Focused Information Criterion and Model Averaging for Large Panels with a Multifactor Error Structure

Shou-Yung Yin* and Chu-An Liu† and Chang-Ching Lin‡

(Preliminary)

February 13, 2016

Abstract

This paper considers model selection and model averaging in panel data models with a multifactor error structure. We propose a focused information criterion and a plug-in averaging estimator for large heterogeneous panels based on the common correlated effects estimation (Pesaran, 2006). We show that the trade-off between bias and variance remains in the asymptotic theory when the rank condition is not satisfied for all submodels. Moreover, adding more regressors could also have positive or negative effects on estimation variance. Since the rank condition is non-testable, some results inferred from degenerate cases would not be useful practically. Our proposed selection criterion and averaging estimators aim to minimize the sample analog of the asymptotic mean squared error and can apply to all cases. Monte Carlo simulations show that both selection and averaging methods compare favorably with other methods.

Keywords: Cross-sectional dependence, Common correlated effects, Focused information criterion, Model averaging, Model selection.

JEL Classification: C23, C51, C52.

*Institute of Economics, Academia Sinica. Email: syyin@econ.sinica.edu.tw

†Institute of Economics, Academia Sinica. Email: caliu@econ.sinica.edu.tw.

‡Department of Economics, National Cheng Kung University. Email: ever@mail.ncku.edu.tw

1 Introduction

In the past decade, there has been an increasing interest in the study of cross-sectional dependence in panel data model. One popular approach to this problem is the common correlated effects (CCE) method proposed by Pesaran (2006). The virtue of the CCE estimation is that it can be easily computed by the least squares regression augmented with cross-sectional averages of the dependent variable and individual regressors. While the asymptotic properties of CCE estimators have been investigated, little work has considered the CCE estimation under model uncertainty.

This paper considers model selection and model averaging in panel data models with a multi-factor error structure. Following Hjort and Claeskens (2003), Claeskens and Hjort (2008), Hansen (2014), and Liu (2015), we study the asymptotic distribution of mean group estimator based on the individual-specific CCE estimators (CCEMG) in a local asymptotic framework where some auxiliary regression coefficients are in a local neighborhood of zero. It is well known that adding more regressors reduces the model bias but causes a large variance in finite sample. The local asymptotic framework has an advantage of yielding the same stochastic order of squared biases and variances. Thus, the asymptotic mean squared error remains finite and provides a good approximation to finite sample mean squared error.

One attractive advantage of the CCEMG estimation is that it continues to work even if the rank condition is not satisfied. We first consider the case where the rank condition is not satisfied for all submodels. Under the local-to-zero assumption, we derive the asymptotic distributions of submodel estimators and show that the trade-off between bias and variance remains in the asymptotic theory. In addition to the bias-variance tradeoff, we find that adding more regressors could have positive or negative effects on estimation variance. This is because that the CCEMG estimator is more efficient when the rank condition is satisfied. Therefore, it could reduce the estimation variance if we include more regressors such that the rank condition is satisfied. The problem is more complicated for some degenerate cases including the case that some submodels satisfy the rank condition and the case has weak factors. The former case implies that we should choose the full model theoretically but the latter one cannot provide any useful guidelines to choose the model. Thus obtaining optimal model seems infeasible because it is not possible to know which kind of case is empirically.

The asymptotic distributions of submodel estimators depend on the rank condition, and this must be taken into account in developing a model selection criterion. However, in practice, it is hard to verify the rank condition. In this paper, we propose a focused information criterion (FIC), which is a sample analog of the asymptotic mean squared error of submodel estimators irrespective of whether the rank condition holds or not. Therefore, the proposed selection criterion can address the bias-variance tradeoff and the positive-negative effect on estimation variance simultaneously.

Building on the idea of FIC, we introduce a frequentist model averaging criterion to select the weights for candidate models and study its properties. We first derive the asymptotic distribution of the averaging estimator with fixed weights, which allows us to characterize the asymptotic mean squared error of the averaging estimator. We then propose a criterion for weight selection and use these estimated weights to construct a plug-in averaging estimator. Similar to the model

selection, the proposed model averaging criterion is an estimate of the asymptotic mean squared error irrespective of whether the rank condition holds or not. Simulation studies show that the proposed model selection and averaging methods generally produce the lower root mean squared error as compared to other data-driven selection criteria.

We now discuss the related literature. There is a large body of literature on large panels with a multifactor error structure. The two main approaches to factor-augmented panel regressions are the correlated common effects estimators and interactive effects estimators. The correlated common effects estimator based on cross-sectional averages has been developed by Pesaran (2006), Kapetanios, Pesaran, and Yamagata (2011), Pesaran and Tosetti (2011), Chudik, Pesaran, and Tosetti (2011), Pesaran, Smith, and Yamagata (2013), and Chudik and Pesaran (2015), while the interactive effects estimator based on principal components has been developed by Stock and Watson (2002), Bai and Ng (2002), Bai (2009), Moon and Weidner (2015a), and Moon and Weidner (2015b); see Kapetanios and Pesaran (2007) and Westerlund and Urbain (2015) for a comparison of these two approaches.

The focused information criterion is introduced by Claeskens and Hjort (2003) for likelihood-based model. In recent years, FIC has been extended to several models, including the Cox hazard regression model (Hjort and Claeskens, 2006), the general semiparametric model (Claeskens and Carroll, 2007), the generalized additive partial linear model (Zhang and Liang, 2011), the Tobin model with a nonzero threshold (Zhang, Wan, and Zhou, 2012), the generalized empirical likelihood estimation (Sueishi, 2013), the generalized method of moments estimation (DiTraglia, 2014), and the propensity score weighted estimation of the treatment effects (Kitagawa and Muris, 2013; Lu, 2015). To our knowledge, the FIC has not been explored before in panel data models with a multifactor error structure.

There is a growing body of literature on frequentist model averaging, including the scores of information criteria (Buckland, Burnham, and Augustin, 1997; Hjort and Claeskens, 2003; Zhang and Liang, 2011; Zhang, Wan, and Zhou, 2012), the adaptive regression by mixing models (Yang, 2000, 2001; Yuan and Yang, 2005), the Mallows' C_p-type averaging estimator (Hansen, 2007, 2009, 2010; Wan, Zhang, and Zou, 2010; Liu and Okui, 2013; Zhang, Zou, and Liang, 2014), the optimal mean squared error averaging (Liang, Zou, Wan, and Zhang, 2011), the jackknife model averaging (Hansen and Racine, 2012; Zhang, Wan, and Zou, 2013; Lu and Su, 2015), and the plug-in averaging averaging (Liu, 2015). However, the existing literature on model averaging in factor-augmented panel regression models is comparatively small. Cheng and Hansen (2015) consider forecast combination based on the Mallows and the leave-h-out cross validation criteria for factor-augmented regression models. Gao, Zhang, Wang, and Zou (2016) propose a leave-subject-our model averaging estimator for panel data models.

The rest of the paper is organized as follows. Section 2 presents the panel data model, the submodel, and the common correlated effects estimator. Section 3 presents the asymptotic framework and assumptions. Section 4 derives the focused information criterion. Section 5 introduces the plug-in averaging estimator. Section 6 presents the results of Monte Carlo experiments, and Section 7 concludes the paper. Proofs and figures are presented in the Appendix.

Notation: For a $k \times k$ matrix $\mathbf{A}$, $\|\mathbf{A}\| = (tr(\mathbf{A}'\mathbf{A}))^{1/2}$ denotes the Euclidean norm, and $\mathbf{A}^-$ denotes its Moore-Penrose generalized inverse.

2 Model and Estimation

Suppose we have observations $(y_{it}, \mathbf{x}_{1it}, \mathbf{x}_{2it})$ for $i = 1, ..., N$ and $t = 1, ..., T$. We consider the following panel data model with a multifactor error structure:

$$y_{it} = \mathbf{x}_{1it}'\boldsymbol{\beta}_{1i} + \mathbf{x}_{2it}'\boldsymbol{\beta}_{2i} + e_{it}, \tag{2.1}$$

$$e_{it} = \boldsymbol{\gamma}_i'\mathbf{f}_t + \varepsilon_{it}, \tag{2.2}$$

where $\mathbf{x}_{1it}$ $(k_1 \times 1)$ and $\mathbf{x}_{2it}$ $(k_2 \times 1)$ are vectors of regressors, $\boldsymbol{\beta}_{1i}$ $(k_1 \times 1)$ and $\boldsymbol{\beta}_{2i}$ $(k_2 \times 1)$ are vectors of unknown coefficients, e_{it} is an error with a multifactor structure, $\boldsymbol{\gamma}_i$ is a $r \times 1$ vector of unobserved factor loadings, $\mathbf{f}_t$ is a $r \times 1$ vector of unobserved common factors so that $\boldsymbol{\gamma}_i'\mathbf{f}_t = \gamma_{i1}f_{1t} + \cdots + \gamma_{ir}f_{rt}$, and ε_{it} is an unobserved idiosyncratic errors. Here, $\mathbf{x}_{1it}$ contain the core regressors that must be included in the model based on theoretical grounds, while $\mathbf{x}_{2it}$ contain the auxiliary regressors that may or may not be included in the model.[1] Let $\boldsymbol{\beta}_i = (\boldsymbol{\beta}_{1i}', \boldsymbol{\beta}_{2i}')'$ and $k = k_1 + k_2$ be the total number of the regressors in the model (2.1).

This model includes the standard fixed-effects model as a special case when $r = 1$, $\mathbf{f}_t = 1$, and $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i. It generalizes the fixed-effects model to allow the interactive-effects between $\boldsymbol{\gamma}_i$ and $\mathbf{f}_t$. The setup is general enough to allow for the unobserved factors $\mathbf{f}_t$ to be correlated with the regressors $\mathbf{x}_{1it}$ and $\mathbf{x}_{2it}$. To allow for this possibility, we follow Pesaran (2006) and assume that

$$\mathbf{x}_{1it} = \boldsymbol{\Gamma}_{1i}'\mathbf{f}_t + \mathbf{v}_{1it}, \tag{2.3}$$

$$\mathbf{x}_{2it} = \boldsymbol{\Gamma}_{2i}'\mathbf{f}_t + \mathbf{v}_{2it}, \tag{2.4}$$

where $\boldsymbol{\Gamma}_{1i}$ and $\boldsymbol{\Gamma}_{2i}$ are $r \times k_1$ and $r \times k_2$ factor loading matrices, and $\mathbf{v}_{1it}$ and $\mathbf{v}_{2it}$ are $k_1 \times 1$ and $k_2 \times 1$ idiosyncratic errors. Let $\mathbf{v}_{it} = (\mathbf{v}_{1it}', \mathbf{v}_{2it}')'$ and assume that $\mathbf{v}_{it}$ follow general covariance stationary processes. In general, $\mathbf{v}_{1it}$ is correlated with $\mathbf{v}_{2it}$, i.e., $\mathrm{Var}(\mathbf{v}_{it})$ is not a diagonal matrix. Hence, the the core regressors $\mathbf{x}_{1it}$ are correlated with the auxiliary regressors $\mathbf{x}_{2it}$ not only due to the presence of the common factors $\mathbf{f}_t$, but also due to the correlation between $\mathbf{v}_{1it}$ and $\mathbf{v}_{2it}$.

We now consider a set of M submodels indexed by $m = 1, ..., M$. Let $\boldsymbol{\Pi}_m$ be the $k_{2m} \times k_2$ selection matrix that selects the included auxiliary regressors. The mth submodel includes all core regressors $\mathbf{x}_{1it}$ and a subset of auxiliary regressors $\mathbf{x}_{2it}$. The mth submodel has $k_m = k_1 + k_{2m}$ regressors, and we use $\mathbf{x}_{mit} = (\mathbf{x}_{1it}', \mathbf{x}_{2it}'\boldsymbol{\Pi}_m')'$ to denote the regressors included in the mth submodel. We do not place any restrictions on the model space. The set of models could be nested or non-nested. If we consider a sequence of nested models, then $M = k_2 + 1$. If we consider all possible subsets of auxiliary regressors, then $M = 2^{k_2}$.

Since the common factors $\mathbf{f}_t$ enter equations (2.2)–(2.4) simultaneously, the estimation of the slope coefficients is nontrivial. We follow Pesaran (2006) and estimate unknown slope coefficients

[1] The core regressors may only include a constant term or even an empty matrix, and the auxiliary regressors can include any nonlinear transformations of the original variables and the interaction terms between the regressors.

by common correlated effects (CCE) estimation. The idea behind the CCE approach is to use the cross-sectional averages to approximate the linear combinations of the unobserved common factors and then estimate slope coefficients by a standard panel regression augmented with these cross-sectional averages.

Let $\mathbf{I}_k$ denote an identity matrix of order k and $\mathbf{0}$ a zero matrix. We first combine equations (2.1)–(2.4) and write the full model as a system of equations

$$\mathbf{h}_{it} = \mathbf{C}_i'\mathbf{f}_t + \mathbf{u}_{it}, \tag{2.5}$$

where $\mathbf{h}_{it} = (y_{it}, \mathbf{x}_{1it}', \mathbf{x}_{2it}')'$, $\mathbf{u}_{it} = (\varepsilon_{it} + \boldsymbol{\beta}_i'\mathbf{v}_{it}, \mathbf{v}_{1it}', \mathbf{v}_{2it}')'$, and

$$\mathbf{C}_i = \begin{bmatrix} \boldsymbol{\gamma}_i & \boldsymbol{\Gamma}_{1i} & \boldsymbol{\Gamma}_{2i} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\beta}_{1i} & \mathbf{I}_{k_1} & \mathbf{0} \\ \boldsymbol{\beta}_{2i} & \mathbf{0} & \mathbf{I}_{k_2} \end{bmatrix}.$$

Similarly, for the submodel m, we have

$$\mathbf{h}_{mit} = \mathbf{C}_{mi}'\mathbf{f}_t + \mathbf{u}_{mit}, \tag{2.6}$$

where $\mathbf{h}_{mit} = (y_{it}, \mathbf{x}_{1it}', \mathbf{x}_{2it}'\boldsymbol{\Pi}_m')'$, $\mathbf{u}_{mit} = (\varepsilon_{it} + \boldsymbol{\beta}_i'\mathbf{v}_{it}, \mathbf{v}_{1it}', \mathbf{v}_{2it}'\boldsymbol{\Pi}_m')'$, and

$$\mathbf{C}_{mi} = \begin{bmatrix} \boldsymbol{\gamma}_i + \boldsymbol{\Gamma}_{2i}(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i} & \boldsymbol{\Gamma}_{1i} & \boldsymbol{\Gamma}_{2i}\boldsymbol{\Pi}_m' \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\beta}_{1i} & \mathbf{I}_{k_1} & \mathbf{0} \\ \boldsymbol{\Pi}_m\boldsymbol{\beta}_{2i} & \mathbf{0} & \mathbf{I}_{k_{2m}} \end{bmatrix}.$$

Define $\bar{\mathbf{Z}}_i = N^{-1}\sum_{i=1}^N \mathbf{Z}_i$ be the cross-sectional average of any variable $\mathbf{Z}_i$. After taking the cross-sectional averages of the equation (2.6), we have

$$\bar{\mathbf{h}}_{mt} = \bar{\mathbf{C}}_m'\mathbf{f}_t + \bar{\mathbf{u}}_{mt}, \tag{2.7}$$

where $\bar{\mathbf{h}}_{mt}$, $\bar{\mathbf{C}}_m$, and $\bar{\mathbf{u}}_{mt}$ are the cross-sectional averages of $\mathbf{h}_{mit}$, $\mathbf{C}_{mi}$, and $\mathbf{u}_{mit}$, respectively. This equation motivates us to use the cross-sectional averages $\bar{\mathbf{h}}_{mt}$ as proxies for unobserved common factors $\mathbf{f}_t$ since $\bar{\mathbf{u}}_{mt} \xrightarrow{p} 0$ as $N \to \infty$ under regularity conditions. Thus, the slope coefficients $\boldsymbol{\beta}_i$ can be consistently estimated by the least squares regression augmented with cross-sectional averages of the dependent variable and individual regressors.

