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

Shou-Yung Yin[†], Chu-An Liu[‡], and Chang-Ching Lin[§]

December 15, 2016

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 analyze the consumer response to gasoline taxes.

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

JEL Classification: C23, C51, C52.

*We thank Jack Porter and Yi-Ting Chen for constructive comments and suggestions. We also thank the conference participants of IAAE 2016, AMES 2016, TER 2016, and EEA-ESEM 2016, and seminar participants of Erasmus University Rotterdam for their discussions and suggestions. All errors remain the authors'.

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

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

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

1 Introduction

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 CCEMG estimation is that the rank condition is not necessary for employing the CCEMG estimator. We first consider a general 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.

Several degenerate cases are discussed, including the case where the rank condition is satisfied for some submodels, the case where all submodels have no asymptotic bias, and the case where the local to zero assumption is imposed on both the cross-sectional means of slope coefficients and the random deviations. However, it is hard to distinguish between cases and to verify if the rank condition holds or not in practice. Therefore, the results inferred from these asymptotic distributions do not provide us a clear 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 both general and degenerate cases. We show that the proposed FIC is an asymptotic unbiased estimator of the AMSE and can be applied to all cases.

Building on the idea of FIC, we introduce a frequentist model averaging criterion to select the weights for candidate models and study its properties. We first derive the asymptotic distribution of the averaging estimator with fixed weights, which allows us to characterize the 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. Simulation studies show that the proposed model selection and averaging methods generally produce lower expected squared error as compared to other methods.

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 (2016), while the interactive effects estimator based on principal components has been developed by Stock and Watson (2002), Bai and Ng (2002), Bai (2009), Moon and Weidner (2015a), and Moon and Weidner (2015b); see Kapetanios and Pesaran (2007) and Westerlund and Urbain (2015) for a comparison of these two approaches.

The focused information criterion is introduced by Claeskens and Hjort (2003) for likelihood-based models. In recent years, FIC has been extended to several models, including the Cox hazard regression model (Hjort and Claeskens, 2006), the general semiparametric model (Claeskens and Carroll, 2007), the generalized additive partial linear model (Zhang and Liang, 2011), the Tobin model with a nonzero threshold (Zhang, Wan, and Zhou, 2012), 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; Hjort and Claeskens, 2003; Zhang and Liang, 2011; Zhang, Wan, and Zhou, 2012), adaptive regression by mixing models (Yang, 2000, 2001; Yuan and Yang, 2005), Mallows' C_p-type averaging (Hansen, 2007, 2009, 2010; Wan, Zhang, and Zou, 2010; Liu and Okui, 2013; Zhang, Zou, and Liang, 2014), optimal mean squared error averaging (Liang, Zou, Wan, and Zhang, 2011), jackknife model averaging (Hansen and Racine, 2012; Zhang, Wan, and Zou, 2013; Lu and Su, 2015), and plug-in averaging (Liu, 2015). 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 panel data model, the submodel, and the common correlated effects estimator. Section 3 presents the asymptotic framework and assumptions. Section 4 derives the focused information criterion. Section 5 introduces the plug-in averaging estimator. Section 6 presents the results of Monte Carlo experiments. Section 7 presents the empirical application and Section 8 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 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. The core regressors may only include a constant term or even an empty matrix, and the auxiliary regressors can include any nonlinear transformations of the original variables and the interaction terms between the regressors. 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., $\text{Var}(\mathbf{v}_{it})$ 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 $\mathbf{\Pi}_m$ 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.

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

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

Similarly, for the submodel m, we have

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

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

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

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. 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.

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

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

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

and the CCE estimator in the submodel m is

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

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

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

The model (2.8) allows the slope coefficients to be heterogeneous over i such that $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i$ with $\boldsymbol{\eta}_i$ being 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.[1]

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 in the submodel m is

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

Remark 2.1. 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.

Remark 2.2. 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., $\widetilde{\boldsymbol{\beta}}_{mi} = (\mathbf{X}_{mi}'\mathbf{M}_h\mathbf{X}_{mi})^{-1}\mathbf{X}_{mi}'\mathbf{M}_h\mathbf{y}_i$. 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.

[1] As an alternative CCEMG estimator for $\boldsymbol{\beta}$, we may consider the common correlated effects pooled estimator proposed by Pesaran (2006). However, to achieve consistency, we need to impose some restrictions on the pooling weights when the rank condition is not satisfied. We therefore do not consider this method in this paper.

