\boldsymbol{\varepsilon}_k'\hat{F}_c\right](F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= \left(\frac{X_i' M_{\hat{F}_c,\hat{F}_g}}{\sqrt{T}}\right)\left[\frac{1}{NT}\sum_{k=1}^{N}\frac{(F_c^0 - \hat{F}_c H_c^{-1})}{\sqrt{T}}\boldsymbol{\lambda}_{c,k}\boldsymbol{\varepsilon}_k'\hat{F}_c\right](F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i}.
\end{aligned}$$

Using

$$\begin{aligned}
&\frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{\lambda}_{c,k}\boldsymbol{\varepsilon}_k'\hat{F}_c \\
&= \frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{\lambda}_{c,k}\boldsymbol{\varepsilon}_k' F_c^0 H_c + \frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{\lambda}_{c,k}\boldsymbol{\varepsilon}_k'(\hat{F}_c - F_c^0 H_c) \\
&= O_p\left(\frac{1}{\sqrt{NT}}\right) + \frac{1}{N^{3/2}}\sum_{k=1}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right) + \frac{1}{N^{1/2}} O_p\left(\frac{1}{\min\{N^{1/2}, T^{1/2}\}}\right) \\
&= \frac{1}{N^{3/2}}\sum_{k=1}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right) + \frac{1}{N^{1/2}} O_p\left(\frac{1}{\min\{N^{1/2}, T^{1/2}\}}\right)
\end{aligned}$$

and

$$\begin{aligned}
&\left(\frac{X_i' M_{\hat{F}_c,\hat{F}_g}}{\sqrt{T}}\right)\left(\frac{F_c^0 - \hat{F}_c H_c^{-1}}{\sqrt{T}}\right)(F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= \frac{1}{N}\sum_{k=1}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right) + O_p\left(\frac{1}{\min\{N^{1/2}, T^{1/2}\}}\right),
\end{aligned}$$

we have

$$\begin{aligned}
&\frac{1}{T} X_i' M_{\hat{F}_c} I_6 (F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= o_p(1)\times O_p(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) + o_p(1)\times\frac{1}{N}\sum_{k\neq i}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right) + O_p\left(\frac{1}{\min\{N, N^{1/2}T^{1/2}\}}\right).
\end{aligned}$$

Next, we have

$$\begin{aligned}
&\frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g} I_7 (F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= \frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g}\left[\frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{\varepsilon}_k \boldsymbol{\lambda}_{c,k}' F_c^{0\prime}\hat{F}_c\right](F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
&= \frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{\lambda}_{c,k}'(\Lambda'\Lambda/N)^{-1}\boldsymbol{\lambda}_{c,i} X_i' M_{\hat{F},\hat{F}_g}\boldsymbol{\varepsilon}_k \\
&= O_p\left(\frac{1}{\sqrt{NT}}\right).
\end{aligned}$$

Defining $E[\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k'] = \Omega_k$, we have

$$\begin{aligned}
&\frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g} I_8 (F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
= \;& \frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g}\left[\frac{1}{NT}\sum_{k=1}^{N}\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k'\hat{F}\right](F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
= \;& \frac{1}{NT^2}\sum_{k=1}^{N}X_i' M_{\hat{F}_c,\hat{F}_g}\Omega_k\hat{F}_c(F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
& + \frac{1}{NT^2}\sum_{k=1}^{N}X_i' M_{\hat{F}_c,\hat{F}_g}(\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k' - \Omega_k)\hat{F}_c(F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
= \;& \frac{1}{NT^2}\sum_{k=1}^{N}X_i' M_{\hat{F}_c,\hat{F}_g}\Omega_k\hat{F}_c(F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
& + O_p\left(\frac{1}{T\sqrt{N}}\right) + \frac{1}{\sqrt{N}}\left[\frac{1}{N}\sum_{k=1}^{N}O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right) + O_p\left(\frac{1}{\min\{N,T\}}\right)\right] \\
& + \frac{1}{\sqrt{NT}} \times \left[\frac{1}{N}\sum_{i=1}^{N}O_p\left(\|\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i\|\right) + O_p\left(\frac{1}{\min\{N^{1/2},T^{1/2}\}}\right)\right],
\end{aligned}$$

which follows from Lemma A3. Next,

$$\begin{aligned}
&\frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g} I_{10} (F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
= \;& \frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g}\left[\frac{1}{NT}\sum_{g=1}^{G}\sum_{k;g_k=g}X_k(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k)\boldsymbol{\lambda}_{j,k}'(F_g^0 - \hat{F}_g H_g^{-1})'\hat{F}_c\right](F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
= \;& \left(\frac{X_i' M_{\hat{F}_c,\hat{F}_g}}{\sqrt{T}}\right)\left[\frac{1}{NT}\sum_{g=1}^{G}\sum_{k;g_k=g}X_k(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k)\boldsymbol{\lambda}_{j,k}'\frac{(F_g^0 - \hat{F}_g H_g^{-1})}{\sqrt{T}}\hat{F}_c\right](F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i}.
\end{aligned}$$

Using the result of Lemma A2, we have

$$\begin{aligned}
&\frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g} I_{10} (F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i} \\
= \;& \left[\frac{1}{N}\sum_{k=1}^{N}O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right) + O_p\left(\frac{1}{\min\{N^{1/2},T^{1/2}\}}\right)\right] \times \left[\frac{1}{N}\sum_{k=1}^{N}O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right)\right] \\
= \;& o_p(1) \times O_p(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) + o_p(1) \times \frac{1}{N}\sum_{k\neq i}O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right),
\end{aligned}$$

where we used $\|\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i\| = o_p(1)$. In a similar manner, we have

$$
\begin{aligned}
& \frac{1}{T} X_i' M_{\hat{F}_c, \hat{F}_g} I_{11} (F_c^{0\prime} \hat{F}_c / T)^{-1} (\Lambda_c' \Lambda_c / N)^{-1} \boldsymbol{\lambda}_{c,i} \\
= \; & o_p(1) \times O_p(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) + o_p(1) \times \frac{1}{N} \sum_{k \neq i}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\| \right).
\end{aligned}
$$

The next term is

$$
\begin{aligned}
& \frac{1}{T} X_i' M_{\hat{F}_c, \hat{F}_g} I_{12} (F_c^{0\prime} \hat{F}_c / T)^{-1} (\Lambda_c' \Lambda_c / N)^{-1} \boldsymbol{\lambda}_{c,i} \\
= \; & \frac{1}{T} X_i' M_{\hat{F}_c, \hat{F}_g} \left[\frac{1}{NT} \sum_{g=1}^{G} \sum_{k; g_k = g} (F_g^0 - \hat{F}_g H_g^{-1}) \boldsymbol{\lambda}_{g,k} \boldsymbol{\varepsilon}_k' \hat{F}_c \right] (F_c^{0\prime} \hat{F}_c / T)^{-1} (\Lambda_c' \Lambda_c / N)^{-1} \boldsymbol{\lambda}_{c,i} \\
= \; & \left(\frac{X_i' M_{\hat{F}_c, \hat{F}_g}}{\sqrt{T}} \right) \left[\frac{1}{NT} \sum_{g=1}^{G} \sum_{k; g_k = g} \left(\frac{F_g^0 - \hat{F}_g H_g^{-1}}{\sqrt{T}} \right) \boldsymbol{\lambda}_{g,k} \boldsymbol{\varepsilon}_k' \hat{F}_c \right] (F_c^{0\prime} \hat{F}_c / T)^{-1} (\Lambda_c' \Lambda_c / N)^{-1} \boldsymbol{\lambda}_{c,i}.
\end{aligned}
$$