In matrix notation, we write the model (2.1)–(2.2) as

$$\mathbf{y}_i = \mathbf{X}_{1i}\boldsymbol{\beta}_{1i} + \mathbf{X}_{2i}\boldsymbol{\beta}_{2i} + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i \tag{2.8}$$

where $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$, $\mathbf{X}_i = (\mathbf{X}_{1i}, \mathbf{X}_{2i})$, $\mathbf{X}_{1i} = (\mathbf{x}_{1i1}, ..., \mathbf{x}_{1iT})'$, $\mathbf{X}_{2i} = (\mathbf{x}_{2i1}, ..., \mathbf{x}_{2iT})'$, $\mathbf{F} = (\mathbf{f}_1, ..., \mathbf{f}_T)'$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$. Let $\bar{\mathbf{H}} = (\bar{\mathbf{y}}, \bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2)$ be the cross-sectional averages of the dependent and independent variables, i.e., $\bar{\mathbf{y}}$, $\bar{\mathbf{X}}_1$, and $\bar{\mathbf{X}}_2$ are the cross-sectional averages of $\mathbf{y}_i$, $\mathbf{X}_{1i}$, and $\mathbf{X}_{2i}$, respectively.

The unconstrained CCE estimator of $\boldsymbol{\beta}_i$ in the full model, i.e., all auxiliary regressors are included in the model, is

$$\hat{\boldsymbol{\beta}}_{ui} = (\mathbf{X}_i'\mathbf{M}_h\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{M}_h\mathbf{y}_i, \tag{2.9}$$

$$\mathbf{M}_h = \mathbf{I}_T - \bar{\mathbf{H}}(\bar{\mathbf{H}}'\bar{\mathbf{H}})^{-}\bar{\mathbf{H}}', \tag{2.10}$$

and the CCE estimator in the submodel m is

$$\hat{\boldsymbol{\beta}}_{mi} = (\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{y}_i, \tag{2.11}$$

$$\mathbf{M}_{hm} = \mathbf{I}_T - \bar{\mathbf{H}}_m(\bar{\mathbf{H}}_m'\bar{\mathbf{H}}_m)^{-}\bar{\mathbf{H}}_m', \tag{2.12}$$

where $\mathbf{X}_{mi} = (\mathbf{X}_{1i}, \mathbf{X}_{2i}\boldsymbol{\Pi}_m')$ and $\bar{\mathbf{H}}_m = (\bar{\mathbf{y}}, \bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2\boldsymbol{\Pi}_m')$.

The model (2.8) allows the slope coefficients to be heterogeneous over i such that $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i$ with $\boldsymbol{\eta}_i$ being i.i.d.. In this paper, the parameter of interest is $\boldsymbol{\beta}$, which is the cross-sectional mean of the unknown slope coefficient $\boldsymbol{\beta}_i$. The unknown parameter $\boldsymbol{\beta}$ can be consistently estimated by a simple average of the individual CCE estimators. The common correlated effects mean group (CCEMG) estimator of $\boldsymbol{\beta}$ in the full model is

$$\hat{\boldsymbol{\beta}}_{\mathrm{MG},u} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_{ui}, \tag{2.13}$$

and the CCEMG estimator in the submodel m is

$$\hat{\boldsymbol{\beta}}_{\mathrm{MG},m} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_{mi}. \tag{2.14}$$

Remark 1. In Pesaran (2006), the cross-sectional average is defined by $\bar{\mathbf{y}} = \sum_{i=1}^{N}\lambda_i\mathbf{y}_i$ with the weights λ_i satisfying (1) $\lambda_i = O(N^{-1})$, (2) $\sum_{i=1}^{N}\lambda_i = 1$, and (3) $\sum_{i=1}^{N}|\lambda_i|$ is bounded. Note that the choice of the weights does not affect the asymptotic distributions of CCE and CCEMG estimators. As suggested by Pesaran (2006), one could use the equal weights when the sample size is reasonably large. Thus, we consider $\lambda_i = 1/N$ in this paper for simplicity.

3 Asymptotic Theory

In this section, we study the limiting distribution of the CCEMG estimator of $\boldsymbol{\beta}$ for the submodel m. In the first subsection, we describe the asymptotic framework and technical assumptions. In the second subsection, we present the asymptotic distribution of the CCEMG estimator in a general case. In the third subsection, we study the asymptotic distribution of the CCEMG estimator in several degenerate cases.

3.1 Assumptions

We now state the assumptions.

Assumption 1. The individual-specific errors ε_{it} and $\mathbf{v}_{it}$ are independently and identically distributed (i.i.d.) across both i and t. For each i, ε_{it} and $\mathbf{v}_{it}$ follow linear stationary processes with absolute summable autocovariances,

$$\varepsilon_{it} = \sum_{\ell=0}^{\infty} a_{i\ell}\zeta_{i,t-\ell}, \quad \mathrm{E}(\varepsilon_{it}) = 0, \quad \mathrm{Var}(\varepsilon_{it}) = \sigma_i^2 \leq \bar{\sigma}^2 < \infty,$$

and

$$\mathbf{v}_{it} = \sum_{\ell=0}^{\infty} \boldsymbol{\alpha}_{i\ell}\boldsymbol{\nu}_{i,t-\ell}, \quad \mathrm{E}(\mathbf{v}_{it}) = \mathbf{0}, \quad \mathrm{Var}(\mathbf{v}_{it}) = \boldsymbol{\Sigma}_i \leq \bar{\boldsymbol{\Sigma}} < \infty,$$

where $\boldsymbol{\Sigma}_i$ is a positive definite matrix, and $(\zeta_{it}, \boldsymbol{\nu}'_{i,t})'$ are $(k+1)\times 1$ vectors of i.i.d. random variables with mean zero, identity covariance matrix, and finite fourth-order cumulants.

Assumption 2. The vector of common factors $\mathbf{f}_t$ is covariance stationary and distributed independently of the individual-specific errors ε_{it} and $\mathbf{v}_{it}$ for all i and t.

Assumption 3. The factor loadings $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Gamma}_i = (\boldsymbol{\Gamma}_{1i}, \boldsymbol{\Gamma}_{2i})$ are i.i.d. across i with fixed means $\boldsymbol{\gamma}$ and $\boldsymbol{\Gamma}$, respectively, and finite variances, and distributed independently of ε_{jt}, $\mathbf{v}_{jt}$, and $\mathbf{f}_t$ for all i, j, and t. For $i = 1, ..., N$,

$$\begin{aligned} \boldsymbol{\gamma}_i &= \boldsymbol{\gamma} + \boldsymbol{\iota}_i, \quad \boldsymbol{\iota}_i \sim \text{i.i.d.}(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\gamma}}), \\ \boldsymbol{\Gamma}_i &= \boldsymbol{\Gamma} + \boldsymbol{\xi}_i, \quad \mathrm{vec}(\boldsymbol{\xi}_i) \sim \text{i.i.d.}(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\Gamma}}), \end{aligned}$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2)$, $\boldsymbol{\xi}_i = (\boldsymbol{\xi}_{1i}, \boldsymbol{\xi}_{2i})$, $\boldsymbol{\Omega}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\Omega}_{\boldsymbol{\Gamma}}$ are symmetric nonnegative definite matrices, and $||\boldsymbol{\Omega}_{\boldsymbol{\gamma}}||$ and $||\boldsymbol{\Omega}_{\boldsymbol{\Gamma}}||$ are bounded.

Assumption 4. The slope coefficient $\boldsymbol{\beta}_{1i}$ follows the random coefficient model

$$\boldsymbol{\beta}_{1i} = \boldsymbol{\beta}_1 + \boldsymbol{\eta}_{1i}, \quad \boldsymbol{\eta}_{1i} \sim \text{i.i.d.}(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\beta}_1}),$$

where $\boldsymbol{\Omega}_{\boldsymbol{\beta}_1}$ is a symmetric nonnegative definite matrix and $||\boldsymbol{\Omega}_{\boldsymbol{\beta}_1}||$ is bounded. The random deviation $\boldsymbol{\eta}_{1i}$ is distributed independently of $\boldsymbol{\gamma}_j$, $\boldsymbol{\Gamma}_j$, ε_{jt}, $\mathbf{v}_{jt}$, and $\mathbf{f}_t$ for all i, j, and t.

Assumption 5. Suppose that $\sqrt{N}\Delta_{NT}^{-1} \to \kappa < \infty$ as $N, T \to \infty$ jointly. The slope coefficient $\boldsymbol{\beta}_{2i}$ follows

$$\boldsymbol{\beta}_{2i} = \Delta_{NT}^{-1}\boldsymbol{\delta}_i, \quad \boldsymbol{\delta}_i = \boldsymbol{\delta} + \boldsymbol{\eta}_{\boldsymbol{\delta},i}, \quad \boldsymbol{\eta}_{\boldsymbol{\delta},i} \sim \text{i.i.d.}(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\delta}}),$$

where $\boldsymbol{\Omega}_{\boldsymbol{\delta}}$ is a symmetric nonnegative definite matrix and $||\boldsymbol{\Omega}_{\boldsymbol{\delta}}||$ is bounded. The random deviation $\boldsymbol{\eta}_{\boldsymbol{\delta},i}$ is distributed independently of $\boldsymbol{\gamma}_j$, $\boldsymbol{\Gamma}_j$, ε_{jt}, $\mathbf{v}_{jt}$, and $\mathbf{f}_t$ for all i, j, and t.

Assumption 1 specifies the individual-specific errors are i.i.d. and imposes some moment conditions. Assumption 2 assumes that the common factors are covariance stationary. Assumptions 3 and 4 impose the random coefficient structure on the factor loadings and the slope coefficients. Assumptions 1–4 are similar to Assumptions 1–4 of Pesaran (2006).

Assumption 5 assumes that $\boldsymbol{\beta}_{2i}$ is local to zero, and it ensures that the asymptotic mean squared error of the submodel estimator remains finite. It is a common technique to analyze the asymptotic and finite sample properties of the model selection and averaging estimator, for example, Hjort and Claeskens (2003), Leeb and Pötscher (2005), Hansen (2014), and Liu (2015). Note that Assumption 5 implies that $\mathrm{E}(\boldsymbol{\beta}_{2i}) = \Delta_{NT}^{-1}\boldsymbol{\delta} \equiv \boldsymbol{\beta}_2$. The assumption says the partial correlations between the dependent variable and the auxiliary regressors are weak for all i, and the correlations will vanish as $T \to \infty$ without specifying the convergence rate. For example, if $\Delta_{NT} = O(T)$, then Assumption 5 holds when $N = O(T^2)$.[2]

3.2 General Case

We now introduce some notations that we will use to characterize the limiting distribution. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2')'$ be the cross-sectional means of the slope coefficients defined in Assumptions 4–5, and $\boldsymbol{\beta}_m = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2'\boldsymbol{\Pi}_m')' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_{2m}')'$. Define $\mathbf{Q}_{mi} = p\lim_{T\to\infty}(T^{-1}\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i)$ and $\boldsymbol{\Sigma}_{mi} = p\lim_{T\to\infty}(T^{-1}\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{F})$, where $\mathbf{M}_{gm} = \mathbf{I}_T - \bar{\mathbf{G}}_m(\bar{\mathbf{G}}_m'\bar{\mathbf{G}}_m)^{-}\bar{\mathbf{G}}_m'$ and $\bar{\mathbf{G}}_m = \mathbf{F}\bar{\mathbf{C}}_m$. Let

$$\mathbf{S}_0 = \begin{bmatrix} \mathbf{0}_{k_1\times k_2} \\ \mathbf{I}_{k_2} \end{bmatrix} \quad \text{and} \quad \mathbf{S}_m = \begin{bmatrix} \mathbf{I}_{k_1} & \mathbf{0}_{k_1\times k_{2m}} \\ \mathbf{0}_{k_2\times k_1} & \boldsymbol{\Pi}_m' \end{bmatrix}$$

be selection matrices of dimension $k \times k_2$ and $k \times (k_1 + k_{2m})$, respectively.

Theorem 1. Suppose that Assumptions 1–5 hold. As $N, T \to \infty$ jointly, we have

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{\mathrm{MG},m} - \boldsymbol{\beta}_m) \xrightarrow{d} \mathbf{A}_m\boldsymbol{\delta} + \mathbf{U}_m + \mathbf{V}_m \sim \mathrm{N}\left(\mathbf{A}_m\boldsymbol{\delta}, \boldsymbol{\Xi}_m\right),$$

where $\mathbf{A}_m = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\boldsymbol{\Sigma}_i\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\kappa$, $\boldsymbol{\Xi}_m = \boldsymbol{\Xi}_{um} + \boldsymbol{\Xi}_{vm}$, $\boldsymbol{\Xi}_{um} = \mathbf{S}_m'\boldsymbol{\Omega}_{\boldsymbol{\beta}}\mathbf{S}_m$, and $\boldsymbol{\Xi}_{vm} = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\boldsymbol{\Sigma}_{mi}\boldsymbol{\Omega}_{\boldsymbol{\gamma}}\boldsymbol{\Sigma}_{mi}'\mathbf{S}_m\left(\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}$, and $\mathbf{U}_m$ and $\mathbf{V}_m$ are two stochastically independent normal random vector with mean zero and covariance matrices $\boldsymbol{\Xi}_{um}$ and $\boldsymbol{\Xi}_{vm}$, respectively.

Proof. See the appendix. □

Theorem 1 presents the asymptotic distribution of the CCEMG estimator for each submodel and implies that the submodel estimate $\hat{\boldsymbol{\beta}}_{\mathrm{MG},m}$ is consistent. As shown in the proof of Theorem 1,

[2] Unlike Hansen (2014), and Liu (2015), which assume Δ_{NT} is equal to $\sqrt{N}$, we allow the rate of convergence in a more general setting as long as the condition $\sqrt{N}\Delta_{NT}^{-1} \to \kappa$ is satisfied.

we can decompose the CCEMG estimator for the mth model as

$$\begin{aligned}\sqrt{N}(\hat{\boldsymbol{\beta}}_{\mathrm{MG},m}-\boldsymbol{\beta}_m) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{S}_m'\boldsymbol{\eta}_i+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{S}_m'\hat{\mathbf{Q}}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\left(\frac{\mathbf{V}_i'\mathbf{M}_{hm}\mathbf{V}_i}{T}\right)\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i}\\ &+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{S}_m'\hat{\mathbf{Q}}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\hat{\boldsymbol{\Sigma}}_{mi}\left(\boldsymbol{\gamma}_i+\boldsymbol{\Gamma}_i\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i}\right)\\ &+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{S}_m'\hat{\mathbf{Q}}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\left(\frac{\boldsymbol{\Gamma}_i'\mathbf{F}'\mathbf{M}_{hm}\mathbf{V}_i}{T}\right)\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i}\\ &+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{S}_m'\hat{\mathbf{Q}}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\boldsymbol{\varepsilon}_i}{T},\end{aligned}$$

where $\hat{\mathbf{Q}}_{mi}=T^{-1}\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i$ and $\hat{\boldsymbol{\Sigma}}_{mi}=T^{-1}\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}$. In the proof, we show that the first term converges to the normal random vector $\mathbf{U}_m$ and the second term converges to the asymptotic bias $\mathbf{A}_m\boldsymbol{\delta}$. The third term converges to the normal random vector $\mathbf{V}_m$ and the last two terms are a small order term.

