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

Shou-Yung Yin†
Department of Economics, National Taipei University

Chu-An Liu‡
Institute of Economics, Academia Sinica

Chang-Ching Lin§
Department of Economics, National Cheng Kung University

August 7, 2018

Abstract

This paper considers model selection and model averaging in panel data models with a multifactor error structure. We investigate the limiting distribution of the common correlated effects estimator (Pesaran, 2006) in a local asymptotic framework and show that the trade-off between bias and variance remains in the asymptotic theory. We then propose a focused information criterion and a plug-in averaging estimator for large heterogeneous panels and examine their theoretical properties. The novel feature of the proposed method is that it aims to minimize the sample analog of the asymptotic mean squared error and can be applied to cases irrespective of whether the rank condition holds or not. Monte Carlo simulations show that both proposed selection and averaging methods generally achieve lower expected squared error than other methods. The proposed methods are applied to examine possible causes that lead to the increasing wage inequality between high-skilled to low-skilled workers in the U.S. manufacturing industries.

Keywords: Cross-sectional dependence, Common correlated effects, Model selection, Plug-in estimators

*We thank the editor, the associate editor, two referees, Yi-Ting Chen, Xu Cheng, and Jack Porter for constructive comments and suggestions. We also thank the seminar participants of Erasmus University Rotterdam, Hitotsubashi University, and Monash University, and conference participants of AMES 2016, TER 2016, EEA-ESEM 2016, EcoSta 2017, IAAE 2017, and IPDC 2017 for their discussions and suggestions. All errors remain the authors'.

†Email: syyin@mail.ntpu.edu.tw.
‡Email: caliu@econ.sinica.edu.tw.
§Email: ever@mail.ncku.edu.tw.

1 Introduction

Panel data models are widely used in economic and statistical research. In the past decade, there has been increasing interest in the study of cross-sectional dependence in panel data models. One popular approach to this problem is the common correlated effects (CCE) approach proposed by Pesaran (2006). The virtue of CCE estimation is that it can be easily computed by 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 CCE estimation under model uncertainty.

This paper considers model selection and model averaging in panel data models with a multifactor error structure. Following Hjort and Claeskens (2003), Hansen (2014), and Liu (2015), we study the asymptotic distribution of the mean group estimator based on the individual-specific CCE estimators in a local asymptotic framework where the cross-sectional means of slope 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 the 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 (AMSE) of the common correlated effects mean group (CCEMG) estimators for all submodels remains finite and provides a good approximation to finite sample mean squared error.

One attractive advantage of CCE-type estimation is that we can still obtain consistency under regular conditions even though the cross-sectional averages of the dependent variable and individual regressors cannot span the space of factors. This case refers to the violation of rank condition of the mean behavior of the factor loading matrix, and it affects the asymptotic variance and makes it larger. We first consider the case where the rank condition is not satisfied for all submodels. Under drifting sequences of parameters, 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 trade-off, we find that adding more regressors could have positive or negative effects on estimation variance. While, in general, adding more regressors causes a larger variance, it could also affect the orthogonal projection matrix and filter out more unobserved common factors.

Hence, a bigger model may have a lower variance than the smaller model.

We next consider the case where the rank condition is satisfied for some submodels. We find that the CCEMG estimator is more efficient in those submodels satisfying the rank condition. However, it is hard to distinguish between cases and to verify whether the rank condition holds or not in practice. Therefore, the results inferred from these asymptotic distributions do not provide us a direct guideline to select the submodel in an empirical study. To address this problem, we propose a focused information criterion (FIC) to select the model for large heterogeneous panels. The proposed FIC aims to minimize the sample analog of AMSE for cases irrespective of whether the rank condition holds or not. We show that the proposed FIC is an asymptotic unbiased estimator of the AMSE and can be applied to both cases.

Building on the idea of FIC, we introduce a frequentist model averaging criterion to select the weights for candidate models and study their properties. We first derive the asymptotic distribution of the averaging estimator with fixed weights, which allows us to characterize the AMSE 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 asymptotic unbiased estimate of the AMSE irrespective of whether the rank condition holds or not.

In simulations, we compare the proposed focused information criterion and the plug-in averaging estimator with other model selection and averaging methods. Simulation studies show that the proposed methods generally produce lower expected squared error than existing methods. As an empirical illustration, we apply the proposed methods to examine possible causes of the increasing wage inequality between high-skilled and low-skilled workers in the U.S. manufacturing industries. We set the focus parameter as the measurement of the strength of the complementarity between upstream and downstream skill bias, and our results show that high-skilled and low-skilled labor are substitutes. To conduct inference after model averaging, we construct simulation-based confidence intervals of the focus parameter, and find that the confidence intervals of focus parameters obtained from FIC and the plug-in averaging estimator are narrower than the narrow model, which suggests a strong effect on increasing wage gap.

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 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), Chudik and Pesaran (2015), and Karabiyik, Reese, and Westerlund (2017), while the interactive effects estimator based on principal components has been developed by Stock and Watson (2002), Bai and Ng (2002), Bai (2009), and Moon and Weidner (2015); 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 models. In recent years, FIC has been extended to several models, including 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), generalized empirical likelihood estimation (Sueishi, 2013), generalized method of moments estimation (DiTraglia, 2016), and propensity score weighted estimation of the treatment effects (Lu, 2015; Kitagawa and Muris, 2016). In this paper, we extend the existing literature on FIC to panel data models in the presence of a multifactor error structure.

There is a growing body of literature on frequentist model averaging, including information criterion weighting (Buckland, Burnham, and Augustin, 1997), adaptive regression by mixing models (Yang, 2001; Yuan and Yang, 2005), Mallows' C_p-type averaging (Hansen, 2007; Liu and Okui, 2013), optimal mean squared error averaging (Liang, Zou, Wan, and Zhang, 2011), jackknife model averaging (Hansen and Racine, 2012; Lu and Su, 2015), and plug-in averaging (Liu, 2015); see Claeskens and Hjort (2008) and Moral-Benito (2015) for a literature review. However, the existing literature on frequentist model averaging in factor-augmented regressions or panel data 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. Paap, Wang, and Zhang (2015) propose an optimal pooling averaging estimator for heterogenous panel data models.

Gao, Zhang, Wang, and Zou (2016) propose a leave-subject-out model averaging estimator for panel data models and demonstrate its asymptotic optimality. To our knowledge, the averaging estimator has not been explored before in panel data models with a multifactor error structure.

The rest of the paper is organized as follows. Section 2 presents the model 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. Section 7 presents the empirical application. An extension discussing the case allowing for correlated slope coefficients and factor loadings is provided in Section 8, and Section 9 concludes the paper. Proofs are presented in the Appendix. Throughout this paper, we employ the following symbols. 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. For an $m \times n$ matrix $\mathbf{B} = (b_{ij})$, $\text{vec}(\mathbf{B}) = [b_{11}, ..., b_{m1}, \ldots, b_{1n}, \ldots, b_{mn}]$.

2 Model and CCE 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 an $r \times 1$ vector of unobserved factor loadings, $\mathbf{f}_t$ is an $r \times 1$ 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 error. 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] The core regressors may only include a constant term or even an empty matrix.

[1] We may follow Su and Jin (2012) and consider a nonparametric form of auxiliary regressors $\mathbf{x}_{2it}$. However, it will be quite challenging to characterize the omitted variable bias of the nonparametric component.

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, $\mathbf{v}_{1it}$ are correlated with $\mathbf{v}_{2it}$, i.e., $\mathrm{Var}(\mathbf{v}_{it})$ is not a diagonal matrix. Hence, 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 $m = 1, ..., M$. The mth submodel includes all core regressors $\mathbf{x}_{1it}$ and $0 \leq k_{2m} \leq k_2$ 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, where $\boldsymbol{\Pi}_m$ is a $k_{2m} \times k_2$ selection matrix that selects the included auxiliary regressors. 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 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.

Such an investigation is beyond the scope of this paper, and thus it is left for future research.

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 $\bar{\mathbf{A}}_t = N^{-1}\sum_{i=1}^N \mathbf{A}_{it}$ as the cross-sectional average of any variable $\mathbf{A}_{it}$. 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.[2] 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 least squares regression augmented with cross-sectional averages of the dependent variable and individual regressors.

[2]In Pesaran (2006), the cross-sectional average is defined by $\bar{\mathbf{h}}_t = \sum_{i=1}^N \lambda_i \mathbf{h}_{it}$ with the weights λ_i that satisfy the conditions: (1) $\lambda_i = O(N^{-1})$, (2) $\sum_{i=1}^N \lambda_i = 1$, and (3) $\sum_{i=1}^N |\lambda_i| < C < \infty$. 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.

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., with all auxiliary regressors included in the model, is

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

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

and the CCE estimator for the submodel m is

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

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

where $\mathbf{X}_{mi} = (\mathbf{X}_{1i}, \mathbf{X}_{2i}\boldsymbol{\Pi}_m')$ and $\bar{\mathbf{H}}_m = (\bar{\mathbf{y}}, \bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2\boldsymbol{\Pi}_m')$.[3] For notation simplicity, we use $\widehat{\boldsymbol{\beta}}_{mi}$ to denote the CCE estimators for the submodel m, that is, $\widehat{\boldsymbol{\beta}}_{1i}$ and $\widehat{\boldsymbol{\beta}}_{2i}$ are the estimators for submodels 1 and 2, but not the estimators for the coefficients $\boldsymbol{\beta}_{1i}$ and $\boldsymbol{\beta}_{2i}$.

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 independent and identically distributed (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, that is, the common correlated effects mean group (CCEMG) estimator.

[3]For the CCE estimator in the submodel m, one may consider using all the cross-sectional averages of the dependent and independent variables as proxies for unobserved common factors, i.e., $\widetilde{\boldsymbol{\beta}}_{mi} = (\mathbf{X}_{mi}'\bar{\mathbf{M}}_h\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\bar{\mathbf{M}}_h\mathbf{y}_i$. Our simulations show that the averaging estimator based on $\widehat{\boldsymbol{\beta}}_{mi}$ has better finite sample performance than the averaging estimator based on $\widetilde{\boldsymbol{\beta}}_{mi}$. Thus, we do not consider $\widetilde{\boldsymbol{\beta}}_{mi}$ in our analysis.

The CCEMG estimator of $\boldsymbol{\beta}$ in the full model is

$$\widehat{\boldsymbol{\beta}}_{\mathrm{MG},f} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_{fi}, \tag{2.13}$$

and the CCEMG estimator for the submodel m is

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

3 Asymptotic Framework

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 local asymptotic framework and technical assumptions. In the second subsection, we present the asymptotic distribution of the CCEMG estimator.

3.1 Assumptions

We now state the assumptions.

Assumption 1. The individual-specific errors ε_{it} and $\mathbf{v}_{js}$ are distributed independently for all i, j, t, and s. 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 $(\zeta_{it}, \boldsymbol{\nu}_{it}')'$ are $(k+1)\times 1$ vectors of i.i.d. random variables with mean zero, identity covariance matrix, and finite fourth-order cumulants, $\boldsymbol{\Sigma}_i$ is a positive definite matrix, and $\bar{\sigma}^2$ and $\bar{\boldsymbol{\Sigma}}$ are constants.

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

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. In particular, for $i = 1, ..., N$,

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

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 coefficients $\boldsymbol{\beta}_i$ follow the random coefficient model

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

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$, $\boldsymbol{\eta}_i = (\boldsymbol{\eta}_{1i}', \boldsymbol{\eta}_{2i}')'$, $\boldsymbol{\Omega}_{\boldsymbol{\beta}}$ is a symmetric nonnegative definite matrix, and $||\boldsymbol{\Omega}_{\boldsymbol{\beta}}||$ is bounded. The random deviations $\boldsymbol{\eta}_i$ are 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 $\sqrt{N}\Delta_{NT}^{-1} \to c < \infty$ as $N, T \to \infty$ jointly. The cross-sectional means of $\boldsymbol{\beta}_{2i}$ follow

$$\boldsymbol{\beta}_2 \equiv \boldsymbol{\beta}_{2,NT} = \Delta_{NT}^{-1}\boldsymbol{\delta},$$

where $\boldsymbol{\delta}$ is an unknown constant vector.

Assumption 6. $\text{Rank}(\bar{\mathbf{C}}_m) \equiv r_m = r \leq k_m + 1$ for the mth model.

Assumption 1 specifies that the individual-specific errors are distributed independently 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). Note that Assumption 4 implies that $\boldsymbol{\beta}_{1i} = \boldsymbol{\beta}_1 + \boldsymbol{\eta}_{1i}$, $\boldsymbol{\eta}_{1i} \sim \text{i.i.d.}(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\beta}_1})$ where $||\boldsymbol{\Omega}_{\boldsymbol{\beta}_1}||$ is bounded.

Assumption 5 assumes that the cross-sectional means of $\boldsymbol{\beta}_{2i}$ are local to zero. The local to zero assumption 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), Hansen (2014), and Liu (2015). Under Assumptions 4–5, it is easy to see that $\mathrm{E}(\boldsymbol{\beta}_{2i}) = \Delta_{NT}^{-1}\boldsymbol{\delta} = \boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_{2i}$ still follow the random coefficient model.[4] This assumption is canonical in the sense that both squared bias and variance have the same order, and it ensures that the asymptotic mean squared error of the submodel estimator remains finite.

The assumption states that the partial correlations between the dependent variable and the auxiliary regressors are weak for all i, and the partial correlations will vanish as $N, T \to \infty$ jointly. Here we do not specify the convergence rate of $\boldsymbol{\beta}_{2,NT}$ but simply let $\boldsymbol{\beta}_{2,NT}$ converge to zero under the condition $\sqrt{N}\Delta_{NT}^{-1} \to c$ as N and T increase. For example, if $\Delta_{NT} = O(T)$, then Assumption 5 holds when $N = O(T^2)$.

Assumption 6 is the rank condition, and it plays a crucial role in CCEMG estimation. Recall that $\bar{\mathbf{C}}_m$ is a matrix of dimension $r \times (k_m + 1)$, and Assumption 6 says that $\bar{\mathbf{C}}_m$ is full rank. This assumption implies that the space spanned by the unknown common factors can be consistently estimated using the cross-sectional averages, and hence it achieves efficiency gain when the rank condition is satisfied. If the rank condition is not satisfied for the mth model, then we have $\mathrm{Rank}(\bar{\mathbf{C}}_m) \equiv r_m < r$. Assumption 6 corresponds to the rank condition (21) of Pesaran (2006).

3.2 Asymptotic Distribution of CCEMG Estimator

We first introduce some notation that we will use to characterize the limiting distribution. Let $\boldsymbol{\beta}_m^* = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2'\boldsymbol{\Pi}_m')' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_{2m}')'$. Define $\mathbf{Q}_{mi} = p\lim_{T\to\infty}(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\mathbf{X}_i)$ and $\boldsymbol{\Sigma}_{mi} = p\lim_{T\to\infty}(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\mathbf{F})$, where $\bar{\mathbf{M}}_{gm} = \mathbf{I}_T - \bar{\mathbf{G}}_m(\bar{\mathbf{G}}_m'\bar{\mathbf{G}}_m)^-\bar{\mathbf{G}}_m'$ 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.

[4]Note that Assumption 5 only imposes the local to zero assumption on the cross-sectional means $\boldsymbol{\beta}_2$. It is possible to impose the local to zero assumption on both the cross-sectional means $\boldsymbol{\beta}_2$ and the random deviations $\boldsymbol{\eta}_{2i}$; see the discussion in Appendix C for more details. It is also possible to impose the homogenous assumption on $\boldsymbol{\beta}_i$, i.e., $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i. However, under this assumption, the convergence rate of the estimate is much faster, and the local to zero assumption would be different. We leave it for future studies.

The following theorem presents the asymptotic distribution of the CCEMG estimator when the rank condition is not satisfied for all submodels.

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

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

where $\boldsymbol{\delta}_c = c \cdot \boldsymbol{\delta}$, $\mathbf{A}_m = \underset{N\to\infty}{p\lim} \frac{1}{N} \sum_{i=1}^N \mathbf{R}_{mi}^{-1} \mathbf{S}_m' \mathbf{Q}_{mi} \mathbf{S}_0 (\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m' \boldsymbol{\Pi}_m)$, $\mathbf{R}_{mi} = \mathbf{S}_m' \mathbf{Q}_{mi} \mathbf{S}_m$, and $\mathbf{U}_m$ and $\mathbf{V}_m$ are two stochastically independent normal random vectors. In particular,

$$\mathbf{U}_m \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Xi}_{um}\right) \quad \text{with} \quad \boldsymbol{\Xi}_{um} = \underset{N\to\infty}{p\lim} \frac{1}{N} \sum_{i=1}^N \mathbf{R}_{mi}^{-1} \mathbf{S}_m' \mathbf{Q}_{mi} \boldsymbol{\Omega}_{\boldsymbol{\beta}} \mathbf{Q}_{mi} \mathbf{S}_m \mathbf{R}_{mi}^{-1},$$

$$\mathbf{V}_m \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Xi}_{vm}\right) \quad \text{with} \quad \boldsymbol{\Xi}_{vm} = \underset{N\to\infty}{p\lim} \frac{1}{N} \sum_{i=1}^N \mathbf{R}_{mi}^{-1} \mathbf{S}_m' \boldsymbol{\Sigma}_{mi} \boldsymbol{\Omega}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{mi}' \mathbf{S}_m \mathbf{R}_{mi}^{-1},$$

and $\boldsymbol{\Xi}_m = \boldsymbol{\Xi}_{um} + \boldsymbol{\Xi}_{vm}$.

Theorem 1 presents the asymptotic normality of the CCEMG estimator for each submodel. This result also implies that the submodel estimate $\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}$ is consistent. Here $\mathbf{A}_m \boldsymbol{\delta}_c$ represents the asymptotic bias of submodel estimators and $\boldsymbol{\Xi}_m$ represents the asymptotic variance. For the full model, it is easy to see that the asymptotic bias is zero since $\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m' \boldsymbol{\Pi}_m = \mathbf{0}$. Furthermore, the asymptotic distribution of the CCEMG estimator in the full model is

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},f} - \boldsymbol{\beta}) \xrightarrow{d} \mathbf{U}_f + \mathbf{V}_f \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Xi}_f\right), \tag{3.1}$$

$$\boldsymbol{\Xi}_f = \boldsymbol{\Omega}_{\boldsymbol{\beta}} + \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \mathbf{Q}_{fi}^{-1} \boldsymbol{\Sigma}_{fi} \boldsymbol{\Omega}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{fi}' \mathbf{Q}_{fi}^{-1}, \tag{3.2}$$

where $\mathbf{Q}_{fi} = \underset{T\to\infty}{p\lim} (T^{-1} \mathbf{X}_i' \bar{\mathbf{M}}_g \mathbf{X}_i)$, $\boldsymbol{\Sigma}_{fi} = \underset{T\to\infty}{p\lim} (T^{-1} \mathbf{X}_i' \bar{\mathbf{M}}_g \mathbf{F})$, $\bar{\mathbf{M}}_g = \mathbf{I}_T - \bar{\mathbf{G}}(\bar{\mathbf{G}}'\bar{\mathbf{G}})^- \bar{\mathbf{G}}'$, and $\bar{\mathbf{G}} = \mathbf{F}\bar{\mathbf{C}}$. The asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{\mathrm{MG},f}$ presented in (3.1) and (3.2) corresponds to Theorem 2 in Pesaran (2006).[5]

Theorem 1 shows that the trade-off between omitted variable bias and estimation variance remains in the asymptotic theory. The asymptotic bias comes from the fact that the

[5] Theorem 2 in Pesaran (2006) is a special case of our theorem 1, and there is no trade-off between bias and variance in Theorem 2 in Pesaran (2006).

core regressors $\mathbf{x}_{1it}$ and the auxiliary regressors $\mathbf{x}_{2it}$ are correlated. Furthermore, the correlation between $\mathbf{x}_{1it}$ and $\mathbf{x}_{2it}$ is due to the common factors $\mathbf{f}_t$ and the correlation between $\mathbf{v}_{1it}$ and $\mathbf{v}_{2it}$. In general, the asymptotic bias of submodel estimators is nonzero. The asymptotic bias $\mathbf{A}_m\boldsymbol{\delta}_c$ is zero if the cross-sectional means of slope coefficients $\boldsymbol{\beta}_{2i}$ are zero, i.e., $\boldsymbol{\beta}_2 = \mathbf{0}$, or the regressors are uncorrelated, i.e., $\mathbf{Q}_{mi}$ is a diagonal matrix.

In addition to the bias-variance trade-off, Theorem 1 also shows that adding more regressors could have positive or negative effects on estimation variance. Note that the asymptotic variance $\boldsymbol{\Xi}_m$ has two components, $\boldsymbol{\Xi}_{um}$ and $\boldsymbol{\Xi}_{vm}$, and the diagonal elements of $\boldsymbol{\Xi}_{um}$ and $\boldsymbol{\Xi}_{vm}$ vary across different submodels. In most cases, the variance term $\boldsymbol{\Xi}_{vm}$ increases when we include more auxiliary regressors. Unlike $\boldsymbol{\Xi}_{vm}$, due to the special structure of the covariance matrix, the variance term $\boldsymbol{\Xi}_{um}$ may decrease with more auxiliary regressors. One clear example is the comparison between the full model and the submodel. For the full model, $\boldsymbol{\Xi}_{uf}$ can be simplified as $\boldsymbol{\Omega}_{\boldsymbol{\beta}}$, which is smaller than $\boldsymbol{\Xi}_{um} = \underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\mathbf{R}_{mi}^{-1}\mathbf{S}_m'\mathbf{Q}_{mi}\boldsymbol{\Omega}_{\boldsymbol{\beta}}\mathbf{Q}_{mi}\mathbf{S}_m\mathbf{R}_{mi}^{-1}$ of any submodel; see the discussion in Appendix C for more details. The intuition behind the negative effect is that additional regressors could filter out extra factors so that the variance term $\boldsymbol{\Xi}_m$ decreases.[6] Since the magnitudes of these two effects are not equal, the total effect of adding more auxiliary regressors on estimation variance could be positive or negative.

We next present the asymptotic distribution of the CCEMG estimator when the rank condition is satisfied for some submodels.

Theorem 2. Suppose that Assumptions 1–6 hold. As $N, T \to \infty$ jointly and $\sqrt{N}/T \to 0$, we have

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

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

As pointed out by Pesaran (2006), the rank condition is not necessary for employing the CCEMG estimator. However, efficiency gains can be achieved when the rank condition

[6]Note that additional regressors could affect the orthogonal projection matrix $\bar{\mathbf{M}}_{gm}$ in both $\mathbf{Q}_{mi}$ and $\boldsymbol{\Sigma}_{mi}$. Thus, it is possible that adding more regressors could have a positive effect on $\boldsymbol{\Xi}_{um}$ or a negative effect on $\boldsymbol{\Xi}_{vm}$ in different submodels.

is satisfied. This is because the effects of unobserved common factors can be efficiently eliminated when the rank condition holds.[7] Compared to Theorem 1, the asymptotic covariance matrix only consists of one term $\boldsymbol{\Xi}_{um}$, and hence the CCEMG estimator for the submodel m is more efficient when the rank condition is satisfied. Moreover, this result is slightly different from the case when the rank condition is not satisfied. This is because of the issue of the different rank behavior between $\mathbf{M}_{hm}$ and $\mathbf{M}_{gm}$ pointed out by Karabiyik, Reese, and Westerlund (2017) when $k_m + 1 > r$.[8]

4 Focused Information Criterion

The traditional model selection approaches such as AIC and BIC select the model based on the global fit of the model. The empirical study, however, tends to focus on one particular parameter instead of assessing the overall properties of the model.

Example 1. Consider the production function: $y_{it} = \theta + \alpha k_{it} + \beta \ell_{it} + \boldsymbol{\gamma}' \mathbf{z}_{it} + \epsilon_{it}$, where y_{it} is the log of output, k_{it} is the log of capital input, ℓ_{it} is the log of labor input, and $\mathbf{z}_{it}$ are other control variables. The parameter of interest may be the labor share of the production function, i.e., the coefficient α, or the measure of the return to scale, i.e., the sum of coefficients α and β; see Claeskens and Hjort (2003, 2008), Hansen (2005), and Chang and DiTraglia (2018) for more motivating examples.

Unlike the traditional model selection approaches, which assess the global fit of the model, we evaluate the model based on the focus parameter $\mu = \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \mu(\boldsymbol{\beta})$, which is a smooth real-valued function of the cross-sectional means of slope coefficients. Let $\mathbf{D}_{\boldsymbol{\beta}} = \partial\mu/\partial\boldsymbol{\beta}$ be 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. Let $\widehat{\mu}_m = \mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m})$

[7] Recall that $\mathbf{Q}_{mi} = \underset{T\to\infty}{p\lim}(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\mathbf{X}_i)$ and $\mathbf{X}_i = \mathbf{F}\boldsymbol{\Gamma}_i + \mathbf{V}_i$, where $\mathbf{V}_i = (\mathbf{v}_{i1}, ..., \mathbf{v}_{iT})'$. Then we can show that $\mathbf{Q}_{mi} = \boldsymbol{\Sigma}_i$ and further simplify $\mathbf{A}_m$ and $\boldsymbol{\Xi}_{um}$ by replacing $\mathbf{Q}_{mi}$ by $\boldsymbol{\Sigma}_i$; see the proof of Theorem 2 for more details.

[8] In Appendix D, we provide numerical results supporting Theorems 1 and 2 based on a simple three-nested-model framework consisting of only three submodels (narrow, middle, and full) under different setups on the variance of factors, correlation between regressors, and number of factors.

denote the submodel estimates. Thus, we aim to select a model with the lowest possible AMSE of $\widehat{\mu}_m$ under the quadratic loss function.

We first derive the asymptotic distribution of the submodel estimator of the focus parameter, and then characterize the AMSE of $\widehat{\mu}_m$. Theorem 1 and the delta method imply the following corollary.

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

$$\begin{aligned}\sqrt{N}(\mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}) - \mu(\boldsymbol{\beta})) \xrightarrow{d} \Lambda_m &= \mathbf{D}'_{\boldsymbol{\beta}}\mathbf{B}_m\boldsymbol{\delta}_c + \mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_m(\mathbf{U}_m + \mathbf{V}_m), \\ &\sim \mathrm{N}\left(\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{B}_m\boldsymbol{\delta}_c, \mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_m\boldsymbol{\Xi}_m\mathbf{S}'_m\mathbf{D}_{\boldsymbol{\beta}}\right),\end{aligned}$$

where $\mathbf{B}_m = \underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{P}_{mi}\mathbf{Q}_{mi} - \mathbf{I}_k\right)\mathbf{S}_0$, $\mathbf{P}_{mi} = \mathbf{S}_m\left(\mathbf{S}'_m\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}'_m$, and $\boldsymbol{\delta}_c$, $\mathbf{U}_m$, $\mathbf{V}_m$, and $\boldsymbol{\Xi}_m$ are defined in Theorem 1.

Corollary 1 shows the asymptotic distribution of $\widehat{\mu}_m$ when the rank condition is not satisfied for all submodels. A direct calculation yields

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

When the rank condition is satisfied for the mth model, the AMSE of $\widehat{\mu}_m$ can be derived by the same approach, that is,

$$\mathrm{AMSE}(\widehat{\mu}_m) = \mathbf{D}'_{\boldsymbol{\beta}}\left(\mathbf{B}_m\boldsymbol{\delta}_c\boldsymbol{\delta}'_c\mathbf{B}'_m + \mathbf{S}_m\boldsymbol{\Xi}_{um}\mathbf{S}'_m\right)\mathbf{D}_{\boldsymbol{\beta}}. \tag{4.2}$$

Ideally, we would choose the model associated with the lowest $\mathrm{AMSE}(\widehat{\mu}_m)$. To use (4.1) and (4.2) for model selection, we need to replace the unknown parameters $\mathbf{D}_{\boldsymbol{\beta}}$, $\mathbf{B}_m$, $\boldsymbol{\Xi}_m$, and $\boldsymbol{\delta}_c$ with the sample analogues. We now follow Claeskens and Hjort (2003) and propose a focused information criterion (FIC) for the large heterogeneous panel data model. The proposed FIC of the mth model is defined as

$$\mathrm{FIC}_m = \widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\left(\widehat{\mathbf{B}}_m\widehat{\boldsymbol{\delta}_c\boldsymbol{\delta}'_c}\widehat{\mathbf{B}}'_m + \mathbf{S}_m\widehat{\boldsymbol{\Xi}}_m\mathbf{S}'_m\right)\widehat{\mathbf{D}}_{\boldsymbol{\beta}}. \tag{4.3}$$

We show that the proposed FIC is an asymptotically unbiased estimator of $\mathrm{AMSE}(\widehat{\mu}_m)$ for both (4.1) and (4.2). Thus, the proposed FIC can be applied to cases irrespective of whether the rank condition holds or not. In practice, we select the model with the lowest

value of FIC_m. Note that the proposed FIC aims to minimize the sample analog of AMSE. Therefore, the model selected by FIC is expected to be the model with the lowest AMSE.