From Lemma A2 and A4, this implies

$$
\begin{aligned}
& \sum_{g=1}^{G} \frac{1}{T^{1/2}} \|\hat{F}_g H_g^{-1} - F_g^0\| \\
= \; & \frac{1}{N} \sum_{i=1}^{N} O_p\left(\|\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i\| \right) + O_p\left(\frac{1}{\min\{\sqrt{N_1}, ..., \sqrt{N_G}, \sqrt{T}\}} \right) \\
= \; & \frac{1}{N} \sum_{i=1}^{N} O_p\left(\|\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i\| \right) + O_p\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T}\}} \right),
\end{aligned}
$$

where we used Assumption E. Thus, we have

$$
\begin{aligned}
& \frac{1}{T} X_i' M_{\hat{F}_c, \hat{F}_g} I_{12} (F_c^{0\prime} \hat{F}_c / T)^{-1} (\Lambda_c' \Lambda_c / N)^{-1} \boldsymbol{\lambda}_{c,i} \\
= \; & \left[\frac{1}{N^{3/2}} \sum_{k=1}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\| \right) + \frac{1}{N^{1/2}} O_p\left(\frac{1}{\min\{N^{1/2}, T^{1/2}\}} \right) \right] \\
& \times \left[\frac{1}{N} \sum_{i=1}^{N} O_p\left(\|\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i\| \right) + O_p\left(\frac{1}{\min\{N^{1/2}, T^{1/2}\}} \right) \right] \\
= \; & o_p(1) \times O_p(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) + o_p(1) \times \frac{1}{N} \sum_{k \neq i}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\| \right) \\
& + o_p(1) \times O_p\left(\frac{1}{\min\{N, N^{1/2} T^{1/2}\}} \right).
\end{aligned}
$$

In a similar manner, we also obtain

$$
\begin{aligned}
& \frac{1}{T} X_i' M_{\hat{F}_c, \hat{F}_g} I_{13} (F_c^{0\prime} \hat{F}_c / T)^{-1} (\Lambda_c' \Lambda_c / N)^{-1} \boldsymbol{\lambda}_{c,i} \\
= \; & o_p(1) \times O_p(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) + o_p(1) \times \frac{1}{N} \sum_{k \neq i}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\| \right) \\
& + o_p(1) \times O_p\left(\frac{1}{\min\{N, N^{1/2} T^{1/2}\}} \right).
\end{aligned}
$$

The next term is

$$
\begin{aligned}
& \frac{1}{T}\left\|X_i' M_{\hat{F}_c,\hat{F}_g} I_{14} (F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i}\right\| \\
= & \left\|\frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g}\left[\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}(F_g^0-\hat{F}_g H_g^{-1})\boldsymbol{\lambda}_{g_i,i}\boldsymbol{\lambda}_{g_i,i}'(F_g^0-\hat{F}_g H_g^{-1})'\hat{F}_c\right]\right. \\
& \left.\times\left(\frac{F_c^{0\prime}\hat{F}_c}{T}\right)^{-1}\left(\frac{\Lambda_c'\Lambda_c}{N}\right)^{-1}\boldsymbol{\lambda}_{c,i}\right\| \\
\leq & \sum_{g=1}^{G}\frac{N_g}{N}\left\|\frac{X_i' M_{\hat{F}_c,\hat{F}_g}}{\sqrt{T}}\right\|\left\|\frac{(F_g^0-\hat{F}_g H_g^{-1})}{\sqrt{T}}\right\|^2\left\|\frac{\Lambda_g'\Lambda_g}{N_g}\right\|^2\left\|\frac{\hat{F}_c}{\sqrt{T}}\right\|\left\|\left(\frac{F_c^{0\prime}\hat{F}_c}{T}\right)^{-1}\right\|\left\|\left(\frac{\Lambda_c'\Lambda_c}{N}\right)^{-1}\right\|\|\boldsymbol{\lambda}_{c,i}\| \\
= & \sum_{g=1}^{G}\frac{N_g}{N}\left[\sum_{k;g_k=g}O_p\left(\|\boldsymbol{\beta}_k-\hat{\boldsymbol{\beta}}_k\|\right)+O_p\left(\frac{1}{\min\{N^{1/2},T^{1/2}\}}\right)\right]^2 \\
= & o_p(1)\times O_p(\hat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i)+o_p(1)\times\frac{1}{N}\sum_{k\neq i}^{N}O_p\left(\|\boldsymbol{\beta}_k-\hat{\boldsymbol{\beta}}_k\|\right)+O_p\left(\frac{1}{\min\{N,T\}}\right),
\end{aligned}
$$

where we used Lemma A2. The next term is

$$
\begin{aligned}
& \frac{1}{T}\left\|X_i' M_{\hat{F}_c,\hat{F}_g} I_{15} (F_c^{0\prime}\hat{F}_c/T)^{-1}(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i}\right\| \\
= & \frac{1}{T}\left\|X_i' M_{\hat{F}_c,\hat{F}_g}\left[\frac{1}{TN}\sum_{g=1}^{G}\sum_{k;g_k=g}(F_g^0-\hat{F}_g H_g^{-1})'\boldsymbol{\lambda}_{g,k}'\boldsymbol{\lambda}_{c,k}'\hat{F}_c\right]\left(\frac{F_c^{0\prime}\hat{F}_c}{T}\right)^{-1}\left(\frac{\Lambda_c'\Lambda_c}{N}\right)^{-1}\boldsymbol{\lambda}_{c,i}\right\| \\
\leq & \frac{1}{\sqrt{T}}\sum_{g=1}^{G}\frac{N_g}{N}\left\|\frac{X_i' M_{\hat{F}_c,\hat{F}_g}}{\sqrt{T}}\right\|\left\|\frac{(F_g^0-\hat{F}_g H_g^{-1})}{\sqrt{T}}\right\|\left\|\frac{\Lambda_g'\Lambda_g}{N_g}\right\|^2\left\|\frac{\hat{F}_c}{\sqrt{T}}\right\|\left\|\left(\frac{F_c^{0\prime}\hat{F}_c}{T}\right)^{-1}\right\|\left\|\left(\frac{\Lambda_c'\Lambda_c}{N}\right)^{-1}\right\|\|\boldsymbol{\lambda}_{c,i}\| \\
\leq & \sum_{g=1}^{G}\frac{N_g}{\sqrt{T}N}\left[\frac{1}{N_g}\sum_{k;g_k=g}O_p\left(\|\boldsymbol{\beta}_k-\hat{\boldsymbol{\beta}}_k\|^2\right)+O_p\left(\frac{1}{\min\{N^{1/2},T^{1/2}\}}\right)\right] \\
= & o_p(1)\times o_p\left(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i\right)+o_p(1)\times\frac{1}{N}\sum_{k\neq i}^{N}O_p\left(\|\boldsymbol{\beta}_k-\hat{\boldsymbol{\beta}}_k\|\right)+O_p\left(\frac{1}{\min\{N^{1/2}T^{1/2},T\}}\right).
\end{aligned}
$$