It is well known that adding more regressors reduces the model bias but causes a large variance in finite sample. Theorem 1 shows the trade-off between omitted variable bias and estimation variance remains in the asymptotic theory. For the full model, the asymptotic bias is zero since $(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)=\mathbf{0}$. For the submodels, the asymptotic bias is zero if $\boldsymbol{\beta}_2=\mathbf{0}$ or $\boldsymbol{\Sigma}_i$ is a diagonal matrix. Note that the asymptotic bias comes from the fact that the core regressors $\mathbf{x}_{1it}$ and the auxiliary regressors $\mathbf{x}_{2it}$ are correlated, and the correlation is due to the common factors $\mathbf{f}_t$ and the correlation between $\mathbf{v}_{1it}$ and $\mathbf{v}_{2it}$.

Corollary 1. Suppose that Assumptions 1–5 hold. Assume that $\boldsymbol{\Sigma}_i$ is a diagonal matrix for all i. As $N,T\to\infty$ jointly, we have

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{\mathrm{MG},m}-\boldsymbol{\beta}_m)\xrightarrow{d}\mathbf{U}_m+\mathbf{V}_m\sim\mathrm{N}\left(\mathbf{0},\boldsymbol{\Xi}_m\right),$$

where $\mathbf{U}_m$, $\mathbf{V}_m$, and $\boldsymbol{\Xi}_m$ are defined in Theorem 1.

Proof. See the appendix. □

Corollary 1 shows that there is no asymptotic bias for the submodel estimate when the covariance matrix of individual-specific errors $\mathbf{v}_{it}$ is a diagonal matrix. The intuition behind Corollary 1 is that we are able to filter the individual-specific regressors by the cross-sectional averages such that the bias from the omitted auxiliary regressors are eliminated when $\mathbf{x}_{1it}$ and $\mathbf{x}_{2it}$ are correlated via the common factors only. This result corresponds to the case of the zero partial correlations between the dependent variable and the auxiliary regressors, i.e. $\boldsymbol{\delta}=\mathbf{0}$. In this case, there is no trade-off between omitted variable bias and estimation variance. Thus it implies that we only have positive-negative effects on estimation variance.

3.3 Degenerate Cases

In this subsection, we study the asymptotic distribution of CCEMG estimator for several special cases.

Assumption 6. (Rank condition) Let $\bar{\mathbf{C}}_m = \frac{1}{N}\sum_{i=1}^{N}\mathbf{C}_{mi}$, where $\mathbf{C}_{mi} = \begin{bmatrix} \boldsymbol{\gamma}_i & \boldsymbol{\Gamma}_i\mathbf{S}_m \end{bmatrix}\begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{S}_m'\boldsymbol{\beta}_i & \mathbf{I}_{k_1+k_{2m}} \end{bmatrix}$, and $\text{Rank}(\bar{\mathbf{C}}_m) = r$.

We replace the original assumption only for the full model with the assumption for each sub-model. This condition plays an important role of proving that whether the factor structure can be eliminated efficiently. In the main result of Pesaran (2006), $\bar{\mathbf{C}}_m$ is constructed by using all core and auxiliary regressors, which means $\bar{\mathbf{C}}_m = \bar{\mathbf{C}}$. Thus the most efficient CCEMG estimator requires that the rank of $\bar{\mathbf{C}}$ should be r, the number of factors. This condition is non-testable, and therefore we cannot know how efficient of these sub-models are. In the following, we highlight two cases that will receive attention, including cases that allows that the correlation of $\mathbf{x}_{1it}$ and $\mathbf{x}_{2it}$ comes from the the factor structure, and allows the correlation also comes from the correlated specified error terms.

Corollary 2. Under assumptions 1–6, we have

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{\text{MG},m} - \boldsymbol{\beta}_m) \xrightarrow{d} \mathbf{A}_m\boldsymbol{\delta} + \mathbf{U}_m \sim \text{N}\left(\mathbf{A}_m\boldsymbol{\delta}, \boldsymbol{\Xi}_{um}\right),$$

where $\mathbf{A}_m$, $\mathbf{U}_m$ and $\boldsymbol{\Xi}_{um}$ are defined in Theorem 1.

Proof. See the appendix. □

According to this result, the trade-off between the bias and variance depends on the rank assumption. More specifically, when the rank condition is satisfied for some sub-models, it implies that the full model satisfies the rank condition as well, and there is no trade-off when adding the auxiliary regressors. Consequently the full model tends achieve the smallest AMSE because there is no bias. While there exists weak factors, a traditional trade-off arises because there is no benefit to eliminate the factor structure when adding useless regressors but these regressors do reduce the bias, and we will discuss it later.

However, the efficiency of CCEMG estimator depends on the hold of rank condition for each submodel. When the rank condition is satisfied, the covariance matrix consists of only one term but two terms when the rank condition is not satisfied. This implies that the CCEMG estimator is more efficient when the rank condition is satisfied. Furthermore, despite the non-random selection matrix $\mathbf{S}_m$, the covariance matrix will asymptotically equivalent for common parameters of interest in these sub-models satisfying the rank condition as $N, T \to \infty$.

For the case when the rank condition is violated, the second term, $\mathbf{V}_m$, is crucial to determine the efficiency in any two sub-models because it changes by a possible benefit from eliminating the additional factors by different auxiliary regressors in $\boldsymbol{\Sigma}_{mi}$ and the cost in $\mathbf{Q}_{mi}$. Note that the Assumption 5 is not a necessary condition when $k_1 + k_{2m} + 1$ is greater than r for each submodel

because the factors might be weak to these core or auxiliary regressors. Let us state the following definitions, for the ℓ-th factor, the factor is strong when the corresponding loading $\Gamma_{1,i\ell}$ and $\Gamma_{2,i\ell}$ obey the following rules,

$$\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}|\Gamma_{1,i\ell}| = c > 0 \text{ and } \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}|\Gamma_{2,i\ell}| = c > 0, \tag{3.1}$$

where c is positive number, a factor is weak when

$$\lim_{N\to\infty} \sum_{i=1}^{N}|\Gamma_{1,i\ell}| = c < \infty \text{ and } \lim_{N\to\infty} \sum_{i=1}^{N}|\Gamma_{2,i\ell}| = c < \infty. \tag{3.2}$$

The above definitions can be found in Chudik, Pesaran, and Tosetti (2011). Following the definition of weak factors, it is easily to obtain both $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}|\Gamma_{1,i\ell}|$ and $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}|\Gamma_{2,i\ell}|$ converges to zero. Accordingly, the existence of weak factors implies that for each submodel,[3]

$$\text{Rank}\left(\frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix}\boldsymbol{\gamma}_i & \boldsymbol{\Gamma}_i\mathbf{S}_m\end{bmatrix}\begin{bmatrix}1 & \mathbf{0}\\ \mathbf{S}_m'\boldsymbol{\beta}_i & \mathbf{I}_{k_1+k_{2m}}\end{bmatrix}\right) \xrightarrow{p} \text{Rank}\left(\begin{bmatrix}\boldsymbol{\gamma} & \boldsymbol{\Gamma}_m\end{bmatrix}\begin{bmatrix}1 & \mathbf{0}\\ \mathbf{S}_m'\boldsymbol{\beta} & \mathbf{I}_{k_1+k_{2m}}\end{bmatrix}\right) = r_m \le r.$$

Given the discussions above, it is clear that adding the auxiliary regressors is not necessarily ensure that the factors can be eliminated efficiently. Therefore, by taking account of this condition, for the special case of two nested models, we can have the following asymptotic property.

Corollary 3. Under assumptions 1–4, for any nested selected models, m_1-th and m_2-th models. Let the regressors selected in the m_1-th model belong into the m_2-th model, and the adding regressor(s) for m_2 is(are) redundant caused by weak factors, then we have the following result for any two nested sub-models,

$$\mathbb{P}\left[\lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{Q}_{m_1i}^{-1}\boldsymbol{\Sigma}_{m_1i}\boldsymbol{\Omega}_{\boldsymbol{\gamma}}\boldsymbol{\Sigma}_{m_1i}'\mathbf{Q}_{m_1i}^{-1}\right]_{\ell\ell} \le \lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{Q}_{m_2i}^{-1}\boldsymbol{\Sigma}_{m_2i}\boldsymbol{\Omega}_{\boldsymbol{\gamma}}\boldsymbol{\Sigma}_{m_2i}'\mathbf{Q}_{m_2i}^{-1}\right]_{\ell\ell}\right] = 1.$$

where $\ell\ell$ denotes the ℓ-th element on the diagonal of the covariance matrix.

Proof. See the appendix. □

[3]For illustration, considering the following example with 5 factors,

$$\bar{\mathbf{x}}_{2t} \xrightarrow{p} \begin{bmatrix}\Gamma_{2,11} & \Gamma_{2,12} & \Gamma_{2,13} & \Gamma_{2,14} & \Gamma_{2,15}\\ \Gamma_{2,21} & 0 & 0 & 0 & 0\\ \Gamma_{2,31} & 0 & 0 & 0 & 0\\ \Gamma_{2,41} & 0 & 0 & 0 & 0\\ \Gamma_{2,51} & 0 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\mathrm{f}_{1,t}\\ \mathrm{f}_{2,t}\\ \mathrm{f}_{3,t}\\ \mathrm{f}_{4,t}\\ \mathrm{f}_{5,t}\end{bmatrix}.$$

This example shows that the rank is less than 5, which implies that $\bar{\mathbf{X}}_2$ cannot span the original 5 factors. Thus, after using the cross-sectional averages of dependent variable and regressors to filter out the factors, the remains are still correlated with $\mathbf{X}_{2i}$.

4 Focused Information Criterion

In this section, we propose a focused information criterion for the panel data model with a multifactor error structure. Let $\mu = \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \mu(\boldsymbol{\beta})$ be a smooth real-valued function. Let $\mathbf{D}_{\boldsymbol{\beta}} = \partial\mu/\partial\boldsymbol{\beta}$ with partial derivatives evaluated at the null points $(\boldsymbol{\beta}_1', \mathbf{0}')'$. Assume that the partial derivatives are continuous in a neighborhood of the null points. Unlike the traditional model selection and model averaging approaches, which assess the global fit of the model, we evaluate the model based on the focus parameter μ. For example, μ may be an individual coefficient or a ratio of two coefficients of regressors. Let $\hat{\mu}_m = \mu(\hat{\boldsymbol{\beta}}_{\mathrm{MG},m})$ denote the submodel estimates. The following theorem shows the asymptotic distribution of $\hat{\mu}_m$.

Theorem 2. Suppose that Assumptions 1–5 hold. As $N, T \to \infty$ jointly, we have

$$\sqrt{N}(\mu(\hat{\boldsymbol{\beta}}_{\mathrm{MG},m}) - \mu(\boldsymbol{\beta})) \xrightarrow{d} \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{B}_m\boldsymbol{\delta} + \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m(\mathbf{U}_m + \mathbf{V}_m) \sim \mathrm{N}\left(\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{B}_m\boldsymbol{\delta}, \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m\boldsymbol{\Xi}_m\mathbf{S}_m'\mathbf{D}_{\boldsymbol{\beta}}\right),$$

where $\mathbf{B}_m = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \left(\mathbf{P}_{mi}\mathbf{Q}_{mi} - \mathbf{I}\right)\mathbf{S}_0\kappa$ and $\mathbf{P}_{mi} = \mathbf{S}_m\left(\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'$.

Proof. See the appendix. □

As can been seen in this theorem, the focused parameters can be expressed by the same random vectors, $\mathbf{U}_m$ and $\mathbf{V}_m$ which converge to two random normal vectors, with a bias term. It can be used to directly calculate the AMSE as the following equation,

$$\mathrm{AMSE}(\mu_m) = \mathbf{D}_{\boldsymbol{\beta}}'\left(\mathbf{B}_m\boldsymbol{\delta}\boldsymbol{\delta}'\mathbf{B}_m' + \mathbf{S}_m\boldsymbol{\Xi}_m\mathbf{S}_m'\right)\mathbf{D}_{\boldsymbol{\beta}}. \tag{4.1}$$

where $\boldsymbol{\Xi}_m$ is defined in Theorem 1 and 1.

The corresponding unknown parameters, $\mathbf{D}_{\boldsymbol{\beta}}$ can be consistently estimated, $\boldsymbol{\Xi}_m$ can be consistently estimated by equation (4.2), and $\mathbf{B}_m\boldsymbol{\delta}$ can be consistently estimated by using the consistent estimates of $\mathbf{P}_{mi}$, $\mathbf{Q}_{mi}$, and asymptotically unbiased estimator, $N^{-1/2}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_{2i} = \sqrt{N}\hat{\boldsymbol{\beta}}_2 = \kappa\hat{\boldsymbol{\delta}}$ under the full model.

Remark 2. $\boldsymbol{\Xi}_m$ can be estimated by

$$\hat{\boldsymbol{\Xi}}_m = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\beta}}_{mi} - \hat{\boldsymbol{\beta}}_{\mathrm{MG},m})(\hat{\boldsymbol{\beta}}_{mi} - \hat{\boldsymbol{\beta}}_{\mathrm{MG},m})' \tag{4.2}$$

in all cases.

5 Plug-In Averaging Estimator

In this section, we propose a frequentist model averaging estimator for the panel data model with a multifactor error structure.

To obtain a robust estimate of $\hat{\mu}$ for taking account of the trade-off between bias and variance, and the presence of common correlated effects, an average estimator of the focus parameter, μ, can

be defined as

$$\bar{\mu}(\mathbf{w}) = \sum_{m=1}^{M} w_m \hat{\mu}_m = \sum_{m=1}^{M} w_m \mu(\hat{\boldsymbol{\beta}}_{\mathrm{MG},m}), \tag{5.1}$$

where $w_m \geq 0$ and $\sum_{m=1}^{M} w_m = 1$ and the weight vector $\mathbf{w} = (w_1, ..., w_M)'$.

Now, we turn to the plug-in averaging estimator under the same framework of focused parameters by a given weight vector $\mathbf{w} = (w_1, ..., w_M)'$. The following theorem shows that the asymptotic properties of the averaging estimator with a fixed weight vector. We then have the following theorem.

Theorem 3. Suppose that Assumptions 1–5 hold. As $N, T \to \infty$ jointly, we have

$$\sqrt{N}(\bar{\mu}(\mathbf{w}) - \mu(\boldsymbol{\beta})) \xrightarrow{d} \mathrm{N}\left(\sum_{m=1}^{M} w_m \mathbf{D}_{\boldsymbol{\beta}}' \mathbf{B}_m \boldsymbol{\delta}, \boldsymbol{\Xi}(\mathbf{w})\right),$$

where $\boldsymbol{\Xi}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{D}_{\boldsymbol{\beta}}' \mathbf{S}_m \boldsymbol{\Xi}_m \mathbf{S}_m' \mathbf{D}_{\boldsymbol{\beta}} + 2 \sum \sum_{m \neq \ell} w_m w_\ell \mathbf{D}_{\boldsymbol{\beta}}' \mathbf{S}_m \boldsymbol{\Xi}_{m\ell} \mathbf{S}_\ell' \mathbf{D}_{\boldsymbol{\beta}}$, and $\boldsymbol{\Xi}_{m\ell} = \mathbb{E}(\mathbf{U}_m \mathbf{U}_\ell') + \mathbb{E}(\mathbf{V}_m \mathbf{V}_\ell')$.

Proof. See the appendix. □

According to this theorem, we can have the following AMSE of the averaging estimator with a fixed weight vector,

$$\mathrm{AMSE}(\mathbf{w}) = \mathbf{w}' \boldsymbol{\Psi} \mathbf{w},$$

where $\boldsymbol{\Psi}_{m\ell} = \mathbf{D}_{\boldsymbol{\beta}}' \left(\mathbf{B}_m \boldsymbol{\delta}\boldsymbol{\delta}' \mathbf{B}_\ell' + \mathbf{S}_m \boldsymbol{\Xi}_{m\ell} \mathbf{S}_\ell'\right) \mathbf{D}_{\boldsymbol{\beta}}$, the (m, ℓ) element of $\boldsymbol{\Psi}$; the optimal weight vector can be obtained by numerically minimizing the AMSE($\mathbf{w}$). Similarly, the consistently estimated parameters are the same as we used in AMSE(μ_m).