We now discuss the estimator for the unknown parameters in (4.1) and (4.2). We first consider the bias part. Define $\widehat{\mathbf{D}}_{\boldsymbol{\beta}} = \partial\mu(\widehat{\boldsymbol{\beta}}_{\text{MG},f})/\partial\boldsymbol{\beta}$, where $\widehat{\boldsymbol{\beta}}_{\text{MG},f}$ is the CCEMG estimate from the full model defined in (2.13). As shown in equations (3.1)–(3.2), $\widehat{\boldsymbol{\beta}}_{\text{MG},f}$ is a consistent estimator of the cross-sectional means $\boldsymbol{\beta}$. Thus, $\widehat{\mathbf{D}}_{\boldsymbol{\beta}}$ is a consistent estimator of $\mathbf{D}_{\boldsymbol{\beta}}$ by the continuous mapping theorem.

For $\mathbf{B}_m$, observe that $\mathbf{B}_m$ is a function of $\mathbf{Q}_{mi}$ and selection matrices. Consider the covariance matrix estimator $\widehat{\mathbf{Q}}_{mi} = T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{X}_i$. In the appendix, we show that $\widehat{\mathbf{Q}}_{mi}$ is a consistent estimator of $\mathbf{Q}_{mi}$. Thus, it follows that $\mathbf{B}_m$ can be consistently estimated by the sample analog $\widehat{\mathbf{B}}_m = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\mathbf{P}}_{mi}\widehat{\mathbf{Q}}_{mi} - \mathbf{I}_k\right)\mathbf{S}_0$.

We next discuss the estimator for the local parameter $\boldsymbol{\delta}_c$. Unlike other unknown parameters, the consistent estimator for the local parameter $\boldsymbol{\delta}_c$ is not available due to the local asymptotic framework. We can, however, construct an asymptotically unbiased estimator of $\boldsymbol{\delta}_c$ by using the estimator from the full model. Let $\widehat{\boldsymbol{\beta}}_{\text{MG},f} = (\widehat{\boldsymbol{\beta}}_{1,f}', \widehat{\boldsymbol{\beta}}_{2,f}')'$ such that $\widehat{\boldsymbol{\beta}}_{2,f} = \mathbf{S}_0'\widehat{\boldsymbol{\beta}}_{\text{MG},f}$. Then the asymptotically unbiased estimator is defined as $\widehat{\boldsymbol{\delta}}_c = \sqrt{N}\widehat{\boldsymbol{\beta}}_{2,f} = N^{-1/2}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_{2,fi}$. From (3.1)–(3.2), we can show that

$$\widehat{\boldsymbol{\delta}}_c = \sqrt{N}\widehat{\boldsymbol{\beta}}_{2,f} \xrightarrow{d} \mathbf{Z}_{\boldsymbol{\delta}} \sim \text{N}\left(\boldsymbol{\delta}_c, \mathbf{S}_0'\boldsymbol{\Xi}_f\mathbf{S}_0\right). \tag{4.4}$$

As shown above, $\widehat{\boldsymbol{\delta}}_c$ is an asymptotically unbiased estimator of $\boldsymbol{\delta}_c$. Therefore, the asymptotically unbiased estimator of $\boldsymbol{\delta}_c\boldsymbol{\delta}_c'$ is

$$\widehat{\boldsymbol{\delta}_c\boldsymbol{\delta}_c'} = \widehat{\boldsymbol{\delta}}_c\widehat{\boldsymbol{\delta}}_c' - \mathbf{S}_0'\widehat{\boldsymbol{\Xi}}_f\mathbf{S}_0, \tag{4.5}$$

where $\widehat{\boldsymbol{\Xi}}_f = \frac{1}{N-1}\sum_{i=1}^{N}(\widehat{\boldsymbol{\beta}}_{fi} - \widehat{\boldsymbol{\beta}}_{\text{MG},f})(\widehat{\boldsymbol{\beta}}_{fi} - \widehat{\boldsymbol{\beta}}_{\text{MG},f})'$ is a consistent estimator of $\boldsymbol{\Xi}_f$ by Corollary 2.

We now consider the variance part. For the covariance matrix $\boldsymbol{\Xi}_m$, we follow Pesaran (2006) and consider the nonparametric covariance matrix estimator

$$\widehat{\boldsymbol{\Xi}}_m = \frac{1}{N-1}\sum_{i=1}^{N}(\widehat{\boldsymbol{\beta}}_{mi} - \widehat{\boldsymbol{\beta}}_{\text{MG},m})(\widehat{\boldsymbol{\beta}}_{mi} - \widehat{\boldsymbol{\beta}}_{\text{MG},m})'. \tag{4.6}$$

The following corollary shows that $\widehat{\boldsymbol{\Xi}}_m$ is a consistent estimator for both cases when Assumption 6 holds and does not hold.

Corollary 2. Suppose that Assumptions 1–5 hold. As $N, T \to \infty$ jointly, we have $\widehat{\boldsymbol{\Xi}}_m \xrightarrow{p} \boldsymbol{\Xi}_m$. Further, if Assumption 6 is satisfied, then $\widehat{\boldsymbol{\Xi}}_m \xrightarrow{p} \boldsymbol{\Xi}_{um}$.

5 Plug-In Averaging Estimator

In this section, we extend the idea of the FIC and propose a plug-in model averaging estimator for the panel data model with a multifactor error structure. Instead of comparing the AMSE of each submodel, we first derive the AMSE of the averaging estimator with fixed weight in a local asymptotic framework. We next use this asymptotic result to characterize the optimal weights of the averaging estimator under the quadratic loss function. We then follow Liu (2015) and propose a plug-in estimator to estimate the infeasible optimal weights.

We now introduce the averaging estimator of the focus parameter μ. Let $w_m \geq 0$ be the weight corresponding to the mth submodel, and $\mathbf{w} = (w_1, ..., w_M)'$ be a weight vector belonging to the weight set $\mathcal{W} = \{w \in [0,1]^M : \sum_{m=1}^M w_m = 1\}$. That is, the weight vector lies in the unit simplex in $\mathbb{R}^M$. The model averaging estimator of μ is defined as

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

Note that the averaging estimator includes the CCEMG estimator in the mth submodel as a special case by setting the weight vector $\mathbf{w}$ to equal the unit weight vector $\mathbf{w}_m^0$ where the mth element is one and others are zeros. The following corollary shows the asymptotic normality of the averaging estimator with fixed weights.

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

$$\sqrt{N}(\widehat{\mu}(\mathbf{w}) - \mu) \xrightarrow{d} \mathrm{N}(\mathbf{D}_{\boldsymbol{\beta}}' \mathbf{B}(\mathbf{w}) \boldsymbol{\delta}_c, \boldsymbol{\Xi}(\mathbf{w})),$$

where

$$\mathbf{B}(\mathbf{w}) = \sum_{m=1}^{M} w_m \left(\underset{N\to\infty}{p\lim} \frac{1}{N} \sum_{i=1}^{N} \left(\mathbf{P}_{mi}\mathbf{Q}_{mi} - \mathbf{I}_k \right) \mathbf{S}_0 \right) = \sum_{m=1}^{M} w_m \mathbf{B}_m,$$

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

$$\boldsymbol{\Xi}_{m\ell} = \underset{N\to\infty}{p\lim} \frac{1}{N} \sum_{i=1}^{N} \mathbf{R}_{mi}^{-1} \mathbf{S}_m' \mathbf{Q}_{mi} \boldsymbol{\Omega}_{\boldsymbol{\beta}} \mathbf{Q}_{\ell i} \mathbf{S}_\ell \mathbf{R}_{\ell i}^{-1} + \underset{N\to\infty}{p\lim} \frac{1}{N} \sum_{i=1}^{N} \mathbf{R}_{mi}^{-1} \mathbf{S}_m' \boldsymbol{\Sigma}_{mi} \boldsymbol{\Omega}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{\ell i}' \mathbf{S}_\ell \mathbf{R}_{\ell i}^{-1}.$$

Corollary 3 shows the asymptotic normality of the averaging estimator with nonrandom weights when the rank condition is not satisfied for all submodels. The asymptotic bias and variance of the averaging estimator are $\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{B}(\mathbf{w})\boldsymbol{\delta}_c$ and $\boldsymbol{\Xi}(\mathbf{w})$, respectively.

This result implies that the AMSE of the averaging estimator $\widehat{\mu}(\mathbf{w})$ is

$$\text{AMSE}(\widehat{\mu}(\mathbf{w})) = \mathbf{w}'\boldsymbol{\Psi}\mathbf{w}, \tag{5.2}$$

where $\boldsymbol{\Psi}$ is an $M \times M$ matrix with the (m, ℓ)th element

$$\boldsymbol{\Psi}_{m\ell} = \mathbf{D}_{\boldsymbol{\beta}}' \left(\mathbf{B}_m \boldsymbol{\delta}_c \boldsymbol{\delta}_c' \mathbf{B}_\ell' + \mathbf{S}_m \boldsymbol{\Xi}_{m\ell} \mathbf{S}_\ell' \right) \mathbf{D}_{\boldsymbol{\beta}}. \tag{5.3}$$

Similarly, when the rank condition is satisfied for all models, the AMSE of $\widehat{\mu}(\mathbf{w})$ can be derived by the same approach, that is, $\text{AMSE}(\widehat{\mu}(\mathbf{w})) = \mathbf{w}'\boldsymbol{\Psi}\mathbf{w}$ with

$$\boldsymbol{\Psi}_{m\ell} = \mathbf{D}_{\boldsymbol{\beta}}' \left(\mathbf{B}_m \boldsymbol{\delta}_c \boldsymbol{\delta}_c' \mathbf{B}_\ell' + \mathbf{S}_m \boldsymbol{\Xi}_{u,m\ell} \mathbf{S}_\ell' \right) \mathbf{D}_{\boldsymbol{\beta}}, \tag{5.4}$$

$$\boldsymbol{\Xi}_{u,m\ell} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbf{R}_{mi}^{-1} \mathbf{S}_m' \mathbf{Q}_{mi} \boldsymbol{\Omega}_{\boldsymbol{\beta}} \mathbf{Q}_{\ell i} \mathbf{S}_\ell \mathbf{R}_{\ell i}^{-1}. \tag{5.5}$$

Since the AMSE of the averaging estimator $\widehat{\mu}(\mathbf{w})$ is linear-quadratic in $\mathbf{w}$, we can minimize the $\text{AMSE}(\widehat{\mu}(\mathbf{w}))$ over $\mathbf{w} \in \mathcal{W}$ and obtain the optimal fixed-weight vector:

$$\mathbf{w}^o = \underset{\mathbf{w}\in\mathcal{W}}{\text{argmin}} \mathbf{w}'\boldsymbol{\Psi}\mathbf{w}. \tag{5.6}$$

Note that when $M = 2$, we have a closed-form solution to (5.6). When $M > 2$, the optimal weight vector can be found numerically via quadratic programming, for which numerical algorithms are available for most programming languages.

The optimal weight vector, however, is infeasible, since $\boldsymbol{\Psi}$ is unknown. We now propose a plug-in estimator to estimate the optimal weights. We first estimate the AMSE of the

averaging estimator by plugging in an asymptotically unbiased estimator of $\mathbf{\Psi}$. We then choose the data-driven weights by minimizing the sample analogue of the AMSE and use these estimated weights to construct the plug-in averaging estimator.

Let $\widehat{\mathbf{\Psi}}$ be a sample analogue of $\mathbf{\Psi}$ with the (m, ℓ)th element

$$\widehat{\mathbf{\Psi}}_{m\ell} = \widehat{\mathbf{D}}'_{\boldsymbol{\beta}} \left(\widehat{\mathbf{B}}_m \widehat{\boldsymbol{\delta}_c \boldsymbol{\delta}'_c} \widehat{\mathbf{B}}'_\ell + \mathbf{S}_m \widehat{\mathbf{\Xi}}_{m\ell} \mathbf{S}'_\ell \right) \widehat{\mathbf{D}}_{\boldsymbol{\beta}}, \tag{5.7}$$

where $\widehat{\boldsymbol{\delta}_c \boldsymbol{\delta}'_c}$ is defined in (4.5) and

$$\widehat{\mathbf{\Xi}}_{m\ell} = \frac{1}{N-1} \sum_{i=1}^{N} (\widehat{\boldsymbol{\beta}}_{mi} - \widehat{\boldsymbol{\beta}}_{\mathrm{MG},m})(\widehat{\boldsymbol{\beta}}_{\ell i} - \widehat{\boldsymbol{\beta}}_{\mathrm{MG},\ell})'. \tag{5.8}$$

The following corollary shows that $\widehat{\mathbf{\Xi}}_{m\ell}$ is a consistent estimator for both cases when Assumption 6 holds and does not hold.

Corollary 4. Suppose that Assumptions 1–5 hold. As $N, T \to \infty$ jointly, we have $\widehat{\mathbf{\Xi}}_{m\ell} \xrightarrow{p} \mathbf{\Xi}_{m\ell}$. Furthermore, if Assumption 6 is satisfied, then $\widehat{\mathbf{\Xi}}_{m\ell} \xrightarrow{p} \mathbf{\Xi}_{u,m\ell}$.

We now define the plug-in averaging estimator. The data-driven weights based on the plug-in estimator are defined as

$$\widehat{\mathbf{w}} = (\widehat{w}_1, ..., \widehat{w}_M)' = \underset{\mathbf{w} \in \mathcal{W}}{\operatorname{argmin}} \mathbf{w}' \widehat{\mathbf{\Psi}} \mathbf{w}, \tag{5.9}$$

where $\mathbf{w}' \widehat{\mathbf{\Psi}} \mathbf{w}$ is an asymptotically unbiased estimator of $\mathbf{w}' \mathbf{\Psi} \mathbf{w}$. Similar to the optimal weight vector, the data-driven weights can also be computed numerically via quadratic programming. The plug-in averaging estimator of μ is defined as

$$\widehat{\mu}(\widehat{\mathbf{w}}) = \sum_{m=1}^{M} \widehat{w}_m \widehat{\mu}_m = \sum_{m=1}^{M} \widehat{w}_m \mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}). \tag{5.10}$$

As mentioned by Hjort and Claeskens (2003), we can also estimate $\mathbf{\Psi}$ by inserting $\widehat{\boldsymbol{\delta}}_c$ for $\boldsymbol{\delta}_c$ directly. Thus, the alternative estimator of $\mathbf{\Psi}_{m\ell}$ is

$$\widetilde{\mathbf{\Psi}}_{m\ell} = \widehat{\mathbf{D}}'_{\boldsymbol{\beta}} \left(\widehat{\mathbf{B}}_m \widehat{\boldsymbol{\delta}}_c \widehat{\boldsymbol{\delta}}_c{}' \widehat{\mathbf{B}}'_\ell + \mathbf{S}_m \widehat{\mathbf{\Xi}}_{m\ell} \mathbf{S}_\ell \right) \widehat{\mathbf{D}}_{\boldsymbol{\beta}}. \tag{5.11}$$

Our simulation shows that both averaging estimators (5.7) and (5.11) have similar finite sample performance. The following corollary presents the asymptotic distribution of the plug-in averaging estimator defined in (5.7)–(5.10).

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

$$\sqrt{N}(\widehat{\mu}(\widehat{\mathbf{w}}) - \mu) \xrightarrow{d} \sum_{m=1}^{M} w_m^* \Lambda_m, \tag{5.12}$$

where Λ_m is defined in Corollary 1, and $\mathbf{w}^* = (w_1^*, ..., w_M^*)' = \underset{\mathbf{w} \in \mathcal{W}}{\text{argmin}} \mathbf{w}' \mathbf{\Psi}^* \mathbf{w}$ and $\mathbf{\Psi}^*$ is an $M \times M$ matrix with the (m, ℓ)th element

$$\mathbf{\Psi}_{m\ell}^* = \mathbf{D}_{\boldsymbol{\beta}}' \left(\mathbf{B}_m (\mathbf{Z}_{\boldsymbol{\delta}} \mathbf{Z}_{\boldsymbol{\delta}}' - \mathbf{S}_0' \mathbf{\Xi}_f \mathbf{S}_0) \mathbf{B}_\ell' + \mathbf{S}_m \mathbf{\Xi}_{m\ell} \mathbf{S}_\ell' \right) \mathbf{D}_{\boldsymbol{\beta}}. \tag{5.13}$$

Unlike the averaging estimator with fixed weights, Corollary 5 shows that the averaging estimator with data-driven weights has a nonstandard limiting distribution. This is because the estimate $\widehat{\boldsymbol{\delta}_c \boldsymbol{\delta}_c'}$ is random in the limit, and hence estimated weights are asymptotically random under the local asymptotic framework. This non-normal nature of the asymptotic distribution of the averaging estimator with data-driven weights is also pointed out by Hjort and Claeskens (2003) as well as Liu (2015).

To conduct inference for the focus parameter μ, we follow Claeskens and Hjort (2008), Lu (2015), and DiTraglia (2016), and consider a simulation-based method to construct the confidence interval. We first replace the unknown parameters in Corollary 5 with the sample analogues. Then, by generating the random draws from Λ_m based on Corollary 1 and $\mathbf{Z}_{\boldsymbol{\delta}}$ based on (4.4) respectively, we can approximate the limiting distribution defined in Corollary 5 to arbitrary precision and use this simulated distribution to conduct inference. The details are described in Appendix E.

6 Simulation Study

6.1 Simulation Setup

We consider the following data generating process:

$$y_{it} = \mathbf{x}_{it}' \boldsymbol{\beta}_i + \boldsymbol{\gamma}_i' \mathbf{f}_t + \varepsilon_{it}, \quad \mathbf{x}_{it} = \mathbf{\Gamma}_i' \mathbf{f}_t + \mathbf{v}_{it},$$
$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i, \quad \boldsymbol{\beta} = d \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{\sqrt{N}} \left(1, \frac{k_2 - 1}{k_2}, \cdots, \frac{1}{k_2} \right) \right)',$$

where $f_{jt} \sim$ i.i.d.N$(0, 1)$, $\gamma_{ij} \sim$ i.i.d.N$(1, 0.25)$, $\Gamma_{ij\ell} \sim$ i.i.d.N$(0.5, 2.25)$, $\varepsilon_{it} \sim$ i.i.d.N$(0, \sigma_\varepsilon^2)$, $\eta_{i\ell} \sim$ i.i.d.N$(0, 0.01)$ for $j = 1, ..., r$ and $\ell = 1, ..., k$, and $\mathbf{v}_{it} = (\mathrm{v}_{1it}, ..., \mathrm{v}_{kit})' \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma}_i)$

where the diagonal elements of $\mathbf{\Sigma}_i$ are σ_{v}^2 and off-diagonal elements are $\rho\sigma_{\mathrm{v}}^2$. The number of regressors is $k = 6$ with two core regressors ($k_1 = 2$) and four auxiliary regressors ($k_2 = 4$). Parameters of σ_ε and σ_{v} are calculated based on the following fractions of the noise over the factor structure with noise evaluated at $d = 2$ and $r = 8$,

$$\tau_\varepsilon = \frac{\mathrm{Var}(\varepsilon_{it})}{\mathrm{Var}(\mathbf{x}_{it}'\boldsymbol{\beta}_i) + \mathrm{Var}(\boldsymbol{\gamma}_i'\mathbf{f}_t) + \mathrm{Var}(\varepsilon_{it})} = 0.15, \quad \tau_{\mathrm{v}} = \frac{\mathrm{Var}(\mathrm{v}_{it\ell})}{\mathrm{Var}(\mathbf{\Gamma}_{i\ell}'\mathbf{f}_t) + \mathrm{Var}(\mathrm{v}_{it\ell})} = 0.35.$$

After calculating σ_ε and σ_{v} based on $d = 2$ and $r = 8$, when we consider different d and r the fraction of noise will change as well. For example, when the true number of factors is less than 8, the fraction of the noise over the factor structure with noise becomes larger and vice versa.

Our parameter of interest is $\mu = \beta_1 + \beta_2$, corresponding to Example 1 mentioned in Section 4. Then we can construct the FIC for all submodels based on equation (4.3). In particular, $\widehat{\mathbf{D}}_{\boldsymbol{\beta}} = (1, 1, 0, 0, 0, 0)'$ and $\widehat{\boldsymbol{\delta}}_c$ is based on equation (4.4).[9] When the values of FIC for all submodels are obtained, we can select the model by FIC, and this model is expected to have the lowest AMSE of $\widehat{\mu} = \widehat{\beta}_1 + \widehat{\beta}_2$ among all submodels. For the plug-in averaging estimator, we do not tend to select the model but lower the AMSE by using the data-driven weights obtained by equation (5.9).

We consider six estimators including the full model, the averaging estimator with equal weights, AIC, BIC, FIC, and the Plug-in averaging estimator, and compare the small sample performance based on the risk.[10]

The risk (the mean squared error of $\widehat{\mu}$) is calculated by the average of $(\widehat{\mu} - \mu)^2$ obtained from each method over $2,500$ replications. We also normalize the risk by dividing by the risk of the infeasible optimal CCEMG estimator, i.e., the minimum of the average of $(\widehat{\mu}_m - \mu)^2$ among all submodels.

[9]The way of calculating $\widehat{\mathbf{D}}_{\boldsymbol{\beta}}$ depends on the choice of μ. For example, if we set $\mu = \beta_1$, then $\widehat{\mathbf{D}}_{\boldsymbol{\beta}} = (1, 0, 0, 0, 0, 0)'$.

[10]The AIC criterion for the mth model is $\mathrm{AIC}_m = \sum_{i=1}^N T\log(\widehat{\sigma}_{mi}^2) + 2N(2k_1 + 2k_{2m} + 1)$, where $\widehat{\sigma}_{mi}^2 = (\widehat{\mathbf{e}}_{mi}'\widehat{\mathbf{e}}_{mi})/(T - 2k_1 - 2k_{2m} - 1)$ and $\widehat{\mathbf{e}}_{mi} = \mathbf{M}_{hm}(\mathbf{y}_i - \mathbf{X}_{mi}\widehat{\boldsymbol{\beta}}_{mi})$ are the residuals of the submodel m. The BIC criterion for the mth model is $\mathrm{BIC}_m = \sum_{i=1}^N T\log(\widehat{\sigma}_{mi}^2) + \log(TN)N(2k_1 + 2k_{2m} + 1)$.

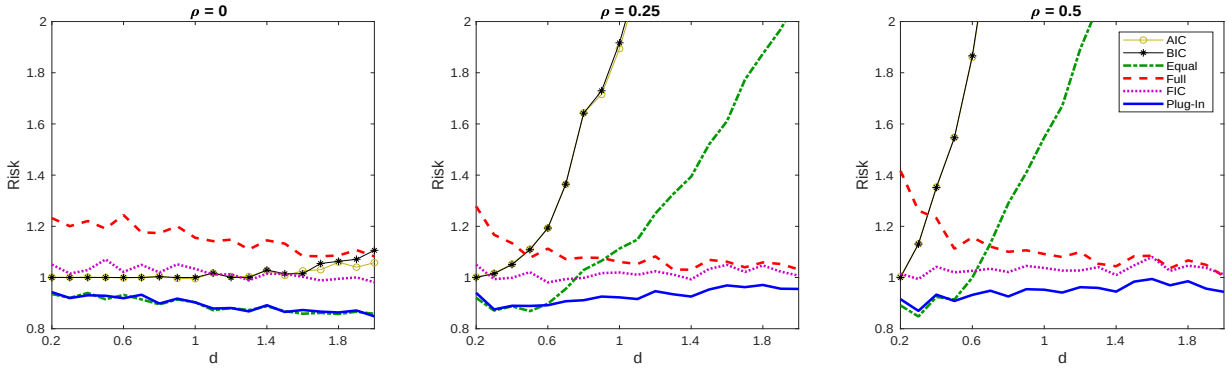

Figure 1: Normalized risk functions for $N = 100$, $T = 50$, and $r = 8$.

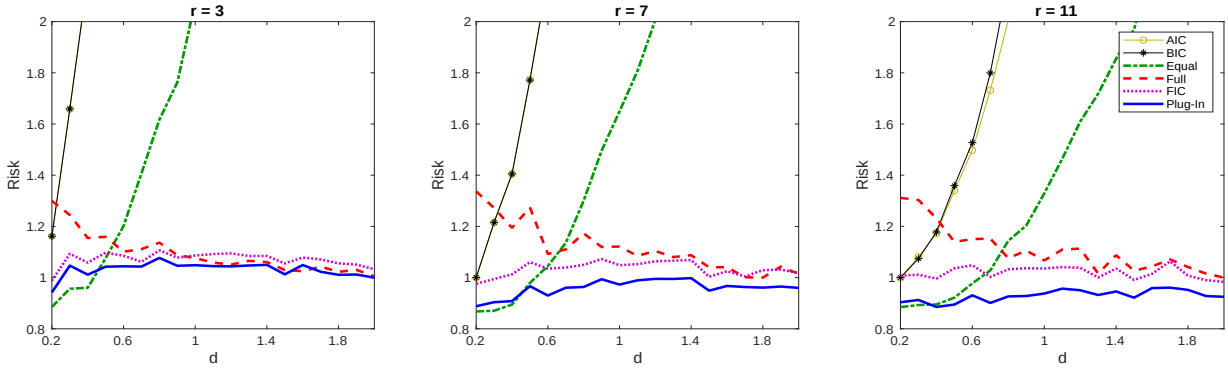

Figure 2: Normalized risk functions for $N = 100$, $T = 50$, and $\rho = 0.5$.

6.2 Simulation Results

The normalized risk functions are displayed in Figures 1–4. We first examine the finite sample performance in a general setting where the rank condition is not satisfied for all submodels. Figure 1 shows the normalized risk for $\rho = 0$, 0.25, and 0.5 in three panels. It is clear that FIC achieves lower normalized risk than Full in all cases, and the normalized risk of FIC is close to that of infeasible optimal model selection. AIC, BIC, and FIC have similar normalized risk for $\rho = 0$. However, both AIC and BIC have quite poor performance for $\rho = 0.25$ and 0.5. The normalized risk of Plug-In and Equal is indistinguishable for $\rho = 0$, but Equal has much larger normalized risk for $\rho = 0.25$ and 0.5. Overall, Plug-In performs well and dominates other estimators in most ranges of the parameter space. Figure 1 also shows that Plug-In achieves lower normalized risk than one, which means that the risk of Plug-In is lower than that of the infeasible best-fitting submodel m.

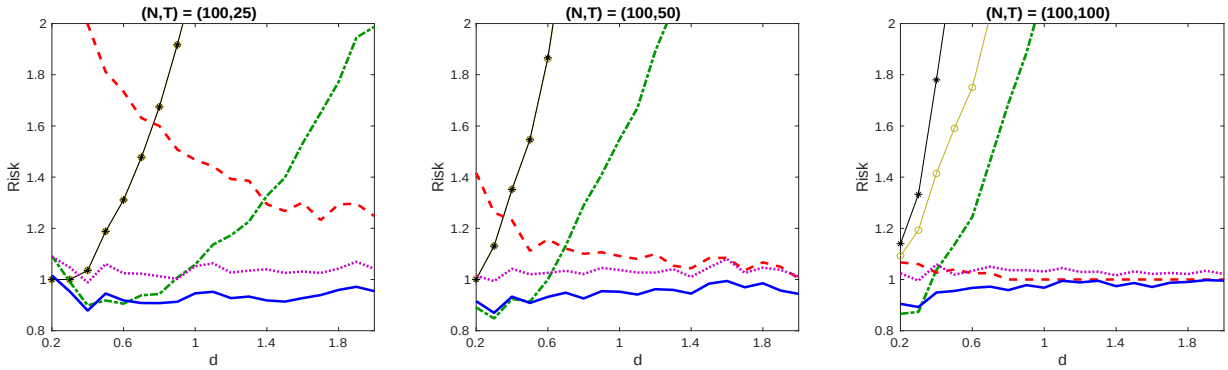

Figure 3: Normalized risk functions for $N = 100$, $r = 8$ and $\rho = 0.5$.