The final term also has the same expression. Summarizing these evaluations, we have

$$
\begin{aligned}
& \frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g} F_c^0\boldsymbol{\lambda}_{c,i} \\
= & \frac{1}{NT}X_i' M_{\hat{F}_c,\hat{F}_g} X_i\boldsymbol{\lambda}_{c,i}'(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i}(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i) \\
& +\frac{1}{NT}\sum_{k\neq i}^{N}X_i' M_{\hat{F}_c,\hat{F}_g} X_k\boldsymbol{\lambda}_{c,k}'(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i}(\boldsymbol{\beta}_k-\hat{\boldsymbol{\beta}}_k) \\
& +o_p(1)\times(\boldsymbol{\beta}_i-\hat{\boldsymbol{\beta}}_i)+o_p(1)\times\frac{1}{N}\sum_{k\neq i}^{N}O_p\left(\|\boldsymbol{\beta}_k-\hat{\boldsymbol{\beta}}_k\|\right) \\
& +O_p\left(\frac{1}{\min\{N,N^{1/2}T^{1/2}\}}\right)+O_p\left(\frac{1}{\min\{N^{1/2}T^{1/2},T\}}\right).
\end{aligned}
$$

Next we evaluate the term

$$\frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g} F_g^0 \boldsymbol{\lambda}_{g,i},$$

where we assumed that i-th security $\boldsymbol{y}_i$ belong to the g-th group. Again, using the result of in the proof of Theorem 1,

$$F_g = \hat{F}_g H_g^{-1} - \left[\sum_{k=1,k\neq 9}^{16} I_k\right] (F_g^{0\prime}\hat{F}_g/T)^{-1}(\Lambda_g'\Lambda_g/N_g)^{-1}$$

we have

$$\frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g} F_g^0 \boldsymbol{\lambda}_{g_i,i} = -\frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g}\left[\sum_{k=1,k\neq 9}^{16} J_k\right](F_g^{0\prime}\hat{F}_g/T)^{-1}(\Lambda_g'\Lambda_g/N_g)^{-1}\boldsymbol{\lambda}_{g_i,i},$$

where we used $M_{\hat{F}_c,\hat{F}_g}\hat{F}_g H_g^{-1} = 0$ and J_k $(k = 1, ..., 16)$ are defined as

$$\begin{aligned}
& \hat{F}_g V_{g,NT} \\
= \;& \frac{1}{N_gT}\sum_{i;g_i=g} X_i(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)' X_i'\hat{F}_g + \frac{1}{N_gT}\sum_{i;g_i=g} X_i(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{g_i,i}F_g^{0\prime}\hat{F}_g \\
& + \frac{1}{N_gT}\sum_{i;g_i=g} X_i(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)\boldsymbol{\varepsilon}_i'\hat{F}_g + \frac{1}{N_gT}\sum_{i;g_i=g} F_c^0\boldsymbol{\lambda}_{c,i}(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)' X_i'\hat{F}_c \\
& + \frac{1}{N_gT}\sum_{i;g_i=g} \boldsymbol{\varepsilon}_i(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)' X_i'\hat{F}_c + \frac{1}{N_gT}\sum_{i;g_i=g} F_c^0\boldsymbol{\lambda}_{c,i}\boldsymbol{\varepsilon}_i'\hat{F}_c + \frac{1}{N_gT}\sum_{i;g_i=g} \boldsymbol{\varepsilon}_{G,i}\boldsymbol{\lambda}_{c,i}'F_c^{0\prime}\hat{F}_c \\
& + \frac{1}{N_gT}\sum_{i;g_i=g} \boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\hat{F}_c + \frac{1}{N_gT}\sum_{i;g_i=g} F_c^0\boldsymbol{\lambda}_{c,i}\boldsymbol{\lambda}_{c,i}'F_c^{0\prime}\hat{F}_c \\
& + \frac{1}{N_gT}\sum_{i;g_i=g} X_i(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)\boldsymbol{\lambda}_{g_i,i}'(F_g^0 - \hat{F}_g H_g^{-1})'\hat{F}_c \\
& + \frac{1}{N_gT}\sum_{i;g_i=g} (F_g^0 - \hat{F}_g H_g^{-1})\boldsymbol{\lambda}_{g_i,i}(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)' X_i'\hat{F}_c \\
& + \frac{1}{N_gT}\sum_{i;g_i=g} (F_g^0 - \hat{F}_g H_g^{-1})\boldsymbol{\lambda}_{g_i,i}\boldsymbol{\varepsilon}_i'\hat{F}_c + \frac{1}{N_gT}\sum_{i;g_i=g} \boldsymbol{\varepsilon}_i\boldsymbol{\lambda}_{g_i,i}'(F_g^0 - \hat{F}_g H_g^{-1})'\hat{F}_c \\
& + \frac{1}{N_gT}\sum_{i;g_i=g} (F_g^0 - \hat{F}_g H_g^{-1})\boldsymbol{\lambda}_{g_i,i}\boldsymbol{\lambda}_{g_i,i}'(F_g^0 - \hat{F}_g H_g^{-1})'\hat{F}_c \\
& + \frac{1}{N_gT}\sum_{i;g_i=g} (F_g^0 - \hat{F}_g H_g^{-1})\boldsymbol{\lambda}_{g_i,i}\boldsymbol{\lambda}_{c,i}'F_c^{0\prime}\hat{F}_c + \frac{1}{N_gT}\sum_{i;g_i=g} F_c^0\boldsymbol{\lambda}_{c,i}\boldsymbol{\lambda}_{g_i,i}'(F_g^0 - \hat{F}_g H_g^{-1})'\hat{F}_c \\
= \;& J_1 + \cdots + J_{16}.
\end{aligned}$$

In the same way of evaluating $X_i' M_{\hat{F}_c,\hat{F}_g} F_c^0 \boldsymbol{\lambda}_{c,i}/T$, we have

$$\begin{aligned}
& \frac{1}{T}X_i' M_{\hat{F}_c,\hat{F}_g} F_g^0 \boldsymbol{\lambda}_{g,i} \\
= \;& \frac{1}{N_gT}X_i' M_{\hat{F}_c,\hat{F}_g} X_i \boldsymbol{\lambda}_{g,i}'(\Lambda_g'\Lambda_g/N_g)^{-1}\boldsymbol{\lambda}_{g,i}(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{N_g T}\sum_{k;g_k=g,k\neq i}^{N_g} X_i' M_{\hat{F}_c,\hat{F}_g} X_k \boldsymbol{\lambda}_{g,k}' (\Lambda_g'\Lambda_g/N_g)^{-1}\boldsymbol{\lambda}_{g,i}(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k) \\
&+o_p(1)\times(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i) + o_p(1)\times \frac{1}{N}\sum_{k\neq i}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right) \\
&+O_p\left(\frac{1}{\min\{N, N^{1/2}T^{1/2}\}}\right) + O_p\left(\frac{1}{\min\{N^{1/2}T^{1/2}, T\}}\right).
\end{aligned}
$$