3 Asymptotic Theory

In this section, we study the limiting distribution of the CCEMG estimator of $\boldsymbol{\beta}$ for the submodel m. In the first subsection, we describe the asymptotic framework and technical assumptions. In the second subsection, we present the asymptotic distribution of the CCEMG estimator in a general case where the rank condition is not satisfied for all submodels. In the third subsection, we study the asymptotic distribution of the CCEMG estimator in several degenerate cases. In the fourth subsection, we provide a numerical comparison in a three-nested-model framework.

3.1 Assumptions

We now state the assumptions.

Assumption 3.1. 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 3.2. 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 3.3. 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 $i = 1, ..., N$,

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

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

Assumption 3.4. 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 3.5. 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 3.6. Rank$(\bar{\mathbf{C}}_m) \equiv r_m = r \leq k_m + 1$ for the mth model.

Assumption 3.1 specifies that the individual-specific errors are distributed independently and imposes some moment conditions. Assumption 3.2 assumes that the common factors are covariance stationary. Assumptions 3.3 and 3.4 impose the random coefficient structure on the factor loadings and the slope coefficients. Assumptions 3.1–3.4 are similar to Assumptions 1–4 of Pesaran (2006). Note that Assumption 3.4 implies that $\boldsymbol{\beta}_{1i} = \boldsymbol{\beta}_1 + \boldsymbol{\eta}_{1i}$, $\boldsymbol{\eta}_{1i} \sim$ i.i.d.$(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\beta}_1})$ where $||\boldsymbol{\Omega}_{\boldsymbol{\beta}_1}||$ is bounded.

Assumption 3.5 assumes that the cross-sectional means of $\boldsymbol{\beta}_{2i}$ are local to zero. Under Assumptions 3.4–3.5, 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. Note that Assumption 3.5 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 $\boldsymbol{\eta}_{2i}$. We will discuss this case in the subsection 3.3.

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), Leeb and Pötscher (2005), Claeskens and Hjort (2008), Hansen (2014), and Liu (2015). This assumption is canonical in the sense that both squared bias and variance have the same order, and it ensures that 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 3.5 holds when $N = O(T^2)$.[2]

Assumption 3.6 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 3.6 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 Rank$(\bar{\mathbf{C}}_m) \equiv r_m < r$. Assumption 3.6 corresponds to the rank condition (21) of Pesaran (2006).

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

3.2 General Case

In this subsection, we study the asymptotic distribution of the CCEMG estimator in a general setting where the rank condition is not satisfied for all submodels. We first introduce some notation that we will use to characterize the limiting distribution. Let $\boldsymbol{\beta}_m = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2'\boldsymbol{\Pi}_m')' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_{2m}')'$ be the cross-sectional means. Define $\mathbf{Q}_{mi} = p\lim_{T\to\infty} (T^{-1}\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i)$ and $\boldsymbol{\Sigma}_{mi} = p\lim_{T\to\infty} (T^{-1}\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{F})$, where $\mathbf{M}_{gm} = \mathbf{I}_T - \bar{\mathbf{G}}_m(\bar{\mathbf{G}}_m'\bar{\mathbf{G}}_m)^{-}\bar{\mathbf{G}}_m'$ and $\bar{\mathbf{G}}_m = \mathbf{F}\bar{\mathbf{C}}_m$. Let

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

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

Theorem 3.1. Suppose that Assumptions 3.1–3.5 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 $\boldsymbol{\delta}_c = c \cdot \boldsymbol{\delta}$, $\mathbf{A}_m = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \mathbf{R}_{mi}\mathbf{Q}_{mi}\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)$, $\mathbf{R}_{mi} = \left(\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'$, and $\mathbf{U}_m$ and $\mathbf{V}_m$ are two stochastically independent normal random vectors. In particular,

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

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

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

Theorem 3.1 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 $\mathbf{Q}_{fi} = p\lim_{T\to\infty}(T^{-1}\mathbf{X}_i'\mathbf{M}_g\mathbf{X}_i)$, $\boldsymbol{\Sigma}_{fi} = p\lim_{T\to\infty}(T^{-1}\mathbf{X}_i'\mathbf{M}_g\mathbf{F})$, $\mathbf{M}_g = \mathbf{I}_T - \bar{\mathbf{G}}(\bar{\mathbf{G}}'\bar{\mathbf{G}})^{-}\bar{\mathbf{G}}'$, 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).