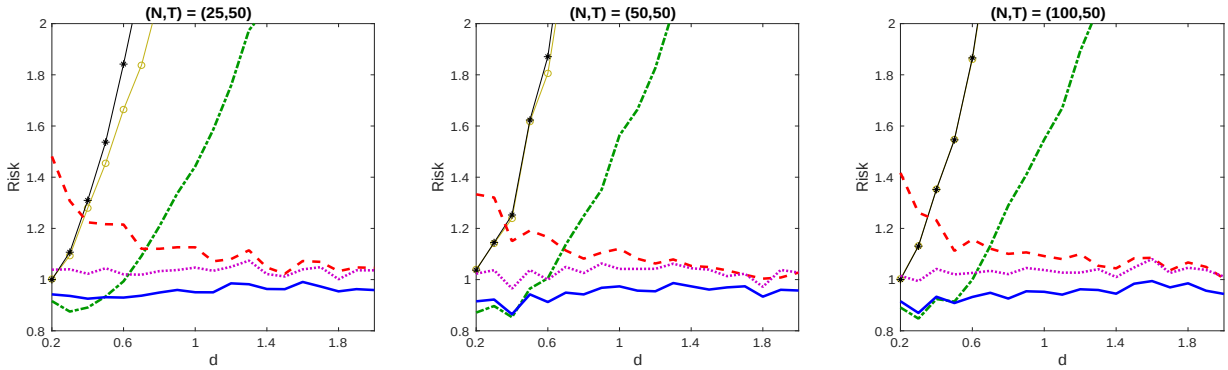

Figure 4: Normalized risk functions for $T = 50$, $r = 8$ and $\rho = 0.5$.

We now examine the normalized risk when the rank condition is satisfied for some submodels in Figure 2. For $r = 3$, the rank condition is satisfied for all models. In this setting, Plug-In, FIC, and Full have similar normalized risk, and they are better than AIC, BIC, and Equal. For $r = 7$, FIC and Full have similar performance for $d > 1$, but FIC has lower normalized risk than Full for $d < 1$. FIC also achieves much lower normalized risk than AIC and BIC in most ranges of the parameter d. In general, Plug-In has better performance than other estimators. The normalized risk of Plug-In and Equal is quite similar for $d < 0.4$. However, the normalized risk of Equal is quite poor relative to that of Plug-In for $d > 0.4$. For $r = 11$, the ranking of estimators is quite similar to that for $r = 7$.

We now examine the effect of the sample size on the normalized risk. Figure 3 shows the normalized risk for a fixed $N = 100$ and for $T = 25$, 50, and 100 in three panels. As the sample size T increases, the normalized risk of most estimators decreases. When $T = 25$ and 50, Plug-In outperforms other estimators in most cases. When $T = 100$, the

normalized risk of Plug-In, FIC, and Full is close to one for larger d and lower than those of AIC, BIC, and Equal in most ranges of the parameter d. Figure 4 shows the normalized risk for a fixed $T = 50$ and for $N = 25$, 50, and 100. Unlike the results shown in Figure 3, the ranking of estimators is quite similar across different sample sizes N. In most cases, Plug-In has much lower normalized risk than other estimators, and the performance of Plug-In is quite robust to different sample sizes.

7 Empirical Example

In this section, we apply the proposed method to examine possible causes of the increasing wage inequality between high-skilled and low-skilled workers in the U.S. manufacturing industries. Particularly, we focus on the effect of intersectoral technology skill complementarity (ITSC), which can be treated as an externality that escalates the wage premium. This cause comes from the idea considered in Voigtländer (2014), i.e. skill upgrading in one sector goes hand in hand with increasing skill demand in many other upstream and downstream sectors. A similar idea of this indirect effect from outsourcing is also suggested in Feenstra and Hanson (1995) and Eslava, Fieler, and Xu (2015).

In extant literature, there has been much debate on possible causes. For example, the inequality has also been increasing in developing countries even though the theory suggests that the increasing demand for goods made by low-skilled labor in developing countries should reduce the inequality.[11] Recently, one important cause that has been addressed considerably is skill-biased technology change (SBTC). It is based on the assumption that capital and skill are complementary, and therefore the increase of capital directly boosts the demand for high-skilled workers and expands wage inequality; see Krusell, Ohanian, Ríos-Rull, and Violante (2000). A review of surveys of wage inequality is available in Kurokawa (2014).

However, it is legitimate to wonder whether the model uncertainty renders the estimates practically when considering all plausible variables together. An added problem is that unobserved common shocks presented in panel data can affect skill intensity and other

[11] Krugman, Cooper, and Srinivasan (1995) pointed out that trade cannot be that important because the share of trade to GDP in some countries is too small.

determinants simultaneously and can lead to biased estimates of marginal effects. To overcome these difficulties, we apply FIC and the plug-in averaging estimator to evaluate the importance of ITSC to wage inequality with the consideration of a multifactor structure in the error term.

We follow a standard SBTC setup, e.g. Ciccone and Peri (2005), and consider a CES production function using two types of labor to produce final goods for each sector i at year t, that is

$$Y_{it} = \left[(B_{it} H_{it})^{\frac{\epsilon-1}{\epsilon}} + L_{it}^{\frac{\epsilon-1}{\epsilon}} \right]^{\frac{\epsilon}{\epsilon-1}} \tag{7.1}$$

where Y_{it} denotes the output, and H_{it} and L_{it} denote the high- and low-skilled workers respectively. ϵ is the elasticity of substitution between two inputs. B_{it} represents SBTC. To address the importance of ITSC, we adopt a similar modification from Voigtländer (2014) and allow $B_{it} = T_{it}\sigma_{it}^{\phi}$, and T_{it} represents a within-sector component, while σ_{it} is the input skill intensity. ϕ measures the strength of the complementarity between upstream and downstream skill bias. Then we can obtain the following relative demand curve for high- and low-skilled workers by combining with cost minimization,

$$\frac{H_{it}}{L_{it}} = \left(T_{it}\sigma_{it}^{\phi} \right)^{\epsilon-1} \left(\frac{w_{L,it}}{w_{H,it}} \right)^{\epsilon}, \tag{7.2}$$

where $\frac{w_{L,it}}{w_{H,it}}$ is the relative wage of low-skilled workers to high-skilled workers. In labor market equilibrium, the above equation implies that the relative wage has the following relationship,

$$\ln\left(\frac{w_{L,it}}{w_{H,it}} \right) = \frac{-\phi(\epsilon-1)}{\epsilon} \ln(\sigma_{it}) + \frac{1}{\epsilon} \ln\left(\frac{H_{it}}{L_{it}} \right) - \frac{\epsilon-1}{\epsilon} \ln(T_{it}). \tag{7.3}$$

According to this relationship, we can rewrite it as a regression framework,

$$\ln\left(\frac{w_{L,it}}{w_{H,it}} \right) = \beta_{1i} \ln(\sigma_{it}) + \beta_{2i} \ln\left(\frac{H_{it}}{L_{it}} \right) + \boldsymbol{\beta}_{3i}' \mathbf{z}_{it} + \boldsymbol{\gamma}_i' \mathbf{f}_t + \varepsilon_{it}. \tag{7.4}$$

Comparing Equations (7.3) and (7.4), we let the within SBTC be a sum of two parts. The first part is captured by observed variables, $\mathbf{z}_{it}$ including real capital equipment per worker (k^{equip}), research and development intensity ($R\&D_{lag}$), sectoral shares of high-technology capital (HT/K) and office, computing, and accounting equipment ($OCAM/K$) and broad (OS^{broad}) and narrow (OS^{narr}) measures of outsourcing used in Voigtländer (2014). The

second part cannot be observed and is represented by a factor structure, $\boldsymbol{\gamma}_i'\mathbf{f}_t$ that can be treated as the common shocks of this simultaneous system, and therefore this specification can also handle the endogenous problem.[12] In addition, we allow for slope heterogeneity and assume $\mathrm{E}(\beta_{1i}, \beta_{2i}, \boldsymbol{\beta}_{3i}) = (\beta_1, \beta_2, \boldsymbol{\beta}_3)$. When all β's can be consistently estimated, $\hat{\epsilon}$ is implied by $\frac{1}{\hat{\beta}_2}$ and $\hat{\phi}$ is implied by calculating $\frac{-\hat{\beta}_1\hat{\beta}_2^{-1}}{(\hat{\beta}_2^{-1}-1)}$.

7.1 Data and Empirical Methodology

The annual data are taken from Voigtländer (2014), available at the Review of Economics and Statistics website. It covers the period from 1958 to 2005 ($T = 48$). Input skill intensity σ_{it} is measured by weighted average share of white-collar workers employed by j and $j \neq i$, and the weights are calculated based on I-O expenditure data.[13]. To ensure the balance in the panel we exclude several sectors having missing values. In total, we have 313 sectors (N) with total sample size $15,024$ (NT).

According to the main regression (7.4), since we are interested in measuring the multiplier effect from σ_{it}, ϕ can be treated as our focus parameter and $\mathbf{z}_{it}$ are treated as auxiliary regressors. That is we can define $\mu = \phi = \frac{-\beta_2\beta_1^{-1}}{(\beta_1^{-1}-1)}$, then by using the results of Corollaries 1 and 3, we can calculate the sample counterpart of the AMSE for FIC and plug-in estimators. In particular,

$$\widehat{\mathbf{D}}_{\boldsymbol{\beta}} = \frac{\partial\mu}{\partial\widehat{\boldsymbol{\beta}}} = \left(-\frac{\hat{\beta}_{2,f}^{-1}}{(\hat{\beta}_{2,f}^{-1}-1)}, \frac{\hat{\beta}_{2,f}}{\hat{\beta}_{2,f}^{2}(\hat{\beta}_{2,f}^{-1}-1)} - \frac{\hat{\beta}_{2,f}}{\hat{\beta}_{2,f}^{3}(\hat{\beta}_{2,f}^{-1}-1)^2}, \mathbf{0}'\right)', \tag{7.5}$$

and $\hat{\beta}_{1,f} = N^{-1}\sum_{i=1}^{N}\hat{\beta}_{1i,f}$ and $\hat{\beta}_{2,f} = N^{-1}\sum_{i=1}^{N}\hat{\beta}_{2i,f}$. $\hat{\beta}_{1i,f}$ and $\hat{\beta}_{2i,f}$ are estimated from CCEMG based on full model specification.

7.2 Empirical Results

We now turn to the results of Equation (7.4) under different model specifications including narrow (without any auxiliary regressor), full, equal weights, AIC, BIC, FIC and plug-in averaging models. The estimation results and the standard error are presented in Table 1. All specifications are based on the CCEMG approach, implying the individual fixed

[12]This unobserved part is treated as the sum of fixed effect and time effect in Ciccone and Peri (2005).

[13]Details of the data description and definition can be founded in Voigtländer (2014).

Table 1: Estimation results

	Narrow	Full	Equal	AIC	BIC	FIC	Plug-In
$\ln(\sigma_{it})$	-0.591**	-0.706**	-0.732**	-0.657**	-0.591**	-0.746**	-0.693**
	(0.225)	(0.237)	(0.213)	(0.242)	(0.225)	(0.223)	(0.201)
$\ln(H_{it}/L_{it})$	0.352**	0.454**	0.410**	0.426**	0.352**	0.410**	0.403**
	(0.014)	(0.014)	(0.013)	(0.014)	(0.014)	(0.014)	(0.013)
k_{it}^{equip}		-1.644**	-0.647**	-1.926**		-0.977*	-0.598**
		(0.567)	(0.264)	(0.475)		(0.656)	(0.298)
$(OCAM/K)_{it}$		-3.291*	-0.270	-1.563*			-0.274
		(1.792)	(0.616)	(1.431)			(0.384)
$(HT/K - OCAM/K)_{it}$		2.280**	1.395**	3.059**			0.842**
		(1.138)	(0.408)	(0.977)			(0.366)
$R\&D_{lag,it}$		1.438*	0.717**				0.761**
		(0.856)	(0.350)				(0.357)
OS_{it}^{narr}		-1.196**	-0.467*			-0.882*	-0.669*
		(0.534)	(0.263)			(0.643)	(0.383)
$(OS^{\text{broad}} - OS^{\text{narr}})_{it}$		0.198	0.033			-0.133	0.036
		(0.387)	(0.135)			(0.277)	(0.166)
μ	0.912	1.292	1.241	1.145	0.912	1.263	1.161

Note: Standard errors are reported in parentheses. The standard error of the plug-in averaging estimator is computed as follows: $\widehat{\boldsymbol{\Xi}}(\widehat{\mathbf{w}}) = \sum_{m=1}^{M} \widehat{w}_m^2 \widehat{\mathbf{D}}_{\boldsymbol{\beta}}' \mathbf{S}_m \widehat{\boldsymbol{\Xi}}_m \mathbf{S}_m' \widehat{\mathbf{D}}_{\boldsymbol{\beta}} + 2\sum\sum_{m\neq\ell} \widehat{w}_m \widehat{w}_\ell \widehat{\mathbf{D}}_{\boldsymbol{\beta}}' \mathbf{S}_m \widehat{\boldsymbol{\Xi}}_{m\ell} \mathbf{S}_\ell' \widehat{\mathbf{D}}_{\boldsymbol{\beta}}$. ** denotes a t-statistic greater than 2, and * denotes a t-statistic between 1 and 2.

effects and unobserved common factors are under control. The coefficients from $\ln(H_{it}/L_{it})$ range from 0.352 to 0.454, which implies that long-run elasticity ranges from 2.20 to 2.841, suggesting that high-skilled and low-skilled labor are substitutes. The coefficients of $\ln(\sigma_{it})$ vary between -0.746 to -0.591, which shows that the ITSC increases the wage inequality. The implied strength of ITSC is reported in the last row of Table 1 and ranges from 0.912 to 1.292. When considering the standard error, we can observe that the standard errors of $\ln(\sigma_{it})$ and $\ln(H_{it}/L_{it})$ of the narrow model are smaller than the full model, and the implied μ's from these two models have the largest difference (0.38) among all considered models. This result raises the concern of model uncertainty.

When model uncertainty is controlled, we can obtain smaller standard errors of $\ln(\sigma_{it})$ and $\ln(H_{it}/L_{it})$ compared to full and narrow models even though the focus parameter is ϕ. For example the standard error of $\ln(\sigma_{it})$ is 0.201 for the plug-in averaging estimator but it is 0.237 for the full model. Similar results can be obtained when we consider FIC. We also construct the 90% confidence interval for FIC and plug-in averaging estimates by one-step simulation-based method discussed in Appendix E.[14] It shows that the 90% confidence intervals of FIC and plug-in averaging estimator of μ are (0.624,1.991) and (0.538,1.772) respectively, which are narrower than the narrow model (0.228,1.596).

For auxiliary regressors, all coefficients have the same signs across different estimation methods, while the magnitudes are quite different. In the full model, we can observe that (k^{equip}) and $OCAM/K$ have larger marginal effects compared to the result from the plug-in estimate, and this supports the notion that an increase in capital per worker can enlarge the wage inequality. We also obtain that narrow outsourcing has a strong effect on wage inequality. However, it is surprising that research and development intensity ($R\&D_{lag}$) and sectoral shares of high-technology capital (HT/K) decrease the wage gap between high-skilled and low-skilled workers, and is different from the idea of Acemoglu (2003). A possible reason is that both ($R\&D_{lag}$) and HT/K provide spillover effects and help unskilled workers improve productivity. But the reason is still unclear, and the interaction of technical change and wage inequality needs to be clarified in order to identify the effects of $R\&D$ on wages of high-skilled and low-skilled workers in future studies.

8 Extension

In this section, we extend our results to the case where slope coefficients and factor loadings are correlated. We relax the independence condition between factor loadings and slope coefficients in Assumptions 3–4 to the following assumption:

Assumption 7. The factor loadings and slope coefficients follow the random coefficient

[14] DiTraglia (2016) and Lu (2015) indicate the one-step simulation-based method performs remarkably well in finite samples.

model: $\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\iota}_i$, $\boldsymbol{\Gamma}_i = \boldsymbol{\Gamma} + \boldsymbol{\xi}_i$, and $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i$, where

$$\begin{bmatrix} \boldsymbol{\eta}_i \\ \text{vec}(\boldsymbol{\xi}_i) \\ \boldsymbol{\iota}_i \end{bmatrix} \sim \text{i.i.d.} \left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Omega}_{\boldsymbol{\beta}} & \boldsymbol{\Omega}_{\boldsymbol{\beta\Gamma}} & \boldsymbol{\Omega}_{\boldsymbol{\beta\gamma}} \\ \boldsymbol{\Omega}_{\boldsymbol{\Gamma\beta}} & \boldsymbol{\Omega}_{\boldsymbol{\Gamma}} & \boldsymbol{\Omega}_{\boldsymbol{\Gamma\gamma}} \\ \boldsymbol{\Omega}_{\boldsymbol{\gamma\beta}} & \boldsymbol{\Omega}_{\boldsymbol{\gamma\Gamma}} & \boldsymbol{\Omega}_{\boldsymbol{\gamma}}, \end{bmatrix} \right)$$

with a finite covariance matrix. The random deviations $\boldsymbol{\tau}_i = (\boldsymbol{\iota}_i, \text{vec}(\boldsymbol{\xi}_i)', \boldsymbol{\eta}_i)$ are distributed independently of $\boldsymbol{\varepsilon}_{jt}$, $\mathbf{v}_{jt}$ and $\mathbf{f}_t$ for all j and t with finite fourth-order cross moments.

However, this relaxation could cause endogeneity because of the fact that $\mathbf{X}_i = \mathbf{F}\boldsymbol{\Gamma}_i + \mathbf{v}_i$. To deal with the endogeneity, we propose a generalized method of moments (GMM) procedure combined with the CCE pooled estimator, hereafter GMM-CCEP for simplicity.[15] We then derive the asymptotic normality of the GMM-CCEP estimator for each submodel and discuss the plug-in averaging estimator in this case. The following assumption is about the spatial error process.

Assumption 8. Let $\boldsymbol{\varepsilon}_t = \boldsymbol{\mathcal{R}}_{\boldsymbol{\varepsilon}}\boldsymbol{\epsilon}_t$ and $\mathbf{v}_t = \boldsymbol{\mathcal{R}}_{\mathbf{v}}\boldsymbol{\upsilon}_t$, where $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\upsilon}_t$ follow linear stationary processes with absolute summable autocovariances, and $\boldsymbol{\mathcal{R}}_{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\mathcal{R}}_{\mathbf{v}}$ are two $N \times N$ matrices with bounded row and column norms.

In Assumption 8, $\boldsymbol{\mathcal{R}}_{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\mathcal{R}}_{\mathbf{v}}$ are used to characterize the spatial correlation measured by geographical or policy and social distance, and it is possible to have $\boldsymbol{\mathcal{R}}_{\boldsymbol{\varepsilon}} = \boldsymbol{\mathcal{R}}_{\mathbf{v}}$. The spatial correlation assumption can also be imposed to analyze the case where factor loadings and slope coefficients are uncorrelated; see Pesaran and Tosetti (2011). It is general enough to cover different specifications of spatial models, for example, a spatial moving average model, spatial autoregressive model, and spatial error component model; see Sarafidis and Wansbeek (2012) for more details.

Let $\mathbf{Z}_{mi}$ be a $T \times k^*$ matrix of instrumental variables constructed from $\{\mathbf{X}_{mj}\}_{j \neq i}$ for the submodel m, and $k^* \geq k_m$. The choice of j is based on the knowledge of spatial weights matrix implied by Assumption 8, that is, we choose the individual j who is the nearest neighbor for individual i based on $\boldsymbol{\mathcal{R}}_{\mathbf{v}}$. We then set $\mathbf{Z}_{mi} = \mathbf{X}_{mj}$ and have $k^* = k_m$. It is possible to choose multiple neighbors as instrumental variables. In this case, k^* is a multiple

[15] Under the slope heterogeneity assumption, we may consider the common correlated effects pooled estimator; see Appendix F for more details.

of k_m; see Appendix G for more details. Note that the proper choice of j can ensure the identification, that is, $\text{Rank}(\mathbf{\Sigma}_{\mathbf{zx},m}\mathbf{S}_m) = k_m$, where $\mathbf{\Sigma}_{\mathbf{zx},m} = \underset{NT\to\infty}{p\lim}\,(NT)^{-1}\mathbf{Z}_{mi}'\mathbf{M}_{gm}\mathbf{X}_i$.

By Assumption 7, the orthogonality condition $\mathrm{E}\left(\boldsymbol{\iota}_i, \boldsymbol{\xi}_i, \boldsymbol{\eta}_i, \boldsymbol{\varepsilon}_i | \mathbf{Z}_{mi}\right) = \mathbf{0}$ holds for all $i \neq j$. Therefore, for the submodel m, we have the following moment condition

$$\mathrm{E}\left(\mathbf{Z}_{mi}'\mathbf{M}_{gm}(\mathbf{F}\mathbf{\Gamma}\boldsymbol{\eta}_i + (\boldsymbol{\xi}_i\boldsymbol{\eta}_i - \overline{\boldsymbol{\xi}\boldsymbol{\eta}}) + \mathbf{V}_i\boldsymbol{\eta}_i + \mathbf{F}\boldsymbol{\iota}_i + \boldsymbol{\varepsilon}_i)\right) = \mathbf{0}, \tag{8.1}$$

where $\mathbf{M}_{gm} = \mathbf{I}_T - \mathbf{F}\mathbf{C}_m\left(\mathbf{C}_m'\mathbf{F}'\mathbf{F}\mathbf{C}_m\right)^{-}\mathbf{C}_m'\mathbf{F}'$ and $\mathbf{C}_m = \mathrm{E}(\bar{\mathbf{C}}_m)$. This moment condition is derived from the following result

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\mathbf{M}_{gm}(\mathbf{F}\mathbf{\Gamma}\boldsymbol{\eta}_i + (\boldsymbol{\xi}_i\boldsymbol{\eta}_i - \overline{\boldsymbol{\xi}\boldsymbol{\eta}}) \\ + \mathbf{V}_i\boldsymbol{\eta}_i + \mathbf{F}\boldsymbol{\iota}_i + \boldsymbol{\varepsilon}_i) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).\end{aligned} \tag{8.2}$$

See Appendix B for the derivation of (8.2).

Note that the moment condition holds only for $\mathbf{M}_{gm}$ but not $\bar{\mathbf{M}}_{gm}$. This is because $\bar{\mathbf{M}}_{gm}$ involves $\bar{\mathbf{C}}_m$, and it is correlated with slope heterogeneity and factor loadings. The equation (8.2) implies that the moment condition can be approximated by $\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$, because the convergence rate of the estimated mean of slope coefficients is established on $\sqrt{N}$. Therefore, using $\bar{\mathbf{M}}_{hm}$ to form the objective function would not change the asymptotic properties of the GMM estimator.

Let $\widehat{\mathbf{W}}_m$ be a $k^* \times k^*$ symmetric and positive definite weighting matrix and $\widehat{\mathbf{W}}_m \xrightarrow{p} \mathbf{W}_m$ as $N, T \to \infty$.[16] The GMM-CCEP estimator for each submodel m minimizes the following objective function

$$N\left(\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}(\mathbf{y}_i - \mathbf{X}_{mi}\boldsymbol{\beta}_m^*)\right)'\widehat{\mathbf{W}}_m\left(\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}(\mathbf{y}_i - \mathbf{X}_{mi}\boldsymbol{\beta}_m^*)\right). \tag{8.3}$$

Thus, the GMM-CCEP estimator for the submodel m is defined as

$$\widehat{\boldsymbol{\beta}}_{\mathrm{GMMP},m} = \left(\mathbf{S}_m'\widehat{\mathbf{\Sigma}}_{\mathbf{zx},m}'\widehat{\mathbf{W}}_m\widehat{\mathbf{\Sigma}}_{\mathbf{zx},m}\mathbf{S}_m\right)^{-1}\left(\mathbf{S}_m'\widehat{\mathbf{\Sigma}}_{\mathbf{zx},m}'\widehat{\mathbf{W}}_m\widehat{\mathbf{\Sigma}}_{\mathbf{zy},m}\right), \tag{8.4}$$

[16]For the choice of $\widehat{\mathbf{W}}_m$, we can conduct a two-step procedure by replacing $\widehat{\mathbf{W}}_m$ by an inverse of a consistent estimate of $\widetilde{\mathbf{\Omega}}_m$ and obtain the efficient GMM estimator for the submodel m. In the first step, we can use either $\widehat{\mathbf{W}}_m = \mathbf{I}$ or $\widehat{\mathbf{W}}_m = \left((NT)^{-1}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{Z}_{mi}\right)^{-1}$.

where $\widehat{\boldsymbol{\Sigma}}_{\mathbf{zx},m} = (NT)^{-1}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_i$ and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{zy},m} = (NT)^{-1}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{y}_i$. Here we use "GMMP" to refer to the GMM-CCEP estimator for simplicity.

We first introduce some notation that we will use to characterize the asymptotic distribution of the GMM-CCEP estimator. Let $\widetilde{\boldsymbol{\Sigma}}_{mi} = (\boldsymbol{\Sigma}_{\mathbf{z},mi}, \boldsymbol{\Sigma}_{\mathbf{zv},mi})$, where $\boldsymbol{\Sigma}_{\mathbf{z},mi} = \underset{T\to\infty}{p\lim} T^{-1}\mathbf{Z}_{mi}'\mathbf{M}_{gm}\mathbf{F}$, and $\boldsymbol{\Sigma}_{\mathbf{zv},mi} = \underset{T\to\infty}{p\lim} T^{-1}\mathbf{Z}_{mi}'\mathbf{M}_{gm}\mathbf{V}_i$. Also, let $\boldsymbol{\Lambda}_i = (\boldsymbol{\phi}_i', \boldsymbol{\eta}_i')'$ and $\mathrm{Var}(\boldsymbol{\Lambda}_i) = \boldsymbol{\Omega}_{\boldsymbol{\Lambda}}$, where $\boldsymbol{\phi}_i = \boldsymbol{\Gamma}\boldsymbol{\eta}_i + (\boldsymbol{\xi}_i\boldsymbol{\eta}_i - \overline{\boldsymbol{\xi}\boldsymbol{\eta}}) + \boldsymbol{\iota}_i$. We next present the asymptotic distribution of the GMM-CCEP estimator in the following theorem.

Theorem 3. Suppose that Assumptions 2, 5, 7 and 8 hold. As $N, T \to \infty$ jointly, we have

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\mathrm{GMMP},m} - \boldsymbol{\beta}_m^*) \xrightarrow{d} \widetilde{\mathbf{A}}_m\boldsymbol{\delta}_c + \mathbf{E}_m \sim \mathrm{N}\left(\widetilde{\mathbf{A}}_m\boldsymbol{\delta}_c, \widetilde{\boldsymbol{\Xi}}_m\right),$$
$$\widetilde{\boldsymbol{\Xi}}_m = \widetilde{\mathbf{R}}_m^{-1}\mathbf{S}_m'\boldsymbol{\Sigma}_{\mathbf{zx},m}'\mathbf{W}_m\widetilde{\boldsymbol{\Omega}}_m\mathbf{W}_m\boldsymbol{\Sigma}_{\mathbf{zx},m}\mathbf{S}_m\widetilde{\mathbf{R}}_m^{-1},$$

where $\widetilde{\mathbf{A}}_m = \widetilde{\mathbf{R}}_m^{-1}\mathbf{S}_m'\widetilde{\mathbf{Q}}_m\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)$, $\widetilde{\mathbf{R}}_m = \mathbf{S}_m'\widetilde{\mathbf{Q}}_m\mathbf{S}_m$, $\widetilde{\mathbf{Q}}_m = \boldsymbol{\Sigma}_{\mathbf{zx},m}'\mathbf{W}_m\boldsymbol{\Sigma}_{\mathbf{zx},m}$, $\widetilde{\boldsymbol{\Omega}}_m = \underset{N\to\infty}{p\lim} N^{-1}\sum_{i=1}^{N}\widetilde{\boldsymbol{\Sigma}}_{mi}\boldsymbol{\Omega}_{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{\Sigma}}_{mi}'$, and $\mathbf{E}_m \sim \mathrm{N}(\mathbf{0}, \widetilde{\boldsymbol{\Xi}}_m)$.

Theorem 3 shows that for each submodel m, the GMM-CCEP estimator is consistent and has asymptotic bias $\widetilde{\mathbf{A}}_m\boldsymbol{\delta}_c$ under a local to zero framework. Compared to Theorem 1, the asymptotic covariance matrix $\widetilde{\boldsymbol{\Xi}}_m$ is more complicated because of the correlation between the slope heterogeneity and factor loadings. Furthermore, we cannot separate $\mathbf{E}_m$ into two independent normal random vectors as shown in Theorem 1.[17] Similarly, the results in Sections 4 and 5 still hold, except that $\mathbf{A}_m$ and $\boldsymbol{\Xi}_m$ are replaced by $\widetilde{\mathbf{A}}_m$ and $\widetilde{\boldsymbol{\Xi}}_m$. Thus, we can construct FIC and the plug-in averaging estimator in the same way as we discuss in Sections 4 and 5; see Appendix G for finite sample comparison of the CCEMG estimator and the GMM-CCEP estimator.