Then, for i that belongs to the g-th group, we have

$$
\begin{aligned}
& \left[\frac{1}{T}\left(X_i' M_{\hat{F}_c,\hat{F}_g} X_i + \Sigma_i(\kappa)\right) + o_p(1)\right]\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) \\
= \; & \frac{1}{NT}\sum_{k\neq i}^{N} X_i' M_{\hat{F}_c,\hat{F}_g} X_k \boldsymbol{\lambda}_{c,k}' (\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i}\sqrt{T}(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k) \\
& +\frac{1}{N_g T}\sum_{k;g_k=g,k\neq i}^{N_g} X_i' M_{\hat{F}_c,\hat{F}_g} X_k \boldsymbol{\lambda}_{g,k}' (\Lambda_g'\Lambda_g/N_g)^{-1}\boldsymbol{\lambda}_{g,i}\sqrt{T}(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k) + \frac{1}{\sqrt{T}} X_i' M_{\hat{F}_c,\hat{F}_g}\boldsymbol{\varepsilon}_i \\
& +o_p(1)\times(\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_i) + o_p(1)\times\frac{\sqrt{T}}{N}\sum_{k\neq i}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right) + o_p(1).
\end{aligned}
$$

Next we study the effect of replacing $M_{\hat{F}_c,\hat{F}_g} = M_{\hat{F}_c} - M_{\hat{F}_c}\hat{F}_g(\hat{F}_g' M_{\hat{F}_c}\hat{F}_g)^{-1}\hat{F}_g' M_{\hat{F}_c}$ in the above equation by $M_{F_c^0,F_g^0}$. We first focus on the first term $M_{\hat{F}_c} = I - \hat{F}_c(\hat{F}_c'\hat{F}_c)^{-1}\hat{F}_c'$ in $M_{\hat{F}_c,\hat{F}_g}$. Using the Lemma A.8 of Bai (2009), we can express the difference between $T^{-1/2}X_i' M_{\hat{F}_c}\boldsymbol{\varepsilon}_i$ and $T^{-1/2}X_i' M_{F_c^0}\boldsymbol{\varepsilon}_i$ as

$$
\begin{aligned}
& \frac{1}{T^{1/2}} X_i' M_{F_c^0}\boldsymbol{\varepsilon}_i - \frac{1}{T^{1/2}} X_i' M_{\hat{F}_c}\boldsymbol{\varepsilon}_i \\
= \; & \frac{1}{T^{3/2}} X_i' F_c^0 \left(\frac{F_c^{0\prime} F_c^0}{T}\right)^{-1} F_c^{0\prime}\boldsymbol{\varepsilon}_i - \frac{1}{T^{3/2}} X_i'\left(\hat{F}_c - F_c^0 H_c + F_c^0 H_c\right)\left(\hat{F}_c - F_c^0 H_c + F_c^0 H_c\right)'\boldsymbol{\varepsilon}_i \\
= \; & \frac{1}{T^{3/2}} X_i' F_c^0 \left[\left(\frac{F_c^{0\prime} F_c^0}{T}\right)^{-1} - H_c H_c'\right] F_c^{0\prime}\boldsymbol{\varepsilon}_i - \frac{1}{T^{3/2}} X_i'\left(\hat{F}_c - F_c^0 H_c\right)\left(\hat{F}_c - F_c^0 H_c\right)'\boldsymbol{\varepsilon}_i \\
& -\frac{1}{T^{3/2}} X_i'\left(\hat{F}_c - F_c^0 H_c\right) H_c' F_c^{0\prime}\boldsymbol{\varepsilon}_i - \frac{1}{T^{3/2}} X_i' F_c^0 H_c\left(\hat{F}_c - F_c^0 H_c\right)'\boldsymbol{\varepsilon}_i \\
= \; & \frac{1}{N}\sum_{k=1}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right) + O_p\left(\frac{T^{1/2}}{\min\{N, T\}}\right).
\end{aligned}
$$

where we used $X_i'(\hat{F}_c - F_c^0 H_c)/T = \frac{1}{N}\sum_{k=1}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right) + O_p(1/\min\{N^{1/2}, T^{1/2}\})$ and $(F_c^{0\prime}F_c^0/T)^{-1} - H_c H_c' = \frac{1}{N}\sum_{k=1}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right) + O_p\left(1/\min\{N, T\}\right)$, which follows from Lemma A7 in Bai (2009). Thus, we have $X_i' M_{F_c^0}\boldsymbol{\varepsilon}_i/T - X_i' M_{\hat{F}_c}\boldsymbol{\varepsilon}_i/T = o_p(T^{-1/2})$. Next, we evaluate the following quantity

$$
\frac{1}{T^{1/2}} X_i' M_{\hat{F}_c}\hat{F}_g(\hat{F}_g' M_{\hat{F}_c}\hat{F}_g)^{-1}\hat{F}_g' M_{\hat{F}_c}\boldsymbol{\varepsilon}_i - \frac{1}{T^{1/2}} X_i' M_{F_c^0} F_g (F_g^{0\prime} M_{F_c^0} F_g^0)^{-1} F_g' M_{F_c^0}\boldsymbol{\varepsilon}_i.
$$

Adding and subtracting terms, we have

$$\begin{aligned}
& M_{\hat{F}_c}\hat{F}_g(\hat{F}_g' M_{\hat{F}_c}\hat{F}_g)^{-1}\hat{F}_g' M_{\hat{F}_c} - M_{F_c^0}F_g^0(F_g^{0\prime} M_{F_c^0}F_g^0)^{-1}F_g^{0\prime} M_{F_c^0} \\
= \; & (M_{\hat{F}_c} - M_{F_c^0})\hat{F}_g(\hat{F}_g' M_{\hat{F}_c}\hat{F}_g)^{-1}\hat{F}_g' M_{\hat{F}_c} + M_{F_c^0}(\hat{F}_g - F_g^0 H_g)(\hat{F}_g' M_{\hat{F}_c}\hat{F}_g)^{-1}\hat{F}_g' M_{\hat{F}_c} \\
& + M_{F_c^0}F_g^0 H_g\left[(\hat{F}_g' M_{\hat{F}_c}\hat{F}_g)^{-1} - ((F_g^0 H_g)' M_{F_c^0}F_g^0 H_g)^{-1}\right]\hat{F}_g' M_{\hat{F}_c} \\
& + M_{F_c^0}F_g^0 H_g((F_g^0 H_g)' M_{F_c^0}F_g^0 H_g)^{-1}(\hat{F}_g - F_g^0 H_g)' M_{\hat{F}_c} \\
& + M_{F_c^0}F_g^0 H_g((F_g^0 H_g)' M_{F_c^0}F_g^0 H_g)^{-1}(F_g^0 H_g)'(M_{\hat{F}_c} - M_{F_c^0}).
\end{aligned}$$

Thus, together with Lemma A2, we have

$$\begin{aligned}
& \frac{1}{T^{1/2}} X_i' M_{\hat{F}_c}\hat{F}_g(\hat{F}_g' M_{\hat{F}_c}\hat{F}_g)^{-1}\hat{F}_g' M_{\hat{F}_c}\boldsymbol{\varepsilon}_i - \frac{1}{T^{1/2}} X_i' M_{F_c^0}F_g^0 H_g((F_g^0 H_g)' M_{F_c^0}F_g^0 H_g)^{-1}(F_g^0 H_g)' M_{F_c^0}\boldsymbol{\varepsilon}_i \\
& = \frac{1}{N}\sum_{k=1}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right) + O_p\left(\frac{\sqrt{T}}{\min\{N,T\}}\right).
\end{aligned}$$