Define $\widehat{\mathbf{Q}}_{mi} = T^{-1}\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i$ and $\widehat{\boldsymbol{\Sigma}}_{mi} = T^{-1}\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{F}$ as the sample analogs of $\mathbf{Q}_{mi}$ and $\boldsymbol{\Sigma}_{mi}$, respectively. As shown in the proof of Theorem 3.1, we can decompose the CCEMG estimator for the mth model as

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}-\boldsymbol{\beta}_m)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{S}_m'\boldsymbol{\eta}_i+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{S}_m'\widehat{\mathbf{Q}}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\widehat{\mathbf{Q}}_{mi}\mathbf{S}_0(\mathbf{I}_{k_2}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2$$
$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{S}_m'\widehat{\mathbf{Q}}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\widehat{\mathbf{Q}}_{mi}\mathbf{S}_0(\mathbf{I}_{k_2}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\eta}_{2i}$$
$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{S}_m'\widehat{\mathbf{Q}}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\widehat{\boldsymbol{\Sigma}}_{mi}\boldsymbol{\gamma}_i$$
$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{S}_m'\widehat{\mathbf{Q}}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\frac{\mathbf{X}_i'\mathbf{M}_{hm}\boldsymbol{\varepsilon}_i}{T}. \tag{3.3}$$

In the proof of Theorem 3.1, we show that the first and third terms of (3.3) together converge to the normal random vector $\mathbf{U}_m$, and the fourth term of (3.3) converges to the normal random vector $\mathbf{V}_m$. Also, the second term converges to the asymptotic bias $\mathbf{A}_m\boldsymbol{\delta}_c$, and the last term is a small order term. From (3.3), we can observe that the normal random vector $\mathbf{U}_m$ comes from the random deviations $\boldsymbol{\eta}_i$, while the normal random vector $\mathbf{V}_m$ comes from the random deviations $\boldsymbol{\iota}_i$. Thus, $\mathbf{U}_m$ and $\mathbf{V}_m$ are independent by Assumptions 3.3 and 3.4.

Theorem 3.1 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 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 auxiliary regressors are uncorrelated with the core regressors. We will discuss this degenerate case in the next subsection.

In addition to the bias-variance trade-off, Theorem 3.1 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, $\boldsymbol{\Xi}_{um}$ can be simplified as $\boldsymbol{\Omega}_{\boldsymbol{\beta}}$, which is smaller than $\boldsymbol{\Xi}_{um}=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbf{R}_{mi}\mathbf{Q}_{mi}\boldsymbol{\Omega}_{\boldsymbol{\beta}}\mathbf{Q}_{mi}'\mathbf{R}_{mi}'$ of any submodel; see Corollary 3.3 in next subsection for more discussions. The intuition behind this negative effect is that additional regressors could filter out extra factors so that the variance term $\boldsymbol{\Xi}_{um}$ decreases.[3] Since the magnitude of these two effects are not equal, the total effect of adding more auxiliary regressors on estimation variance could be positive or negative.

[3]Note that additional regressors could affect the orthogonal projection matrix $\mathbf{M}_{gm}$ 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.

3.3 Degenerate Cases

In this subsection, 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 3.1. Suppose that Assumptions 3.1–3.5 hold. Assume that $\boldsymbol{\Sigma}_i$ is a diagonal matrix for all i. As $N, T \to \infty$ jointly, we have

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

where $\mathbf{U}_m$, $\mathbf{V}_m$, and $\boldsymbol{\Xi}_m$ are defined in Theorem 3.1.

Corollary 3.1 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 3.1 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 discuss 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.

Corollary 3.2. Suppose that Assumptions 3.1–3.6 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 3.1.

Corollary 3.2 presents the asymptotic distribution of the CCEMG estimator of the mth model when the rank condition is satisfied. 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 is satisfied. This is because the effects of unobserved common factors can be efficiently eliminated when the rank condition holds.[4] Compared to Theorem 3.1, the asymptotic

[4]Recall that $\mathbf{Q}_{mi} = p\lim_{T\to\infty}(T^{-1}\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i)$ and $\mathbf{X}_i = \mathbf{F}\boldsymbol{\Gamma}_i + \mathbf{V}_i$, where $\mathbf{V}_i = (\mathbf{v}_{i1}, ..., \mathbf{v}_{iT})'$. Thus, we can further simplify $\boldsymbol{\Xi}_{um}$ as $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \widetilde{\mathbf{R}}_{mi}\boldsymbol{\Sigma}_i\boldsymbol{\Omega}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}_i'\widetilde{\mathbf{R}}_{mi}'$, where $\widetilde{\mathbf{R}}_{mi} = \left(\mathbf{S}_m'\boldsymbol{\Sigma}_i\mathbf{S}_m\right)^{-1}\mathbf{S}_m'$. Therefore, all the components related to the unobserved common factors in the covariance matrix can be efficiently eliminated.