9 Conclusion

In this paper, we extend the existing literature on frequentist model averaging to a panel data framework with a multifactor error structure. We follow Pesaran (2006) and estimate the cross-sectional means of unknown slope coefficients by common correlated effects mean

[17] When the slope coefficients and factor loadings are uncorrelated as we discuss in Section 3.1, one can show that the GMM-CCEP estimator is still valid and $\mathbf{E}_m$ is the sum of two independent random vectors.

group estimators and study the limiting distributions of all submodel estimators in a local asymptotic framework. We then propose a focused information criterion and a plug-in averaging estimator for large heterogeneous panels and study the asymptotic properties in a local asymptotic framework. Our proposed selection criterion and averaging estimators aim to minimize the sample analog of the asymptotic mean squared error and can be applied to cases irrespective of whether the rank condition holds or not. Our Monte Carlo simulations show that the proposed estimators have satisfactory expected squared error compared to other methods.

References

Acemoglu, D. (2003). Patterns of skill premia. *The Review of Economic Studies 70*(2), 199–230.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4), 1229–1279.

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

Buckland, S., K. Burnham, and N. Augustin (1997). Model selection: An integral part of inference. *Biometrics 53*, 603–618.

Chang, M. and F. J. DiTraglia (2018). A generalized focused information criterion for GMM. *Journal of Applied Econometrics 33*(3), 378–397.

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*(2), 393 – 420.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal 14*(1), C45–C90.

Ciccone, A. and G. Peri (2005). Long-run substitutability between more and less educated workers: Evidence from US states, 1950–1990. *Review of Economics and statistics 87*(4), 652–663.

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.

Claeskens, G. and N. L. Hjort (2008). *Model Selection and Model Averaging*. Cambridge: Cambridge University Press.

DiTraglia, F. (2016). Using invalid instruments on purpose: Focused moment selection and averaging for GMM. *Journal of Econometrics 195*, 187–208.

Eslava, M., A. C. Fieler, and D. Y. Xu (2015). (Indirect) input linkages. *American Economic Review 105*(5), 662–66.

Feenstra, R. C. and G. H. Hanson (1995). Foreign investment, outsourcing and relative wages. *National Bureau of Economic Research No. w5121*.

Gao, Y., X. Zhang, S. Wang, and G. Zou (2016). Model averaging based on leave-subject-out cross-validation. *Journal of Econometrics 192*(1), 139–151.

Hansen, B. E. (2005). Challenges for econometric model selection. *Econometric Theory 21*(1), 60–68.

Hansen, B. E. (2007). Least squares model averaging. *Econometrica 75*, 1175–1189.

Hansen, B. E. (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.

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 G. D. A. Phillips and E. Tzavalis (Eds.), *The Refinement of Econometric Estimation and Test Procedures*, pp. 239–281. Cambridge University Press.

Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics 160*(2), 326 – 348.

Karabiyik, H., S. Reese, and J. Westerlund (2017). On the role of the rank condition in CCE estimation of factor-augmented panel regressions. *Journal of Econometrics 197*(1), 60–64.

Kitagawa, T. and C. Muris (2016). Model averaging in semiparametric estimation of treatment effects. *Journal of Econometrics 193*, 271–289.

Krugman, P., R. N. Cooper, and T. Srinivasan (1995). Growing world trade: Causes and consequences. *Brookings Papers on Economic Activity 1995*(1), 327–377.

Krusell, P., L. E. Ohanian, J.-V. Ríos-Rull, and G. L. Violante (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica 68*(5), 1029–1053.

Kurokawa, Y. (2014). A survey of trade and wage inequality: Anomalies, resolutions and new trends. *Journal of Economic Surveys 28*(1), 169–193.

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 and Economic Statistics 33*, 506–522.

Lu, X. and L. Su (2015). Jackknife model averaging for quantile regressions. *Journal of Econometrics 188*(1), 40–58.

Moon, H. R. and M. Weidner (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica 83*(4), 1543–1579.

Moral-Benito, E. (2015). Model averaging in economics: An overview. *Journal of Economic Surveys 29*(1), 46–75.

Paap, R., W. Wang, and X. Zhang (2015). To pool or not to pool: What is a good strategy? Working paper.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica 74*(4), 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*(2), 94 – 115.

Pesaran, M. H. and E. Tosetti (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics 161*(2), 182 – 202.

Sarafidis, V. and T. Wansbeek (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews 31*(5), 483–531.

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*(460), 1167–1179.

Su, L. and S. Jin (2012). Sieve Estimation of Panel Data Models with Cross Section Dependence. *Journal of Econometrics 169*(1), 34–47.

Sueishi, N. (2013). Generalized empirical likelihood-based focused information criterion and model averaging. *Econometrics 1*, 141–156.

Voigtländer, N. (2014). Skill bias magnified: Intersectoral linkages and white-collar labor demand in US manufacturing. *Review of Economics and Statistics 96*(3), 495–513.

Westerlund, J. and J.-P. Urbain (2015). Cross-sectional averages versus principal components. *Journal of Econometrics 185*(2), 372–377.

Yang, Y. (2001). Adaptive regression by mixing. *Journal of the American Statistical Association 96*(454), 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 and Economic Statistics 30*, 132–142.

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

This appendix contains six parts. Appendix A contains the proofs of the main theorems and corollaries. Appendix B provides supplementary lemmas and their proofs. Appendix C provides the discussion of several special cases. Appendix D provides numerical results to illustrate the phenomena claimed in Theorems 1 and 2. Appendix E describes the details of the simulation-based method for constructing the confidence interval of the averaging estimator. Appendix F discusses the CCE pooled estimator. Appendix G details the GMM-CCEP estimator when factor loadings and slope heterogeneity are correlated.

A Proofs of Theorems and Corollaries

Proof of Theorem 1. Note that $\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_{mi}$. We first decompose the CCE estimator in the submodel m. Observe that $\mathbf{X}_{mi} = (\mathbf{X}_{1i}, \mathbf{X}_{2i}\boldsymbol{\Pi}_m') = \mathbf{X}_i\mathbf{S}_m$, $\boldsymbol{\beta}_{mi} = (\boldsymbol{\beta}_{1i}', \boldsymbol{\beta}_{2i}'\boldsymbol{\Pi}_m')' = \mathbf{S}_m'\boldsymbol{\beta}_i$, and $\mathbf{X}_{2i} = \mathbf{X}_i\mathbf{S}_0$. By some algebra, it follows that

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{mi} &= (\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{y}_i \\
&= (\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}(\mathbf{X}_{1i}\boldsymbol{\beta}_{1i} + \mathbf{X}_{2i}\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\boldsymbol{\beta}_{2i} + \mathbf{X}_{2i}(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i} + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i) \\
&= (\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_{mi}\boldsymbol{\beta}_{mi} \\
&\quad + (\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}(\mathbf{X}_{2i}(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i} + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta}_{mi} + (\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}(\mathbf{X}_i\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i} + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i). \qquad \text{(A.1)}
\end{aligned}
$$

By Assumption 4, we have $\boldsymbol{\beta}_{mi} = \mathbf{S}_m'\boldsymbol{\beta}_i = \mathbf{S}_m'(\boldsymbol{\beta} + \boldsymbol{\eta}_i)$. Then we have

$$
\begin{aligned}
\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m} - \boldsymbol{\beta}_m^*) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\widehat{\boldsymbol{\beta}}_{mi} - \boldsymbol{\beta}_{mi} + \boldsymbol{\beta}_{mi} - \boldsymbol{\beta}_m^*) \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{S}_m'\boldsymbol{\eta}_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_i\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 \\
&\quad + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_i\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\eta}_{2i} \\
&\quad + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{F}\boldsymbol{\gamma}_i
\end{aligned}
$$

$$
\begin{aligned}
&+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\mathbf{X}_{mi}' \bar{\mathbf{M}}_{hm} \mathbf{X}_{mi})^{-1} \mathbf{X}_{mi}' \bar{\mathbf{M}}_{hm} \boldsymbol{\varepsilon}_i \\
&\equiv I_1 + I_2 + I_3 + I_4 + I_5. \qquad \text{(A.2)}
\end{aligned}
$$

We consider the first and third terms of (A.2). Observe that

$$
\begin{aligned}
\mathbf{S}_0(\mathbf{I}_{k_2} - \mathbf{\Pi}_m' \mathbf{\Pi}_m)\boldsymbol{\eta}_{2i} &= \mathbf{S}_0(\mathbf{I}_{k_2} - \mathbf{\Pi}_m' \mathbf{\Pi}_m)\mathbf{S}_0' \boldsymbol{\eta}_i \\
&= \begin{bmatrix} \mathbf{0}_{k_1} & \mathbf{0}_{k_1 \times k_2} \\ \mathbf{0}_{k_2 \times k_1} & \mathbf{I}_{k_2} - \mathbf{\Pi}_m' \mathbf{\Pi}_m \end{bmatrix} \boldsymbol{\eta}_i \\
&= (\mathbf{I}_k - \mathbf{S}_m \mathbf{S}_m') \boldsymbol{\eta}_i. \qquad \text{(A.3)}
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
I_1 + I_3 &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{S}_m' \boldsymbol{\eta}_i + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\mathbf{X}_{mi}' \bar{\mathbf{M}}_{hm} \mathbf{X}_{mi})^{-1} \mathbf{X}_{mi}' \bar{\mathbf{M}}_{hm} \mathbf{X}_i (\mathbf{I}_k - \mathbf{S}_m \mathbf{S}_m') \boldsymbol{\eta}_i \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\mathbf{S}_m' + (\mathbf{X}_{mi}' \bar{\mathbf{M}}_{hm} \mathbf{X}_{mi})^{-1} \mathbf{X}_{mi}' \bar{\mathbf{M}}_{hm} \mathbf{X}_i (\mathbf{I}_k - \mathbf{S}_m \mathbf{S}_m') \right) \boldsymbol{\eta}_i \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\mathbf{S}_m' + (\mathbf{S}_m' (T^{-1} \mathbf{X}_i' \bar{\mathbf{M}}_{hm} \mathbf{X}_i) \mathbf{S}_m)^{-1} \mathbf{S}_m' (T^{-1} \mathbf{X}_i' \bar{\mathbf{M}}_{hm} \mathbf{X}_i)(\mathbf{I}_k - \mathbf{S}_m \mathbf{S}_m') \right) \boldsymbol{\eta}_i \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Big(\mathbf{S}_m' + (\mathbf{S}_m' (T^{-1} \mathbf{X}_i' \bar{\mathbf{M}}_{hm} \mathbf{X}_i) \mathbf{S}_m)^{-1} \mathbf{S}_m' (T^{-1} \mathbf{X}_i' \bar{\mathbf{M}}_{hm} \mathbf{X}_i) \\
&\quad - (\mathbf{S}_m' (T^{-1} \mathbf{X}_i' \bar{\mathbf{M}}_{hm} \mathbf{X}_i) \mathbf{S}_m)^{-1} (\mathbf{S}_m' (T^{-1} \mathbf{X}_i' \bar{\mathbf{M}}_{hm} \mathbf{X}_i) \mathbf{S}_m) \mathbf{S}_m' \Big) \boldsymbol{\eta}_i \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\mathbf{S}_m' (T^{-1} \mathbf{X}_i' \bar{\mathbf{M}}_{hm} \mathbf{X}_i) \mathbf{S}_m)^{-1} \mathbf{S}_m' (T^{-1} \mathbf{X}_i' \bar{\mathbf{M}}_{hm} \mathbf{X}_i) \boldsymbol{\eta}_i \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\mathbf{S}_m' (T^{-1} \mathbf{X}_i' \bar{\mathbf{M}}_{gm} \mathbf{X}_i) \mathbf{S}_m)^{-1} \mathbf{S}_m' (T^{-1} \mathbf{X}_i' \bar{\mathbf{M}}_{gm} \mathbf{X}_i) \boldsymbol{\eta}_i + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),
\end{aligned}
$$

where the last equality holds by Lemma B.1 (i). By Assumption 4, as $N, T \to \infty$ jointly, we have

$$
I_1 + I_3 \xrightarrow{d} \mathbf{U}_m \sim \mathrm{N}(\mathbf{0}, \mathbf{\Xi}_{um}), \qquad \text{(A.4)}
$$

where

$$
\begin{aligned}
\mathbf{\Xi}_{um} &= p\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} (\mathbf{S}_m' \mathbf{Q}_{mi} \mathbf{S}_m)^{-1} \mathbf{S}_m' \mathbf{Q}_{mi} \mathbf{\Omega}_{\boldsymbol{\beta}} \mathbf{Q}_{mi} \mathbf{S}_m (\mathbf{S}_m' \mathbf{Q}_{mi} \mathbf{S}_m)^{-1} \\
&= p\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathbf{R}_{mi}^{-1} \mathbf{S}_m' \mathbf{Q}_{mi} \mathbf{\Omega}_{\boldsymbol{\beta}} \mathbf{Q}_{mi} \mathbf{S}_m \mathbf{R}_{mi}^{-1},
\end{aligned}
$$

and $\mathbf{R}_{mi} = \mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_m$.

We next consider the second term of (A.2). By Assumption 5 and Lemma B.1 (i), we have

$$
\begin{aligned}
I_2 &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m)^{-1}\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_0(\mathbf{I}_{k_2}-\mathbf{\Pi}_m'\mathbf{\Pi}_m)\sqrt{N}\Delta_{NT}^{-1}\boldsymbol{\delta}, \\
&= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\mathbf{X}_i)\mathbf{S}_m)^{-1}\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\mathbf{X}_i)\mathbf{S}_0(\mathbf{I}_{k_2}-\mathbf{\Pi}_m'\mathbf{\Pi}_m)\sqrt{N}\Delta_{NT}^{-1}\boldsymbol{\delta} \\
&\quad + O_p\left(\frac{1}{\Delta_{NT}\sqrt{N}}\right) + O_p\left(\frac{1}{\Delta_{NT}\sqrt{T}}\right), \\
&\xrightarrow{p} \underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}(\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_m)^{-1}\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_0(\mathbf{I}_{k_2}-\mathbf{\Pi}_m'\mathbf{\Pi}_m)c\boldsymbol{\delta} = \mathbf{A}_m\boldsymbol{\delta}_c, \qquad \text{(A.5)}
\end{aligned}
$$

where $\mathbf{A}_m = \underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\mathbf{R}_{mi}^{-1}\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_0(\mathbf{I}_{k_2}-\mathbf{\Pi}_m'\mathbf{\Pi}_m)$ and $\boldsymbol{\delta}_c = c\cdot\boldsymbol{\delta}$.

We now consider the fourth term of (A.2). Let $\bar{\boldsymbol{\gamma}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\gamma}_i$ and $\bar{\boldsymbol{\iota}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\iota}_i$. By Assumption 3, we have

$$
\begin{aligned}
I_4 &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m)^{-1}\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F})\boldsymbol{\gamma}_i \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m)^{-1}\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F})\left(\bar{\boldsymbol{\gamma}}+(\boldsymbol{\iota}_i-\bar{\boldsymbol{\iota}})\right) \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\mathbf{X}_i)\mathbf{S}_m)^{-1}\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\mathbf{F})\left(\boldsymbol{\iota}_i-\bar{\boldsymbol{\iota}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),
\end{aligned}
$$

where the last equality holds by Lemma B.1 (i), (iii), and (iv). Therefore, by Assumption 3, as $N, T \to \infty$ jointly, we have

$$
I_4 \xrightarrow{d} \mathbf{V}_m \sim \mathrm{N}(\mathbf{0}, \mathbf{\Xi}_{vm}), \qquad \text{(A.6)}
$$

where

$$
\begin{aligned}
\mathbf{\Xi}_{vm} &= \underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}(\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_m)^{-1}\mathbf{S}_m'\mathbf{\Sigma}_{mi}\mathbf{\Omega}_{\boldsymbol{\gamma}}\mathbf{\Sigma}_{mi}'\mathbf{S}_m(\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_m)^{-1} \\
&= \underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\mathbf{R}_{mi}^{-1}\mathbf{S}_m'\mathbf{\Sigma}_{mi}\mathbf{\Omega}_{\boldsymbol{\gamma}}\mathbf{\Sigma}_{mi}'\mathbf{S}_m\mathbf{R}_{mi}^{-1}.
\end{aligned}
$$

Note that $\boldsymbol{\eta}_i$ are distributed independently of $\boldsymbol{\iota}_i$ across i by Assumption 4. Thus, $\mathbf{U}_m$ and $\mathbf{V}_m$ are two stochastically independent normal random vectors.

For the last term of (A.2), note that ε_{it} is independent of $\mathbf{v}_{1it}$, $\mathbf{v}_{2it}$ and $\mathbf{f}_t$ for all i and t by Assumptions 1 and 2. It follows that $\mathrm{E}(I_5) = 0$ and

$$\begin{aligned} I_5 &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m)^{-1}\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\boldsymbol{\varepsilon}_i) \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\mathbf{X}_i)\mathbf{S}_m)^{-1}\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\boldsymbol{\varepsilon}_i) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \\ &= O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \end{aligned} \tag{A.7}$$

where the second equality holds by Lemma B.1 (i) and (ii), and the last equality holds by the fact that $T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\boldsymbol{\varepsilon}_i = O_p\left(T^{-1/2}\right)$. Combining (A.4), (A.5), (A.6), and (A.7), we have

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m} - \boldsymbol{\beta}_m^*) \xrightarrow{d} \mathbf{A}_m\boldsymbol{\delta}_c + \mathbf{U}_m + \mathbf{V}_m \sim \mathrm{N}\left(\mathbf{A}_m\boldsymbol{\delta}_c, \boldsymbol{\Xi}_{um} + \boldsymbol{\Xi}_{vm}\right).$$

This completes the proof. □

Proof of Theorem 2. We first consider the case where $k_m + 1 > r$ when Assumption 6 holds. Using the results from Lemma B.2 and the fact that $T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{X}_i = O_p(1)$, the fourth and fifth terms of (A.2) become

$$I_4 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m)^{-1}\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F})\boldsymbol{\gamma}_i = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),$$

and

$$\begin{aligned} I_5 &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m)^{-1}\mathbf{S}_m' T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\boldsymbol{\varepsilon}_i \\ &= \frac{1}{N} \sum_{i=1}^{N} (\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m)^{-1}\mathbf{S}_m' \frac{\sqrt{N}}{T}\mathbf{X}_i'\mathbf{M}_{f^0}\boldsymbol{\varepsilon}_i \\ &\quad + O_p(T^{-1}N^{1/2}) + O_p(N^{-1/2}) + O_p(T^{-1/2}) \\ &= O_p(T^{-1}N^{1/2}) + O_p(N^{-1/2}) + O_p(T^{-1/2}). \end{aligned}$$

We now consider the case in which $k_m + 1 = r$ when Assumption 6 holds. In this case, we have $\bar{\mathbf{M}}_{gm} = \mathbf{I}_T - \bar{\mathbf{G}}_m(\bar{\mathbf{G}}_m'\bar{\mathbf{G}}_m)^{-}\bar{\mathbf{G}}_m' = \mathbf{I}_T - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-}\mathbf{F}' = \mathbf{M}_f$. Thus, it follows that $\mathbf{M}_{gm}\mathbf{F} = \mathbf{M}_f\mathbf{F} = \mathbf{0}$. Therefore, we have

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

Similarly, the fourth and fifth terms of (A.2) become

$$
\begin{aligned}
I_4 &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m)^{-1}\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F})\boldsymbol{\gamma}_i \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m)^{-1}\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\mathbf{M}_f\mathbf{F})\boldsymbol{\gamma}_i + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \\
&= O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
I_5 &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m)^{-1}\mathbf{S}_m' T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\boldsymbol{\varepsilon}_i \\
&= \frac{1}{N} \sum_{i=1}^{N} (\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m)^{-1}\mathbf{S}_m' \frac{\sqrt{N}}{T}\mathbf{X}_i'\mathbf{M}_f\boldsymbol{\varepsilon}_i + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \\
&= O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}
$$

Also, the remaining terms of (A.2) have the same order of limits as shown in the proof of Theorem 1. Note that when Assumption 6 holds, we have

$$
\begin{aligned}
T^{-1}\mathbf{X}_i'\mathbf{M}_f\mathbf{X}_i &= T^{-1}\mathbf{V}_i'\mathbf{M}_f\mathbf{V}_i \\
&= T^{-1}\mathbf{V}_i'\mathbf{V}_i - (T^{-1}\mathbf{V}_i'\mathbf{F})\left(T^{-1}\mathbf{F}'\mathbf{F}\right)^{-}(T^{-1}\mathbf{F}'\mathbf{V}_i) \\
&= T^{-1}\mathbf{V}_i'\mathbf{V}_i + O_p\left(T^{-1}\right),
\end{aligned}
$$

where the last equality holds by the facts that $T^{-1}\mathbf{V}_i'\mathbf{F} = O_p(T^{-1/2})$ and $T^{-1}\mathbf{F}'\mathbf{F} = O_p(1)$. Thus, we have $\mathbf{Q}_{mi} = p\lim_{T\to\infty}(\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\mathbf{X}_i) = \boldsymbol{\Sigma}_i$. This completes the proof. □

Proof of Theorem 3. For each submodel m, we can decompose the GMM-CCEP estimator as follows,

$$
\begin{aligned}
&\sqrt{N}\left(\hat{\boldsymbol{\beta}}_{\text{GMM-P},m} - \boldsymbol{\beta}_m^*\right) \\
&= \mathbf{q}_{NT}\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_{mi}\boldsymbol{\eta}_{mi} \\
&\quad + \mathbf{q}_{NT}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_i\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\sqrt{N}\boldsymbol{\beta}_2 \\
&\quad + \mathbf{q}_{NT}\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_i\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\eta}_{2i}
\end{aligned}
$$

$$
\begin{aligned}
&+ \mathbf{q}_{NT}\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\mathbf{Z}'_{mi}\bar{\mathbf{M}}_{hm}\mathbf{F}\boldsymbol{\gamma}_i \\
&+ \mathbf{q}_{NT}\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\mathbf{Z}'_{mi}\bar{\mathbf{M}}_{hm}\boldsymbol{\varepsilon}_i \\
&\equiv L_1 + L_2 + L_3 + L_4 + L_5, \qquad \text{(A.8)}
\end{aligned}
$$

where $\mathbf{q}_{NT} = \frac{\left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}'_{mi}\bar{\mathbf{M}}_{hm}\mathbf{X}_{mi}\right)'\left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}'_{mi}\bar{\mathbf{M}}_{hm}\mathbf{Z}_{mi}\right)^{-1}}{\left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}'_{mi}\bar{\mathbf{M}}_{hm}\mathbf{X}_{mi}\right)'\left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}'_{mi}\bar{\mathbf{M}}_{hm}\mathbf{Z}_{mi}\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}'_{mi}\bar{\mathbf{M}}_{hm}\mathbf{X}_{mi}\right)}$. The first equality follows the same steps used in proving Theorems 3.1. Combining the results from the facts that $\boldsymbol{\Sigma}_{\mathbf{zx},m} = p\lim_{NT\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}'_{mi}\bar{\mathbf{M}}_{hm}\mathbf{X}_i$ and $\mathbf{Q}_{\mathbf{z},m} = p\lim_{NT\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}'_{mi}\bar{\mathbf{M}}_{hm}\mathbf{Z}_{mi}$ and Assumption 5, we can first derive that

$$
L_2 \xrightarrow{p} (\mathbf{S}'_m\mathbf{Q}_{\mathbf{xzx},m}\mathbf{S}_m)^{-1}\mathbf{S}'_m\mathbf{Q}_{\mathbf{xzx},m}\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\boldsymbol{\delta}_c,
$$

where $\mathbf{Q}_{\mathbf{xzx},m} = \boldsymbol{\Sigma}'_{\mathbf{zx},m}\mathbf{Q}^{-1}_{\mathbf{z},m}\boldsymbol{\Sigma}_{\mathbf{zx},m}$. As for L_1, L_3 and L_4, using Lemma B.3, it follows that

$$
\begin{aligned}
&L_1 + L_3 + L_4 \\
&= \mathbf{q}_{NT}\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\mathbf{Z}'_{mi}\mathbf{M}_{gm}\left(\mathbf{F}(\boldsymbol{\Gamma}\boldsymbol{\eta}_i + (\boldsymbol{\xi}_i\boldsymbol{\eta}_i - \overline{\boldsymbol{\xi}\boldsymbol{\eta}}) + (\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}})) + \mathbf{V}_i\boldsymbol{\eta}_i\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \\
&= \mathbf{q}_{NT}\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\mathbf{Z}'_{mi}\mathbf{M}_{gm}\mathbf{F}\boldsymbol{\phi}_i + \mathbf{q}_{NT}\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\mathbf{Z}'_{mi}\mathbf{M}_{gm}\mathbf{V}_i\boldsymbol{\eta}_i + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \\
&\xrightarrow{d} \mathbf{E}_{\text{GMM-P},m} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Xi}_{\text{GMM-P},m}),
\end{aligned}
$$

where $\boldsymbol{\phi}_i = \boldsymbol{\Gamma}\boldsymbol{\eta}_i + (\boldsymbol{\xi}_i\boldsymbol{\eta}_i - \overline{\boldsymbol{\xi}\boldsymbol{\eta}}) + (\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}})$, and $\boldsymbol{\phi}_i$ and $\boldsymbol{\eta}_i$ are independent of $\mathbf{Z}_{mi}$. In particular,

$$
\begin{aligned}
\boldsymbol{\Xi}_{\text{GMM-P},m} =& (\mathbf{S}'_m\mathbf{Q}_{\mathbf{xzx},m}\mathbf{S}_m)^{-1}\mathbf{S}'_m\boldsymbol{\Sigma}'_{\mathbf{zx},m}\mathbf{Q}^{-1}_{\mathbf{z},m}(\ddot{\boldsymbol{\Lambda}}_{\boldsymbol{\phi},m} + \ddot{\boldsymbol{\Lambda}}_{\boldsymbol{\beta},m} \\
&+ \ddot{\boldsymbol{\Lambda}}_{\boldsymbol{\phi\beta},m} + \ddot{\boldsymbol{\Lambda}}'_{\boldsymbol{\phi\beta},m})\mathbf{Q}^{-1}_{\mathbf{z},m}\boldsymbol{\Sigma}_{\mathbf{zx},m}\mathbf{S}_m(\mathbf{S}'_m\mathbf{Q}_{\mathbf{xzx},m}\mathbf{S}_m)^{-1},
\end{aligned}
$$

where

$$
\begin{aligned}
\ddot{\boldsymbol{\Lambda}}_{\boldsymbol{\phi},m} &= p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{\mathbf{z},mi}\boldsymbol{\Omega}_{\boldsymbol{\phi}}\boldsymbol{\Sigma}'_{\mathbf{z},mi}, \\
\ddot{\boldsymbol{\Lambda}}_{\boldsymbol{\beta},m} &= p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{\mathbf{zv},mi}\boldsymbol{\Omega}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}'_{\mathbf{zv},mi}, \\
\ddot{\boldsymbol{\Lambda}}_{\boldsymbol{\phi\beta},m} &= p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{\mathbf{z},mi}\boldsymbol{\Omega}_{\boldsymbol{\phi\beta}}\boldsymbol{\Sigma}'_{\mathbf{zv},mi},
\end{aligned}
$$

where $\boldsymbol{\Omega}_{\boldsymbol{\beta}} = \mathrm{Var}(\boldsymbol{\eta}_i)$, $\boldsymbol{\Omega}_{\boldsymbol{\phi}} = \mathrm{Var}(\boldsymbol{\phi}_i)$ and $\boldsymbol{\Omega}_{\boldsymbol{\phi\beta}} = \mathrm{Cov}(\boldsymbol{\phi}_i, \boldsymbol{\eta}_i)$ which exist and are bounded implied by Assumption 7. $L_5 = O_p(N^{-1/2}) + O_p(T^{-1/2})$ and the proof is similar to proving I_5. □

Proof of Corollary 1. Define $\boldsymbol{\beta}_{2m}^c = \{\boldsymbol{\beta}_2 : \beta_{2j} \notin \boldsymbol{\beta}_{2m},\ for\ j = 1, ..., k_2\}$. That is, $\boldsymbol{\beta}_{2m}^c$ is the set of parameters β_{2j} that are not included in submodel m. Hence, we can write $\mu(\boldsymbol{\beta}) = \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \boldsymbol{\beta}_{2m}^c)$ and $\mu(\boldsymbol{\beta}_m^*) = \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \mathbf{0})$. By a standard Taylor series expansion, it follows that

$$\begin{aligned}
\mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \boldsymbol{\beta}_{2m}^c) &= \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \mathbf{0}) + \frac{\partial \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \mathbf{0})}{\partial \boldsymbol{\beta}_{2m}^c}' \boldsymbol{\beta}_{2m}^c + O\left(\frac{1}{\Delta_{NT}^2}\right) \\
&= \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \mathbf{0}) + \frac{\partial \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \mathbf{0})}{\partial \boldsymbol{\beta}_2}' (\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m' \boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 + O\left(\frac{1}{\Delta_{NT}^2}\right) \\
&= \mu(\boldsymbol{\beta}_m^*) + \mathbf{D}_{\boldsymbol{\beta}_2}'(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m' \boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 + O\left(\frac{1}{\Delta_{NT}^2}\right). \qquad \text{(A.9)}
\end{aligned}$$