These investigations provides

$$\frac{1}{\sqrt{T}} X_i' M_{\hat{F}_c,\hat{F}_g}\boldsymbol{\varepsilon}_i - \frac{1}{\sqrt{T}} X_i' M_{F_c^0,F_g^0}\boldsymbol{\varepsilon}_i = \frac{1}{N}\sum_{k=1}^{N} O_p\left(\|\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k\|\right) + O_p\left(\frac{\sqrt{T}}{\min\{N,T\}}\right).$$

Using $X_i' M_{\hat{F}_c,\hat{F}_g} X_i - X_i' M_{F_c^0,F_g^0} X_i = o_p(1)$, we finally have

$$\begin{aligned}
& \left[\frac{1}{T}\left(X_i' M_{F_c^0,F_g^0} X_i + \Sigma_i(\kappa)\right)\right]\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) \\
= \; & \frac{1}{NT}\sum_{k\neq i}^{N} X_i' M_{F_c^0,F_g^0} X_k \boldsymbol{\lambda}_{c,k}'(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i}\sqrt{T}(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k) \\
& + \frac{1}{N_g T}\sum_{k;g_k=g,k\neq i}^{N_g} X_i' M_{F_c^0,F_g^0} X_k \boldsymbol{\lambda}_{g,k}'(\Lambda_g'\Lambda_g/N_g)^{-1}\boldsymbol{\lambda}_{g,i}\sqrt{T}(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k) \\
& + \frac{1}{\sqrt{T}} X_i' M_{F_c^0,F_g^0}\boldsymbol{\varepsilon}_i + o_p(1). \qquad (21)
\end{aligned}$$

For notational simplicity, we define

$$\begin{aligned}
\boldsymbol{\eta}_i &= D_i^{-1}\frac{1}{\sqrt{T}} X_i' M_{F_c^0,F_g^0}\boldsymbol{\varepsilon}_i, \\
\Gamma_{c,ik} &= D_i^{-1}\left(\frac{1}{T} X_i' M_{F_c^0,F_g^0} X_k\right)\boldsymbol{\lambda}_{c,k}'(\Lambda_c'\Lambda_c/N)^{-1}\boldsymbol{\lambda}_{c,i}, \\
\Gamma_{g,ik} &= D_i^{-1}\left(\frac{1}{T} X_i' M_{F_c^0,F_g^0} X_k\right)\boldsymbol{\lambda}_{g,k}'(\Lambda_g'\Lambda_g/N_g)^{-1}\boldsymbol{\lambda}_{g,i},
\end{aligned}$$

where $D_i \equiv D_i(F_c^0, F_g^0, \kappa) = T^{-1}\left(X_i' M_{F_c^0,F_g^0} X_i + \Sigma_i(\kappa)\right)$. Then the expression (21) becomes

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) = \frac{1}{N}\sum_{k\neq i}^{N}\Gamma_{c,ik}\sqrt{T}(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k) + \frac{1}{N_g}\sum_{k;g_k=g,k\neq i}^{N_g}\Gamma_{g,ik}\sqrt{T}(\boldsymbol{\beta}_k - \hat{\boldsymbol{\beta}}_k) + \boldsymbol{\eta}_i + o_p(1),$$

which is equivalent to

$$\sqrt{T}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \frac{1}{N}\Gamma\sqrt{T}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})+\boldsymbol{\eta}+o_p\left(1\right),$$

where $\hat{\boldsymbol{\beta}}=(\hat{\boldsymbol{\beta}}_1,...,\hat{\boldsymbol{\beta}}_N')'$, $\boldsymbol{\beta}=(\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_N')'$, $\boldsymbol{\eta}=(\boldsymbol{\eta}_1',...,\boldsymbol{\eta}_N')'$,

$$\Gamma=\begin{pmatrix} 0 & \Gamma_{c,12}+\delta(g_1,g_2)\frac{N}{N_{g_2}}\Gamma_{g_1,12} & \cdots & \Gamma_{c,1N}+\delta(g_1,g_N)\frac{N}{N_{g_1}}\Gamma_{g_1,1N} \\ \Gamma_{c,21}+\delta(g_2,g_1)\frac{N}{N_{g_2}}\Gamma_{g_2,21} & 0 & \cdots & \Gamma_{c,2N}+\delta(g_2,g_N)\frac{N}{N_{g_2}}\Gamma_{g_2,2N} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{c,N1}+\delta(g_N,g_1)\frac{N}{N_{g_N}}\Gamma_{g_N,N1} & \Gamma_{c,N2}+\delta(g_N,g_2)\frac{N}{N_{g_N}}\Gamma_{g_N,N2} & \cdots & 0 \end{pmatrix},$$

where $\delta(g_i,g_j)=1$ if $g_i=g_j$ and $\delta(g_i,g_j)=0$ otherwise. This part of analysis is similar to Song (2013). In summary, $\sqrt{T}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})=(I+\frac{1}{N}\Gamma)\boldsymbol{\eta}+o_p(1)$, where we used $(I-\frac{1}{N}\Gamma)^{-1}=I+\frac{1}{N}\Gamma+o_p(1)$, it can be shown that the higher order terms are negligible. This implies that for each i, we have

$$\begin{aligned}\sqrt{T}(\hat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i) &= D_i^{-1}\frac{1}{\sqrt{T}}X_i'M_{F_c^0,F_g^0}\boldsymbol{\varepsilon}_i+\left[\frac{1}{N}\sum_{k\neq i}^{N}\Gamma_{c,ik}D_k^{-1}\frac{1}{\sqrt{T}}X_k'M_{F_c^0,F_g^0}\boldsymbol{\varepsilon}_k\right] \\ &\quad+\left[\frac{1}{N_g}\sum_{k;g_k=g,k\neq i}^{N_g}\Gamma_{g,ik}D_k^{-1}\frac{1}{\sqrt{T}}X_k'M_{F_c^0,F_g^0}\boldsymbol{\varepsilon}_k\right]+o_p\left(1\right) \\ &\equiv D_i^{-1}\frac{1}{\sqrt{T}}X_i'M_{F_c^0,F_g^0}\boldsymbol{\varepsilon}_i+\frac{1}{N}\sum_{k\neq i}^{N}\frac{1}{\sqrt{T}}A_{c,ik}\boldsymbol{\varepsilon}_k+\frac{1}{N_g}\sum_{k;g_k=g,k\neq i}^{N_g}\frac{1}{\sqrt{T}}A_{g,ik}\boldsymbol{\varepsilon}_k+o_p\left(1\right).\end{aligned}$$

where $A_{c,ik}$ denotes $\Gamma_{c,ik}D_k^{-1}X_k'M_{F_c^0,F_g^0}$, and $A_{g,ik}$ is similarly defined. Under cross-sectional independence, i.e., $E[\varepsilon_{it}\varepsilon_{jt}]=0$ $(i\neq j)$, the averages of independent terms $A_{c,ik}\boldsymbol{\varepsilon}_k$ and $A_{g,ik}\boldsymbol{\varepsilon}_k$ converge to zero. Thus, the first term $D_i^{-1}\frac{1}{\sqrt{T}}X_i'M_{F_c^0,F_g^0}\boldsymbol{\varepsilon}_i$ becomes the only leading term.