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.

When the rank condition is satisfied for at least one submodel, Corollary 3.2 implies that the variance of the full model estimator is $\boldsymbol{\Omega}_{\boldsymbol{\beta}}$. We now compare the variance of the full model estimator, $\boldsymbol{\Omega}_{\boldsymbol{\beta}}$, and the variance of the submodel m, $\boldsymbol{\Xi}_{um} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbf{R}_{mi}\mathbf{Q}_{mi}\boldsymbol{\Omega}_{\boldsymbol{\beta}}\mathbf{Q}'_{mi}\mathbf{R}'_{mi}$, in the following corollary.

Corollary 3.3. For $j = 1, ..., k_1$, we have

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

where $[\mathbf{A}]_{jj}$ is the jth diagonal element of the matrix $\mathbf{A}$, and $\boldsymbol{\Xi}_{um}$ is defined in Theorem 3.1.

Corollary 3.3 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. However, it is hard to verify if the rank condition holds or not in practice. Therefore, the result inferred from Corollaries 3.2–3.3 does not provide us a clear guideline to select the submodel in an empirical study. The empirical method of choosing the submodel is described in the next section.

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

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

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 3.4. Suppose that Assumptions 3.1–3.4 and 3.5′ hold. As $N, T \to \infty$ jointly, we have

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

where $\mathbf{A}_m$, $\boldsymbol{\delta}_c$, $\mathbf{V}_m$, and $\boldsymbol{\Xi}_{vm}$ are defined in Theorem 3.1, and $\widetilde{\mathbf{U}}_m \sim \text{N}(\mathbf{0}, \mathbf{S}'_m\widetilde{\boldsymbol{\Omega}}_{\boldsymbol{\beta}}\mathbf{S}_m)$ 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 3.4 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 3.5′, 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 $\widetilde{\mathbf{U}}_m \sim \text{N}(\mathbf{0}, \mathbf{S}'_m\widetilde{\boldsymbol{\Omega}}_{\boldsymbol{\beta}}\mathbf{S}_m)$, 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.