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

By the result from Theorem 1, and the delta method, we have

$$\sqrt{N}(\mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}) - \mu(\boldsymbol{\beta}_m^*)) \xrightarrow{d} \mathbf{D}_{\boldsymbol{\beta}_m^*}' \left(\mathbf{A}_m \boldsymbol{\delta}_c + \mathbf{U}_m + \mathbf{V}_m\right). \qquad \text{(A.10)}$$

Combining (A.9) and (A.10), it follows that

$$\begin{aligned}
\sqrt{N}(\mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}) - \mu(\boldsymbol{\beta})) &= \sqrt{N}(\mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}) - \mu(\boldsymbol{\beta}_m^*)) - \sqrt{N}(\mu(\boldsymbol{\beta}) - \mu(\boldsymbol{\beta}_m^*)) \\
&\xrightarrow{d} \mathbf{D}_{\boldsymbol{\beta}_m^*}'(\mathbf{A}_m \boldsymbol{\delta}_c + \mathbf{U}_m + \mathbf{V}_m) - \mathbf{D}_{\boldsymbol{\beta}_2}'(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m' \boldsymbol{\Pi}_m)\boldsymbol{\delta}_c. \qquad \text{(A.11)}
\end{aligned}$$

Using the facts $\mathbf{S}_0 \boldsymbol{\Pi}_m' = \mathbf{S}_m(\mathbf{0}_{k_1 \times k_{2m}}', \mathbf{I}_{k_{2m}})'$, $\mathbf{D}_{\boldsymbol{\beta}_m^*}' = \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m$, $\mathbf{D}_{\boldsymbol{\beta}_2}' = \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_0$ and $\mathbf{A}_m = \underset{N\to\infty}{p\lim} \frac{1}{N} \sum_{i=1}^N (\mathbf{S}_m' \mathbf{Q}_{mi} \mathbf{S}_m)^{-1} \mathbf{S}_m' \mathbf{Q}_{mi} \mathbf{S}_0 (\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m' \boldsymbol{\Pi}_m)$, we can obtain

$$\begin{aligned}
&\mathbf{D}_{\boldsymbol{\beta}_m^*}' \mathbf{A}_m \boldsymbol{\delta}_c - \mathbf{D}_{\boldsymbol{\beta}_2}'(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m' \boldsymbol{\Pi}_m)\boldsymbol{\delta}_c \\
&= \underset{N\to\infty}{p\lim} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{D}_{\boldsymbol{\beta}}' \mathbf{S}_m \left(\mathbf{S}_m' \mathbf{Q}_{mi} \mathbf{S}_m\right)^{-1} \mathbf{S}_m' \mathbf{Q}_{mi} \mathbf{S}_0 - \mathbf{D}_{\boldsymbol{\beta}}' \mathbf{S}_0 \right) \boldsymbol{\delta}_c \\
&\quad - \underset{N\to\infty}{p\lim} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{D}_{\boldsymbol{\beta}}' \mathbf{S}_m \left(\mathbf{S}_m' \mathbf{Q}_{mi} \mathbf{S}_m\right)^{-1} \mathbf{S}_m' \mathbf{Q}_{mi} \mathbf{S}_0 \boldsymbol{\Pi}_m' \boldsymbol{\Pi}_m - \mathbf{D}_{\boldsymbol{\beta}}' \mathbf{S}_0 \boldsymbol{\Pi}_m' \boldsymbol{\Pi}_m \right) \boldsymbol{\delta}_c \\
&= \underset{N\to\infty}{p\lim} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{D}_{\boldsymbol{\beta}}' \mathbf{S}_m \mathbf{R}_{mi}^{-1} \mathbf{S}_m' \mathbf{Q}_{mi} \mathbf{S}_0 - \mathbf{D}_{\boldsymbol{\beta}}' \mathbf{S}_0 \right) \boldsymbol{\delta}_c \\
&= \mathbf{D}_{\boldsymbol{\beta}}' \mathbf{B}_m \boldsymbol{\delta}_c, \qquad \text{(A.12)}
\end{aligned}$$

where $\mathbf{B}_m = \underset{N\to\infty}{p\lim} \frac{1}{N} \sum_{i=1}^N \mathbf{B}_{mi}$ and $\mathbf{B}_{mi} = (\mathbf{S}_m \mathbf{R}_{mi}^{-1} \mathbf{S}_m' \mathbf{Q}_{mi} - \mathbf{I}_k)\mathbf{S}_0$. Consequently, equation (A.11) becomes

$$\begin{aligned}
\sqrt{N}(\mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}) - \mu(\boldsymbol{\beta})) &\xrightarrow{d} \mathbf{D}_{\boldsymbol{\beta}}'(\mathbf{B}_m \boldsymbol{\delta}_c + \mathbf{S}_m \mathbf{U}_m + \mathbf{S}_m \mathbf{V}_m) \\
&= \Lambda_m \sim \mathrm{N}(\mathbf{D}_{\boldsymbol{\beta}}' \mathbf{B}_m \boldsymbol{\delta}_c, \mathbf{D}_{\boldsymbol{\beta}}' \mathbf{S}_m \boldsymbol{\Xi}_m \mathbf{S}_m' \mathbf{D}_{\boldsymbol{\beta}}).
\end{aligned}$$

When $\boldsymbol{\Sigma}_i$ is a diagonal matrix for all i, $\mathbf{A}_m$ is $o_p(1)$ as $N, T \to \infty$, which is shown in Corollary C.1. Therefore, the bias term only involves $\mathbf{D}'_{\boldsymbol{\beta}_2}(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\boldsymbol{\delta}_c = \mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\boldsymbol{\delta}_c$. When Assumption 6 holds, $\mathbf{V}_m = o_p(1)$ as $N, T \to \infty$ and $\sqrt{N}/T \to 0$ as we discussed in Theorem 2. This completes the proof. □

Proof of Corollary 2. The argument is similar to the proof of Corollary 4, and we omit it for brevity. □

Proof of Corollary 3. By Assumptions 1–5 and the result from Theorem 1, we have

$$\begin{aligned}\sqrt{N}(\mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}) - \mu(\boldsymbol{\beta}^*_m)) &= \widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\mathbf{S}_m\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m} - \boldsymbol{\beta}^*_m) + o_p(1)\\ &= \widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big(\mathbf{S}_m\widehat{\mathbf{R}}^{-1}_{mi}\mathbf{S}'_m\widehat{\mathbf{Q}}_{mi}\boldsymbol{\eta}_i + \mathbf{S}_m\widehat{\mathbf{R}}^{-1}_{mi}\mathbf{S}_m\widehat{\boldsymbol{\Sigma}}_{mi}(\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}})\\ &\quad + \mathbf{S}_m\widehat{\mathbf{A}}_{mi}\Delta^{-1}_{NT}\boldsymbol{\delta}\Big) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),\end{aligned}$$

where $\widehat{\mathbf{A}}_{mi} = (\mathbf{S}'_m\widehat{\mathbf{Q}}_{mi}\mathbf{S}_m)^{-1}\mathbf{S}'_m\widehat{\mathbf{Q}}_{mi}\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)$ and $\widehat{\mathbf{R}}_{mi} = \mathbf{S}'_m\widehat{\mathbf{Q}}_{mi}\mathbf{S}_m$. Therefore, by (A.9) and the fact that the weights are non-random, it follows that

$$\begin{aligned}&\sqrt{N}\left(\widehat{\mu}(\mathbf{w}) - \mu(\boldsymbol{\beta})\right)\\ &= \sum_{m=1}^{M} w_m\sqrt{N}\left(\mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}) - \mu(\boldsymbol{\beta}^*_m) + \mu(\boldsymbol{\beta}^*_m) - \mu(\boldsymbol{\beta})\right)\\ &= \sum_{m=1}^{M} w_m\widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{S}_m\widehat{\mathbf{R}}^{-1}_{mi}\mathbf{S}'_m\widehat{\mathbf{Q}}_{mi}\boldsymbol{\eta}_i + \sum_{m=1}^{M} w_m\widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{S}_m\widehat{\mathbf{R}}^{-1}_{mi}\mathbf{S}'_m\widehat{\boldsymbol{\Sigma}}_{mi}(\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}})\\ &\quad + \sum_{m=1}^{M} w_m\widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\frac{1}{N}\sum_{i=1}^{N}\mathbf{S}_m\widehat{\mathbf{A}}_{mi}\sqrt{N}\Delta^{-1}_{NT}\boldsymbol{\delta}\\ &\quad - \sum_{m=1}^{M} w_m\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\sqrt{N}\Delta^{-1}_{NT}\boldsymbol{\delta} + O\left(\frac{1}{\Delta_{NT}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)\\ &\equiv J_1 + J_2 + J_3 + J_4 + O\left(\frac{1}{\Delta_{NT}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).\end{aligned}$$

For J_1, by Assumption 4, as $N, T \to \infty$ jointly, we have

$$J_1 = \widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\sum_{m=1}^{M} w_m\mathbf{S}_m\widehat{\mathbf{R}}^{-1}_{mi}\mathbf{S}'_m\widehat{\mathbf{Q}}_{mi}\right)\boldsymbol{\eta}_i \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Xi}_u(\mathbf{w})) \equiv \mathbf{U}(\mathbf{w}), \tag{A.13}$$

where

$$\boldsymbol{\Xi}_u(\mathbf{w}) = \underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\mathbf{D}_{\boldsymbol{\beta}}'\left(\sum_{m=1}^{M}w_m\mathbf{S}_m\mathbf{R}_{mi}^{-1}\mathbf{S}_m'\mathbf{Q}_{mi}\right)\boldsymbol{\Omega}_{\boldsymbol{\beta}}\left(\sum_{m=1}^{M}w_m\mathbf{Q}_{mi}\mathbf{S}_m\mathbf{R}_{mi}^{-1}\mathbf{S}_m'\right)\mathbf{D}_{\boldsymbol{\beta}}.$$

Note that $\boldsymbol{\Xi}_u(\mathbf{w})$ can be rewritten as

$$\begin{aligned}\boldsymbol{\Xi}_u(\mathbf{w}) &= \underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\mathbf{D}_{\boldsymbol{\beta}}'\sum_{m=1}^{M}w_m^2\mathbf{S}_m\mathbf{R}_{mi}^{-1}\mathbf{S}_m'\mathbf{Q}_{mi}\boldsymbol{\Omega}_{\boldsymbol{\beta}}\mathbf{Q}_{mi}\mathbf{S}_m\mathbf{R}_{mi}^{-1}\mathbf{S}_m'\mathbf{D}_{\boldsymbol{\beta}}\\ &\quad+2\underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\mathbf{D}_{\boldsymbol{\beta}}'\sum\sum_{m\neq\ell}w_mw_\ell\mathbf{S}_m\mathbf{R}_{mi}^{-1}\mathbf{S}_m'\mathbf{Q}_{mi}\boldsymbol{\Omega}_{\boldsymbol{\beta}}\mathbf{Q}_{\ell i}\mathbf{S}_\ell\mathbf{R}_{\ell i}^{-1}\mathbf{S}_\ell'\mathbf{D}_{\boldsymbol{\beta}}\\ &= \sum_{m=1}^{M}w_m^2\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m\boldsymbol{\Xi}_{um}\mathbf{S}_m'\mathbf{D}_{\boldsymbol{\beta}}+2\sum\sum_{m\neq\ell}w_mw_\ell\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m\boldsymbol{\Xi}_{u,m\ell}\mathbf{S}_\ell'\mathbf{D}_{\boldsymbol{\beta}}.\end{aligned}$$

Similarly, by Assumption 3, as $N,T\to\infty$ jointly, we have

$$J_2 = \widehat{\mathbf{D}}_{\boldsymbol{\beta}}'\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\sum_{m=1}^{M}w_m\mathbf{S}_m\widehat{\mathbf{R}}_{mi}^{-1}\mathbf{S}_m'\widehat{\boldsymbol{\Sigma}}_{mi}\right)(\boldsymbol{\iota}_i-\bar{\boldsymbol{\iota}})\xrightarrow{d}\mathrm{N}(\mathbf{0},\boldsymbol{\Xi}_v(\mathbf{w}))\equiv\mathbf{V}(\mathbf{w}),\qquad\text{(A.14)}$$

where

$$\boldsymbol{\Xi}_v(\mathbf{w}) = \sum_{m=1}^{M}w_m^2\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m\boldsymbol{\Xi}_{vm}\mathbf{S}_m'\mathbf{D}_{\boldsymbol{\beta}}+2\sum\sum_{m\neq\ell}w_mw_\ell\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m\boldsymbol{\Xi}_{v,m\ell}\mathbf{S}_\ell'\mathbf{D}_{\boldsymbol{\beta}}.$$

For J_3 and J_4, by Assumption 5, we have

$$\begin{aligned}J_3+J_4 &= \sum_{m=1}^{M}w_m\left(\widehat{\mathbf{D}}_{\boldsymbol{\beta}}'\frac{1}{N}\sum_{i=1}^{N}\mathbf{S}_m\widehat{\mathbf{A}}_{mi}-\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\right)\sqrt{N}\Delta_{NT}^{-1}\boldsymbol{\delta}\\ &= \sum_{m=1}^{M}w_m\left(\mathbf{D}_{\boldsymbol{\beta}}'\underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\mathbf{S}_m\mathbf{A}_{mi}-\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\right)\boldsymbol{\delta}_c\\ &= \sum_{m=1}^{M}w_m\mathbf{D}_{\boldsymbol{\beta}}'\underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\mathbf{B}_{mi}\boldsymbol{\delta}_c\\ &= \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{B}(\mathbf{w})\boldsymbol{\delta}_c.\end{aligned}\qquad\text{(A.15)}$$

where $\mathbf{B}(\mathbf{w})=\sum_{m=1}^{M}w_m\mathbf{B}_m$. The second equality can be shown by applying the result from equation (A.12). Since $\boldsymbol{\eta}_i$ and $\boldsymbol{\iota}_i$ are mutually independent by Assumption 4, $\mathbf{U}(\mathbf{w})$ and $\mathbf{V}(\mathbf{w})$ are two stochastically independent normal random vectors. Combining (A.13)–(A.15), we have

$$\sqrt{N}\left(\widehat{\mu}(\mathbf{w})-\mu(\boldsymbol{\beta})\right)\xrightarrow{d}\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{B}(\mathbf{w})\boldsymbol{\delta}_c+\mathbf{V}(\mathbf{w})+\mathbf{U}(\mathbf{w})\sim\mathrm{N}(\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{B}(\mathbf{w})\boldsymbol{\delta}_c,\boldsymbol{\Xi}_u(\mathbf{w})+\boldsymbol{\Xi}_v(\mathbf{w})).$$

This completes the proof. □

Proof of Corollary 4. From (A.1), we have

$$\widehat{\boldsymbol{\beta}}_{mi} - \boldsymbol{\beta}_{mi} = (\mathbf{X}'_{mi}\bar{\mathbf{M}}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}'_{mi}\bar{\mathbf{M}}_{hm}(\mathbf{X}_i\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i} + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i).$$

Recall that $\mathbf{X}_{mi} = \mathbf{X}_i\mathbf{S}_m$, $\boldsymbol{\beta}_{mi} = \mathbf{S}'_m(\boldsymbol{\beta} + \boldsymbol{\eta}_i)$, and $\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\iota}_i$. Then it follows that

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{mi} - \boldsymbol{\beta}^*_m &= \widehat{\boldsymbol{\beta}}_{mi} - \boldsymbol{\beta}_{mi} + \boldsymbol{\beta}_{mi} - \boldsymbol{\beta}^*_m \\
&= \mathbf{S}'_m\boldsymbol{\eta}_i + \left(\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m\right)^{-1}\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\boldsymbol{\beta}_{2i} \\
&\quad + \left(\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m\right)^{-1}\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{hm}\mathbf{F})(\bar{\boldsymbol{\gamma}} + (\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}})) \\
&\quad + \left(\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m\right)^{-1}\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{hm}\boldsymbol{\varepsilon}_i) \\
&\equiv K_1 + K_2 + K_3 + K_4.
\end{aligned}$$

By Assumptions 4 and 5, we have

$$\begin{aligned}
K_1 + K_2 &= \mathbf{S}'_m\boldsymbol{\eta}_i + \left(\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m\right)^{-1}\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\Delta_{NT}^{-1}\boldsymbol{\delta} \\
&\quad + \left(\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m\right)^{-1}\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{hm}\mathbf{X}_i)(\mathbf{I}_k - \mathbf{S}_m\mathbf{S}'_m)\boldsymbol{\eta}_i \\
&= \left(\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{gm}\mathbf{X}_i)\mathbf{S}_m\right)^{-1}\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{gm}\mathbf{X}_i)\boldsymbol{\eta}_i \\
&\quad + O_p\left(\frac{1}{\Delta_{NT}}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) \\
&\equiv \boldsymbol{\Phi}_{1imT} + O_p\left(\frac{1}{\Delta_{NT}}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \qquad \text{(A.16)}
\end{aligned}$$

where $\boldsymbol{\Phi}_{1imT} = \left(\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{gm}\mathbf{X}_i)\mathbf{S}_m\right)^{-1}\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{gm}\mathbf{X}_i)\boldsymbol{\eta}_i$, the first equality holds by (A.3), and the second equality holds by Lemma B.1 (i) and (iv) and the fact that $T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{hm}\mathbf{X}_i = O_p(1)$.

For K_3, we have

$$\begin{aligned}
K_3 &= \left(\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m\right)^{-1}\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{hm}\mathbf{F})(\bar{\boldsymbol{\gamma}} + (\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}})) \\
&= \left(\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{gm}\mathbf{X}_i)\mathbf{S}_m\right)^{-1}\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{gm}\mathbf{F})(\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}}) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) \\
&\equiv \boldsymbol{\Phi}_{2imT} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \qquad \text{(A.17)}
\end{aligned}$$

where $\boldsymbol{\Phi}_{2imT} = \left(\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{gm}\mathbf{X}_i)\mathbf{S}_m\right)^{-1}\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{gm}\mathbf{F})(\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}})$.

For K_4, we have

$$\begin{aligned}
K_4 &= \left(\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{hm}\mathbf{X}_i)\mathbf{S}_m\right)^{-1}\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{hm}\boldsymbol{\varepsilon}_i) \\
&= \left(\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{gm}\mathbf{X}_i)\mathbf{S}_m\right)^{-1}\mathbf{S}'_m(T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}_{gm}\boldsymbol{\varepsilon}_i) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)
\end{aligned}$$

$$= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{T}}\right), \tag{A.18}$$

where the last equality holds because $T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\boldsymbol{\varepsilon}_i = O_p(T^{-1/2})$.

Combining (A.16), (A.17), and (A.18), it follows that

$$\widehat{\boldsymbol{\beta}}_{mi} - \boldsymbol{\beta}_m^* = \boldsymbol{\Phi}_{1imT} + \boldsymbol{\Phi}_{2imT} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O\left(\frac{1}{\Delta_{NT}}\right) \tag{A.19}$$

and

$$\frac{1}{N}\sum_{i=1}^{N}(\widehat{\boldsymbol{\beta}}_{mi} - \boldsymbol{\beta}_m^*) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Phi}_{1imT} + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Phi}_{2imT} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O\left(\frac{1}{\Delta_{NT}}\right). \tag{A.20}$$

By subtracting equation (A.19) from equation (A.20), we obtain

$$\widehat{\boldsymbol{\beta}}_{mi} - \widehat{\boldsymbol{\beta}}_{\mathrm{MG},m} = \left(\boldsymbol{\Phi}_{1imT} - \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Phi}_{1imT}\right) + \left(\boldsymbol{\Phi}_{2imT} - \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Phi}_{2imT}\right) + o_p(1).$$

By Assumptions 3 and 4, $\boldsymbol{\Phi}_{1imT}$ and $\boldsymbol{\Phi}_{2imT}$ are mutually independent. Thus, we have

$$\frac{1}{N-1}\mathrm{E}\left[\sum_{i=1}^{N}\left[\left(\widehat{\boldsymbol{\beta}}_{mi} - \widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}\right)\left(\widehat{\boldsymbol{\beta}}_{\ell i} - \widehat{\boldsymbol{\beta}}_{\mathrm{MG},\ell}\right)'\right]\right] = \boldsymbol{\Xi}_{m\ell} + o_p(1).$$

This completes the proof. □

Proof of Corollary 5. We first show the limiting distribution of $\widehat{\boldsymbol{\Psi}}_{m\ell}$. By Equation (3.1), we have $\widehat{\boldsymbol{\beta}}_{\mathrm{MG},f} \xrightarrow{p} \boldsymbol{\beta}$, which implies that $\widehat{\mathbf{D}}_{\boldsymbol{\beta}} \xrightarrow{p} \mathbf{D}_{\boldsymbol{\beta}}$. By Lemma B.1 (i) and the continuous mapping theorem, we have $\widehat{\mathbf{B}}_m \xrightarrow{p} \mathbf{B}_m$, Also, by Corollary 4, we have $\widehat{\boldsymbol{\Xi}}_{m\ell} \xrightarrow{p} \boldsymbol{\Xi}_{m\ell}$. Recall that $\widehat{\boldsymbol{\delta}}_c \xrightarrow{d} \mathbf{Z}_{\boldsymbol{\delta}} \sim \mathrm{N}(\boldsymbol{\delta}_c, \mathbf{S}_0'\boldsymbol{\Xi}_f\mathbf{S}_0)$. Then by the application of Slutsky's theorem, it follows that

$$\begin{aligned}\widehat{\boldsymbol{\Psi}}_{m\ell} &= \widehat{\mathbf{D}}_{\boldsymbol{\beta}}'\left(\widehat{\mathbf{B}}_m\widehat{\boldsymbol{\delta}_c\boldsymbol{\delta}_c'}\widehat{\mathbf{B}}_\ell' + \mathbf{S}_m\widehat{\boldsymbol{\Xi}}_{m\ell}\mathbf{S}_\ell'\right)\widehat{\mathbf{D}}_{\boldsymbol{\beta}} \\ &\xrightarrow{d} \mathbf{D}_{\boldsymbol{\beta}}'\left(\mathbf{B}_m(\mathbf{Z}_{\boldsymbol{\delta}}\mathbf{Z}_{\boldsymbol{\delta}}' - \mathbf{S}_0'\boldsymbol{\Xi}_f\mathbf{S}_0)\mathbf{B}_\ell' + \mathbf{S}_m\boldsymbol{\Xi}_{m\ell}\mathbf{S}_\ell'\right)\mathbf{D}_{\boldsymbol{\beta}} = \boldsymbol{\Psi}_{m\ell}^*.\end{aligned}$$

We next show the limiting distribution of $\mathbf{w}'\widehat{\boldsymbol{\Psi}}\mathbf{w}$. Using the results from the proof of Corollary 3 and the Cramer-Wold Theorem, we can show that

$$\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{R}_{1i}^{-1}\mathbf{S}_m'\mathbf{Q}_{1i}\boldsymbol{\eta}_i\right)', ..., \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{R}_{Mi}^{-1}\mathbf{S}_m'\mathbf{Q}_{Mi}\boldsymbol{\eta}_i\right)'\right)' \xrightarrow{d} (\mathbf{U}_1', ..., \mathbf{U}_M')'. \tag{A.21}$$

Similarly, we have

$$\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{R}_{1i}^{-1}\mathbf{S}_m'\boldsymbol{\Sigma}_{1i}(\boldsymbol{\iota}_i-\bar{\boldsymbol{\iota}})\right)',...,\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{R}_{Mi}^{-1}\mathbf{S}_m'\boldsymbol{\Sigma}_{Mi}(\boldsymbol{\iota}_i-\bar{\boldsymbol{\iota}})\right)'\right)' \xrightarrow{d} (\mathbf{V}_1',...,\mathbf{V}_M')'. \tag{A.22}$$

Note that $\mathbf{Z}_{\boldsymbol{\delta}}=\boldsymbol{\delta}_c+\mathbf{S}_0'\mathbf{U}_f+\mathbf{S}_0'\mathbf{V}_f$ by Equations (3.1) and (4.4). Since all of $\boldsymbol{\Psi}_{m\ell}^*$ can be expressed in terms of the normal random vectors $\mathbf{U}_m$ and $\mathbf{V}_m$, there is joint convergence in distribution of all $\widehat{\boldsymbol{\Psi}}_{m\ell}$ to $\boldsymbol{\Psi}_{m\ell}^*$. Hence, it follows that $\mathbf{w}'\widehat{\boldsymbol{\Psi}}\mathbf{w}\xrightarrow{d}\mathbf{w}'\boldsymbol{\Psi}^*\mathbf{w}$. Therefore, by Theorem 3.2.2 of Van der Vaart and Wellner (1996) or Theorem 2.7 of Kim and Pollard (1990), the minimizer $\widehat{\mathbf{w}}$ converges in distribution to the minimizer of $\mathbf{w}'\boldsymbol{\Psi}^*\mathbf{w}$, which is $\mathbf{w}^*$.

We now show the asymptotic distribution of the plug-in averaging estimator. Note that both Λ_m and w_m^* can be expressed in terms of the normal random vectors $\mathbf{U}_m$ and $\mathbf{V}_m$. Thus, there is joint convergence in distribution of all $\widehat{\mu}_m$ and $\widehat{w}_m$. Thus, it follows that

$$\sqrt{N}(\mu(\widehat{\mathbf{w}})-\mu(\boldsymbol{\beta}))=\sum_{m=1}^{M}\widehat{w}_m\sqrt{N}\left(\mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m})-\mu(\boldsymbol{\beta})\right)\xrightarrow{d}\sum_{m=1}^{M}w_m^*\Lambda_m.$$

This completes the proof. □

B Supplementary Lemmas and Their Proofs

Lemma B.1. Suppose that Assumptions 1–5 hold. Then we have

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

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

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

$$(iv)\quad \frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F}}{T}\bar{\boldsymbol{\gamma}}=O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right).$$

$$(v)\quad \frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F}}{T}\bar{\boldsymbol{\Gamma}}_2(\mathbf{I}_{k_2}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2=O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right).$$

Proof of Lemma B.1. The proofs of (i)-(iii) follow the proof structure adopted in Pesaran (2006). We highlight steps in the proofs that differ from Pesaran (2006), while steps

that are similar to Pesaran (2006) are sketched. Recall that for the submodel m, we have the following system of equations

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

After taking the cross-sectional averages under the equal weights, we have $\bar{\mathbf{h}}_{mt} = \bar{\mathbf{C}}_m' \mathbf{f}_t + \bar{\mathbf{u}}_{mt}$. Stacking all observations over t, we have

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

We show (i)-(iii) by establishing that $\mathrm{E}||\bar{\mathbf{u}}_{mt}||^2 = O(N^{-1})$ with Assumption 5. Note that the local to zero assumption is imposed on $\boldsymbol{\beta}_2$ only. Thus, by Lemma 1 of Pesaran (2006), we have $\mathrm{Var}(N^{-1} \sum_{i=1}^N \varepsilon_{it}) = O(N^{-1})$, $\mathrm{Var}(N^{-1} \sum_{i=1}^N \mathbf{v}_{1it}) = O(N^{-1})$, $\mathrm{Var}(N^{-1} \sum_{i=1}^N \mathbf{\Pi}_m \mathbf{v}_{2it}) = O(N^{-1})$, and $\mathrm{Var}(N^{-1} \sum_{i=1}^N \boldsymbol{\beta}_{1i}' \mathbf{v}_{1it}) = O(N^{-1})$.