Even under cross-sectional dependence and serial correlation, we can show that the averages of $A_{c,ik}\boldsymbol{\varepsilon}_k$ and $A_{g,ik}\boldsymbol{\varepsilon}_k$ still converge to zero. In fact, from $\|\Gamma_{c,ik}D_k^{-1}\|\leq C$ and $\|\Gamma_{g,ik}D_k^{-1}\|\leq C$, for some $C<\infty$, it is enough to show that

$$\frac{1}{N\sqrt{T}}\sum_{k\neq i}^{N}X_k'M_{F_c^0,F_g^0}\boldsymbol{\varepsilon}_k=o_p(1), \quad \text{and} \quad \frac{1}{N_g\sqrt{T}}\sum_{k;g_k=g,k\neq i}^{N_g}X_k'M_{F_c^0,F_g^0}\boldsymbol{\varepsilon}_k=o_p(1).$$

Consider the first term. The expected value of its second moment is

$$E\left[\frac{1}{N^2T}\sum_{k\neq i}^{N}\sum_{\ell\neq i}^{N}X_k'M_{F_c^0,F_g^0}\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_\ell'M_{F_c^0,F_g^0}X_\ell\right]=\frac{1}{N^2T}\sum_{k\neq i}^{N}\sum_{\ell\neq i}^{N}X_k'M_{F_c^0,F_g^0}\Omega_{k\ell}M_{F_c^0,F_g^0}X_\ell$$

which converges to zero from Assumption G. Thus the first term is $o_p(1)$. Similarly, the second term is also $o_p(1)$. Summarizing these results, we finally have

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i)=D_i(F_c^0,F_g^0,\kappa)^{-1}\frac{1}{\sqrt{T}}X_i'M_{F_c^0,F_g^0}\boldsymbol{\varepsilon}_i+o_p\left(1\right),$$

The right hand side does not depend on estimated quantities. By Assumption F,

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) \to N(0, D_i(F_c^0, F_g^0)^{-1} J_i(F_c^0, F_g^0) D_i(F_c^0, F_g^0)^{-1}).$$

where $D_i(F_c^0, F_g^0)$ and $J_i(F_c^0, F_g^0)$ are defined in Theorem 2. Practical estimation procedure for $D_i(F_c^0, F_g^0)$ and $J_i(F_c^0, F_g^0)$ is discussed in Section 5.

Next, we prove the variable selection consistency $P(\hat{\boldsymbol{\beta}}_{i20} = \mathbf{0}) \to 1$ as $N, T \to \infty$. This part is almost identical to the proof of Fan and Li (2001). It is sufficient to show that with probability tending to 1 as $N, T \to \infty$, for some small $\delta_{N,T} = C/\sqrt{T}$ with a constant C, and for each element of $\boldsymbol{\beta}_{i2} = (\beta_{i21}, ..., \beta_{i2,p_i-q_i})$, we have

$$\frac{\partial S_{NT}(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_N, F_c, F_1, ..., F_S)}{\partial \beta_{i2k}} > 0 \quad (0 < \beta_{i2k} < \delta_{N,T}),$$
$$\frac{\partial S_{NT}(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_N, F_c, F_1, ..., F_S)}{\partial \beta_{i2k}} < 0 \quad (-\delta_{N,T} < \beta_{i2k} < 0),$$

for $k = 1, ..., p_i - q_i$. Let $X_{i,2}$ be the set of $p_i - q_i$ columns of X_i, corresponding to $\boldsymbol{\beta}_{i2}$. So, $X_{i,2}$ is a $T \times (p_i - q_i)$ dimensional matrix. Consider the first derivative of $S_{NT}(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_N, F_c, F_1, ..., F_S)/(NT)$ with respect to $\boldsymbol{\beta}_{i2} = (\beta_{i21}, ..., \beta_{i2,p_i-q_i})$,

$$\begin{aligned}
& \frac{1}{NT} \cdot \frac{\partial S_{NT}(\hat{\boldsymbol{\beta}}_1, ..., \hat{\boldsymbol{\beta}}_N, \hat{F}_c, \hat{F}_1, ..., \hat{F}_S)}{\partial \boldsymbol{\beta}_{i2}} \\
= \; & -\frac{2}{NT} \cdot \sum_{j=1}^{S} \sum_{i;g_i=j} X_{i,2}'(\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_c\hat{\boldsymbol{\lambda}}_{c,i} - \hat{F}_{g_i}\hat{\boldsymbol{\lambda}}_{g_i,i}) + \frac{\partial p_{\kappa,\gamma}\left(|\hat{\boldsymbol{\beta}}_{i2}|\right)}{\partial \boldsymbol{\beta}_{i2}} \\
= \; & -\frac{2}{NT} \sum_{j=1}^{S} \sum_{i;g_i=j} X_{i,2}'X_i(\boldsymbol{\beta}_i^0 - \hat{\boldsymbol{\beta}}_i) - \frac{2}{NT} \sum_{j=1}^{S} \sum_{i;g_i=j} X_{i,2}'(F_c^0\boldsymbol{\lambda}_{c,i}^0 - \hat{F}_c\hat{\boldsymbol{\lambda}}_{c,i}) \\
& + \frac{2}{NT} \sum_{j=1}^{S} \sum_{i;g_i=j} X_{i,2}'(F_{g_i}^0\boldsymbol{\lambda}_{g_i,i}^0 - \hat{F}_{g_i}\hat{\boldsymbol{\lambda}}_{g_i,i}) + \frac{2}{NT} \sum_{j=1}^{S} \sum_{i;g_i=j} X_{i,2}'\boldsymbol{\varepsilon}_i + \frac{\partial p_{\kappa,\gamma}\left(|\hat{\boldsymbol{\beta}}_{i2}|\right)}{\partial \boldsymbol{\beta}_{i2}} \\
= \; & I_1 + I_2 + I_3 + I_4 + \frac{\partial p_{\kappa,\gamma}\left(|\hat{\boldsymbol{\beta}}_{i2}|\right)}{\partial \boldsymbol{\beta}_{i2}},
\end{aligned}$$

where $\partial p_{\kappa,\gamma}\left(|\hat{\boldsymbol{\beta}}_{i2}|\right)/\partial \boldsymbol{\beta}_{i2}$ is a $(p_i - q_i) \times 1$ vector containing elements $p'_{\kappa,\gamma}\left(|\hat{\beta}_{i2k}|\right)\mathrm{sign}(\hat{\beta}_{i2k})$ for $k = 1, ..., p_i - q_i$.