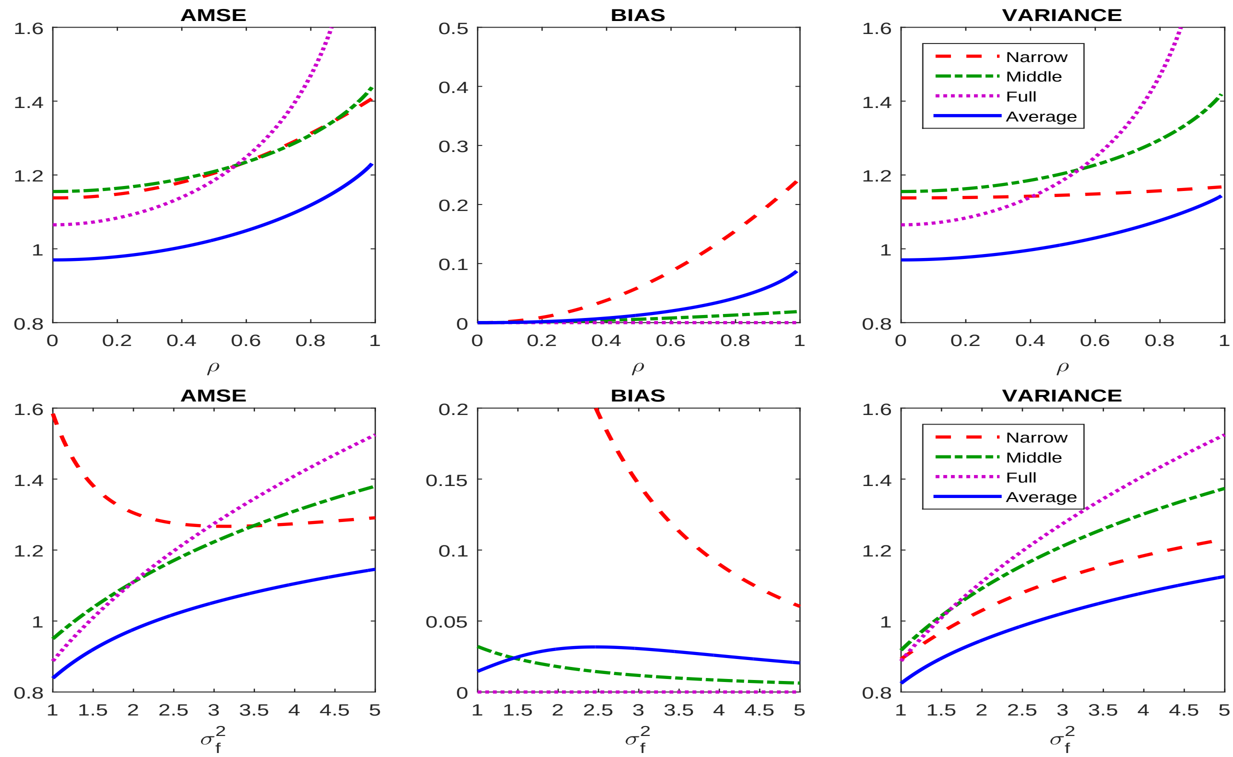

Figure 1: The AMSE, asymptotic squared bias, and asymptotic variance of $\sqrt{N}\left(\widehat{\beta}_1 - \beta_1\right)$ of submodel estimators and the averaging estimator in a three-nested-model framework. The situation is that of $r = 8$ and $d = 1$. The upper three panels correspond to $\sigma_f^2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

3.4 Numerical Comparison in a Three-Nested-Model Framework

In this subsection, we illustrate the bias-variance trade-off in a simple three-nested-model framework based on the model (2.1)–(2.2). The model specification is $k_1 = 1$, $k_2 = 2$, $M = 3$, $\boldsymbol{\beta}_1 = 1$, and $\boldsymbol{\delta} = d \cdot (2.5, 0.75)'$. The narrow model includes no auxiliary regressor. The middle model includes the first auxiliary regressor. The full model includes both auxiliary regressors. We set $\mathrm{E}(\mathbf{f}_t\mathbf{f}_t') = \sigma_f^2\mathbf{I}_r$ and $\mathrm{Var}(\mathbf{v}_{it}) = \boldsymbol{\Sigma}_i$, where the diagonal elements of Σ_i are $\sqrt{r}$, and off-diagonal elements are $\rho\sqrt{r}$ for all i.[5]

Figure 1 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. It is clear that the best submodel, which has the lowest AMSE, varies with ρ and σ_f^2 in the upper and

[5]We set $\boldsymbol{\beta}_i \sim$ i.i.d.N$(\boldsymbol{\beta}, \boldsymbol{\Omega}_{\boldsymbol{\beta}})$, $\boldsymbol{\gamma}_i \sim$ i.i.d.N$(\boldsymbol{\gamma}, \boldsymbol{\Omega}_{\boldsymbol{\gamma}})$, and $\boldsymbol{\Gamma}_i \sim$ 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 compute $\mathbf{A}_m$ and $\boldsymbol{\Xi}_m$ by using 10,000 random samples. Note that $\mathrm{E}(\mathbf{f}_t\mathbf{f}_t') = \sigma_f^2\mathbf{I}_r$. Then we have $\mathbf{Q}_{mi} = p\lim_{T\to\infty}(T^{-1}\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i) = \boldsymbol{\Sigma}_i + \sigma_f^2(\boldsymbol{\Gamma}_i'\boldsymbol{\Gamma}_i - \boldsymbol{\Gamma}_i'\bar{\mathbf{C}}_m(\bar{\mathbf{C}}_m'\bar{\mathbf{C}}_m)^-\bar{\mathbf{C}}_m'\boldsymbol{\Gamma}_i)$, and $\boldsymbol{\Sigma}_{mi} = p\lim_{T\to\infty}(T^{-1}\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{F}) = \sigma_f^2(\boldsymbol{\Gamma}_i' - \boldsymbol{\Gamma}_i'\bar{\mathbf{C}}_m(\bar{\mathbf{C}}_m'\bar{\mathbf{C}}_m)^-\bar{\mathbf{C}}_m')$.