Remark 3. $\boldsymbol{\Xi}_{m\ell}$ can be estimated by

$$\hat{\boldsymbol{\Xi}}_{m\ell} = \frac{1}{N-1} \sum_{i=1}^{N} (\hat{\boldsymbol{\beta}}_{mi} - \hat{\boldsymbol{\beta}}_{\mathrm{MG},m})(\hat{\boldsymbol{\beta}}_{\ell i} - \hat{\boldsymbol{\beta}}_{\mathrm{MG},\ell})' \tag{5.2}$$

in all cases. The inference can be found in the Appendix.

Theorem 4. Suppose that Assumptions 1–5 hold. Let

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}' \hat{\boldsymbol{\Psi}} \mathbf{w},$$

where $\hat{\boldsymbol{\Psi}}_{m\ell} = \mathbf{D}_{\boldsymbol{\beta}}' \left(\hat{\mathbf{B}}_m \hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}' \hat{\mathbf{B}}_\ell' + \mathbf{S}_m \hat{\boldsymbol{\Xi}}_{m\ell} \mathbf{S}_\ell\right) \mathbf{D}_{\boldsymbol{\beta}}$, and $\hat{\Psi}_{m\ell}$ is the (m, ℓ) element of $\hat{\boldsymbol{\Psi}}$. Then we have

$$\hat{\mathbf{w}} \xrightarrow{d} \mathbf{w}^* = \arg\min_{\mathbf{w}} \mathbf{w}' \boldsymbol{\Psi} \mathbf{w},$$

and

$$\sqrt{N}(\bar{\mu}(\hat{\mathbf{w}}) - \mu(\boldsymbol{\beta})) \xrightarrow{d} \sqrt{N}(\bar{\mu}(\mathbf{w}^*) - \mu(\boldsymbol{\beta})).$$

Proof. The proof is similar to that of Theorem 3 of Liu (2015). □

6 Simulation Study

In this section, we use the simulation to investigate the small sample property. Basically, the main regression follows a linear panel data model with the factor structure

$$y_{it} = \sum_{j=1}^{k_1} x_{1,jit}\beta_{1ji} + \sum_{k=1}^{k_2} x_{2,kit}\beta_{2ki} + \mathbf{f}_t'\boldsymbol{\gamma}_i + \varepsilon_{it}. \tag{6.1}$$

where the regression coefficients are determined by the following rule,

$$\begin{aligned}
\beta_{1j} &= \frac{1}{3}, \quad \beta_{1ji} = \beta_{1j} + \mathrm{N}(0,1) \quad \forall j \\
\delta_k &= (k_2 - k + 1)/k_2, \quad k = 1, ..., k_2, \text{ and } \delta_{ki} = \delta_k + \mathrm{N}(0,1) \quad \forall \ell \\
\beta_{2ki} &= \delta_{ki}/\sqrt{N}.
\end{aligned}$$

Particularly, we let

$$\begin{aligned}
\mathbf{x}_{1it} &= \mathbf{f}_t'\boldsymbol{\Gamma}_{1i} + \mathbf{v}_{1it} \\
\mathbf{x}_{2it} &= \mathbf{f}_t'\boldsymbol{\Gamma}_{2i} + \mathbf{v}_{2it}.
\end{aligned}$$

All factors, $\mathbf{f}_t$ are generated from the normal distribution $N(0, \sigma_f^2)$, where $\sigma_f = \{0.5, 1, 1.5\}$ to control the importance of factors. The corresponding factor loadings $\boldsymbol{\gamma}_i$ are generated from the normal distribution $N(0.5, 1)$. The same setting is used for $\boldsymbol{\Gamma}_{1i}$ and $\boldsymbol{\Gamma}_{2i}$. The error term for the main regression follows $\sqrt{r} \times \mathrm{N}(0,1)$, which is similar to the setup used in Cheng and Hansen (2014). The error term of core regressors and auxiliary regressors follow $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{12})$, where the diagonal elements of $\boldsymbol{\Sigma}_{12}$ are $\sqrt{r}$ and off-diagonal elements are $\sqrt{r} \times \rho$. The number of core regressors is 3 whereas we consider three numbers of auxiliary regressors which are 3, 4 and 5, respectively. We also consider the following three different specifications,

DGP1 $\rho = 0$ which implies that the correlation of $\mathbf{x}_{1it}$ and $\mathbf{x}_{2it}$ comes from the factor structure.

DGP2 $\rho = 0.5$ which implies that the correlation of $\mathbf{x}_{1it}$ and $\mathbf{x}_{2it}$ comes from the factor structure and the error term $\mathbf{v}_{1it}$ and $\mathbf{v}_{2it}$.

DGP3 $\rho = 0.5$ and $\boldsymbol{\Gamma}_{2i}$ is a sparse matrix, that is only first three rows of $\boldsymbol{\Gamma}_{2i}$ follows $N(0.5, 1)$ while the rest are 0, which implies that all auxiliary regressors are contributed by the first three factors.

In the simulations, we consider three types of estimators: (1) optimal model selection estimators including AIC (Akaike information criterion) and BIC (Bayesian information criterion), (2) averaging estimators including S-AIC (smoothed Akaike information criterion model averaging estimator), S-BIC (smoothed Bayesian information criterion model averaging estimator), simple average estimator and Plug-in averaging estimator, and (3) naive models including narrow model

and full model. The criteria of AIC and BIC for each model are defined as the following:

$$\mathrm{AIC}_m = \sum_{i=1}^{N} T \log(\hat{\sigma}_i^2(m)) + 2N(2k_1 + 2k_{2m} + 1),$$
$$\mathrm{BIC}_m = \sum_{i=1}^{N} T \log(\hat{\sigma}_i^2(m)) + \log(TN)N(2k_1 + 2k_{2m} + 1),$$

where $\hat{\sigma}_i^2(m) = (\hat{\varepsilon}_i(m)'\hat{\varepsilon}_i(m))/(T - 2k_1 - 2k_{2m} - 1)$ and $\hat{\varepsilon}_i(m)$ is the residual from the sub-model m. Accordingly, the optimal model is chosen by finding the smallest value of the criterion used. The S-AIC and S-BIC are based on the values of exponential AIC and exponential BIC and suggest assigning the weights for each candidate model. The weights is defined as $\hat{w}_m^{\mathrm{AIC}} = \exp\left(-\frac{1}{2}\mathrm{AIC}_m\right)/\sum_{j=1}^{M}\exp\left(-\frac{1}{2}\mathrm{AIC}_j\right)$ for S-AIC and $\hat{w}_m^{\mathrm{BIC}} = \exp\left(-\frac{1}{2}\mathrm{BIC}_m\right)/\sum_{j=1}^{M}\exp\left(-\frac{1}{2}\mathrm{BIC}_j\right)$ for S-BIC respectively.

The results are summarized by evaluating the normalized risk(expected squared errors) calculated by averaging across 1,000 replications in figures 1-3 for the sample size $(N, T) = (50, 50)$ and figures 4-6 for the sample size $(N, T) = (100, 50)$, respectively. Three different numbers of auxiliary regressors are considered and plotted in figures 1-3, and also in figures 4-6. In each figure, we consider three different DGPs mentioned above and plot the normalized risks against different numbers of factors, 3–12. In sum, we can find that when the factors are more important (larger σ of $\mathbf{F}$), the more divergent normalized risks we have in these four estimators. When considering different numbers of factors with DGP1 (no correlation between $\mathbf{v}_{1it}$ and $\mathbf{v}_{2it}$), see part (a) of figures 1 and 4, the figure shows that when the number of factors increases, the full model dominates other estimators except the number of factors is larger and the σ of $\mathbf{F}$ is larger while the plug-in averaging estimator has a smaller normalized risk under this situation. This is consistent with our results that when the number of factors is larger than the sum of p and q, we have two opposite effects on variance. The simple average estimator shares a similar pattern with the plug-in averaging estimator.

7 Conclusion

Many studies have revealed the importance of taking account of the trade-off between bias and model uncertainty by adopting model average approach. However, such research focus only on one dimensional data (time-series or cross-sectional data). In this paper, we fill this gap by extending the model average to the panel data framework with multifactor error structure. The novel feature of our estimator is that it avoids the cost of inappropriately selecting the useless auxiliary regressors which are not helpful for eliminating factors, and takes the possible bias caused by ignoring regressors at the same time. Our Monte Carlo simulations show that the proposed estimator has satisfactory and stable normalized risk compared with other estimators and it is robust against different settings of the number of factors and of the variance of factors.

Appendix

A Proofs

Lemma 1. Let m_1 and m_2 are the two selected sets of regressors and are nested to each other. The following results hold when the regressor set of m_1 belongs into the set of m_2.

$$\left(\mathbf{X}'_{m_1 i}\mathbf{M}_{hm}\mathbf{X}_{m_1 i}\right)^{-1}_{\ell\ell} \leq \left(\mathbf{X}'_{m_2 i}\mathbf{M}_{hm}\mathbf{X}_{m_2 i}\right)^{-1}_{\ell\ell}.$$

$\ell\ell$ denotes the ℓ-th element of diagonal.

Proof. We let

$$\mathbf{X}_{m_2 i} = \begin{bmatrix} \mathbf{X}_{m_1 i} & \mathbf{X}_{m_2 \backslash m_1 i} \end{bmatrix},$$

where $\mathbf{X}_{m_2 \backslash m_1 i}$ denotes that the set of regressors included in model m_2 excludes the regressors included in model m_1, and for any annihilator matrix $\mathbf{M}_{hm}$, $m = 1, ..., M$, we have

$$\begin{aligned}\left(\mathbf{X}'_{m_2 i}\mathbf{M}_{hm}\mathbf{X}_{m_2 i}\right)^{-1} &= \left(\begin{bmatrix} \mathbf{X}'_{m_1 i} \\ \mathbf{X}'_{m_2 \backslash m_1 i} \end{bmatrix} \mathbf{M}_{hm} \begin{bmatrix} \mathbf{X}_{m_1 i} & \mathbf{X}_{m_2 \backslash m_1 i} \end{bmatrix}\right)^{-1} \\ &= \begin{bmatrix} \mathbf{X}'_{m_1 i}\mathbf{M}_{hm}\mathbf{X}_{m_1 i} & \mathbf{X}'_{m_1 i}\mathbf{M}_{hm}\mathbf{X}_{m_2 \backslash m_1 i} \\ \mathbf{X}'_{m_2 \backslash m_1 i}\mathbf{M}_{hm}\mathbf{X}_{m_1 i} & \mathbf{X}'_{m_2 \backslash m_1 i}\mathbf{M}_{hm}\mathbf{X}_{m_2 \backslash m_1 i} \end{bmatrix}^{-1}.\end{aligned}$$

By using the simple algebra, we can obtain

$$\left(\mathbf{X}'_{m_2 i}\mathbf{M}_{hm}\mathbf{X}_{m_2 i}\right)^{-1} = \begin{bmatrix} \tilde{\mathbb{A}}_{11}^{-1} + \tilde{\mathbb{A}}_{11}^{-1}\tilde{\mathbb{A}}_{12}\tilde{\mathbb{B}}\tilde{\mathbb{A}}_{21}\tilde{\mathbb{A}}_{11}^{-1} & -\tilde{\mathbb{A}}_{11}^{-1}\tilde{\mathbb{A}}_{12}\tilde{\mathbb{B}} \\ -\tilde{\mathbb{B}}\tilde{\mathbb{A}}_{21}\tilde{\mathbb{A}}_{11}^{-1} & \tilde{\mathbb{B}} \end{bmatrix},$$

where $\tilde{\mathbb{A}}_{11} = \mathbf{X}'_{m_1 i}\mathbf{M}_{hm}\mathbf{X}_{m_1 i}$, $\tilde{\mathbb{A}}_{12} = \mathbf{X}'_{m_1 i}\mathbf{M}_{hm}\mathbf{X}_{m_2 \backslash m_1 i}$ and $\tilde{\mathbb{B}} = \left(\mathbf{X}'_{m_2 \backslash m_1 i}\mathbf{M}_{hm}\mathbf{X}_{m_2 \backslash m_1 i} - \tilde{\mathbb{A}}_{21}\tilde{\mathbb{A}}_{11}^{-1}\tilde{\mathbb{A}}_{12}\right)^{-1}$. Since $\tilde{\mathbb{B}}$ is positive-semidefinite matrix, the upper-left block of matrix $\left(\mathbf{X}'_{m_2 i}\mathbf{M}_{hm}\mathbf{X}_{m_2 i}\right)^{-1}$ has the following property,

$$\tilde{\mathbb{A}}_{11}^{-1} + \tilde{\mathbb{A}}_{11}^{-1}\tilde{\mathbb{A}}_{12}\tilde{\mathbb{B}}\tilde{\mathbb{A}}_{21}\tilde{\mathbb{A}}_{11}^{-1} \geq \tilde{\mathbb{A}}_{11}^{-1},$$

which completes the proof of the second result.

□

Lemma 2. Suppose that Assumptions 1–5 hold.

1. $\frac{\mathbf{X}'_i\mathbf{M}_{hm}\mathbf{X}_i}{T} = \frac{\mathbf{X}'_i\mathbf{M}_{gm}\mathbf{X}_i}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)$.
2. $\frac{\mathbf{X}'_i\mathbf{M}_{hm}\boldsymbol{\varepsilon}_i}{T} = \frac{\mathbf{X}'_i\mathbf{M}_{gm}\boldsymbol{\varepsilon}_i}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)$.
3. $\frac{\mathbf{V}'_i\mathbf{M}_{hm}\mathbf{V}_i}{T} = \frac{\mathbf{V}'_i\mathbf{M}_{gm}\mathbf{V}_i}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)$.

4. $\frac{\mathbf{F}'\mathbf{M}_{hm}\mathbf{V}_i}{T} = \frac{\mathbf{F}'\mathbf{M}_{gm}\mathbf{V}_i}{T} + O_p\left(\frac{1}{\sqrt{T}}\right)$.

5. $\frac{\sqrt{N}\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}}{T}\left(\bar{\boldsymbol{\gamma}} + \bar{\boldsymbol{\Gamma}}_2(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2\right) = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$.