Term I_4 is $O_p(1/\sqrt{NT})$. Together with the result of Theorem 1, we know that $\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0 = O_p(1/\sqrt{T})$. Thus, the first term I_1 is $O_p(1/\sqrt{T})$. The third term, I_3 is

$$\begin{aligned}
& \frac{1}{NT} \sum_{j=1}^{S} \sum_{i;g_i=j} X_{i,2}'(F_{g_i}^0\boldsymbol{\lambda}_{g_i,i}^0 - \hat{F}_{g_i}\hat{\boldsymbol{\lambda}}_{g_i,i}) \\
= \; & \frac{1}{NT} \sum_{j=1}^{S} \sum_{i;g_i=j} X_{i,2}'(F_{g_i}^0 - \hat{F}_{g_i})\boldsymbol{\lambda}_{g_i,i}^0 + \frac{1}{NT} \sum_{j=1}^{S} \sum_{i;g_i=j} X_{i,2}'\hat{F}_{g_i}(\boldsymbol{\lambda}_{g_i,i}^0 - \hat{\boldsymbol{\lambda}}_{g_i,i}),
\end{aligned}$$

which is $O_p(1/\min\{N,T\})$. The second term I_2 is also $O_p(1/\min\{N,T\})$. Thus

$$\begin{aligned}&\frac{\partial S_{NT}(\hat{\boldsymbol{\beta}}_1,...,\hat{\boldsymbol{\beta}}_N,\hat{F}_c,\hat{F}_1,...,\hat{F}_S)}{\partial \beta_{i2k}}\\ =\;& T\cdot\kappa\left[\frac{1}{\kappa}p'_{\kappa,\gamma}\left(|\hat{\beta}_{i2k}|\right)\mathrm{sign}(\hat{\beta}_{i2k})+O_p\left(1/(\sqrt{T}\cdot\kappa)\right)\right].\end{aligned}$$

Because $\frac{1}{\kappa}p'_{\kappa,\gamma}\left(|\hat{\beta}_{i2k}|\right)>0$ and $1/(\sqrt{T}\kappa)\to 0$, the sign of $\hat{\beta}_{i2k}$ determines the sign of $\partial S_{NT}(\hat{\boldsymbol{\beta}}_1,...,\hat{\boldsymbol{\beta}}_N,\hat{F}_c,\hat{F}_1,...,\hat{F}_S)/\partial\beta_{i2k}$. Hence, this result implies the sign claim. This completes the proof.

A6. Proof of Theorem 3

We divide the proof of Theorem 3 into two steps. In Step 1, we develop an estimator of the expected mean squared error, which can be used to select the number of predictors $\boldsymbol{x}$ under no factor structure. However, we still need an additional penalty term that penalizes the model complexity caused by the factor structures. Thus, Step 2 modifies the model selection criterion to select the number of factors.

Step 1: We decompose the bias b as

$$b=B_1+B_2+B_3,$$

where

$$\begin{aligned}B_1 &= E_y\Bigg[\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\left\|\boldsymbol{y}_i-X_i\hat{\boldsymbol{\beta}}_i-\hat{F}_c\hat{\boldsymbol{\lambda}}_{c,i}-\hat{F}_{g_i}\hat{\boldsymbol{\lambda}}_{g_i,i}\right\|^2\\ &\quad-\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\left\|\boldsymbol{y}_i-X_i\boldsymbol{\beta}_i^0-\hat{F}_c\hat{\boldsymbol{\lambda}}_{c,i}-\hat{F}_{g_i}\hat{\boldsymbol{\lambda}}_{g_i,i}\right\|^2\Bigg],\\ B_2 &= E_y\Bigg[\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\left\|\boldsymbol{y}_i-X_i\boldsymbol{\beta}_i^0-\hat{F}_c\hat{\boldsymbol{\lambda}}_{c,i}-\hat{F}_{g_i}\hat{\boldsymbol{\lambda}}_{g_i,i}\right\|^2\\ &\quad-E_z\Bigg[\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\left\|\boldsymbol{z}_i-X_i\boldsymbol{\beta}_i^0-\hat{F}_c\hat{\boldsymbol{\lambda}}_{c,i}-\hat{F}_{g_i}\hat{\boldsymbol{\lambda}}_{g_i,i}\right\|^2\Bigg]\Bigg],\\ B_3 &= E_y\Bigg[E_z\Bigg[\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\left\|\boldsymbol{z}_i-X_i\boldsymbol{\beta}_i^0-\hat{F}_c\hat{\boldsymbol{\lambda}}_{c,i}-\hat{F}_{g_i}\hat{\boldsymbol{\lambda}}_{g_i,i}\right\|^2\Bigg]\\ &\quad-E_z\Bigg[\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\left\|\boldsymbol{z}_i-X_i\hat{\boldsymbol{\beta}}_i-\hat{F}_c\hat{\boldsymbol{\lambda}}_{c,i}-\hat{F}_{g_i}\hat{\boldsymbol{\lambda}}_{g_i,i}\right\|^2\Bigg]\Bigg].\end{aligned}$$

Expectations $E_y[\cdot]$ and $E_z[\cdot]$ are taken with respect to the joint distribution of $\{\boldsymbol{y}_1,...,\boldsymbol{y}_N\}$ and $\{\boldsymbol{z}_1,...,\boldsymbol{z}_N\}$ given the predictors and factor structures.

We denote $\boldsymbol{\beta}=(\boldsymbol{\beta}_1',...,\boldsymbol{\beta}_N')'$ and

$$\ell_y(\boldsymbol{\beta},F_c,F_1,...,F_G,\Lambda_1,...,\Lambda_G)=\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\left\|\boldsymbol{y}_i-X_i\boldsymbol{\beta}_i-F_c\boldsymbol{\lambda}_{c,i}-F_g\boldsymbol{\lambda}_{g_i,i}\right\|^2,$$

$$\ell_z(\boldsymbol{\beta}, F_c, F_1, ..., F_G, \Lambda_1, ..., \Lambda_G) = E_z\left[\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\|\boldsymbol{z}_i - X_i\boldsymbol{\beta}_i - F_c\boldsymbol{\lambda}_{c,i} - F_g\boldsymbol{\lambda}_{g_i,i}\|^2\right].$$

Now, we first evaluate B_1. Noting that

$$\frac{\partial}{\partial\boldsymbol{\beta}}\left\{\ell_y(\boldsymbol{\beta}, \hat{F}_c, \hat{F}_1, ..., \hat{F}_G, \hat{\Lambda}_1, ..., \hat{\Lambda}_G) + N^{-1}\sum_{i=1}^{N}p_{\kappa,\gamma}(|\boldsymbol{\beta}_i|)\right\}\Bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \mathbf{0},$$

the Taylor expansion of $\ell_y(\boldsymbol{\beta}^0, \hat{F}_c, \hat{F}_1, ..., \hat{F}_G, \hat{\Lambda}_1, ..., \hat{\Lambda}_G)$ around $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1', ..., \hat{\boldsymbol{\beta}}_N')'$ gives

$$\begin{aligned}
&\ell_y(\boldsymbol{\beta}^0, \hat{F}_c, \hat{F}_1, ..., \hat{F}_G, \hat{\Lambda}_1, ..., \hat{\Lambda}_G) \\
= \;& \ell_y(\hat{\boldsymbol{\beta}}, \hat{F}_c, \hat{F}_1, ..., \hat{F}_G, \hat{\Lambda}_1, ..., \hat{\Lambda}_G) - N^{-1}\sum_{i=1}^{N}\partial p_{\kappa,\gamma}(|\hat{\boldsymbol{\beta}}_i|)/\partial\boldsymbol{\beta}_i'(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0) \\
& + \frac{1}{2}\times\frac{1}{NT}\times\sum_{i=1}^{N}\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0)'K_i\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0) + o_p\left(T^{-1}\right),
\end{aligned}$$

where $K_i = 2X_i'X_i/T$. Thus

$$B_1 = -N^{-1}\sum_{i=1}^{N}\frac{\partial p_{\kappa,\gamma}(|\hat{\boldsymbol{\beta}}_i|)}{\partial\boldsymbol{\beta}_i'}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0) + \frac{1}{2}\frac{1}{NT}\sum_{i=1}^{N}\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0)'K_i\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0) + o_p(T^{-1}),$$