For $\mathrm{Var}(N^{-1} \sum_{i=1}^N \boldsymbol{\beta}_{2i}' \mathbf{v}_{2it})$, note that

$$\begin{aligned} \mathrm{Var}\left(\frac{1}{N} \sum_{i=1}^N \boldsymbol{\beta}_{2i}' \mathbf{v}_{2it} \right) &= \frac{1}{N^2} \mathrm{Var}\left(\sum_{i=1}^N (\boldsymbol{\beta}_2 + \boldsymbol{\eta}_{2i})' \mathbf{v}_{2it} \right) \\ &= \frac{1}{N^2} \sum_{i=1}^N \mathrm{E}(\boldsymbol{\beta}_2' \mathbf{\Sigma}_i \boldsymbol{\beta}_2) + \frac{1}{N^2} \mathrm{E}(\boldsymbol{\eta}_{2i}' \mathbf{\Sigma}_i \boldsymbol{\eta}_{2i}) \\ &\leq \frac{1}{N^2 \Delta_{NT}^2} \boldsymbol{\delta}' \boldsymbol{\delta} \sum_{i=1}^N \lambda_{\max}(\mathbf{\Sigma}_i) + \frac{1}{N^2} \sum_{i=1}^N \mathrm{E}(\boldsymbol{\eta}_{2i}' \boldsymbol{\eta}_{2i}) \lambda_{\max}(\mathbf{\Sigma}_i) \\ &= O\left(\frac{1}{N} \right) + O\left(\frac{1}{N \Delta_{NT}^2} \right), \end{aligned}$$

where $\lambda_{\max}(\mathbf{\Sigma}_i)$ denotes the largest eigenvalue of $\mathbf{\Sigma}_i$. Also, we have

$$\begin{aligned} &\mathrm{Cov}\left(\frac{1}{N} \sum_{i=1}^N \varepsilon_{it} + \frac{1}{N} \sum_{i=1}^N \boldsymbol{\beta}_{1i}' \mathbf{v}_{1it} + \frac{1}{N} \sum_{i=1}^N \boldsymbol{\beta}_{2i}' \mathbf{v}_{2it}, \frac{1}{N} \sum_{i=1}^N \mathbf{v}_{1it} \right) \\ &= \frac{1}{N^2} \sum_{i=1}^N \mathrm{E}(\boldsymbol{\beta}_{1i}' \mathbf{v}_{1it} \mathbf{v}_{1it}' + \boldsymbol{\beta}_{2i}' \mathbf{v}_{2it} \mathbf{v}_{1it}') \\ &= O\left(\frac{1}{N} \right) + O\left(\frac{1}{N \Delta_{NT}} \right), \end{aligned}$$

where the last equality holds because the random deviation of $\boldsymbol{\beta}_2$ is independent of $\mathbf{v}_{1it}$ and $\mathbf{v}_{2it}$. Similarly, we have

$$\mathrm{Cov}\left(N^{-1} \sum_{i=1}^N \varepsilon_{it} + N^{-1} \sum_{i=1}^N \boldsymbol{\beta}_{1i}' \mathbf{v}_{1it} + N^{-1} \sum_{i=1}^N \boldsymbol{\beta}_{2i}' \mathbf{v}_{2it}, N^{-1} \sum_{i=1}^N \mathbf{v}_{2it} \right) = O\left(\frac{1}{N} \right) + O\left(\frac{1}{N \Delta_{NT}} \right).$$

Combining these results, we have $\mathrm{Var}(\bar{\mathbf{u}}_{mt}) = O(N^{-1})$ no matter whether Δ_{NT}^{-1} is $O(1)$ or Δ_{NT}^{-1} converges to zero at any rate. Therefore, we have $\mathrm{E}||\bar{\mathbf{u}}_{mt}||^2 = O(N^{-1})$. Thus, by Lemmas 2 and 3 of Pesaran (2006) with the above results, we have (i)-(iii).

We now show (iv) and (v). Note that $\bar{\mathbf{M}}_{gm} = \mathbf{I}_T - \bar{\mathbf{G}}_m(\bar{\mathbf{G}}_m'\bar{\mathbf{G}}_m)^{-}\bar{\mathbf{G}}_m'$ and $\bar{\mathbf{M}}_{gm}\bar{\mathbf{G}}_m = \mathbf{0}$. Then, we have $\bar{\mathbf{M}}_{gm}\mathbf{F}(\bar{\boldsymbol{\gamma}} + \overline{\boldsymbol{\Gamma}_2\boldsymbol{\beta}_2} + \overline{\boldsymbol{\Gamma}_1\boldsymbol{\beta}_1}) = \mathbf{0}$, $\bar{\mathbf{M}}_{gm}\mathbf{F}\bar{\boldsymbol{\Gamma}}_1 = \mathbf{0}$, and $\bar{\mathbf{M}}_{gm}\mathbf{F}\bar{\boldsymbol{\Gamma}}_2\boldsymbol{\Pi}_m' = \mathbf{0}$. Also,

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

By Assumption 5 and the above results, we have

$$\begin{aligned}
&\frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F}}{T}(\bar{\boldsymbol{\gamma}} + \overline{\boldsymbol{\Gamma}_2\boldsymbol{\beta}_2} + \overline{\boldsymbol{\Gamma}_1\boldsymbol{\beta}_1}) \\
&= \frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F}}{T}\left(\bar{\boldsymbol{\gamma}} + \bar{\boldsymbol{\Gamma}}_2\boldsymbol{\beta}_2 + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{2i}\boldsymbol{\eta}_{2i} + \bar{\boldsymbol{\Gamma}}_1\boldsymbol{\beta}_1 + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{1i}\boldsymbol{\eta}_{1i}\right) \\
&= \frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F}}{T}\left(\bar{\boldsymbol{\gamma}} + \bar{\boldsymbol{\Gamma}}_2(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 + \bar{\boldsymbol{\Gamma}}_2\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m\boldsymbol{\beta}_2 + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{2i}\boldsymbol{\eta}_{2i} + \bar{\boldsymbol{\Gamma}}_1\boldsymbol{\beta}_1 + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{1i}\boldsymbol{\eta}_{1i}\right) \\
&= \frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F}}{T}\left(\bar{\boldsymbol{\gamma}} + \bar{\boldsymbol{\Gamma}}_2(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{2i}\boldsymbol{\eta}_{2i} + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{1i}\boldsymbol{\eta}_{1i}\right) \\
&= O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

Since $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{2i}\boldsymbol{\eta}_{2i} = O_p\left(N^{-1/2}\right)$ and $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{1i}\boldsymbol{\eta}_{1i} = O_p(N^{-1/2})$, it follows that

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

By a similar argument, we have

$$\frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F}}{T}\bar{\boldsymbol{\Gamma}}_2(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).$$

This completes the proof. □

Lemma B.2. Suppose that Assumptions 1–6 hold and $k_m + 1 > r$. Then we have

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

$$(ii) \quad \frac{\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\boldsymbol{\varepsilon}_i}{T} = \frac{\mathbf{X}_i'\mathbf{M}_{f^0}\boldsymbol{\varepsilon}_i}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).$$

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

Proof of Lemma B.2. We follow the same strategy used in Karabiyik, Reese, and Westerlund (2017). For the submodel m, we can partition $\bar{\mathbf{H}}_m$ as

$$\bar{\mathbf{H}}_m = [\mathbf{F}\bar{\mathbf{C}}_{m,r}, \mathbf{F}\bar{\mathbf{C}}_{m,-r}] + [\bar{\mathbf{U}}_{m,r}, \bar{\mathbf{U}}_{m,-r}],$$

where $\bar{\mathbf{C}}_{m,r}$ is full rank. We further define

$$\mathbf{J}_m = [\mathbf{J}_{m,r}, \mathbf{J}_{m,-r}] = \begin{bmatrix} \bar{\mathbf{C}}_{m,r}^{-1} & -\bar{\mathbf{C}}_{m,r}^{-1}\bar{\mathbf{C}}_{m,-r} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \quad \text{and} \quad \mathbf{D}_N = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sqrt{N}\mathbf{I} \end{bmatrix},$$

where $\mathbf{D}_N$ is a normalization matrix.

Multiplying $\bar{\mathbf{H}}_m$ with $\mathbf{J}_m$ and $\mathbf{D}_N$, we have

$$\begin{aligned} \widetilde{\mathbf{H}}_m &= \bar{\mathbf{H}}_m\mathbf{J}_m\mathbf{D}_N \\ &= \mathbf{F}\bar{\mathbf{C}}_m\mathbf{J}_m\mathbf{D}_N + \bar{\mathbf{U}}_m\mathbf{J}_m\mathbf{D}_N \\ &= [\mathbf{F}, \mathbf{0}] + [\bar{\mathbf{U}}_{m,r}\bar{\mathbf{C}}_{m,r}^{-1}, \sqrt{N}\bar{\mathbf{U}}_{m,-r} - \sqrt{N}\bar{\mathbf{U}}_{m,r}\bar{\mathbf{C}}_{m,r}^{-1}\bar{\mathbf{C}}_{m,-r}] \\ &= \mathbf{F}^0 + \widetilde{\mathbf{U}}_m. \end{aligned}$$

The following facts will be used through this part frequently, including $\mathbf{M}_{hm} = \mathbf{M}_{\widetilde{h}m}$, $\mathbf{M}_{\widetilde{h}m} = \mathbf{I} - \widetilde{\mathbf{H}}_m(\widetilde{\mathbf{H}}_m'\widetilde{\mathbf{H}}_m)^{-}\widetilde{\mathbf{H}}_m'$, $\mathbf{M}_{f^0} = \mathbf{I} - \mathbf{F}^0(\mathbf{F}^{0\prime}\mathbf{F}^0)^{-}\mathbf{F}^{0\prime}$, $\widetilde{\mathbf{U}}_{m,-r} = \sqrt{N}\bar{\mathbf{U}}_m\mathbf{J}_{m,-r}$, $\widetilde{\mathbf{U}}_{m,r} = \bar{\mathbf{U}}_m\mathbf{J}_{m,r}$, and the following results used in Karabiyik, Reese, and Westerlund (2017)

$$\begin{aligned} ||T^{-1}\bar{\mathbf{U}}_m'\widetilde{\mathbf{H}}_m|| &= O_p(N^{-1/2}), \\ ||T^{-1}\mathbf{V}_i'\widetilde{\mathbf{H}}_m|| &= O_p(T^{-1/2}) + O_p(N^{-1/2}), \\ ||T^{-1}\widetilde{\mathbf{H}}_m'\boldsymbol{\varepsilon}_i|| &= O_p(T^{-1/2}) + O_p(N^{-1/2}), \\ ||(T^{-1}\widetilde{\mathbf{H}}_m'\widetilde{\mathbf{H}}_m)^{-} - \boldsymbol{\Sigma}_{f^0}{}^{-}|| &= O_p(T^{-1/2}) + O_p(N^{-1/2}). \end{aligned}$$

where

$$\boldsymbol{\Sigma}_{f^0} = \begin{bmatrix} T^{-1}\mathbf{F}'\mathbf{F} & \mathbf{0} \\ \mathbf{0} & T^{-1}\widetilde{\mathbf{U}}_{m,-r}'\widetilde{\mathbf{U}}_{m,-r} \end{bmatrix}.$$

Also, by an argument similar to Pesaran (2006), we have

$$\begin{aligned} ||T^{-1}\mathbf{V}_i'\bar{\mathbf{U}}_m|| &= O_p(N^{-1}) + O_p((NT)^{-1/2}), \\ ||T^{-1}\bar{\mathbf{U}}_m'\bar{\mathbf{U}}_m|| &= O_p(N^{-1}), \end{aligned}$$

$$||T^{-1}\bar{\mathbf{V}}_i'\mathbf{F}|| = O_p(T^{-1/2}),$$
$$||T^{-1}\mathbf{F}'\bar{\mathbf{U}}_m|| = O_p((NT)^{-1/2}).$$

Therefore, (ii) becomes

$$\begin{aligned}
T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\boldsymbol{\varepsilon}_i &= T^{-1}\mathbf{X}_i'\mathbf{M}_{\tilde{h}m}\boldsymbol{\varepsilon}_i \\
&= T^{-1}(\mathbf{V}_i' - \boldsymbol{\Gamma}_i\bar{\mathbf{C}}_m'^{-}\bar{\mathbf{U}}_m')\mathbf{M}_{\tilde{h}m}\boldsymbol{\varepsilon}_i \\
&= T^{-1}(\mathbf{V}_i' - \boldsymbol{\Gamma}_i\bar{\mathbf{C}}_m'^{-}\bar{\mathbf{U}}_m')\boldsymbol{\varepsilon}_i - T^{-1}(\mathbf{V}_i' - \boldsymbol{\Gamma}_i\bar{\mathbf{C}}_m'^{-}\bar{\mathbf{U}}_m')\mathbf{P}_{f^0}\boldsymbol{\varepsilon}_i \\
&\quad + T^{-1}(\mathbf{V}_i' - \boldsymbol{\Gamma}_i\bar{\mathbf{C}}_m'^{-}\bar{\mathbf{U}}_m')(\mathbf{M}_{\tilde{h}m} - \mathbf{M}_{f^0})\boldsymbol{\varepsilon}_i \\
&= T^{-1}\mathbf{X}_i'\mathbf{M}_{f^0}\boldsymbol{\varepsilon}_i + T^{-1}\mathbf{V}_i'(\mathbf{M}_{\tilde{h}m} - \mathbf{M}_{f^0})\boldsymbol{\varepsilon}_i - \boldsymbol{\Gamma}_i\bar{\mathbf{C}}_m'^{-}\bar{\mathbf{U}}_m'(\mathbf{M}_{\tilde{h}m} - \mathbf{M}_{f^0})\boldsymbol{\varepsilon}_i,
\end{aligned}$$

where $\bar{\mathbf{C}}_m'^{-} = \bar{\mathbf{C}}_m'(\bar{\mathbf{C}}_m\bar{\mathbf{C}}_m')^{-1}$. We can further decompose $T^{-1}\mathbf{V}_i'(\mathbf{M}_{\tilde{h}m} - \mathbf{M}_{f^0})\boldsymbol{\varepsilon}_i$ as follows

$$\begin{aligned}
T^{-1}\mathbf{V}_i'(\mathbf{M}_{\tilde{h}m} - \mathbf{M}_{f^0})\boldsymbol{\varepsilon}_i &= T^{-1}\mathbf{V}_i'\widetilde{\mathbf{U}}_{m,-r}\big(T^{-1}\widetilde{\mathbf{U}}_{m,-r}'\widetilde{\mathbf{U}}_{m,-r}\big)^{-}T^{-1}\widetilde{\mathbf{U}}_{m,-r}'\boldsymbol{\varepsilon}_i \\
&\quad + T^{-1}\mathbf{V}_i'\widetilde{\mathbf{U}}_{m,r}\big(T^{-1}\mathbf{F}'\mathbf{F}\big)^{-}T^{-1}\widetilde{\mathbf{U}}_{m,r}'\boldsymbol{\varepsilon}_i \\
&\quad + T^{-1}\mathbf{V}_i'\widetilde{\mathbf{U}}_{m,r}\big(T^{-1}\mathbf{F}'\mathbf{F}\big)^{-}T^{-1}\mathbf{F}'\boldsymbol{\varepsilon}_i \\
&\quad + T^{-1}\mathbf{V}_i'\mathbf{F}\big(T^{-1}\mathbf{F}'\mathbf{F}\big)^{-}T^{-1}\widetilde{\mathbf{U}}_{m,r}'\boldsymbol{\varepsilon}_i \\
&\quad + T^{-1}\mathbf{V}_i'\widetilde{\mathbf{H}}_m\big(\big(T^{-1}\widetilde{\mathbf{H}}_m'\widetilde{\mathbf{H}}_m\big)^{-} - \boldsymbol{\Sigma}_{f^0}{}^{-}\big)T^{-1}\widetilde{\mathbf{H}}_m'\boldsymbol{\varepsilon}_i.
\end{aligned}$$

Investigating each term of the above equation, we have

$$\begin{aligned}
&||T^{-1}\mathbf{V}_i'\widetilde{\mathbf{U}}_{m,-r}\big(T^{-1}\widetilde{\mathbf{U}}_{m,-r}'\widetilde{\mathbf{U}}_{m,-r}\big)^{-}T^{-1}\widetilde{\mathbf{U}}_{m,-r}'\boldsymbol{\varepsilon}_i|| \\
&\quad \le N \underbrace{||T^{-1}\mathbf{V}_i'\bar{\mathbf{U}}_m||}_{O_p(N^{-1})+O_p((NT)^{-1/2})} \underbrace{||\big(T^{-1}\widetilde{\mathbf{U}}_{m,-r}'\widetilde{\mathbf{U}}_{m,-r}\big)^{-}||}_{O_p(1)} \underbrace{||T^{-1}\bar{\mathbf{U}}_m'\boldsymbol{\varepsilon}_i||}_{O_p(N^{-1})+O_p((NT)^{-1/2})} \underbrace{||\mathbf{J}_{m,-r}||^2}_{O_p(1)} \\
&\quad = O_p(N^{-1}) + O_p(T^{-1}) + O_p((NT)^{-1/2}), \\
&||T^{-1}\mathbf{V}_i'\widetilde{\mathbf{U}}_{m,r}\left(T^{-1}\mathbf{F}'\mathbf{F}\right)^{-}T^{-1}\widetilde{\mathbf{U}}_{m,r}'\boldsymbol{\varepsilon}_i|| \le ||T^{-1}\mathbf{V}_i'\bar{\mathbf{U}}_m||\,||\left(T^{-1}\mathbf{F}'\mathbf{F}\right)^{-}||\,||T^{-1}\bar{\mathbf{U}}_m'\boldsymbol{\varepsilon}_i||\,||\mathbf{J}_{m,r}||^2 \\
&\quad = O_p(N^{-2}) + O_p((NT)^{-1}) + O_p(N^{-3/2}T^{-1/2}), \\
&||T^{-1}\mathbf{V}_i'\widetilde{\mathbf{U}}_{m,r}\left(T^{-1}\mathbf{F}'\mathbf{F}\right)^{-}T^{-1}\mathbf{F}'\boldsymbol{\varepsilon}_i|| \le ||T^{-1}\mathbf{V}_i'\bar{\mathbf{U}}_m||\,||\left(T^{-1}\mathbf{F}'\mathbf{F}\right)^{-}||\,||T^{-1}\mathbf{F}'\boldsymbol{\varepsilon}_i||\,||\mathbf{J}_{m,r}|| \\
&\quad = O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}), \\
&||T^{-1}\mathbf{V}_i'\mathbf{F}\left(T^{-1}\mathbf{F}'\mathbf{F}\right)^{-}T^{-1}\widetilde{\mathbf{U}}_{m,r}'\boldsymbol{\varepsilon}_i|| \le ||T^{-1}\mathbf{V}_i'\mathbf{F}||\,||\left(T^{-1}\mathbf{F}'\mathbf{F}\right)^{-}||\,||T^{-1}\bar{\mathbf{U}}_m'\boldsymbol{\varepsilon}_i||\,||\mathbf{J}_{m,r}|| \\
&\quad = O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}), \\
&||T^{-1}\mathbf{V}_i'\widetilde{\mathbf{H}}_m\left(\left(T^{-1}\widetilde{\mathbf{H}}_m'\widetilde{\mathbf{H}}_m\right)^{-} - \boldsymbol{\Sigma}_{f^0}{}^{-}\right)T^{-1}\widetilde{\mathbf{H}}_m'\boldsymbol{\varepsilon}_i|| \\
&\quad \le ||\mathbf{V}_i'\widetilde{\mathbf{H}}_m||\,||\left(T^{-1}\widetilde{\mathbf{H}}_m'\widetilde{\mathbf{H}}_m\right)^{-} - \boldsymbol{\Sigma}_{f^0}{}^{-}||\,||T^{-1}\widetilde{\mathbf{H}}_m'\boldsymbol{\varepsilon}_i|| \\
&\quad = O_p(N^{-3/2}) + O_p(T^{-3/2}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}).
\end{aligned}$$

Therefore, it follows that $T^{-1}\mathbf{V}_i'(\mathbf{M}_{\widetilde{h}m} - \mathbf{M}_{f^0})\boldsymbol{\varepsilon}_i = O_p(N^{-1}) + O_p(T^{-1}) + O_p((NT)^{-1/2})$. By a similar argument, we have

$$\begin{aligned}
T^{-1}\bar{\mathbf{U}}_m'(\mathbf{M}_{\widetilde{h}m} - \mathbf{M}_{f^0})\bar{\mathbf{U}}_m &= O_p(N^{-1}),\\
T^{-1}\mathbf{V}_i'(\mathbf{M}_{\widetilde{h}m} - \mathbf{M}_{f^0})\bar{\mathbf{U}}_m &= O_p(N^{-1}) + O_p((NT)^{-1/2}),\\
T^{-1}\boldsymbol{\varepsilon}_i'(\mathbf{M}_{\widetilde{h}m} - \mathbf{M}_{f^0})\bar{\mathbf{U}}_m &= O_p(N^{-1}) + O_p((NT)^{-1/2}).
\end{aligned}$$

Then we can obtain the second result:

$$T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\boldsymbol{\varepsilon}_i = T^{-1}\mathbf{X}_i'\mathbf{M}_{f^0}\boldsymbol{\varepsilon}_i + O_p(N^{-1}) + O_p(T^{-1}) + O_p((NT)^{-1/2}).$$

In the same way, we can have that (i) and (iii) become

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_{mi} &= T^{-1}\mathbf{X}_i'\mathbf{M}_{\widetilde{h}m}\mathbf{F} = O_p(N^{-1}) + O_p((NT)^{-1/2}),\\
\widehat{\mathbf{Q}}_{mi} &= T^{-1}\mathbf{X}_i'\mathbf{M}_{\widetilde{h}m}\mathbf{X}_i = T^{-1}\mathbf{V}_i'\mathbf{M}_{f^0}\mathbf{V}_i + O_p(N^{-1}) + O_p(T^{-1}) + O_p((NT)^{-1/2})\\
&= T^{-1}\mathbf{V}_i'\mathbf{V}_i + O_p(N^{-1}) + O_p(T^{-1}) + O_p((NT)^{-1/2}).
\end{aligned}$$

□

Lemma B.3. Suppose that Assumptions 2, 5, and 7 hold. As $N, T \to \infty$ jointly, we have

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}(\mathbf{X}_i\boldsymbol{\eta}_i + \mathbf{F}\boldsymbol{\iota}_i + \boldsymbol{\varepsilon}_i)\\
&\quad= \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\mathbf{M}_{gm}(\mathbf{F}\boldsymbol{\Gamma}\boldsymbol{\eta}_i + (\boldsymbol{\xi}_i\boldsymbol{\eta}_i - \overline{\boldsymbol{\xi}\boldsymbol{\eta}}) + \mathbf{V}_i\boldsymbol{\eta}_i + \mathbf{F}\boldsymbol{\iota}_i + \boldsymbol{\varepsilon}_i) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).
\end{aligned}$$

Proof of Lemma B.3. Let $\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}(\mathbf{X}_i\boldsymbol{\eta}_i+\mathbf{F}\boldsymbol{\iota}_i+\boldsymbol{\varepsilon}_i) = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_i\boldsymbol{\eta}_i + \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{F}\boldsymbol{\iota}_i + \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\boldsymbol{\varepsilon}_i = C_1 + C_2 + C_3$. Using the fact that $\mathbf{X}_i = \mathbf{F}\boldsymbol{\Gamma}_i + \mathbf{V}_i$, we can have

$$C_1 = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_i\boldsymbol{\eta}_i = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{F}\boldsymbol{\Gamma}_i\boldsymbol{\eta}_i + \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{V}_i\boldsymbol{\eta}_i = C_{11} + C_{12}.$$

We first consider C_{11}. By Assumption 7, we can have $\boldsymbol{\Gamma}_i\boldsymbol{\eta}_i = \boldsymbol{\Gamma}\boldsymbol{\eta}_i + \boldsymbol{\xi}_i\boldsymbol{\eta}_i$. Then

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{F}\boldsymbol{\Gamma}_i\boldsymbol{\eta}_i =& \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{F}\boldsymbol{\Gamma}\boldsymbol{\eta}_i + \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{F}\boldsymbol{\xi}_i\boldsymbol{\eta}_i\\
=& \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{F}\boldsymbol{\Gamma}\boldsymbol{\eta}_i + \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{F}\left(\boldsymbol{\xi}_i\boldsymbol{\eta}_i - \overline{\boldsymbol{\xi}\boldsymbol{\eta}}\right) + O_p\left(\frac{1}{N}\right)\\
=& \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{gm}\mathbf{F}\boldsymbol{\Gamma}\boldsymbol{\eta}_i + \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{gm}\mathbf{F}\left(\boldsymbol{\xi}_i\boldsymbol{\eta}_i - \overline{\boldsymbol{\xi}\boldsymbol{\eta}}\right)
\end{aligned}$$

$$+ O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).$$

The second equality holds because $N^{-1}\sum_{i=1}^{N}\mathbf{Z}_{mi}\bar{\mathbf{M}}_{hm}\mathbf{F}\overline{\boldsymbol{\xi}\boldsymbol{\eta}} = O_p(N^{-1})$ comes from the fact that $\mathbf{Z}_{mi} = (\mathbf{X}_{j^{(i)}}\mathbf{S}_m, \mathbf{X}_{j'^{(i)}}\mathbf{S}_m)$, and the third equality holds by using Lemma B.1 (iii). To understand the property of the first and second terms of the last equation, we focus on the lth row of $\mathbf{Z}'_{mi}$, denoted by $\mathbf{z}^{l'}_{mi} = (z^l_{mi,1}, ..., z^l_{mi,T})'$, and let $\boldsymbol{\Delta}_m = \bar{\mathbf{M}}_{gm} - \mathbf{M}_{gm} = O_p(N^{-1/2})$. Since $p\lim_{T\to\infty} T^{-1}\mathbf{Z}'_{mi}\mathbf{M}_{gm}\mathbf{F} = O(1)$, we can investigate the following term by assuming T is fixed.

$$\begin{aligned}
\mathrm{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\mathbf{z}^{l'}_{mi}\boldsymbol{\Delta}_m\mathbf{F}\boldsymbol{\Gamma}\boldsymbol{\eta}_i\right] =& \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathrm{E}\left[\Delta_{m,st}\sum_{i=1}^{N}\boldsymbol{\eta}'_i\boldsymbol{\Gamma}'\mathbf{f}_t z^l_{mi,s}\right] \\
\leq& \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\sqrt{\mathrm{E}\left[\Delta^2_{m,st}\right]\mathrm{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}'_i\boldsymbol{\Gamma}'\mathbf{f}_t z^l_{mi,s}\right)^2\right]} \\
=& \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\sqrt{O\left(\frac{1}{N}\right)O\left(\frac{1}{N}\right)} = O\left(\frac{1}{N}\right).
\end{aligned}$$

The second term of C_{11}, C_{12}, C_2, and C_3 can be proved in the same strategy. Then we complete the proof that

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}'_{mi}\bar{\mathbf{M}}_{hm}(\mathbf{X}_i\boldsymbol{\eta}_i + \mathbf{F}\boldsymbol{\iota}_i + \boldsymbol{\varepsilon}_i) \\
=& \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}'_{mi}\mathbf{M}_{gm}(\mathbf{F}\boldsymbol{\Gamma}\boldsymbol{\eta}_i + \mathbf{F}(\boldsymbol{\xi}_i\boldsymbol{\eta}_i - \overline{\boldsymbol{\xi}\boldsymbol{\eta}}) + \mathbf{V}_i\boldsymbol{\eta}_i + \mathbf{F}\boldsymbol{\iota}_i + \boldsymbol{\varepsilon}_i) \\
&+ O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).
\end{aligned}$$

□

C Special Cases

In this section, we study the asymptotic distribution of the CCEMG estimator for several special cases. We first consider the case where the correlation of the core regressors $\mathbf{x}_{1it}$ and the auxiliary regressors $\mathbf{x}_{2it}$ only comes from the factor structure. Recall that $\mathrm{Var}(\mathbf{v}_{it}) = \boldsymbol{\Sigma}_i$. Thus, the individual-specific errors $\mathbf{v}_{1it}$ and $\mathbf{v}_{2it}$ are uncorrelated when $\boldsymbol{\Sigma}_i$ is a diagonal matrix.

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

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

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

Corollary C.1 shows that the submodel estimate has no asymptotic bias when the presence of the common factors is the only source of the correlation between $\mathbf{x}_{1it}$ and $\mathbf{x}_{2it}$ in equations (2.3)–(2.4). The intuition behind Corollary C.1 is that we are able to filter the common factors by the cross-sectional averages such that the bias from the omitted auxiliary regressors is eliminated when $\mathbf{x}_{1it}$ and $\mathbf{x}_{2it}$ are correlated via the common factors only. In this case, there is no trade-off between omitted variable bias and estimation variance, and we only have positive or negative effects on estimation variance.

We next consider the case where the rank condition is satisfied for some submodels. Note that when the rank condition is satisfied for the mth model, the larger model that contains all the regressors in the mth model also satisfies the rank condition. Thus, when the rank condition is satisfied for at least one submodel, it implies that the full model satisfies the rank condition as well. Therefore, Theorem 2 implies that the variance of the full model estimator is $\mathbf{\Xi}_f = \mathbf{\Xi}_{uf} = \mathbf{\Omega}_{\boldsymbol{\beta}}$ when the rank condition is satisfied for at least one submodel.

We now compare the variance of the full model estimator, $\mathbf{\Omega}_{\boldsymbol{\beta}}$, and the variance of the submodel m, $\mathbf{\Xi}_{um} = \underset{N\to\infty}{p\lim} \frac{1}{N}\sum_{i=1}^{N} \mathbf{R}_{mi}^{-1}\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{\Omega}_{\boldsymbol{\beta}}\mathbf{Q}_{mi}\mathbf{S}_m\mathbf{R}_{mi}^{-1}$, in the following corollary.