Proof. The results 1–4 from lemma 2 can be obtained by using Lemmas 2 and 3 from Pesaran (2006). For the last result, recall the matrix representation of each submodel m,

$$\begin{bmatrix} y_{it} \\ \mathbf{x}_{1it} \\ \boldsymbol{\Pi}_m\mathbf{x}_{2it} \end{bmatrix} = \begin{bmatrix} 1 & \boldsymbol{\beta}_{1i}' & \boldsymbol{\beta}_{2i}'\boldsymbol{\Pi}_m' \\ \mathbf{0} & \mathbf{I}_{k_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{k_{2m}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma}_i' + \boldsymbol{\beta}_{2i}'(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\Gamma}_{2i}' \\ \boldsymbol{\Gamma}_{1i}' \\ \boldsymbol{\Pi}_m\boldsymbol{\Gamma}_{2i}' \end{bmatrix} \mathbf{f}_t + \begin{bmatrix} \varepsilon_{it} + \boldsymbol{\beta}_{1i}'\mathbf{v}_{1it} + \boldsymbol{\beta}_{2i}'\mathbf{v}_{2it} \\ \mathbf{v}_{1it} \\ \boldsymbol{\Pi}_m\mathbf{v}_{2it} \end{bmatrix}.$$

After taking the cross-sectional averages under the equal weight, we have

$$\bar{\mathbf{h}}_{mt} = \bar{\mathbf{C}}_m'\mathbf{f}_t + \bar{\mathbf{u}}_{mt}.$$

By stacking all observations over t, we have

$$\begin{aligned}[\bar{\mathbf{y}}, \bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2\boldsymbol{\Pi}_m] &= \mathbf{F}[\bar{\boldsymbol{\gamma}} + \overline{\boldsymbol{\Gamma}_2\boldsymbol{\beta}_2} + \overline{\boldsymbol{\Gamma}_1\boldsymbol{\beta}_1}, \bar{\boldsymbol{\Gamma}}_1, \bar{\boldsymbol{\Gamma}}_2\boldsymbol{\Pi}_m'] + \bar{\mathbf{U}}_m \\ &= \mathbf{F}\bar{\mathbf{C}}_m + \bar{\mathbf{U}}_m = \bar{\mathbf{G}}_m + \bar{\mathbf{U}}_m\end{aligned}$$

and since $\mathbf{M}_{gm} = \mathbf{I}_T - \bar{\mathbf{G}}_m(\bar{\mathbf{G}}_m'\bar{\mathbf{G}}_m)^{-}\bar{\mathbf{G}}_m'$, using the similar result from Pesaran (2006), we have

$$\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}}{T} = \frac{\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{F}}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).$$

When Assumption 6 holds, $\mathbf{M}_{gm}\mathbf{F} = \mathbf{0}$, but still have $\mathbf{M}_{gm}\bar{\mathbf{G}}_m = \mathbf{0}$ even Assumption 6 is violated. By the fact $\mathbf{M}_{gm}\bar{\mathbf{G}}_m = \mathbf{0}$, we can obtain

$$\mathbf{M}_{gm}\mathbf{F}(\bar{\boldsymbol{\gamma}} + \overline{\boldsymbol{\Gamma}_2\boldsymbol{\beta}_2} + \overline{\boldsymbol{\Gamma}_1\boldsymbol{\beta}_1}) = \mathbf{0}, \quad \mathbf{M}_{gm}\mathbf{F}\bar{\boldsymbol{\Gamma}}_1 = \mathbf{0}, \quad \mathbf{M}_{gm}\mathbf{F}\bar{\boldsymbol{\Gamma}}_2\boldsymbol{\Pi}_m' = \mathbf{0},$$

then

$$\begin{aligned}\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}}{T}(\bar{\boldsymbol{\gamma}} + \overline{\boldsymbol{\Gamma}_2\boldsymbol{\beta}_2} + \overline{\boldsymbol{\Gamma}_1\boldsymbol{\beta}_1}) &= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \\ \frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}}{T}\bar{\boldsymbol{\Gamma}}_1 &= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \\ \frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}}{T}\bar{\boldsymbol{\Gamma}}_2\boldsymbol{\Pi}_m' &= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).\end{aligned}$$

The above results imply that

$$\begin{aligned}&\frac{\sqrt{N}\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}}{T}(\bar{\boldsymbol{\gamma}} + \overline{\boldsymbol{\Gamma}_2\boldsymbol{\beta}_2} + \overline{\boldsymbol{\Gamma}_1\boldsymbol{\beta}_1}) \\ &= \frac{\sqrt{N}\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}}{T}\left(\bar{\boldsymbol{\gamma}} + \bar{\boldsymbol{\Gamma}}_2\boldsymbol{\beta}_2 + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{2i}\Delta_{NT}\boldsymbol{\eta}_{\delta,i} + \bar{\boldsymbol{\Gamma}}_1\boldsymbol{\beta}_1 + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{1i}\boldsymbol{\eta}_{1i}\right) \\ &= \frac{\sqrt{N}\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}}{T}\left(\bar{\boldsymbol{\gamma}} + \bar{\boldsymbol{\Gamma}}_2(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 + \bar{\boldsymbol{\Gamma}}_2\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\boldsymbol{\beta}_2 + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{2i}\Delta_{NT}^{-1}\boldsymbol{\eta}_{\delta,i} + \bar{\boldsymbol{\Gamma}}_1\boldsymbol{\beta}_1 + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{1i}\boldsymbol{\eta}_{1i}\right) \\ &= \frac{\sqrt{N}\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}}{T}\left(\bar{\boldsymbol{\gamma}} + \bar{\boldsymbol{\Gamma}}_2(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{2i}\Delta_{NT}^{-1}\boldsymbol{\eta}_{\delta,i} + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{1i}\boldsymbol{\eta}_{1i}\right) \\ &= O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),\end{aligned}$$

where $\boldsymbol{\beta}_2 = \Delta_{NT}^{-1}\boldsymbol{\delta}$. Since $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{2i}\Delta_{NT}\boldsymbol{\eta}_{\delta,i} = O_p(1/N)$ and $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{1i}\boldsymbol{\eta}_{1i} = O_p(1/\sqrt{N})$, which in turn leads to

$$\frac{\sqrt{N}\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}}{T}\left(\bar{\boldsymbol{\gamma}} + \bar{\boldsymbol{\Gamma}}_2(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2\right) = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).$$

□

Proof of Theorem 1. Note that $\mathbf{X}_{mi} = (\mathbf{X}_{1i}, \mathbf{X}_{2i}\boldsymbol{\Pi}_m') = \mathbf{X}_i\mathbf{S}_m$. We first rewrite $\hat{\boldsymbol{\beta}}_{mi}$ as follows,

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{mi} =&(\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{y}_i\\
=&(\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\mathbf{M}_{hm}(\mathbf{X}_{1i}\boldsymbol{\beta}_{1i} + \mathbf{X}_{2i}\boldsymbol{\beta}_{2i} + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i)\\
=&(\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\mathbf{M}_{hm}(\mathbf{X}_{1i}\boldsymbol{\beta}_{1i} + \mathbf{X}_{2i}\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\boldsymbol{\beta}_{2i} + \mathbf{X}_{2i}\boldsymbol{\beta}_{2i} - \mathbf{X}_{2i}\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\boldsymbol{\beta}_{2i} + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i)\\
=&(\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\mathbf{M}_{hm}(\mathbf{X}_{1i}\boldsymbol{\beta}_{1i} + \mathbf{X}_{2i}\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\boldsymbol{\beta}_{2i}) + (\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{X}_{2i}(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i}\\
&+ (\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{F}\boldsymbol{\gamma}_i + (\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\mathbf{M}_{hm}\boldsymbol{\varepsilon}_i\\
=&\boldsymbol{\beta}_{mi} + (\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{X}_{2i}(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i} + (\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{F}\boldsymbol{\gamma}_i\\
&+ (\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\mathbf{M}_{hm}\boldsymbol{\varepsilon}_i\\
=&\boldsymbol{\beta}_{mi} + (\mathbf{S}_m'\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i\mathbf{S}_m)^{-1}\mathbf{S}_m'\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i} + (\mathbf{S}_m'\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i\mathbf{S}_m)^{-1}\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{F}\boldsymbol{\gamma}_i\\
&+ (\mathbf{S}_m'\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i\mathbf{S}_m)^{-1}\mathbf{S}_m'\mathbf{X}_i'\mathbf{M}_{hm}\boldsymbol{\varepsilon}_i,
\end{aligned}$$

then since $\hat{\boldsymbol{\beta}}_{\mathrm{MG},m} = N^{-1}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_{mi}$ and $\mathbf{X}_i = \mathbf{F}\boldsymbol{\Gamma}_i + \mathbf{V}_i$ along with the above decomposition, we can further have

$$\begin{aligned}
&\sqrt{N}\left(\hat{\boldsymbol{\beta}}_{\mathrm{MG},m} - \boldsymbol{\beta}_m\right)\\
&\quad= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_{mi} - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\beta}_m\\
&\quad= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{S}_m'\boldsymbol{\eta}_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\left(\frac{\mathbf{V}_i'\mathbf{M}_{hm}\mathbf{V}_i}{T}\right)\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i}\\
&\qquad+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}}{T}\left(\boldsymbol{\gamma}_i + \boldsymbol{\Gamma}_i\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i}\right)\\
&\qquad+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\left(\frac{\boldsymbol{\Gamma}_i'\mathbf{F}'\mathbf{M}_{hm}\mathbf{V}_i}{T}\right)\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i}\\
&\qquad+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\boldsymbol{\varepsilon}_i}{T}\\
&\quad=: I_1 + I_2 + I_3 + I_4 + I_5 \qquad \text{(A.1)}
\end{aligned}$$

For simplicity, we first discuss the first and fifth terms from (A.1). As for I_1, by the random coefficients specification from Assumption 4, we can easily have

$$I_1 = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{S}_m'\boldsymbol{\eta}_i \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{S}_m'\boldsymbol{\Omega}_{\boldsymbol{\beta}}\mathbf{S}_m) = \mathbf{U}_m. \qquad \text{(A.2)}$$

Consider I_5, since ε_{it} is independent of $\mathbf{v}_{1it}$, $\mathbf{v}_{2it}$ and $\mathbf{f}_t$ for all i, t, and using the fact $T^{-1}\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i = O_p(1)$ implied by Lemma 21. Moreover, since $\mathbb{E}(I_5) = \mathbf{0}$ and

$$\mathbf{S}_m' \frac{\mathbf{X}_i'\mathbf{M}_{hm}\boldsymbol{\varepsilon}_i}{T} = O_p\left(\frac{1}{\sqrt{T}}\right),$$

we have

$$I_5 = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{A.3}$$

Next, we show the asymptotic properties of the second term from (A.1). Since the result stated by Lemma 23, we have $T^{-1}\mathbf{V}_i'\mathbf{M}_{hm}\mathbf{V}_i = T^{-1}\mathbf{V}_i'\mathbf{M}_{gm}\mathbf{V}_i + O_p(N^{-1}) + O_p((NT)^{-1/2})$. In particular

$$\frac{\mathbf{V}_i'\mathbf{M}_{gm}\mathbf{V}_i}{T} = \frac{\mathbf{V}_i'\mathbf{V}_i}{T} - \frac{\mathbf{V}_i'\mathbf{F}\bar{\mathbf{C}}_m}{T}\left(\frac{\bar{\mathbf{C}}_m'\mathbf{F}'\mathbf{F}\bar{\mathbf{C}}_m}{T}\right)^{-}\frac{\bar{\mathbf{C}}_m'\mathbf{F}'\mathbf{V}_i}{T} = \frac{\mathbf{V}_i'\mathbf{V}_i}{T} + O_p\left(\frac{1}{T}\right).$$

The last equality holds because $T^{-1}(\mathbf{F}'\mathbf{V}_i)$ is $O_p(T^{-1/2})$ by the facts $\mathbf{v}_{it}$ is independent of $\mathbf{f}_t$, $\mathbf{f}_t$ is covariance stationary, and $\bar{\mathbf{C}}_m$ is bounded. Let $\boldsymbol{\eta}_{\boldsymbol{\delta}^c,i} = (\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\eta}_{\boldsymbol{\delta},i}$ representing a $(k_1 + k_2) \times 1$ vector consisting of the random deviations of $\boldsymbol{\delta}_i$ from the excluded auxiliary regressors at the corresponding positions and zero otherwise. Therefore, by the assumption that $\boldsymbol{\eta}_{\boldsymbol{\delta}^c,i}$ is independent of the covariance matrix constructed by selected regressors, we can firstly have

$$\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\mathbf{S}_0\boldsymbol{\eta}_{\boldsymbol{\delta}^c,i} = O_p\left(\frac{1}{\sqrt{N}}\right),$$

then I_2 follows that

$$\begin{aligned}
&\frac{\sqrt{N}}{N}\sum_{i=1}^{N}\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{V}_i'\mathbf{M}_{hm}\mathbf{V}_i}{T}\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i} \\
=&\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\sqrt{N}\Delta_{NT}^{-1}(\boldsymbol{\delta} + \boldsymbol{\eta}_{\boldsymbol{\delta},i}) \\
&+ O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) \\
\xrightarrow{p}& \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\boldsymbol{\Sigma}_i\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\kappa\boldsymbol{\delta} = \mathbf{A}_m\boldsymbol{\delta},
\end{aligned} \tag{A.4}$$

where $\sqrt{N}\Delta_{NT}^{-1} \to \kappa$ by Assumption 5.

Now, we consider I_3,

$$\begin{aligned}
I_3 =& \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\frac{\sqrt{N}}{T}\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{F}\left(\bar{\boldsymbol{\gamma}} + \bar{\boldsymbol{\Gamma}}_2(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 + (\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}}) + \right. \\
&\left.\bar{\boldsymbol{\Gamma}}_2(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\Delta_{NT}^{-1}\boldsymbol{\eta}_{\delta,i} + (\boldsymbol{\xi}_{2i} - \bar{\boldsymbol{\xi}}_2)(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 + (\boldsymbol{\xi}_{2i} - \bar{\boldsymbol{\xi}}_2)(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\Delta_{NT}^{-1}\boldsymbol{\eta}_{\delta,i}\right) \\
=& \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\frac{\sqrt{N}}{T}\mathbf{S}_m'\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{F}(\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}}) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \\
\xrightarrow{d}& \mathrm{N}(\mathbf{0}, \boldsymbol{\Xi}_{vm}) = \mathbf{V}_m.
\end{aligned} \tag{A.5}$$

The result holds because of the result from Lemma 25, $\sqrt{N}T^{-1}\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}\left(\bar{\boldsymbol{\gamma}}+\bar{\boldsymbol{\Gamma}}_2(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2\right)=O_p\left(N^{-1/2}\right)+O_p\left(T^{-1/2}\right)$ along with the fact $\sqrt{N}\boldsymbol{\beta}_2=O(1)$ and $(\boldsymbol{\xi}_{2i}-\bar{\boldsymbol{\xi}}_2)(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)$ is independent of selected regressors. In particular, similar to equation (60) in Pesaran (2006), we can have $\boldsymbol{\Xi}_{vm}=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\boldsymbol{\Sigma}_{mi}\boldsymbol{\Omega}_{\boldsymbol{\gamma}}\boldsymbol{\Sigma}_{mi}'\mathbf{S}_m\left(\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}$.