For small κ, the expected value of $N^{-1}\sum_{i=1}^{N}\partial p_{\kappa,\gamma}(|\hat{\boldsymbol{\beta}}_i|)/\partial\boldsymbol{\beta}_i'(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0)$ is of order $o(1/T)$. This follows from the following expansion

$$\frac{\partial p_{\kappa,\gamma}(|\hat{\boldsymbol{\beta}}_i|)}{\partial\boldsymbol{\beta}_i'}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0) = \frac{\partial p_{\kappa,\gamma}(|\boldsymbol{\beta}_i^0|)}{\partial\boldsymbol{\beta}_i'}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0) + (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0)'\frac{\partial^2 p_{\kappa,\gamma}(|\boldsymbol{\beta}_i^0|)}{\partial\boldsymbol{\beta}_i\partial\boldsymbol{\beta}_i'}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0)$$

where we ignore the higher order terms. The expected value of the first term on the right is approximately zero, and the expected value of the second term is $O(1/T)$ times the second derivative. But the second derivative, under small κ, is $o(1)$. This gives the desirable result.

The covariance matrix of $\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i^0)$ is given by

$$R(F_c^0, F_g^0) = D_i(F_c^0, F_g^0)^{-1}J_i(F_c^0, F_g^0)D_i(F_c^0, F_g^0)^{-1}.$$

Thus, by replacing the expectation $E_y[\cdot]$ with the empirical distribution, we have an estimator of B_1 as

$$B_1 = \frac{1}{2NT}\sum_{i=1}^{N}\mathrm{tr}\left[K_iR(F_c^0, F_g^0)\right] + o(T^{-1}).$$

It can be shown that B_2 is dominated by B_1 and B_3. Using the same augment for the evaluation of B_1, we have $B_3 = \frac{1}{2NT}\sum_{i=1}^{N}\mathrm{tr}\left[K_iR(F_c^0, F_g^0)\right] + o(T^{-1})$. Finally, summing up all terms, the bias, contributed by the observable structure $X_i\hat{\boldsymbol{\beta}}_i$, becomes

$$\frac{1}{NT}\sum_{i=1}^{N}\mathrm{tr}\left[K_iR(F_c^0, F_g^0)\right] + o(T^{-1}).$$

Therefore, the expected mean squared error can be approximated by

$$\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}^{N_g}\left\|\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_c\hat{\boldsymbol{\lambda}}_{c,i} - \hat{F}_{g_i}\hat{\boldsymbol{\lambda}}_{g_i,i}\right\|^2 + \frac{1}{NT}\sum_{i=1}^{N}\mathrm{tr}\left[K_iR(F_c^0, F_g^0)\right],$$

where $\frac{1}{NT}\sum_{i=1}^{N}\mathrm{tr}\{K_iR(F_c^0, F_g^0)\}$ is the bias term, contributed by the estimated observable structure $X_i\hat{\boldsymbol{\beta}}$. The penalty on the estimated factor structures will be investigated in Step 2.

Step 2: Under no factor structure, the approximated model evaluation criterion, developed in Step 1, can be used for selecting the regularization parameter κ. However, we still need an additional penalty term that penalizes the model complexity caused by the factor structures. Thus, the final model evaluation criterion for evaluating k global factor model with k_j local factors ($j = 1, ..., S$) has the form

$$\begin{aligned}&\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\left\|\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i - \hat{F}_c\hat{\boldsymbol{\lambda}}_{c,i} - \hat{F}_{g_i}\hat{\boldsymbol{\lambda}}_{g_i,i}\right\|^2 + \frac{1}{NT}\sum_{i=1}^{N}\mathrm{tr}\left[K_iR(F_c^0, F_g^0)\right]\\&+k\times h(T,N,N_1,...,N_G) + \sum_{g=1}^{G}k_g\times h_g(T,N,N_1,...,N_G)\end{aligned}$$

our goal is to find a penalty function to consistently estimate the true number of factors.

This step for consistently selecting the number of factors uses a similar augment as in Bai (2009). First, we focus on the selection of the number of global factors k given the true number of local factors $r_1, ..., r_G$. We assume that $r \le k$, where k is given number of factors in the estimation process. Under $r \le k$, we have $\hat{\boldsymbol{\beta}}_i(k) - \boldsymbol{\beta}_i = O_p(1/\sqrt{T})$, where the script k indicates k factor models are estimated. Then it is shown that

$$\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i(k) - \hat{F}_c(k)'\hat{\boldsymbol{\lambda}}_{c,i} - \hat{F}_{g_i}(r_{g_i})\hat{\boldsymbol{\lambda}}_{j,i} = \boldsymbol{\varepsilon}_i + O_p\left(T^{-1/2}\right) + O_p\left(N^{-1/2}\right),$$

which implies that

$$\begin{aligned}&\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\left\|\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i(k) - \hat{F}_c(k)\hat{\boldsymbol{\lambda}}_{c,i} - \hat{F}_{g_i}(r_{g_i})\hat{\boldsymbol{\lambda}}_{j,i}\right\|^2\\&-\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\left\|\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i(r) - \hat{F}_c(r)\hat{\boldsymbol{\lambda}}_{c,i} - \hat{F}_{g_i}(r_{g_i})\hat{\boldsymbol{\lambda}}_{g_i,i}\right\|^2 = O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{N}\right).\end{aligned}$$

If $k < r$, it can be shown that for some $c > 0$, not depending on N and T,

$$\begin{aligned}&\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\left\|\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i(k) - \hat{F}_c(k)\hat{\boldsymbol{\lambda}}_{c,i} - \hat{F}_{g_i}(r_{g_i})\hat{\boldsymbol{\lambda}}_{g_i,i}\right\|^2\\&-\frac{1}{NT}\sum_{g=1}^{G}\sum_{i;g_i=g}\left\|\boldsymbol{y}_i - X_i\hat{\boldsymbol{\beta}}_i(r) - \hat{F}_c(r)\hat{\boldsymbol{\lambda}}_{c,i} - \hat{F}_{g_i}(r_{g_i})\hat{\boldsymbol{\lambda}}_{g_i,i}\right\|^2 > c.\end{aligned}$$

These results imply that a penalty function that converges to zero but is of greater magnitude than $O_p(1/T) + O_p(1/N)$ will lead to consistent estimation of the number of factors. The function $h(T, N, N_1, ..., N_G) = (\frac{T+N}{TN}) \log(TN)$ satisfies these conditions. The penalty term for selecting the number of group-specific factors is similarly derived. This completes the proof of Theorem 3.