Corollary C.2. Suppose that Assumptions 1–6 hold. For $j = 1, ..., k_1$, we have

$$[\mathbf{\Omega}_{\boldsymbol{\beta}}]_{jj} \leq [\mathbf{\Xi}_{um}]_{jj},$$

where $[\mathbf{A}]_{jj}$ is the jth diagonal element of the matrix $\mathbf{A}$, and $\mathbf{\Xi}_{um}$ is defined in Theorem 1.

Corollary C.2 shows that the variance of the core regressor in the full model is smaller than that in any submodel when the rank condition is satisfied. Since the full model has no asymptotic bias and has smaller asymptotic variance than any submodel, we should prefer the full model when the rank condition is satisfied for at least one submodel.

We now consider the case where we impose the local to zero assumption on both the cross-sectional means $\boldsymbol{\beta}_2$ and the random deviations $\boldsymbol{\eta}_{2i}$.

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

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

where $\boldsymbol{\delta}$ is an unknown constant vector, $\boldsymbol{\Omega}_{\boldsymbol{\delta}}$ is a symmetric nonnegative definite matrix, and $||\boldsymbol{\Omega}_{\boldsymbol{\delta}}||$ is bounded.

Corollary C.3. Suppose that Assumptions 1–4 and 5′ hold. As $N, T \to \infty$ jointly, we have

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m} - \boldsymbol{\beta}_m^*) \xrightarrow{d} \mathbf{A}_m\boldsymbol{\delta}_c + \widetilde{\mathbf{U}}_m + \mathbf{V}_m \sim \mathrm{N}\left(\mathbf{A}_m\boldsymbol{\delta}_c, \mathbf{S}_m'\widetilde{\boldsymbol{\Omega}}_{\boldsymbol{\beta}}\mathbf{S}_m + \boldsymbol{\Xi}_{vm}\right),$$

where $\mathbf{A}_m$, $\boldsymbol{\delta}_c$, $\mathbf{V}_m$, and $\boldsymbol{\Xi}_{vm}$ are defined in Theorem 1, and $\widetilde{\mathbf{U}}_m \sim \mathrm{N}(\mathbf{0}, \mathbf{S}_m'\widetilde{\boldsymbol{\Omega}}_{\boldsymbol{\beta}}\mathbf{S}_m)$ where $\widetilde{\boldsymbol{\Omega}}_{\boldsymbol{\beta}}$ is a block diagonal matrix with two blocks $\boldsymbol{\Omega}_{\boldsymbol{\beta}_1}$ and $\mathbf{0}_{k_2\times k_2}$.

Corollary C.3 presents the asymptotic distribution of the CCEMG estimator for each submodel when we impose the local to zero assumption on both the cross-sectional means and the random deviations. Under Assumption 5′, the limits of the second, fourth, and fifth terms of the equation (3.3) remain the same. However, the first term of (3.3) converges to $\widetilde{\mathbf{U}}_m \sim \mathrm{N}(\mathbf{0}, \mathbf{S}_m'\widetilde{\boldsymbol{\Omega}}_{\boldsymbol{\beta}}\mathbf{S}_m)$, and the third term of (3.3) becomes a small order term. Therefore, the random deviations from the slope coefficients $\boldsymbol{\beta}_{2i}$ have no effect on the asymptotic distribution of the CCEMG estimator.

Proof of Corollary C.1. From the proof of Theorem 1, it is easy to see that the limits of I_1, I_3, I_4, and I_5 of (A.2) have the same order of limits when $\boldsymbol{\Sigma}_i$ is a diagonal matrix. Thus, we only need to consider I_2. Let $\bar{\boldsymbol{\xi}}_2 = \frac{1}{N}\sum_{i=1}^N \boldsymbol{\xi}_{2i}$. By Assumption 3 and the fact that $\mathbf{X}_i\mathbf{S}_0 = \mathbf{X}_{2i} = \mathbf{F}\boldsymbol{\Gamma}_{2i} + \mathbf{V}_{2i}$, we have

$$\begin{aligned}
&\sqrt{N}(T^{-1}\mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_i\mathbf{S}_0)(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 \\
&\quad = \sqrt{N}\left(T^{-1}\mathbf{S}_m'\mathbf{X}_i'\bar{\mathbf{M}}_{hm}(\mathbf{F}\boldsymbol{\Gamma}_{2i} + \mathbf{V}_{2i})\right)(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 \\
&\quad = \sqrt{N}\left(T^{-1}\mathbf{S}_m'\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F}(\bar{\boldsymbol{\Gamma}}_2 + \boldsymbol{\xi}_{2i} - \bar{\boldsymbol{\xi}}_2)\right)(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 \\
&\qquad + \left(T^{-1}\mathbf{S}_m'\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{V}_{2i}\right)(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\sqrt{N}\Delta_{NT}^{-1}\boldsymbol{\delta} \\
&\quad = \left(T^{-1}\mathbf{S}_m'\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F}(\boldsymbol{\xi}_{2i} - \bar{\boldsymbol{\xi}}_2)\right)(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\sqrt{N}\Delta_{NT}^{-1}\boldsymbol{\delta} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),
\end{aligned}$$

where the last equality holds by both Lemma B.1 (v) and the facts that $\sqrt{N}(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F})\bar{\boldsymbol{\Gamma}}_2(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 = O_p(N^{-1/2}) + O_p(T^{-1/2})$, $\left(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{V}_{2i}\right)(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m) = \left(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\mathbf{V}_{2i}\right)(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m) + O_p(N^{-1}) + O_p((NT)^{-1/2})$, and $\left(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\mathbf{V}_{2i}\right)(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m) = O_p(T^{-1/2})$. Therefore, by Lemma B.1 (i) and (iii), the second term of (A.2) is

$$I_2 = \frac{1}{N}\sum_{i=1}^N \left(\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\mathbf{X}_i)\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\left(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_{gm}\mathbf{F}(\boldsymbol{\xi}_{2i} - \bar{\boldsymbol{\xi}}_2)\right)(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\sqrt{N}\Delta_{NT}^{-1}\boldsymbol{\delta}$$

$$
\begin{aligned}
&+ O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \\
&= O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),
\end{aligned}
$$

where the last equality holds by the fact that $(\boldsymbol{\xi}_{2i} - \bar{\boldsymbol{\xi}}_2)(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)$ is independent of $\mathbf{S}_m'\mathbf{X}_i$ for all i. This completes the proof. □

Proof of Corollary C.2. When the rank condition is satisfied for the mth model, the asymptotic variance of the mth model is

$$
\boldsymbol{\Xi}_{um} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left(\mathbf{S}_m'\boldsymbol{\Sigma}_i\mathbf{S}_m\right)^{-1} \mathbf{S}_m'\boldsymbol{\Sigma}_i\boldsymbol{\Omega}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}_i'\mathbf{S}_m \left(\mathbf{S}_m'\boldsymbol{\Sigma}_i\mathbf{S}_m\right)^{-1}.
$$

Since $\boldsymbol{\Omega}_{\boldsymbol{\beta}}^{-1} \geq 0$, we have $\boldsymbol{\Omega}_{\boldsymbol{\beta}}^{-1} = \mathbf{G}\mathbf{G}'$. Let $\mathbf{G}_m = \mathbf{S}_m'\mathbf{G}$ and $\mathbf{H}_i = \mathbf{G}^{-1}\boldsymbol{\Sigma}_i'\mathbf{S}_m$. Thus, we have

$$
\begin{aligned}
&\mathbf{S}_m'\boldsymbol{\Sigma}_i\mathbf{S}_m(\mathbf{S}_m'\boldsymbol{\Sigma}_i\boldsymbol{\Omega}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}_i'\mathbf{S}_m)^{-1}\mathbf{S}_m'\boldsymbol{\Sigma}_i\mathbf{S}_m \\
&\quad = \mathbf{S}_m'\mathbf{G}\mathbf{G}^{-1}\boldsymbol{\Sigma}_i\mathbf{S}_m(\mathbf{S}_m'\boldsymbol{\Sigma}_i\mathbf{G}'^{-1}\mathbf{G}^{-1}\boldsymbol{\Sigma}_i'\mathbf{S}_m)^{-1}\mathbf{S}_m'\boldsymbol{\Sigma}_i\mathbf{G}'^{-1}\mathbf{G}'\mathbf{S}_m \\
&\quad = \mathbf{G}_m\mathbf{H}_i(\mathbf{H}_i'\mathbf{H}_i)^{-1}\mathbf{H}_i'\mathbf{G}_m'.
\end{aligned}
$$

We now compare the diagonal elements of $\boldsymbol{\Omega}_{\boldsymbol{\beta}}$ and $\boldsymbol{\Xi}_{um}$. Let $\mathbf{e}_j$ be a selection vector where the jth element is one and others are zeros. Let $\mathbf{e}_{jm} = \mathbf{S}_m'\mathbf{e}_j$. For $1 \leq j \leq k_1$, the jth diagonal elements are $\mathbf{e}_j'\boldsymbol{\Omega}_{\boldsymbol{\beta}}^{-1}\mathbf{e}_j = \mathbf{e}_{jm}'\mathbf{S}_m'\mathbf{G}\mathbf{G}'\mathbf{S}_m\mathbf{e}_{jm} = \mathbf{e}_{jm}'\mathbf{G}_m\mathbf{G}_m'\mathbf{e}_{jm}$ and $\mathbf{e}_{jm}'\mathbf{G}_m\mathbf{H}_i(\mathbf{H}_i'\mathbf{H}_i)^{-1}\mathbf{H}_i'\mathbf{G}_m'\mathbf{e}_{jm}$ for the full model and mth submodel, respectively. Note that $\mathbf{e}_{jm}'\mathbf{G}_m\mathbf{G}_m'\mathbf{e}_{jm} - \mathbf{e}_{jm}'\mathbf{G}_m\mathbf{H}_i(\mathbf{H}_i'\mathbf{H}_i)^{-1}\mathbf{H}_i'\mathbf{G}_m'\mathbf{e}_{jm} \geq 0$, which implies that the variance of the core regressor in the full model is smaller than those in other submodels. This completes the proof. □

Proof of Corollary C.3. From the proof of Theorem 1, it is easy to see that the limits of I_2, I_4, and I_5 of (A.2) remain the same because we still can obtain $\mathrm{E}||\bar{\mathbf{u}}_{mt}||^2 = O(N^{-1})$ when Assumption 5′ holds. Also, by a similar argument, we can show that I_1 converges to $\widetilde{\mathbf{U}}_m \sim \mathrm{N}(\mathbf{0}, \mathbf{S}_m'\boldsymbol{\Omega}_{\boldsymbol{\beta}}\mathbf{S}_m)$ where $\boldsymbol{\Omega}_{\boldsymbol{\beta}}$ is a block diagonal matrix with two blocks $\boldsymbol{\Omega}_{\boldsymbol{\beta}_1}$ and $\mathbf{0}_{k_2\times k_2}$, and I_3 converges to zero instead of a non-degenerated distribution. This completes the proof. □

D Numerical Comparison in a Three-Nested-Model Framework

In this subsection, we consider a simple three-nested-model framework based on the models (1)–(2) to illustrate the bias-variance trade-off shown in Theorems 1 and 2. The model specification is $k_1 = 1$, $k_2 = 2$, $M = 3$, $\boldsymbol{\beta}_1 = 1$, and $\boldsymbol{\delta}_c = d \cdot (2.5, 0.75)'$. The narrow model includes no auxiliary regressor. The middle model includes the first auxiliary regressor. The full model includes both auxiliary regressors.

The numerical setup is as follows: $\boldsymbol{\beta}_i \sim \text{i.i.d.N}(\boldsymbol{\beta}, \boldsymbol{\Omega}_{\boldsymbol{\beta}})$, $\boldsymbol{\gamma}_i \sim \text{i.i.d.N}(\boldsymbol{\gamma}, \boldsymbol{\Omega}_{\boldsymbol{\gamma}})$, and $\boldsymbol{\Gamma}_i \sim \text{i.i.d.N}(\boldsymbol{\Gamma}, \boldsymbol{\Omega}_{\boldsymbol{\Gamma}})$, where $\boldsymbol{\gamma} = \mathbf{1}_{r\times 1}$, $\boldsymbol{\Omega}_{\boldsymbol{\beta}} = 0.5 \cdot \mathbf{I}_k$, $\boldsymbol{\Omega}_{\boldsymbol{\gamma}} = 3 \cdot \mathbf{I}_r$, and $\boldsymbol{\Omega}_{\boldsymbol{\Gamma}}$ is an identity matrix. For $r \geq k_1 + k_2$, $\boldsymbol{\Gamma} = [\mathbf{I}_{(k_1+k_2)} \quad \mathbf{1}_{(k_1+k_2)\times(r-k_1+k_2)}]'$, and for $r < k_1 + k_2$, $\boldsymbol{\Gamma} = [\mathbf{I}_r \quad \mathbf{1}_{r\times(k_1+k_2-r)}]'$. We set $\text{E}(\mathbf{f}_t\mathbf{f}_t') = \sigma_f^2\mathbf{I}_r$ and $\text{Var}(\mathbf{v}_{it}) = \boldsymbol{\Sigma}_i$, where the diagonal elements of Σ_i are $\sqrt{r}$, and off-diagonal elements are $\rho\sqrt{r}$ for all i. Since $\text{E}(\mathbf{f}_t\mathbf{f}_t') = \sigma_f^2\mathbf{I}_r$, we have $\mathbf{Q}_{mi} = \boldsymbol{\Sigma}_i + \sigma_f^2(\boldsymbol{\Gamma}_i'\boldsymbol{\Gamma}_i - \boldsymbol{\Gamma}_i'\bar{\mathbf{C}}_m(\bar{\mathbf{C}}_m'\bar{\mathbf{C}}_m)^-\bar{\mathbf{C}}_m'\boldsymbol{\Gamma}_i)$, and $\boldsymbol{\Sigma}_{mi} = \sigma_f^2(\boldsymbol{\Gamma}_i' - \boldsymbol{\Gamma}_i'\bar{\mathbf{C}}_m(\bar{\mathbf{C}}_m'\bar{\mathbf{C}}_m)^-\bar{\mathbf{C}}_m')$. We compute $\mathbf{A}_m$ and $\boldsymbol{\Xi}_m$ by using $10,000$ random samples.

Figure A1 shows the asymptotic mean squared error (AMSE), asymptotic squared bias, and asymptotic variance of $\sqrt{N}(\widehat{\beta}_1 - \beta_1)$ of the narrow model estimator, the middle model estimator, the full model estimator, and the averaging estimator in a three-nested-model framework. The averaging estimator is calculated based on equations (5.1) and (5.6); see Section 5 for more details. It is clear that the best submodel, which has the lowest AMSE, varies with ρ and σ_f^2 in the upper and lower panels, respectively. Compared with the three submodels, the averaging estimator has much lower AMSE in most ranges of the parameter space. Examining the bias and variance in both panels, we find that the averaging estimator achieves a much lower AMSE by introducing a small bias and simultaneously obtaining a large variance reduction.

The upper three panels of Figure A1 show that both bias and variance terms are increasing with ρ. When $\rho = 0$, $\boldsymbol{\Sigma}_i$ is a diagonal matrix, and all three submodels have no asymptotic bias but different variance. The lower three panels of Figure A1 show that the bias term of the submodel estimators is decreasing with σ_f^2, while the variance term is increasing with σ_f^2.

Besides the bias-variance trade-off, the upper and lower variance panels of Figure A1 also demonstrate the positive or negative effects on estimation variance when adding more auxiliary regressors. The upper panel shows that the full model has the smallest variance for $\rho \leq 0.4$, while the lower variance panel shows that the narrow model has the smaller variance in most of the range of σ_f^2.

Figure A2 shows the AMSE of $\sqrt{N}(\widehat{\beta}_1 - \beta_1)$ of the narrow model estimator, the middle model estimator, the full model estimator, and the averaging estimator in a three-nested-

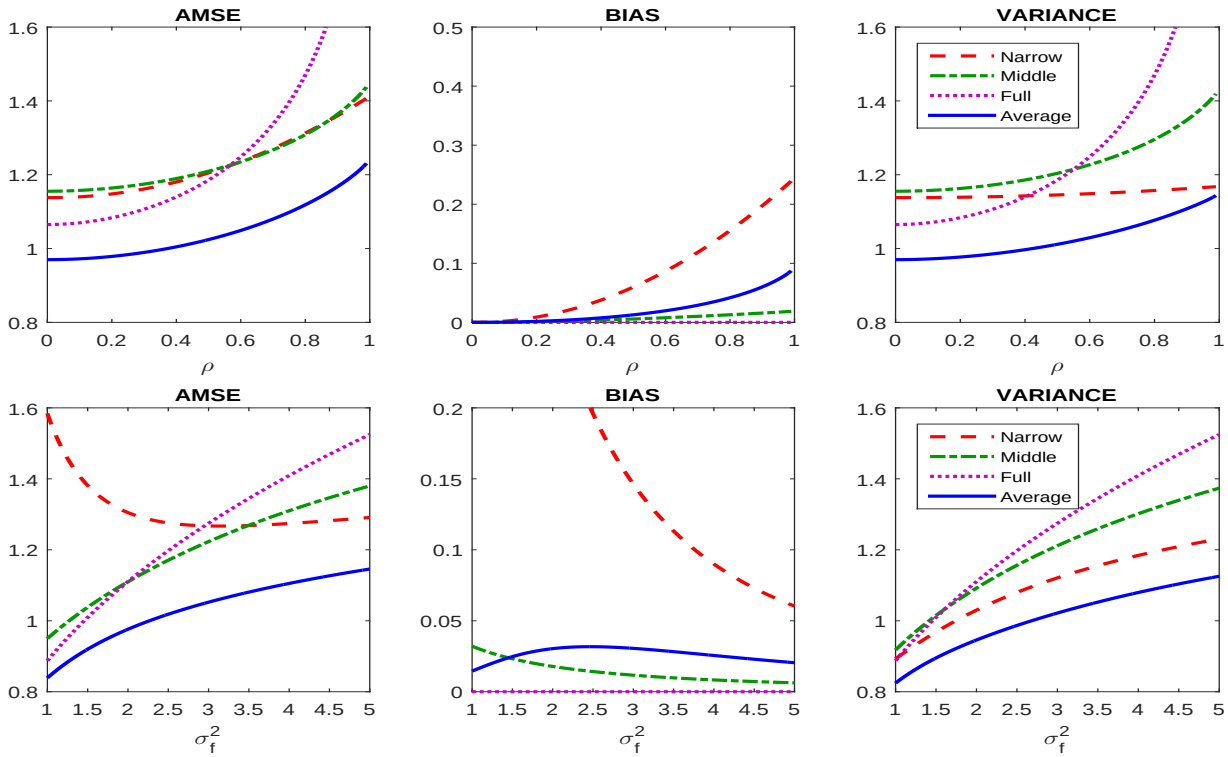

Figure A1: The AMSE, asymptotic squared bias, and asymptotic variance of $\sqrt{N}\left(\widehat{\beta}_1 - \beta_1\right)$ of submodel estimators and the averaging estimator in a three-nested-model framework. The situation is that of $r = 8$ and $d = 1$. The upper three panels correspond to $\sigma_f^2 = 3.5$, and the lower three panels correspond to $\rho = 0.70$.

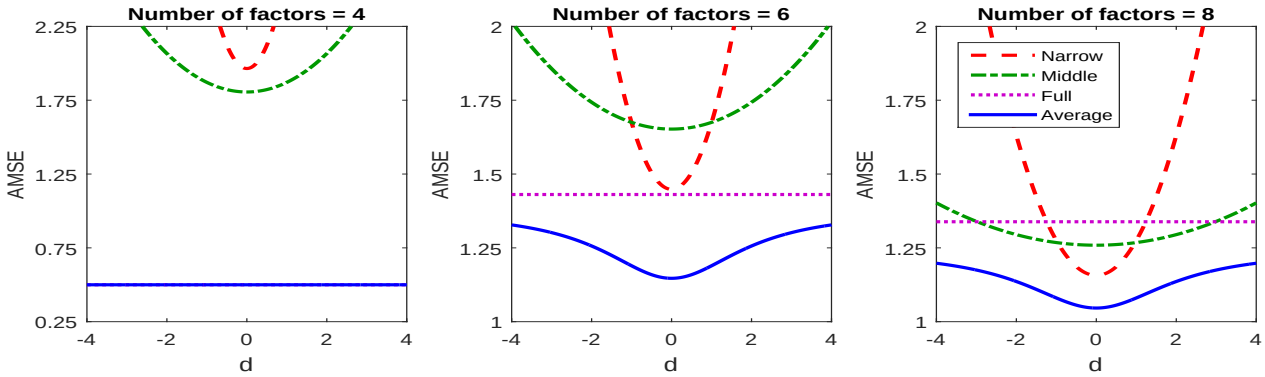

Figure A2: The AMSE of $\sqrt{N}\left(\widehat{\beta}_1 - \beta_1\right)$ of submodel estimators and the averaging estimator in a three-nested-model framework. The situation is that of $\rho = 0.8$ and $\sigma_f^2 = 3.5$. The three panels correspond to $r = 4$, 6, and 8, respectively.

model framework for $r = 4$, 6, and 8, respectively. The three panels show that the best submodel varies with d and r. For $r = 4$, the rank condition is satisfied for the full model only. For $r = 6$ and 8, the rank condition is not satisfied for any submodel.

We first consider the case where the rank condition is satisfied for some submodels. According to Theorem 2, the full model has no bias and the smallest variance. In this case, we should prefer the full model for all values of d. The left panel demonstrates that the averaging estimator assigns the whole weight to the full model. The left panel also shows that the negative effect dominates the positive effect on the estimation variance for $d = 0$ since the narrow model has the largest asymptotic variance.

We next consider the case where the rank condition is not satisfied for any submodel. When d is small, the omitted variable bias is relatively small, and we should prefer the narrow model. On the other hand, when d is larger, we should prefer the full model; see the right panel. However, due to the positive or negative effects on estimation variance, the larger model could have smaller variance and then smaller AMSE; see the middle panel.

E Simulation-Based Confidence Intervals

In this section, we detail the procedure of constructing the confidence interval for the averaging estimator. In the proof of Corollary 5, we have shown that

$$\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\widehat{\mathbf{R}}_{1i}^{-1}\mathbf{S}_1'\widehat{\mathbf{Q}}_{1i}\boldsymbol{\eta}_i\right)', ..., \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\widehat{\mathbf{R}}_{Mi}\widehat{\mathbf{Q}}_{Mi}^{-1}\mathbf{S}_M'\boldsymbol{\eta}_i\right)'\right)' \xrightarrow{d} (\mathbf{U}_1', ..., \mathbf{U}_M')',$$

and

$$\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\widehat{\mathbf{R}}_{1i}^{-1}\mathbf{S}_1'\widehat{\boldsymbol{\Sigma}}_{1i}(\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}})\right)', ..., \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\widehat{\mathbf{R}}_{Mi}^{-1}\mathbf{S}_M'\widehat{\boldsymbol{\Sigma}}_{Mi}(\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}})\right)'\right)' \xrightarrow{d} (\mathbf{V}_1', ..., \mathbf{V}_M')'.$$

Then, we can have the following results immediately,

$$\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\mathbf{D}}_{\boldsymbol{\beta}}'\mathbf{S}_1\widehat{\mathbf{R}}_{1i}^{-1}\mathbf{S}_1'\widehat{\mathbf{Q}}_{1i}\boldsymbol{\eta}_i, ..., \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\mathbf{D}}_{\boldsymbol{\beta}}'\mathbf{S}_{M-1}\widehat{\mathbf{R}}_{M-1i}^{-1}\mathbf{S}_{M-1}'\widehat{\mathbf{Q}}_{M-1i}\boldsymbol{\eta}_i, \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\widehat{\mathbf{R}}_{Mi}^{-1}\mathbf{S}_M'\widehat{\mathbf{Q}}_{Mi}\boldsymbol{\eta}_i\right)'\right)'$$
$$\xrightarrow{d} (\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_1\mathbf{U}_1, ..., \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_{M-1}\mathbf{U}_{M-1}, \mathbf{U}_M')' \sim \mathrm{N}_{M-1+k}(\mathbf{0}, \boldsymbol{\Omega}_u),$$

where the (m, ℓ)th element of the upper-left $(M-1, M-1)$ block matrix of $\boldsymbol{\Omega}_u$ is $\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m\boldsymbol{\Xi}_{u,m\ell}\mathbf{S}_\ell'\mathbf{D}_{\boldsymbol{\beta}}$, the mth column of the lower-left $(k, M-1)$ block matrix is $\boldsymbol{\Xi}_{u,Mm}\mathbf{S}_m'\mathbf{D}_{\boldsymbol{\beta}}$, and the lower-right (k, k) block matrix is $\boldsymbol{\Xi}_{uf} = \boldsymbol{\Xi}_{uM}$. Similarly,

$$\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\mathbf{D}}_{\boldsymbol{\beta}}'\mathbf{S}_1\widehat{\mathbf{R}}_{1i}^{-1}\mathbf{S}_1'\widehat{\boldsymbol{\Sigma}}_{1i}(\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}})\right.$$

$$, ..., \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\mathbf{S}_{M-1}\widehat{\mathbf{R}}^{-1}_{M-1i}\mathbf{S}'_{M-1}\widehat{\boldsymbol{\Sigma}}_{M-1i}(\boldsymbol{\iota}_i-\bar{\boldsymbol{\iota}}), \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\widehat{\mathbf{R}}^{-1}_{Mi}\mathbf{S}'_{M}\widehat{\boldsymbol{\Sigma}}_{Mi}(\boldsymbol{\iota}_i-\bar{\boldsymbol{\iota}})\right)'\Bigg)'$$
$$\xrightarrow{d}(\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_1\mathbf{V}_1, ..., \mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_{M-1}\mathbf{V}_{M-1}, \mathbf{V}'_M)' \sim \mathrm{N}_{M-1+k}(\mathbf{0}, \boldsymbol{\Omega}_v),$$

where the (m,ℓ)th element of the upper-left $(M-1, M-1)$ block matrix of $\boldsymbol{\Omega}_v$ is $\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_m\boldsymbol{\Xi}_{v,m\ell}\mathbf{S}'_\ell\mathbf{D}_{\boldsymbol{\beta}}$, the mth column of the lower-left $(k, M-1)$ block matrix is $\boldsymbol{\Xi}_{v,Mm}\mathbf{S}'_m\mathbf{D}_{\boldsymbol{\beta}}$, and the lower-right (k,k) block matrix is $\boldsymbol{\Xi}_{vf} = \boldsymbol{\Xi}_{vM}$. Then we can establish the two-step confidence interval as the following steps:

1. Construct the $(1-\nu)\times 100\%$ confidence region based on the asymptotic distribution $\mathbf{Z}_{\boldsymbol{\delta}}$ for $\boldsymbol{\delta}$.

2. Independently generate J draws based on $\boldsymbol{\mathcal{U}}^j \sim \mathrm{N}_{M-1+k}(\mathbf{0}, \widehat{\boldsymbol{\Omega}})$, where $\widehat{\boldsymbol{\Omega}}$ is the consistent estimate of $\boldsymbol{\Omega} = \boldsymbol{\Omega}_u + \boldsymbol{\Omega}_v$ by applying the result of Corollary 4.

3. For each $\boldsymbol{\delta}^*$ from step 1, based on $\boldsymbol{\mathcal{U}}^j = \left(\mathcal{U}^j_1, ..., \mathcal{U}^j_{M-1}, \boldsymbol{\mathcal{U}}^{j\,\prime}_k\right)'$, we can obtain the weights $\mathbf{w}^j$ via minimizing $\mathbf{w}'\boldsymbol{\Psi}^j\mathbf{w}$, where $\boldsymbol{\Psi}^j_{m\ell} = \widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\widehat{\mathbf{B}}_m((\boldsymbol{\delta}^* + \mathbf{S}'_0\boldsymbol{\mathcal{U}}^j_k)(\boldsymbol{\delta}^* + \mathbf{S}'_0\boldsymbol{\mathcal{U}}^j_k)' - \mathbf{S}'_0\widehat{\boldsymbol{\Xi}}_f\mathbf{S}_0)\widehat{\mathbf{B}}'_\ell\widehat{\mathbf{D}}_{\boldsymbol{\beta}} + \widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\mathbf{S}_m\widehat{\boldsymbol{\Xi}}_{m\ell}\mathbf{S}'_\ell\widehat{\mathbf{D}}_{\boldsymbol{\beta}}$, and yield $\Lambda^j(\boldsymbol{\delta}^*) = \sum_{m=1}^{M} w^j_m\widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\widehat{\mathbf{B}}_m\boldsymbol{\delta}^* + \sum_{m=1}^{M-1} w^j_m\mathcal{U}^j_m + w^j_M\widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\mathbf{S}_M\boldsymbol{\mathcal{U}}^j_k$ for each $j = 1, ..., J$.