As for term I_4, by using the result from lemma 24, we have $T^{-1}(\boldsymbol{\Gamma}_i'\mathbf{F}'\mathbf{M}_{hm}\mathbf{V}_i)=T^{-1}(\boldsymbol{\Gamma}_i'\mathbf{F}'\mathbf{M}_{gm}\mathbf{V}_i)+O_p(T^{-1/2})$. Furthermore, since $\mathbf{v}_{it}$ is independent of $\mathbf{f}_t$ for all t and $\mathbf{G}_m=\mathbf{F}\bar{\mathbf{C}}_m$, which implies that

$$\frac{\mathbf{F}'\mathbf{M}_{gm}\mathbf{V}_i}{T}=\frac{\mathbf{F}'\mathbf{V}_i}{T}-\frac{\mathbf{F}'\mathbf{F}}{T}\bar{\mathbf{C}}_m\left(\bar{\mathbf{C}}_m'\frac{\mathbf{F}'\mathbf{F}}{T}\bar{\mathbf{G}}_m\right)^{-}\bar{\mathbf{C}}_m'\frac{\mathbf{F}'\mathbf{V}_i}{T}=O_p\left(\frac{1}{\sqrt{T}}\right).$$

The result holds because $\bar{\mathbf{C}}_m$ is bounded and $T^{-1}\mathbf{F}'\mathbf{V}_i=O_p(T^{-1/2})$. Since $\sqrt{N}\boldsymbol{\beta}_{2i}$ is $O(1)$,

$$\begin{aligned}
I_4&=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\left(\frac{\boldsymbol{\Gamma}_i'\mathbf{F}'\mathbf{M}_{hm}\mathbf{V}_i}{T}\right)\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i}\\
&=\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\left(\frac{\boldsymbol{\Gamma}_i'\mathbf{F}'\mathbf{M}_{gm}\mathbf{V}_i}{T}\right)\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\sqrt{N}\boldsymbol{\beta}_{2i}+O_p\left(\frac{1}{\sqrt{T}}\right)\\
&=O_p\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}\tag{A.6}$$

This completes the proof by combining equations (A.2)–(A.6). □

Proof of Corollary 1. Recall equation (A.1) and notice that terms I_1, I_3–I_5 are not influenced by the hold of the assumption on $\boldsymbol{\Sigma}_i$ whereas I_2 is. Using the result derived from equation (A.4) and for notation simplicity we let $\mathbf{V}_{im}=\mathbf{V}_i\mathbf{S}_m$ and $\mathbf{V}_{i2}^c=\mathbf{V}_i\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)$, we have

$$\begin{aligned}
&\frac{\sqrt{N}}{N}\sum_{i=1}^{N}\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{V}_i'\mathbf{M}_{gm}\mathbf{V}_i}{T}\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i}\\
=&\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{V}_i'\mathbf{V}_i}{T}\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\sqrt{N}\boldsymbol{\beta}_{2i}+O_p\left(\frac{1}{T}\right)\\
=&\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\frac{\mathbf{V}_{im}'\mathbf{V}_{i2}^c}{T}\sqrt{N}\boldsymbol{\beta}_{2i}+O_p\left(\frac{1}{T}\right)\\
=&O_p\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

The last equality holds because the covariance matrix of $\mathrm{Var}(\mathbf{v}_{it})=\boldsymbol{\Sigma}_i$ is a diagonal matrix, which implies that $T^{-1}\mathbf{V}_{im}'\mathbf{V}_{i2}^c$ is $O_p\left(T^{-1/2}\right)$. Then I_2 is $O_p\left(T^{-1/2}\right)$, this completes the proof. □

Proof of Corollary 2. Recall equation (A.1), since the hold of rank condition (Assumption 5) only has the influence on eliminating the factors, only I_2 and I_3 are affected. When Assumption 5 holds, we have $\mathbf{M}_{gm}=\mathbf{I}_T-\bar{\mathbf{G}}_m(\bar{\mathbf{G}}_m'\bar{\mathbf{G}}_m)^{-}\bar{\mathbf{G}}_m'=\mathbf{I}_T-\mathbf{F}(\mathbf{F}'\mathbf{F})^{-}\mathbf{F}'=\mathbf{M}_{\mathbf{f}}$. Then $\mathbf{M}_{\mathbf{f}}\mathbf{F}=\mathbf{0}$ follows. Therefore, following the above result and a similar argument from equations (A.4) and (A.5), I_3 and I_4 are $o_p(1)$. We complete the proof. □

Proof of Corollary 3. For the m_1-th submodel we let

$$\bar{\mathbf{C}}'_{m_1} = \begin{bmatrix} \bar{\boldsymbol{\gamma}}' + \overline{\boldsymbol{\Gamma}'_{2i}\boldsymbol{\beta}'_{2i}} + \overline{\boldsymbol{\Gamma}'_{1i}\boldsymbol{\beta}'_{1i}} \\ \bar{\boldsymbol{\Gamma}}'_1 \\ \boldsymbol{\Pi}_{m_1}\bar{\boldsymbol{\Gamma}}'_2 \end{bmatrix}.$$

The adding regressor(s) is(are) redundant because they are weakly contributed by the factors (the condition is stated in equation (3.1)), which implies that

$$(\boldsymbol{\Pi}'_{m_2}\boldsymbol{\Pi}_{m_2} - \boldsymbol{\Pi}'_{m_1}\boldsymbol{\Pi}_{m_1})\bar{\boldsymbol{\Gamma}}'_2 \xrightarrow{p} \mathbf{0}_{(k_2\times r)},$$

as $N \to \infty$. Accordingly, we have

$$\bar{\mathbf{C}}'_{m_2} \xrightarrow{p} \begin{bmatrix} \bar{\mathbf{C}}'_{m_1} \\ \mathbf{0} \end{bmatrix}$$

for the m_2-th submodel. Thus

$$\begin{aligned}
\mathbf{M}_{m_2q} &= \mathbf{I}_T - \bar{\mathbf{G}}_{m_2}(\bar{\mathbf{G}}'_{m_2}\bar{\mathbf{G}}_{m_2})^{-}\bar{\mathbf{G}}'_{m_2} \\
&\xrightarrow{p} \mathbf{I}_T - \mathbf{F}\begin{bmatrix} \bar{\mathbf{C}}_{m_1} & \mathbf{0} \end{bmatrix}\left(\begin{bmatrix} \bar{\mathbf{C}}'_{m_1} \\ \mathbf{0} \end{bmatrix}\mathbf{F}'\mathbf{F}\begin{bmatrix} \bar{\mathbf{C}}_{m_1} & \mathbf{0} \end{bmatrix}\right)^{-}\begin{bmatrix} \bar{\mathbf{C}}'_{m_1} \\ \mathbf{0} \end{bmatrix}\mathbf{F}' \\
&= \mathbf{I}_T - \bar{\mathbf{G}}_{m_1}(\bar{\mathbf{G}}'_{m_1}\bar{\mathbf{G}}_{m_1})^{-}\bar{\mathbf{G}}'_{m_1} \\
&= \mathbf{M}_{m_1q}.
\end{aligned}$$

Together with the result form Lemma 1, $\left(\mathbf{X}'_{m_1i}\mathbf{M}_{hm}\mathbf{X}_{m_1i}\right)^{-1}_{\ell\ell} \leq \left(\mathbf{X}'_{m_2i}\mathbf{M}_{hm}\mathbf{X}_{m_2i}\right)^{-1}_{\ell\ell}$, the proof is done. □

Proof of Theorem 4.1. Let $\boldsymbol{\beta}^c_{2m}$ is the rest of selected $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_{2m}$,

$$\begin{aligned}
\mu(\boldsymbol{\beta}) =& \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \boldsymbol{\beta}^c_{2m}) = \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \mathbf{0}) + \frac{\partial\mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \mathbf{0})'}{\partial\boldsymbol{\beta}^c_{2m}}(\boldsymbol{\beta}^c_{2m}) + O\left(\frac{1}{\Delta^2_{NT}}\right) \\
=& \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \mathbf{0}) + \frac{\partial\mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \mathbf{0})'}{\partial\boldsymbol{\beta}_2}(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 + O\left(\frac{1}{\Delta^2_{NT}}\right) \\
=& \mu(\boldsymbol{\beta}_m) + \mathbf{D}'_{\boldsymbol{\beta}_2}(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 + O\left(\frac{1}{\Delta^2_{NT}}\right).
\end{aligned} \tag{A.7}$$

Then we have $\mu(\boldsymbol{\beta}) - \mu(\boldsymbol{\beta}_m) = \mathbf{D}'_{\boldsymbol{\beta}_2}(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 + O\left(\Delta^{-2}_{NT}\right)$.

By using delta method and the result from Theorem 1, under the Assumptions used in Theorem 1, we can obtain

$$\sqrt{N}\left(\mu(\hat{\boldsymbol{\beta}}_{\mathrm{MG},m}) - \mu(\boldsymbol{\beta}_m)\right) \xrightarrow{d} \mathbf{D}'_{\boldsymbol{\beta}_m}\left(\mathbf{A}_m\boldsymbol{\delta} + \mathbf{U}_m + \mathbf{V}_m\right). \tag{A.8}$$

Combining equation (A.8) and the result of equation (A.7), we have

$$
\begin{aligned}
&\sqrt{N}\left(\mu(\hat{\boldsymbol{\beta}}_{\mathrm{MG},m})-\mu(\boldsymbol{\beta})\right)=\sqrt{N}\left(\mu(\hat{\boldsymbol{\beta}}_{\mathrm{MG},m})-\mu(\boldsymbol{\beta}_m)\right)-\sqrt{N}\Big(\mu(\boldsymbol{\beta})-\mu(\boldsymbol{\beta}_m)\Big)\\
&\xrightarrow{d}\mathbf{D}'_{\boldsymbol{\beta}_m}(\mathbf{A}_m\boldsymbol{\delta}+\mathbf{U}_m+\mathbf{V}_m)-\mathbf{D}'_{\boldsymbol{\beta}_2}(\mathbf{I}_{k_2}-\boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\kappa\boldsymbol{\delta}\\
&=\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_m(\mathbf{U}_m+\mathbf{V}_m)+\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_m\lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{S}'_m\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}'_m\mathbf{Q}_{mi}\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\kappa\boldsymbol{\delta}\right)\\
&\quad-\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_0(\mathbf{I}_{k_2}-\boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\kappa\boldsymbol{\delta}\\
&=\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_m(\mathbf{U}_m+\mathbf{V}_m)+\left[\lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_m\left(\mathbf{S}'_m\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}'_m\mathbf{Q}_{mi}\mathbf{S}_0-\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_0\right)\right]\\
&\quad\times(\mathbf{I}_{k_2}-\boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\kappa\boldsymbol{\delta}\\
&=\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_m(\mathbf{U}_m+\mathbf{V}_m)+\left[\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_m\left(\mathbf{S}'_m\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}'_m\mathbf{Q}_{mi}\mathbf{S}_0-\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_0\right)\right]\kappa\boldsymbol{\delta}\\
&\quad-\left[\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_m\left(\mathbf{S}'_m\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}'_m\mathbf{Q}_{mi}\mathbf{S}_0-\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_0\right)\boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m\right]\kappa\boldsymbol{\delta}\\
&=\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_m(\mathbf{U}_m+\mathbf{V}_m)+\left[\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{P}_{mi}\mathbf{Q}_{mi}\mathbf{S}_0-\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_0\right)\right]\kappa\boldsymbol{\delta}\\
&\quad-\left[\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{P}_{mi}\mathbf{Q}_{mi}\mathbf{S}_0-\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_0\right)\boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m\right]\kappa\boldsymbol{\delta}\\
&=\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_m(\mathbf{U}_m+\mathbf{V}_m)+\left[\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{D}'_{\boldsymbol{\beta}}(\mathbf{P}_{mi}\mathbf{Q}_{mi}-\mathbf{I}_{k_1+k_2})\mathbf{S}_0\right)\right]\kappa\boldsymbol{\delta}\\
&=\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_m(\mathbf{U}_m+\mathbf{V}_m)+\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{B}_m\boldsymbol{\delta}\sim\mathrm{N}(\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{B}_m\boldsymbol{\delta},\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_m\boldsymbol{\Xi}_m\mathbf{S}'_m\mathbf{D}_{\boldsymbol{\beta}}),
\end{aligned}
$$

where $\mathbf{P}_{mi}=\mathbf{S}_m\left(\mathbf{S}'_m\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}'_m$ and $\mathbf{B}_m=\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}\left((\mathbf{P}_{mi}\mathbf{Q}_{mi}-\mathbf{I}_{k_1+k_2})\mathbf{S}_0\right)\kappa$ for simplicity. When Assumption 6 holds, $\mathbf{V}_m=o_p(1)$ as we discussed in Corollary 2 and when the $\boldsymbol{\Sigma}_i$ is a diagonal matrix for all i, $\mathbf{B}_m=o_p(1)$ as we discussed in Corollary 1. We complete the proof. □

Proof of Theorem 4.2. Let $\Lambda_j = \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_j(\mathbf{U}_j + \mathbf{V}_j) + \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{B}_j\boldsymbol{\delta}$, $j = m, \ell$, we have

$$\begin{aligned}
\mathrm{Cov}(\Lambda_m, \Lambda_\ell) =& \mathbb{E}\left[\left(\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m(\mathbf{U}_m + \mathbf{V}_m) + \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{B}_m\boldsymbol{\delta} - \mathbb{E}\left(\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m(\mathbf{U}_m + \mathbf{V}_m) + \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{B}_m\boldsymbol{\delta}\right)\right)\right.\\
&\left.\times\left(\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_\ell(\mathbf{U}_\ell + \mathbf{V}_\ell) + \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{B}_\ell\boldsymbol{\delta} - \mathbb{E}\left(\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_\ell(\mathbf{U}_\ell + \mathbf{V}_\ell) + \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{B}_\ell\boldsymbol{\delta}\right)\right)\right]\\
=& \mathbb{E}\left[\left(\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m(\mathbf{U}_m + \mathbf{V}_m)\right)\left(\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_\ell(\mathbf{U}_\ell + \mathbf{V}_\ell)\right)'\right]\\
=& \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m\mathbb{E}(\mathbf{U}_m\mathbf{U}_\ell')\mathbf{S}_\ell'\mathbf{D}_{\boldsymbol{\beta}} + \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m\mathbb{E}(\mathbf{V}_m\mathbf{V}_m')\mathbf{S}_\ell'\mathbf{D}_{\boldsymbol{\beta}}\\
=& \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m(\mathbb{E}(\mathbf{U}_m\mathbf{U}_\ell') + \mathbb{E}(\mathbf{V}_m\mathbf{V}_\ell'))\mathbf{S}_\ell'\mathbf{D}_{\boldsymbol{\beta}}\\
=& \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m\boldsymbol{\Xi}_{m\ell}\mathbf{S}_\ell'\mathbf{D}_{\boldsymbol{\beta}}.
\end{aligned}$$

□

Inference of equation (5.2). Recall that,

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{mi} - \boldsymbol{\beta}_{mi} =& \left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i}\\
&+ \left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}}{T}(\bar{\boldsymbol{\gamma}} + \boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}})\\
&+ \left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\boldsymbol{\varepsilon}_i}{T},
\end{aligned}$$

We can also rewrite the above equation as

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{mi} - \boldsymbol{\beta}_m =& (\boldsymbol{\beta}_{mi} - \boldsymbol{\beta}_m) + \left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}(\mathbf{F}\boldsymbol{\Gamma}_i + \mathbf{V}_i)}{T}\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i}\\
&+ \left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}}{T}(\bar{\boldsymbol{\gamma}} + \boldsymbol{\eta}_i - \bar{\boldsymbol{\eta}})\\
&+ \left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\boldsymbol{\varepsilon}_i}{T}\\
=:& I_1 + I_2 + I_3 + I_4.
\end{aligned}$$

As for I_2, similarly, we can decompose I_2 into

$$\begin{aligned}
I_{21} =& \left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}}{T}\boldsymbol{\Gamma}_i\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i},\\
I_{22} =& \left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{V}_i}{T}\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i}.
\end{aligned}$$

By the fact $T^{-1}(\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i) = T^{-1}(\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i) + O_p(N^{-1}) + O_p((NT)^{-1/2}) \xrightarrow{p} \mathbf{Q}_{mi} = O_p(1)$, $I_{22} \xrightarrow{p} \mathbf{0}$ irrespective of whether the assumption on $\boldsymbol{\Sigma}_i$ holds or not because $\boldsymbol{\beta}_{2i}$ is $O(\Delta_{NT}^{-1}) \to \mathbf{0}$ by Assumption 5.

Consider I_{21} and I_3,

$$
\begin{aligned}
I_{21}+I_3 = &\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\frac{1}{T}\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{F}\left(\bar{\boldsymbol{\gamma}}+\boldsymbol{\Gamma}_{2i}(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i}+(\boldsymbol{\iota}_i-\bar{\boldsymbol{\iota}})+\right.\\
&\left.\bar{\boldsymbol{\Gamma}}_2(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\Delta_{NT}^{-1}\boldsymbol{\eta}_{\delta,i}+(\boldsymbol{\xi}_{2i}-\bar{\boldsymbol{\xi}}_2)(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2+(\boldsymbol{\xi}_{2i}-\bar{\boldsymbol{\xi}}_2)(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\Delta_{NT}^{-1}\boldsymbol{\eta}_{\delta,i}\right)\\
=&\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\frac{1}{T}\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{F}(\boldsymbol{\iota}_i-\bar{\boldsymbol{\iota}})+o_p\left(1\right).
\end{aligned}
$$

The last equality holds because of the result from Lemma 1, $T^{-1}\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}\left(\bar{\boldsymbol{\gamma}}+\bar{\boldsymbol{\Gamma}}_2(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2\right)=O_p\left(N^{-1}\right)+O_p\left((NT)^{-1/2}\right)$ and $\Delta_{NT}^{-1}\to 0$. $I_{21}+I_3$ is $o_p(1)$ when Asssumption 6 holds. Similar arguments is used by the fact $\mathbf{M}_{gm}=\mathbf{M}_{\mathbf{f}}$ and $\mathbf{M}_{\mathbf{f}}\mathbf{F}=\mathbf{0}$, therefore we skip it. Last, I_4 is $O_p(T^{-1/2})$ because $\boldsymbol{\varepsilon}_{it}$ is independent of $\mathbf{f}_t$ and $\mathbf{v}_{it}$ for all i,t. Then we have

$$
\hat{\boldsymbol{\beta}}_{mi}-\boldsymbol{\beta}_{mi}=(\boldsymbol{\beta}_{mi}-\boldsymbol{\beta}_m)+\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\frac{1}{T}\mathbf{X}_{mi}'\mathbf{M}_{hm}\mathbf{F}(\boldsymbol{\iota}_i-\bar{\boldsymbol{\iota}})+o_p(1). \tag{A.9}
$$

Using the result from Theorem 1, we have

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{\mathrm{MG},m}-\boldsymbol{\beta}_m=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_{mi}-\boldsymbol{\beta}_m+\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\frac{1}{T}\mathbf{S}_m'\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{F}(\boldsymbol{\iota}_i-\bar{\boldsymbol{\iota}})\\
&+O_p\left(\frac{1}{N}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right).
\end{aligned} \tag{A.10}
$$

By subtracting equation (A.9) from equation (A.10) and let $\boldsymbol{\chi}_{mi}=\left(\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i}{T}\mathbf{S}_m\right)^{-1}\frac{1}{T}\mathbf{S}_m'\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{F}(\boldsymbol{\iota}_i-\bar{\boldsymbol{\iota}})$ we can obtain

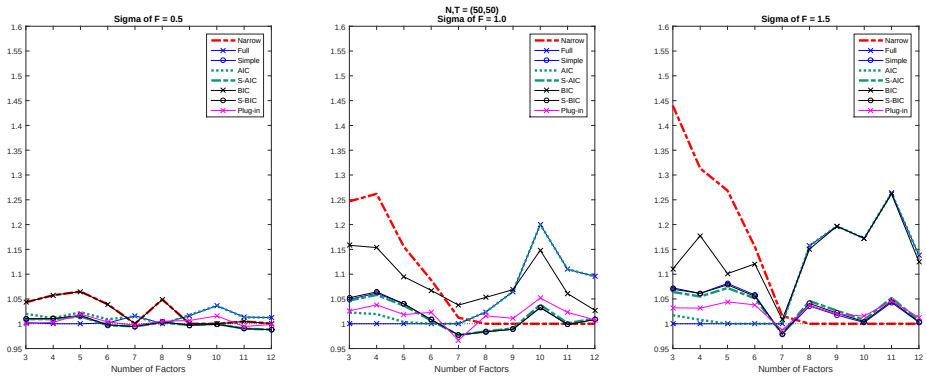
$$
\hat{\boldsymbol{\beta}}_{mi}-\hat{\boldsymbol{\beta}}_{\mathrm{MG},m}=\left(\boldsymbol{\beta}_{mi}-\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\beta}_{mi}\right)+\left(\boldsymbol{\chi}_{mi}-\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\chi}_{mi}\right)+o_p\left(1\right).
$$

By assumptions that $\boldsymbol{\beta}_i$ and $\boldsymbol{\chi}_{mi}$ are mutually independent, then

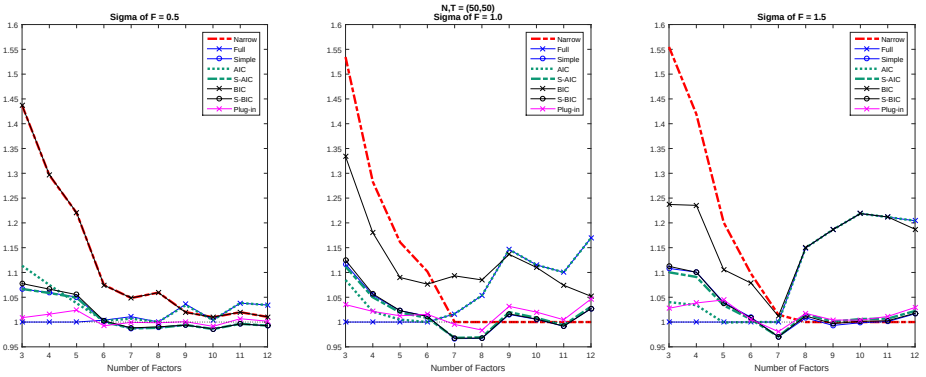
$$
\begin{aligned}
&\frac{1}{N-1}\sum_{i=1}^{N}\mathbb{E}\left\{\left[\left(\boldsymbol{\beta}_{mi}-\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\beta}_{mi}\right)+\left(\boldsymbol{\chi}_{mi}-\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\chi}_{mi}\right)\right]\right.\\
&\left.\times\left[\left(\boldsymbol{\beta}_{\ell i}-\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\beta}_{\ell i}\right)+\left(\boldsymbol{\chi}_{\ell i}-\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\chi}_{\ell i}\right)\right]'\right\}\\
=&\boldsymbol{\Xi}_{m\ell}+o_p\left(1\right).
\end{aligned}
$$

□

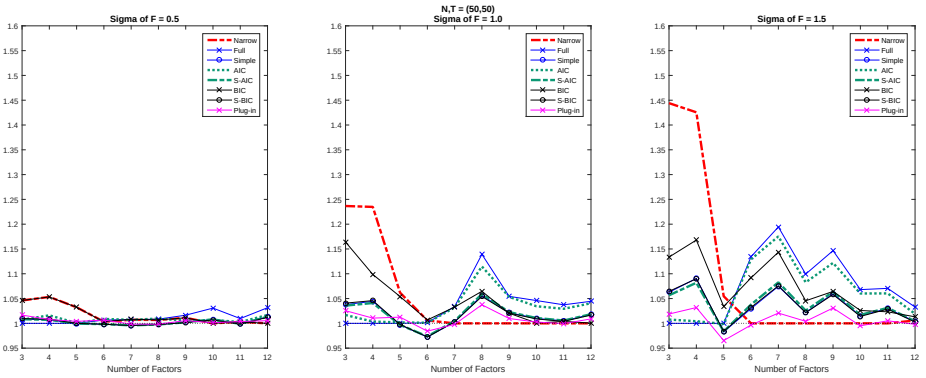
B Figures



(a) DGP 1

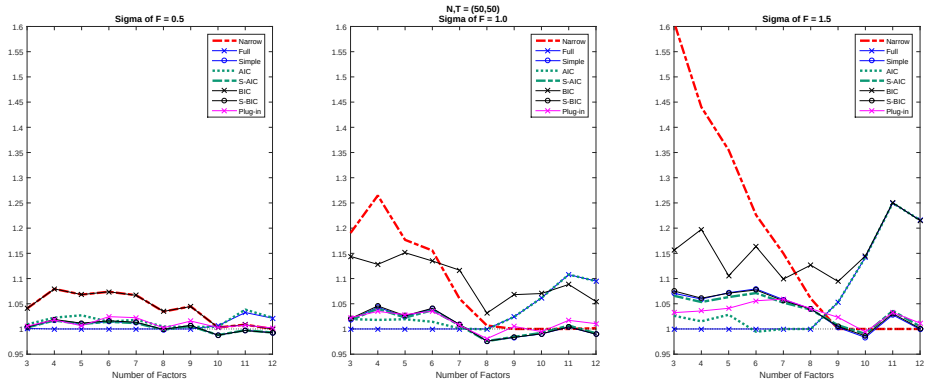


(b) DGP 2

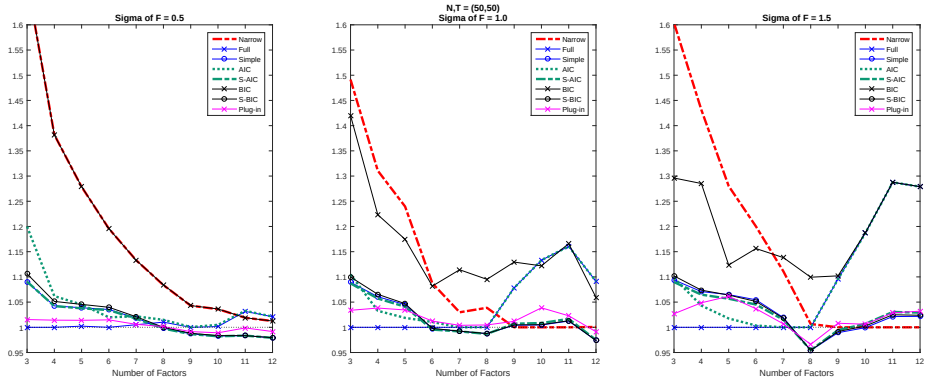


(c) DGP 3

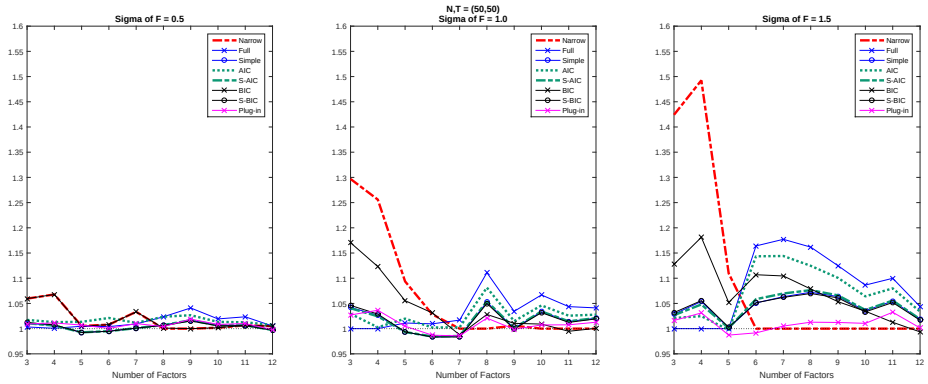
Figure 1: The number of auxiliary regressors is 3 ($L = 3$) and $(N, T) = (50, 50)$



(a) DGP 1

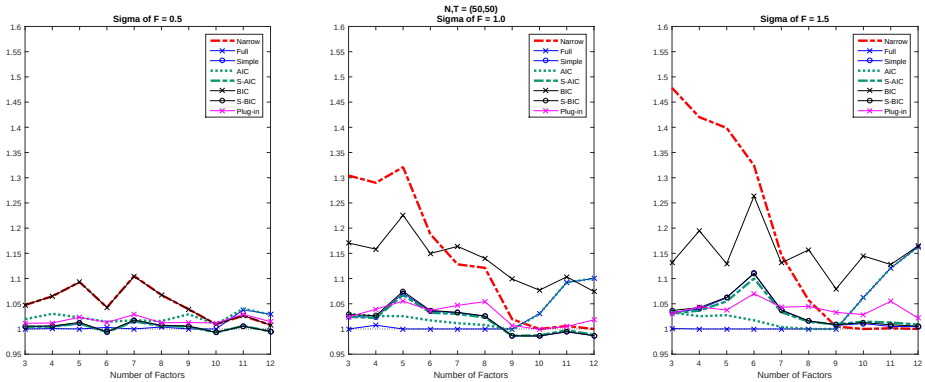


(b) DGP 2

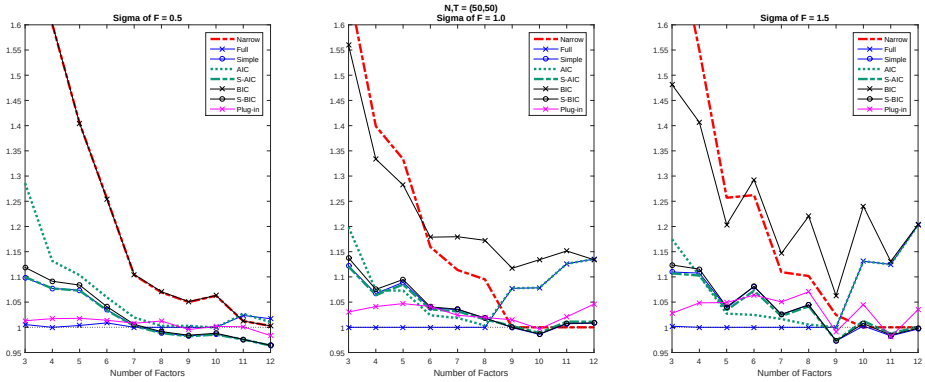


(c) DGP 3

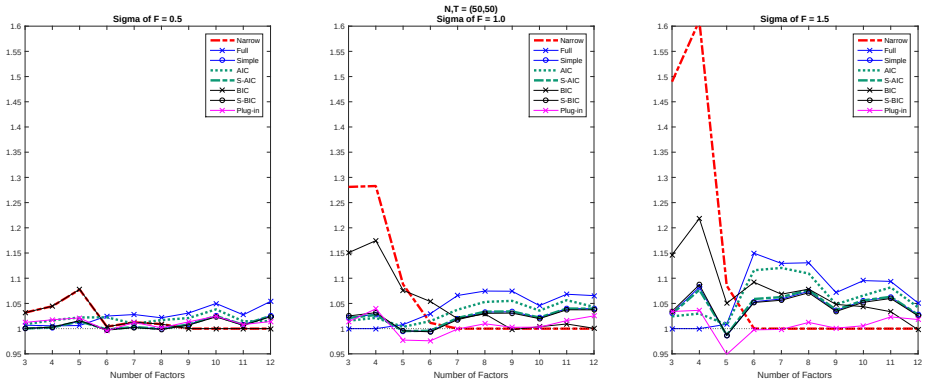
Figure 2: The number of auxiliary regressors is 4 ($L = 4$) and $(N, T) = (50, 50)$



(a) DGP 1

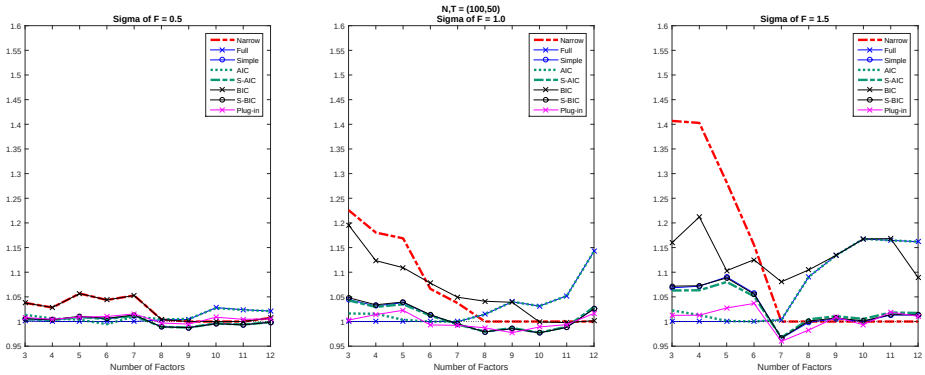


(b) DGP 2

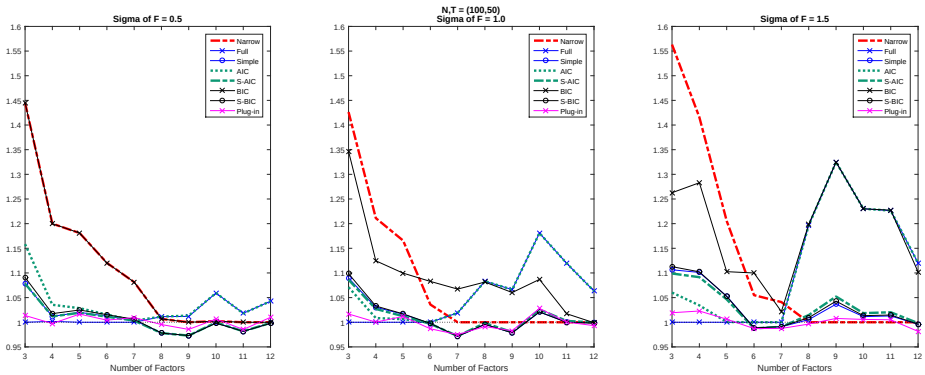


(c) DGP 3

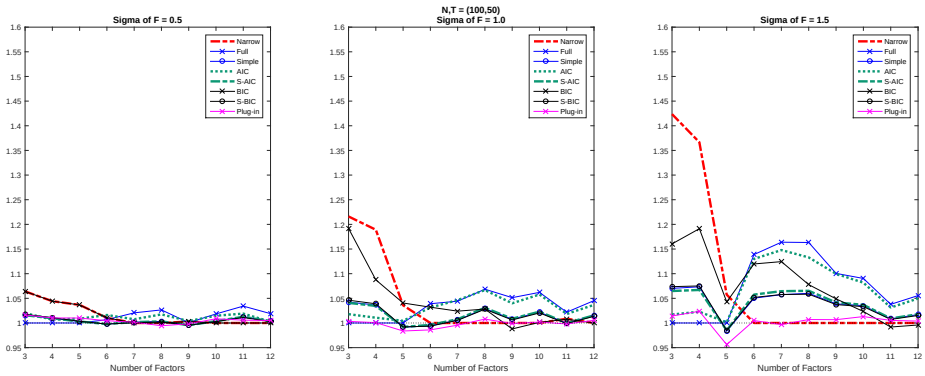
Figure 3: The number of auxiliary regressors is 5 ($L = 5$) and $(N, T) = (50, 50)$