4. By using $\{\Lambda^j(\boldsymbol{\delta}^*)\}_{j=1}^J$, calculate $\widehat{\Lambda}_{(\alpha/2)}(\boldsymbol{\delta}^*)$ and $\widehat{\Lambda}_{(1-\alpha/2)}(\boldsymbol{\delta}^*)$ such that $P(\widehat{\Lambda}_{(\alpha/2)}(\boldsymbol{\delta}^*) \leq \Lambda(\boldsymbol{\delta}^*) \leq \widehat{\Lambda}_{(1-\alpha/2)}(\boldsymbol{\delta}^*)) = 1-\alpha$.

5. Set $\widehat{\Lambda}_{\min}(\boldsymbol{\delta}^*) = \min_{\boldsymbol{\delta}^*}\widehat{\Lambda}_{(\alpha/2)}(\boldsymbol{\delta}^*)$ and $\widehat{\Lambda}_{\max}(\boldsymbol{\delta}^*) = \max_{\boldsymbol{\delta}^*}\widehat{\Lambda}_{(1-\alpha/2)}(\boldsymbol{\delta}^*)$.

6. Construct the confidence interval: $\mathrm{CI} = (\widehat{\mu} - \widehat{\Lambda}_{\max}(\boldsymbol{\delta}^*)/\sqrt{N}, \widehat{\mu} - \widehat{\Lambda}_{\min}(\boldsymbol{\delta}^*)/\sqrt{N})$.

Under the assumptions of Corollary 5,the two-step confidence interval can have asymptotic coverage probability no less than $(1-(\nu+\alpha))$ as $J, N, T \to \infty$. We can also use the one-step simulation-based method by giving $\widehat{\boldsymbol{\delta}}$ and construct the confidence interval by using $\widehat{\Lambda}_{(\alpha/2)}(\widehat{\boldsymbol{\delta}})$ and $\widehat{\Lambda}_{(1-\alpha/2)}(\widehat{\boldsymbol{\delta}})$ from step 3. While a two-step simulation-based method is theoretically valid, it is time-consuming. Based on our simulations, the coverage rate of the one-step method is close to the nominal size. Therefore, in our empirical study, the confidence interval we provided is based on the one-step method. The simulation result is available upon request.

F CCE Pooled Estimator

In this section, we briefly discuss the CCE pooled (CCEP) estimator under a local to zero framework. Basically we will introduce the estimator, the asymptotic properties of

submodels, and how to construct the FIC and plug-in averaging estimator. The proofs for this section are quite similar to those for the CCEMG estimator and are available upon request. Throughout this section, we use "P" to refer to the CCEP estimator. Consider

$$\widehat{\boldsymbol{\beta}}_{\mathrm{P},f} = \left(\sum_{i=1}^{N} \mathbf{X}_i'\bar{\mathbf{M}}_h\mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i'\bar{\mathbf{M}}_h\mathbf{y}_i,$$

and

$$\widehat{\boldsymbol{\beta}}_{\mathrm{P},m} = \left(\sum_{i=1}^{N} \mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_{mi}\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{y}_i,$$

for the full model and submodel m, respectively.[1] Using similar decomposition, we can easily obtain the following result:

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

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

where $\boldsymbol{\delta}_c = c \cdot \boldsymbol{\delta}$. In particular

$$\begin{aligned}
&\mathbf{A}_{\mathrm{P},m} = \mathbf{R}_m^{-1}\mathbf{S}_m'\mathbf{Q}_m\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m),\\
&\mathbf{U}_{\mathrm{P},m} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Xi}_{\mathrm{P},um}\right) \quad \text{with} \quad \boldsymbol{\Xi}_{\mathrm{P},um} = \mathbf{R}_m^{-1}\mathbf{S}_m'\boldsymbol{\Lambda}_{m,\boldsymbol{\beta}}\mathbf{S}_m\mathbf{R}_m^{-1},\\
&\mathbf{V}_{\mathrm{P},m} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Xi}_{\mathrm{P},vm}\right) \quad \text{with} \quad \boldsymbol{\Xi}_{\mathrm{P},vm} = \mathbf{R}_m^{-1}\mathbf{S}_m'\boldsymbol{\Lambda}_{m,\boldsymbol{\gamma}}\mathbf{S}_m\mathbf{R}_m^{-1},
\end{aligned}$$

where $\mathbf{R}_m = \mathbf{S}_m'\mathbf{Q}_m\mathbf{S}_m$, $\boldsymbol{\Lambda}_{m,\boldsymbol{\beta}} = \underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\mathbf{Q}_{mi}\boldsymbol{\Omega}_{\boldsymbol{\beta}}\mathbf{Q}_{mi}$ and $\boldsymbol{\Lambda}_{m,\boldsymbol{\gamma}} = \underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{mi}\boldsymbol{\Omega}_{\boldsymbol{\gamma}}\boldsymbol{\Sigma}_{mi}'$.

This result is quite similar to that for the CCEMG estimator, consisting of three components, the bias, $\mathbf{A}_{\mathrm{P},m}$, and two stochastic independent random vectors, $\mathbf{U}_{\mathrm{P},m}$ and $\mathbf{V}_{\mathrm{P},m}$ for submodel m. Therefore, under the full model specification, the bias, $\mathbf{A}_{\mathrm{P},f} = \mathbf{0}$, and the covariance matrix $\boldsymbol{\Xi}_{\mathrm{P},f}$ becomes $\mathbf{Q}_f^{-1}\boldsymbol{\Lambda}_{f,\boldsymbol{\beta}}\mathbf{Q}_f^{-1} + \mathbf{Q}_f^{-1}\boldsymbol{\Lambda}_{f,\boldsymbol{\gamma}}\mathbf{Q}_f^{-1}$.[2] When the rank condition, Assumption 6, is satisfied, the factor structure can be eliminated asymptotically, therefore $\mathbf{V}_{\mathrm{P},m} = o_p(1)$, $\mathbf{Q}_{mi} = \boldsymbol{\Sigma}_i$ and $\mathbf{Q}_m = \underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_i$. Then the implication can be obtained, which is those submodels satisfying the rank condition are in general more efficient.

It is natural to apply the result from Theorem E.1 to construct the focus information criterion (FIC) and plug-in averaging estimator by replacing $\widehat{\mathbf{D}}_{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\delta}}_c$ with those estimated based on the full model CCEP estimator, $\widehat{\mathbf{B}}_{\mathrm{P},m} = \left(\widehat{\mathbf{P}}_m\widehat{\mathbf{Q}}_m - \mathbf{I}_k\right)\mathbf{S}_0$ with $\widehat{\mathbf{B}}_m$, and

[1] For the CCEP estimator, the choice of weights to calculate $\bar{\mathbf{y}}, \bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2$ should be the same as the pooling weights to obtain consistency and normality, e.g. equal weights, N^{-1}.

[2] $\mathbf{Q}_{fi} = \underset{T\to\infty}{p\lim}(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_g\mathbf{X}_i)$, $\boldsymbol{\Sigma}_{fi} = \underset{T\to\infty}{p\lim}(T^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_g\mathbf{F})$, $\bar{\mathbf{M}}_g = \mathbf{I}_T - \bar{\mathbf{G}}(\bar{\mathbf{G}}'\bar{\mathbf{G}})^{-}\bar{\mathbf{G}}'$, and $\bar{\mathbf{G}} = \mathbf{F}\bar{\mathbf{C}}$.

$\widehat{\boldsymbol{\Xi}}_{\mathrm{P},m\ell}$ with $\widehat{\boldsymbol{\Xi}}_{m\ell}$ in equations (4.6) and (5.9).[3] Then we can either use FIC to select the model or obtain the optimal weights for the plug-in averaging estimator via minimizing the asymptotically unbiased AMSE. In addition, the covariance matrix $\widehat{\boldsymbol{\Xi}}_{\mathrm{P},m\ell}$ can be estimated consistently by the following estimator,

$$\widehat{\boldsymbol{\Xi}}_{\mathrm{P},m\ell} = \left(\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{R}}_{mi}\right)^{-1}\left(\frac{1}{N-1}\sum_{i=1}^{N}\widehat{\mathbf{R}}_{mi}(\widehat{\boldsymbol{\beta}}_{mi}-\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m})(\widehat{\boldsymbol{\beta}}_{\ell i}-\widehat{\boldsymbol{\beta}}_{\mathrm{MG},\ell})'\widehat{\mathbf{R}}_{\ell i}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{R}}_{\ell i}\right)^{-1}.$$

G GMM-CCEP Estimator

In this section, we detail the procedure of GMM-CCEP estimation used to deal with the case with correlated loadings and slope coefficients. Two parts are addressed including (i) the choice of instrumental variables; (ii) a simple simulation to compare the performance between the plug-in averaging approach based on CCEMG, CCEP and GMM-CCEP estimators.

Choice of Instrumental Variables: One difficulty for GMM is to find proper instruments to construct a moment condition under this framework. Considering the random coefficients specification assumed in Assumption 7, it is natural to choose $\mathbf{Z}_i = \mathbf{X}_j$ when individual i is considered with a proper j, and $j \neq i$. The reason is that $\boldsymbol{\Gamma}_j$ is independent of $\boldsymbol{\eta}_i$, $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\gamma}_i$.

The choice of j is based on the knowledge of the spatial weights matrix from Assumption 8 and it is used to ensure that $\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_j'\mathbf{M}_{gm}\mathbf{X}_i$, satisfying the identification condition. Without loss of generality, we can assume the spatial weights matrix is formed by the rook's type contiguity. Under this assumption, for each individual i, we can find at least two adjacent neighbors if we only focus on first-order spatial lag, $j^{(i)}, j^{'(i)} \in \{1, ..., N\}$ and $j^{(i)} \neq j^{'(i)}$. These neighbors can form two sets $\{j^{(i)} : i = 1, ..., N\}$ and $\{j^{'(i)} : i = 1, ..., N\}$ for all individuals. In particular, $|\{j^{(i)} : i = 1, ..., N\}| = |\{j^{'(i)} : i = 1, ..., N\}| = O(N)$. Then we can have $\mathbf{Z}_{mi} = (\mathbf{X}_{j^{(i)}}\mathbf{S}_m, \mathbf{X}_{j^{'(i)}}\mathbf{S}_m)$, a $T \times (2k_m)$ matrix representing instrumental variables.[4] Then these instrumental variables satisfy the following moment condition for submodel m.

$$\mathrm{E}\left(\mathbf{Z}_{mi}'\mathbf{M}_{gm}(\mathbf{F}\boldsymbol{\Gamma}\boldsymbol{\eta}_i + (\boldsymbol{\xi}_i\boldsymbol{\eta}_i - \overline{\boldsymbol{\xi}\boldsymbol{\eta}}) + \mathbf{V}_i\boldsymbol{\eta}_i + \mathbf{F}\boldsymbol{\iota}_i + \boldsymbol{\varepsilon}_i)\right) = \mathbf{0}, \tag{G.1}$$

where $\overline{\boldsymbol{\xi}\boldsymbol{\eta}} = N^{-1}\sum_{i=1}^{N}\boldsymbol{\xi}_i\boldsymbol{\eta}_i$.

[3] $\widehat{\mathbf{Q}}_m = (NT)^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{X}_i$.

[4] It is also valid by using $\mathbf{Z}_i = (\mathbf{X}_{j^{(i)}}, \mathbf{X}_{j^{'(i)}})$ as instrumental variables. Considering different choice of instrumental variables to construct the moment condition for each submodel is also possible.

Simulation: We consider the following data generating process:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}_i + \sum_{j=1}^{r} \gamma_{ij} f_{jt} + \varepsilon_{it}, \quad x_{it\ell} = \sum_{j=1}^{r} \Gamma_{ij\ell} f_{jt} + \mathrm{v}_{it\ell},\ i = 1, ..., N,\ t = 1, ..., T,\ \ell = 1, ..., k,$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i, \quad \boldsymbol{\beta} = d\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{\sqrt{N}}\left(1, \frac{k_2 - 1}{k_2}, \cdots, \frac{1}{k_2}\right)\right)',$$

where $\mathbf{x}_{it} = (x_{it1}, ..., x_{itk})'$, $\boldsymbol{\beta}_i = (\beta_{i1}, ..., \beta_{ik})'$ and $\boldsymbol{\eta}_i = (\eta_{i1}, ..., \eta_{ik})'$. To characterize the spatial correlation, we let the idiosyncratic errors ε_{it} and v_{it} be generated by

$$\varepsilon_{it} = \delta_\varepsilon \sum_{l=1}^{N} w_{il} \varepsilon_{lt} + \epsilon_{it},$$
$$\mathrm{v}_{it\ell} = \delta_{\mathrm{v}} \sum_{l=1}^{N} w_{il} \mathrm{v}_{lt\ell} + \upsilon_{itl},$$

where δ_ε and δ_{v} denote the spatial autoregressive coefficients. $w_{il} = \mathbb{W}_{(i,l)}$, the (i,l)th element of spatial weights matrix $\mathbb{W}$ driven by a rook-type matrix where the Euclidean distance between two units is less than or equal to a certain value. The weights matrix is normalized such that the sum for each row is one. We follow Kelejian and Prucha (2007) and Pesaran and Tosetti (2011) and assume that units are located on rectangular grid locations.[5]

To characterize the correlation between γ_{ij}, $\Gamma_{ij\ell}$ and $\beta_{i\ell}$, we assume that

$$\gamma_{ij} = \mu_\gamma + \iota_{ij},\ j = 1, ..., r$$
$$\Gamma_{ij\ell} = \mu_\Gamma + \rho_\Gamma(\gamma_{ij} - \mu_\gamma)\frac{\sigma_\Gamma}{\sigma_\gamma} + (1 - \rho_\Gamma^2)^{1/2}\xi_{ij},\ j = 1, ..., r$$
$$\boldsymbol{\beta}_{i\ell} = \boldsymbol{\beta} + \rho_\beta(\gamma_{i1} - \mu_\gamma)\frac{\sigma_\beta}{\sigma_\gamma} + +(1 - \rho_\beta^2)^{1/2}\eta_i,$$

where $\mu_\gamma = 1.4$, $\mu_\Gamma = 0.5$, $\iota_{ij} \sim$ i.i.d.N$(0, \sigma_\gamma^2)$, $\xi_{ij} \sim$ i.i.d.N$(0, \sigma_\Gamma^2)$, $\eta_i \sim$ i.i.d.N$(0, \sigma_\beta^2)$, and $\sigma_\gamma = 1.4$, $\sigma_\Gamma = 0.7$ and $\sigma_\beta = 0.1$. To control the degree of the correlation between loadings and slope heterogeneity, we let $\rho_\Gamma = \rho_\beta = 0.5$. To determine parameters σ_ϵ and σ_υ, we follow equations

$$\tau_\epsilon = \frac{\mathrm{Var}(\epsilon_{it})}{\mathrm{Var}(\sum_{\ell=1}^{k}\sum_{j=1}^{8}\Gamma_{ij\ell}f_{jt}\beta_{il}^*) + \mathrm{Var}(\sum_{\ell=1}^{k}\upsilon_{it\ell}\beta_{i\ell}^*) + \mathrm{Var}(\sum_{j=1}^{r}\gamma_{ij}f_{jt}) + \mathrm{Var}(\epsilon_{it})},$$
$$\tau_\upsilon = \frac{\mathrm{Var}(\upsilon_{it\ell})}{\mathrm{Var}(\sum_{j=1}^{8}\Gamma_{ij\ell}f_{jt}) + \mathrm{Var}(\upsilon_{it\ell})},$$

[5]For example, a grid at locations (r, s), for $r = 1, ..., m1$; $s = 1, 2, ..., m2$, such that $N = m1m2$. For a given value of N we set $m1$ and $m2$ such that these are integer numbers and $|m1 - m2|$ is minimized.

and let $\tau_\epsilon = 0.35$ and $\tau_\upsilon = 0.35$. $\boldsymbol{\beta}_i^* = (\beta_{i1}^*, ..., \beta_{ik}^*)'$ under $d = 2$ and the equality holds only when $r = 8$.[6]

The parameter d is varied on a grid between 0.2 and 2. The number of regressors is $k = 6$ with two core regressors ($k_1 = 2$) and four auxiliary regressors ($k_2 = 4$). We consider all possible submodels, that is, the number of models is $M = 16$. We only show the case when $N = T = 100$ with varied d, and further consider the case with different $\rho_{\boldsymbol{\gamma}}$ varied on a grid between 0.2 and 0.8 with fixed $d = 0.6$. The plug-in averaging estimator (Plug-In) based on CCEMG, CCEP and GMM-CCEP approaches and full model GMM-CCEP estimator is considered. The instrumental variables are chosen by the Euclidean distance between units which is less than or equal to 3, which implies that we consider a spatial lag order to be 3. The number of those units is restricted by 6 for each individual. In GMM-CCEP estimate, instead of using a two-step procedure, for simplicity, we use $\widehat{\mathbf{W}}_m = \left((NT)^{-1}\sum_{i=1}^N \mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{Z}_{mi}\right)^{-1}$.

Our parameter of interest is $\mu = \beta_1 + \beta_2$, that is, the sum of cross-sectional means of the slope coefficients for core regressors. We follow Hansen (2007) and compare these estimators based on the risk (expected squared error). The risk is calculated by averaging across $2,500$ random samples. We normalize the risk by dividing by the risk of the infeasible optimal estimator, that is, the risk of the best-fitting submodel m.

In this part, Figures A3–A4 show the normalized risk functions based on Plug-In(MG), Plug-In(P) and Plug-In(GMM) estimators. The main purpose of the comparison is to emphasize the importance of controlling the correlated slope heterogeneity and factor loadings.[7] In Figure A3, the normalized risk is shown for $r = 3$, 7, 11, and 15 in four panels. For $r = 3$, the rank condition is satisfied for all models. We can observe that Plug-In(MG) and Plug-In(P) outperform the Plug-In(GMM) and full model based on the GMM approach. This suggests that the theoretical result that Plug-In(MG) and Plug-In(P) would not suffer from the correlation between the slope heterogeneity and factor loadings when the rank condition is satisfied and they are more efficient than the GMM method. When $r = 7$, only the full model satisfies the rank condition. Plug-In(MG) and Plug-In(P) have larger normalized risk compared to Plug-In(GMM). Plug-In(GMM) and Full(GMM) have similar performance for $d > 1$, but Plug-In(GMM) has a lower normalized risk than Full for $d < 1$. Similar patterns are observed when $r = 11$ and $r = 15$.

Next, we consider the effect of different degrees of correlation between $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\gamma}_i$ when $N, T = 100$. We can observe that when $\rho_{\boldsymbol{\Gamma}}$ increases, the normalized risk of Plug-In(MG) and Plug-In(P) becomes large quickly. When $\rho_{\boldsymbol{\Gamma}} < 0.3$ and $\rho_{\boldsymbol{\Gamma}} < 0.35$, Plug-In(MG) and Plug-In(MG) have lower normalized risks than Plug-In(GMM) respectively. In addition,

[6] This fraction ignores the correlation between $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Gamma}_i$.

[7] Due to the limitation of space, we do not show the results based on different sample size.

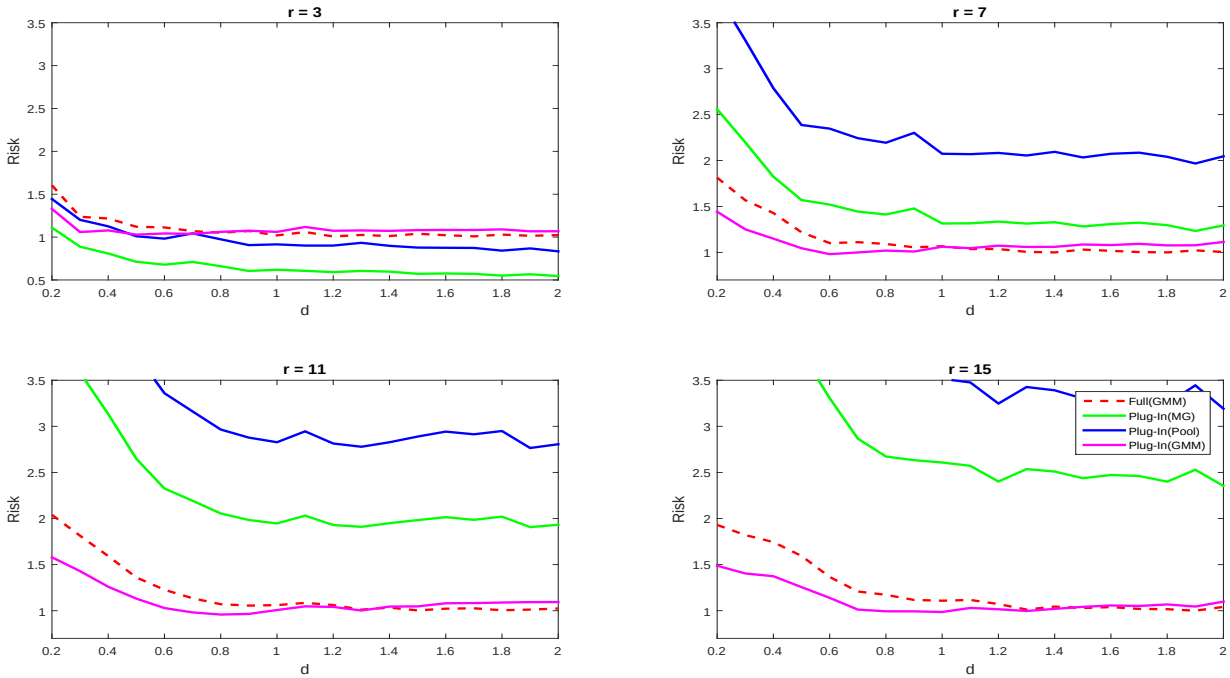

Figure A3: Normalized risk functions for $N, T = 100$, $\rho = 0.5$, $\rho_{\mathbf{\Gamma}} = 0.5$, $\rho_{\boldsymbol{\beta}} = 0.5$.

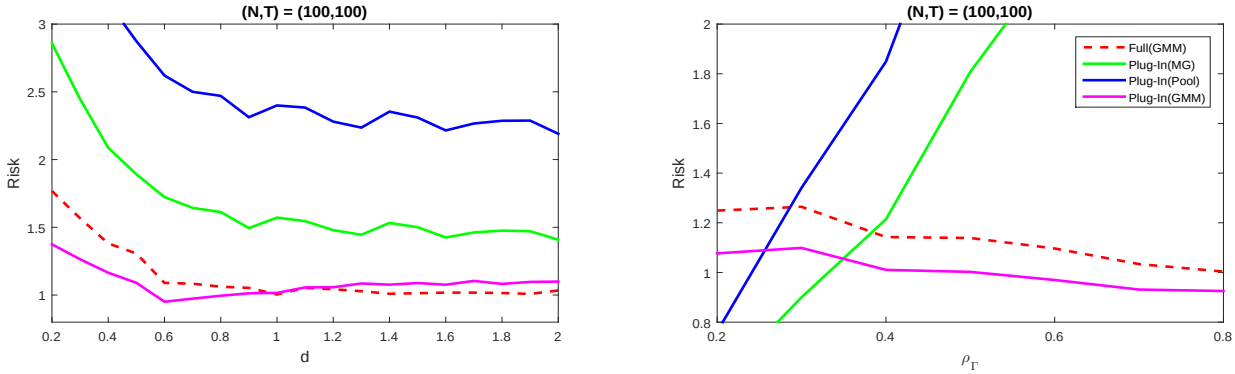

Figure A4: Normalized risk functions for $\rho = 0.5$, $\rho_{\boldsymbol{\beta}} = 0.5$, $d = 0.6$ (right).

Plug-In(GMM) in general has a smaller normalized risk than Full(GMM).

Estimator of $\widetilde{\Xi}_{m\ell}$ for GMM-CCEP Estimation: Note that $\widetilde{\Xi}_{m\ell}$ can be consistently estimated by the following estimator,

$$\begin{aligned}\widehat{\widetilde{\Xi}}_{m\ell} =& \widehat{\widetilde{\mathbf{R}}}_m^{-1}\mathbf{S}_m'\widehat{\widetilde{\boldsymbol{\Sigma}}}_{\mathbf{zx},m}'\widehat{\mathbf{W}}_m\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{y}_i-\frac{1}{T}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_i\widehat{\boldsymbol{\beta}}_{\text{GMMP},f}\right)\\ &\times\left(\frac{1}{T}\mathbf{Z}_{\ell i}'\bar{\mathbf{M}}_{h\ell}\mathbf{y}_i-\frac{1}{T}\mathbf{Z}_{\ell i}'\bar{\mathbf{M}}_{h\ell}\mathbf{X}_i\widehat{\boldsymbol{\beta}}_{\text{GMMP},f}\right)'\widehat{\mathbf{W}}_\ell\widehat{\boldsymbol{\Sigma}}_{\mathbf{zx},\ell}\widehat{\widetilde{\mathbf{R}}}_\ell^{-1}. \qquad \text{(G.2)}\end{aligned}$$

Proof of equation (G.2). We first can observe that

$$\begin{aligned}&\frac{1}{T}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{y}_i-\frac{1}{T}\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\mathbf{X}_i\widehat{\boldsymbol{\beta}}_{\text{GMM-P},f}\\ &\quad=\frac{1}{T}\left(\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}(\mathbf{X}_i(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_{\text{GMM-P},f})+\mathbf{X}_i\boldsymbol{\eta}_i+\mathbf{F}\boldsymbol{\gamma}_i+\boldsymbol{\varepsilon}_i)\right)\\ &\quad=\frac{1}{T}\left(\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}(\mathbf{F}\boldsymbol{\phi}_i+\mathbf{V}_i\boldsymbol{\eta}_i)\right)+O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)\\ &=\equiv\boldsymbol{\Sigma}_{\mathbf{z},mi}\boldsymbol{\phi}_i+\boldsymbol{\Sigma}_{\mathbf{zv},mi}\boldsymbol{\eta}_i+o_p(1),\end{aligned}$$

as $N,T\to\infty$. The second equality holds by $\frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F}}{T}\left(\bar{\boldsymbol{\gamma}}+\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i\boldsymbol{\eta}_i\right)=O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)$, the fact $\widehat{\boldsymbol{\beta}}_{\text{GMM-P},f}$ is $\sqrt{N}$ consistent and $T^{-1}(\mathbf{Z}_{mi}'\bar{\mathbf{M}}_{hm}\boldsymbol{\varepsilon}_i)=O_p(N^{-1})+O_p(T^{-1/2})$.[8] Letting $\widetilde{\boldsymbol{\Sigma}}_{mi}=(\boldsymbol{\Sigma}_{\mathbf{z},mi},\boldsymbol{\Sigma}_{\mathbf{zv},mi})$ and $\boldsymbol{\Lambda}_i=(\boldsymbol{\Phi}_i',\boldsymbol{\eta}_i')'$, then we can show that

$$\frac{1}{N}\sum_{i=1}^{N}\widetilde{\boldsymbol{\Sigma}}_{mi}\boldsymbol{\Lambda}_i\boldsymbol{\Lambda}_i'\widetilde{\boldsymbol{\Sigma}}_{mi}'\xrightarrow{p}\underset{N\to\infty}{p\lim}\sum_{i=1}^{N}\frac{1}{N}\widetilde{\boldsymbol{\Sigma}}_{mi}\boldsymbol{\Omega}_{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{\Sigma}}_{mi}'.$$

The result holds because $\boldsymbol{\Lambda}_i$ is independent of $\mathbf{Z}_{mi}$ and $\mathbf{V}_i$. With consistent estimates of $\boldsymbol{\Sigma}_{\mathbf{zx},m}$ and $\mathbf{W}_m$, we complete the proof. □

References

Hansen, B. E. (2007). Least squares model averaging. *Econometrica 75*, 1175–1189.

Karabiyik, H., S. Reese, and J. Westerlund (2017). On the role of the rank condition in CCE estimation of factor-augmented panel regressions. *Journal of Econometrics 197*(1), 60–64.

[8] $\frac{\sqrt{N}\mathbf{X}_i'\bar{\mathbf{M}}_{hm}\mathbf{F}}{T}\left(\bar{\boldsymbol{\gamma}}+\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i\boldsymbol{\eta}_i\right)=O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)$ can be shown by using a similar argument for proving Lemma B.1(v) when Assumption 7 holds.

Kelejian, H. H. and I. R. Prucha (2007). HAC estimation in a spatial framework. *Journal of Econometrics 140*(1), 131–154.

Kim, J. and D. Pollard (1990). Cube root asymptotics. *The Annals of Statistics 18*, 191–219.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica 74*(4), 967–1012.

Pesaran, M. H. and E. Tosetti (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics 161*(2), 182 – 202.

Van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes*. New York, NY: Springer.