(a) DGP 1

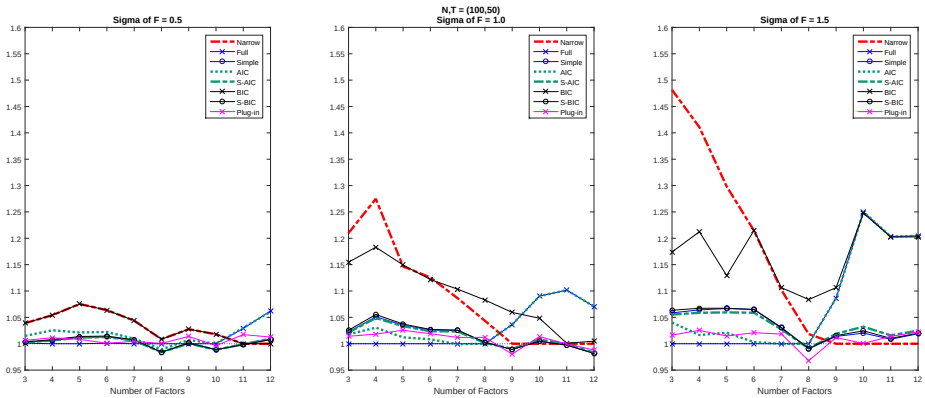


(b) DGP 2

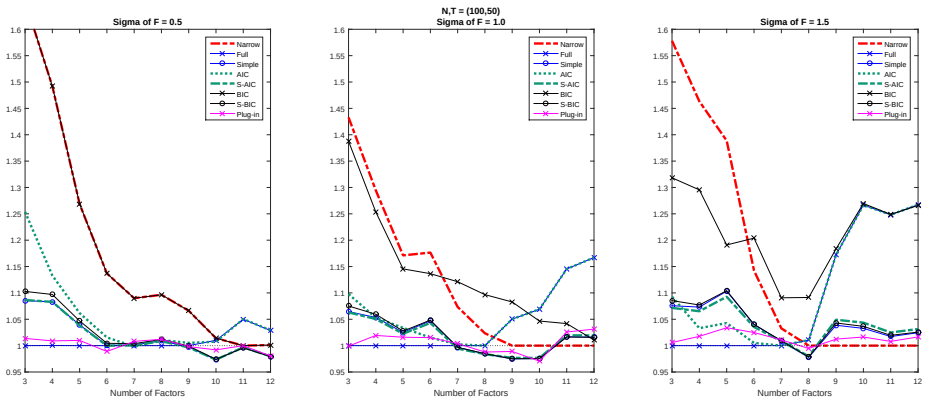


(c) DGP 3

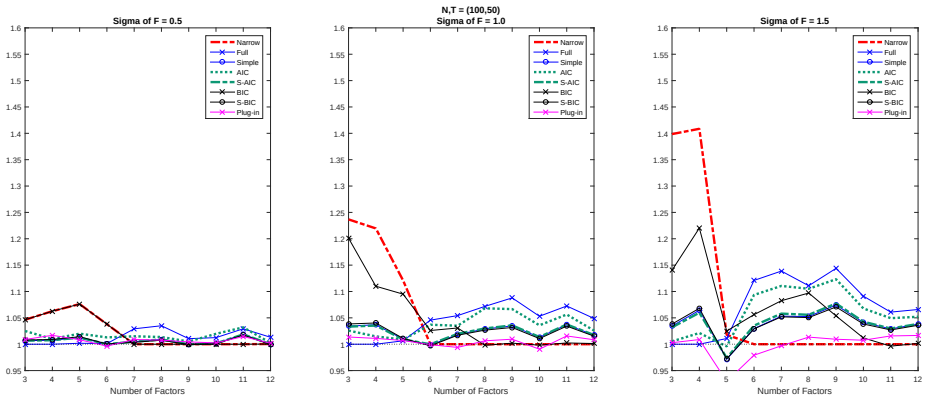
Figure 4: The number of auxiliary regressors is 3 ($L = 3$) and $(N, T) = (100, 50)$



(a) DGP 1

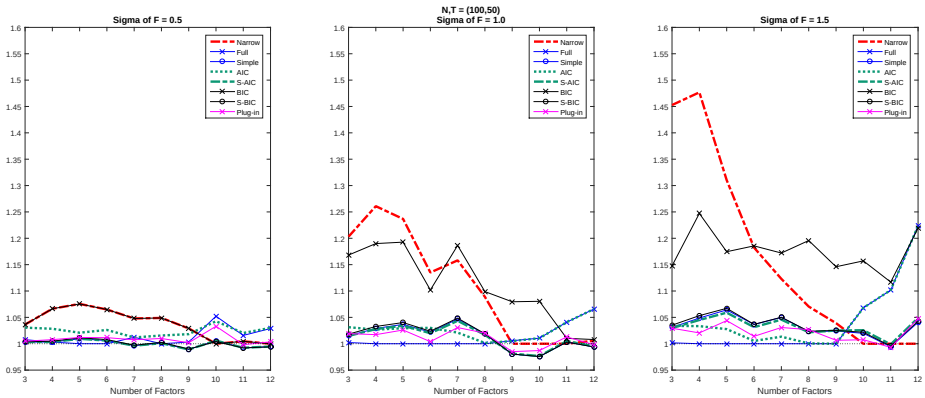


(b) DGP 2

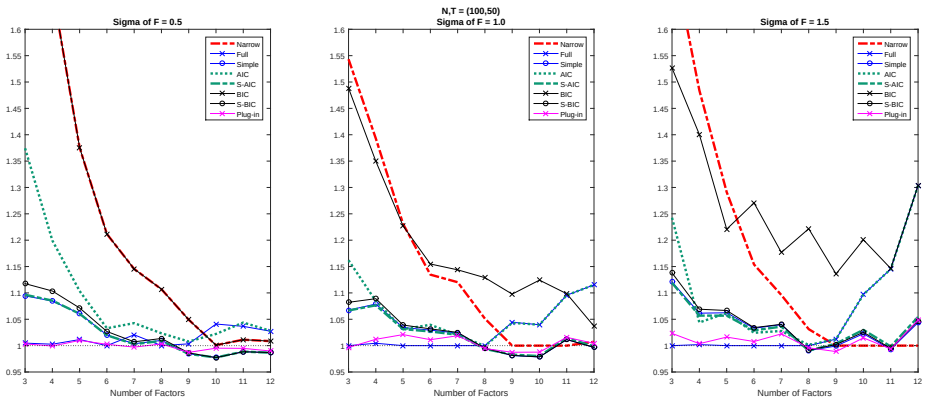


(c) DGP 3

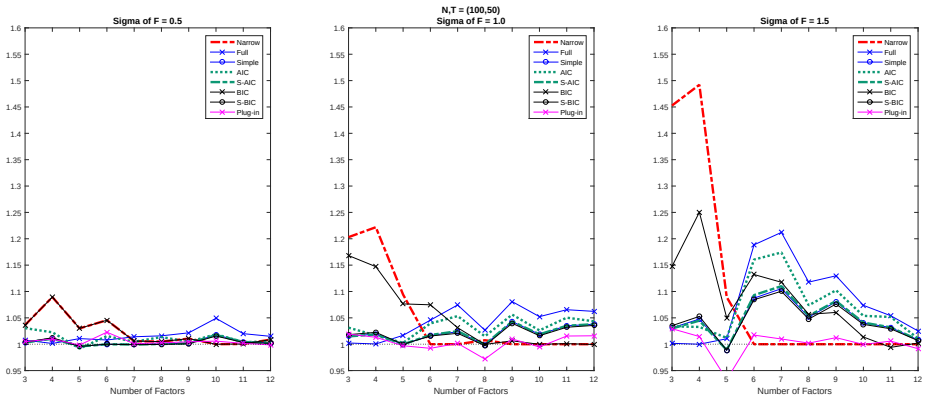
Figure 5: The number of auxiliary regressors is 4 ($L = 4$) and $(N, T) = (100, 50)$



(a) DGP 1



(b) DGP 2



(c) DGP 3

Figure 6: The number of auxiliary regressors is 5 ($L = 5$) and $(N, T) = (100, 50)$

References

Bai, J. (2009): "Panel Data Models with Interactive Fixed Effects," *Econometrica*, 77, 1229–1279.

Bai, J. and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

Buckland, S., K. Burnham, and N. Augustin (1997): "Model Selection: An Integral Part of Inference," *Biometrics*, 53, 603–618.

Cheng, X. and B. E. Hansen (2015): "Forecasting with Factor-Augmented Regression: A Frequentist Model Averaging Approach," *Journal of Econometrics*, 186, 280–293.

Chudik, A. and M. H. Pesaran (2015): "Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors," *Journal of Econometrics*, 188, 393 – 420, heterogeneity in Panel Data and in Nonparametric Analysis in honor of Professor Cheng Hsiao.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011): "Weak and Strong Cross-Section Dependence and Estimation of Large Panels," *The Econometrics Journal*, 14, C45–C90.

Claeskens, G. and R. J. Carroll (2007): "An Asymptotic Theory for Model Selection Inference in General Semiparametric Problems," *Biometrika*, 94, 249–265.

Claeskens, G. and N. L. Hjort (2003): "The Focused Information Criterion," *Journal of the American Statistical Association*, 98, 900–916.

——— (2008): *Model Selection and Model Averaging*, Cambridge: Cambridge University Press.

DiTraglia, F. (2014): "Using Invalid Instruments on Purpose: Focused Moment Selection and Averaging for GMM," Working Paper, University of Pennsylvania.

Gao, Y., X. Zhang, S. Wang, and G. Zou (2016): "Model Averaging Based on Leave-Subject-Out Cross-Validation," *Journal of Econometrics*, 192, 139–151.

Hansen, B. E. (2007): "Least Squares Model Averaging," *Econometrica*, 75, 1175–1189.

——— (2009): "Averaging Estimators for Regressions with a Possible Structural Break," *Econometric Theory*, 25, 1498–1514.

——— (2010): "Averaging Estimators for Autoregressions with a Near Unit Root," *Journal of Econometrics*, 158, 142–155.

——— (2014): "Model Averaging, Asymptotic Risk, and Regressor Groups," *Quantitative Economics*, 5, 495–530.

Hansen, B. E. and J. Racine (2012): "Jackknife Model Averaging," *Journal of Econometrics*, 167, 38–46.

Hjort, N. L. and G. Claeskens (2003): "Frequentist Model Average Estimators," *Journal of the American Statistical Association*, 98, 879–899.

——— (2006): "Focused Information Criteria and Model Averaging for the Cox Hazard Regression Model," *Journal of the American Statistical Association*, 101, 1449–1464.

Kapetanios, G. and M. H. Pesaran (2007): "Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results with an Application to Modelling Asset Returns," in *The Refinement of Econometric Estimation and Test Procedures*, ed. by G. D. A. Phillips and E. Tzavalis, Cambridge University Press, 239–281, cambridge Books Online.

Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011): "Panels with Non-Stationary Multifactor Error Structures," *Journal of Econometrics*, 160, 326 – 348.

Kitagawa, T. and C. Muris (2013): "Covariate Selection and Model Averaging in Semiparametric Estimation of Treatment Effects," Cemmap Working Paper.

Leeb, H. and B. Pötscher (2005): "Model Selection and Inference: Facts and Fiction," *Econometric Theory*, 21, 21–59.

Liang, H., G. Zou, A. Wan, and X. Zhang (2011): "Optimal Weight Choice for Frequentist Model Average Estimators," *Journal of the American Statistical Association*, 106, 1053–1066.

Liu, C.-A. (2015): "Distribution Theory of the Least Squares Averaging Estimator," *Journal of Econometrics*, 186, 142–159.

Liu, Q. and R. Okui (2013): "Heteroscedasticity-Robust C_p Model Averaging," *The Econometrics Journal*, 16, 463–472.

Lu, X. (2015): "A Covariate Selection Criterion for Estimation of Treatment Effects," *Journal of Business & Economic Statistics*, 33, 506–522.

Lu, X. and L. Su (2015): "Jackknife Model Averaging for Quantile Regressions," *Journal of Econometrics*, 188, 40–58.

Moon, H. R. and M. Weidner (2015a): "Dynamic Linear Panel Regression Models with Interactive Fixed Effects," Forthcoming. *Econometric Theory*.

——— (2015b): "Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects," *Econometrica*, 83, 1543–1579.

Pesaran, M. H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74, 967–1012.

Pesaran, M. H., L. V. Smith, and T. Yamagata (2013): "Panel Unit Root Tests in the Presence of a Multifactor Error Structure," *Journal of Econometrics*, 175, 94 – 115.

Pesaran, M. H. and E. Tosetti (2011): "Large Panels with Common Factors and Spatial Correlation," *Journal of Econometrics*, 161, 182 – 202.

Stock, J. H. and M. W. Watson (2002): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American statistical association*, 97, 1167–1179.

Sueishi, N. (2013): "Generalized Empirical Likelihood-Based Focused Information Criterion and Model Averaging," *Econometrics*, 1, 141–156.

Wan, A., X. Zhang, and G. Zou (2010): "Least Squares Model Averaging by Mallows Criterion," *Journal of Econometrics*, 156, 277–283.

Westerlund, J. and J.-P. Urbain (2015): "Cross-Sectional Averages versus Principal Components," *Journal of Econometrics*, 185, 372–377.

Yang, Y. (2000): "Combining Different Procedures for Adaptive Regression," *Journal of Multivariate Analysis*, 74, 135–161.

——— (2001): "Adaptive Regression by Mixing," *Journal of the American Statistical Association*, 96, 574–588.

Yuan, Z. and Y. Yang (2005): "Combining Linear Regression Models: When and How?" *Journal of the American Statistical Association*, 100, 1202–1214.

Zhang, X. and H. Liang (2011): "Focused Information Criterion and Model Averaging for Generalized Additive Partial Linear Models," *The Annals of Statistics*, 39, 174–200.

Zhang, X., A. T. Wan, and S. Z. Zhou (2012): "Focused Information Criteria, Model Selection, and Model Averaging in a Tobit Model with a Nonzero Threshold," *Journal of Business & Economic Statistics*, 30, 132–142.

Zhang, X., A. T. Wan, and G. Zou (2013): "Model Averaging by Jackknife Criterion in Models with Dependent Data," *Journal of Econometrics*, 174, 82–94.

Zhang, X., G. Zou, and H. Liang (2014): "Model Averaging and Weight Choice in Linear Mixed-Effects Models," *Biometrika*, 101, 205–218.