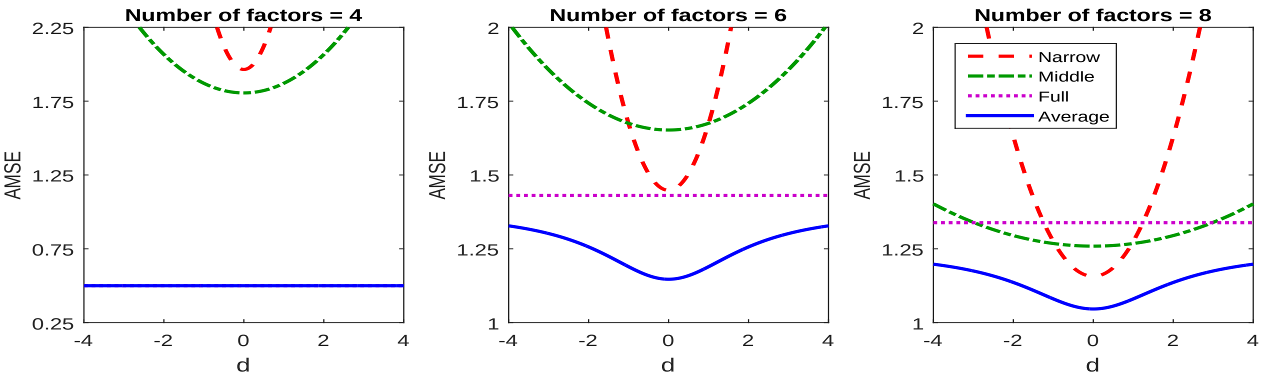

Figure 2: The AMSE of $\sqrt{N}\big(\widehat{\beta}_1 - \beta_1\big)$ of submodel estimators and the averaging estimator in a three-nested-model framework. The situation is that of $\rho = 0.8$ and $\sigma_f^2 = 3.5$. The three panels correspond to $r = 4$, 6, and 8, respectively.

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 1 show that both bias and variance terms are increasing with ρ. When $\rho = 0$, $\mathbf{\Sigma}_i$ is a diagonal matrix. It is easy to see that all three submodels have no asymptotic bias but different variance, which is consistent with the theoretical result in Corollary 3.1. The lower three panels of Figure 1 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 1 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 2 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-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 all submodels.

We first consider the case where the rank condition is satisfied for some submodels. According to Corollaries 3.2–3.3, 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 all submodels. 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.

4 Focused Information Criterion

In this section, we propose a focused information criterion (FIC) for the panel data model with a multifactor error structure. The parameter of interest is $\mu = \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \mu(\boldsymbol{\beta})$, which is a smooth real-valued function of the cross-sectional means of slope coefficients. Unlike the traditional model selection approaches, which assess the global fit of the model, we evaluate the model based on the focus parameter μ. For example, μ could be the average effect of some regressor on a dependent variable, or the sum of 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 $\widehat{\mu}_m = \mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m})$ denote the submodel estimates. We first study the asymptotic distribution of the submodel estimator of the focus parameter. Theorem 3.1 and the delta method imply the following theorem.

Theorem 4.1. Suppose that Assumptions 3.1–3.5 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 $\mathbf{B}_m = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{P}_{mi}\mathbf{Q}_{mi} - \mathbf{I}_k\right)\mathbf{S}_0$, $\mathbf{P}_{mi} = \mathbf{S}_m\left(\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'$, and $\boldsymbol{\delta}_c$, $\mathbf{U}_m$, $\mathbf{V}_m$, and $\boldsymbol{\Xi}_m$ are defined in Theorem 3.1.

Theorem 4.1 shows the asymptotic distribution of $\widehat{\mu}_m$ for the general case. A direct calculation yields

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

The AMSE of $\widehat{\mu}_m$ for the degenerate cases can be derived by the same approach. For example, when the rank condition is satisfied for the mth model, we have

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

or when $\boldsymbol{\Sigma}_i$ is a diagonal matrix for all i, we have[6]

$$\mathrm{AMSE}(\widehat{\mu}_m) = \mathbf{D}_{\boldsymbol{\beta}}'\left(\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\delta}_c\boldsymbol{\delta}_c'(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\mathbf{S}_0' + \mathbf{S}_m\boldsymbol{\Xi}_m\mathbf{S}_m'\right)\mathbf{D}_{\boldsymbol{\beta}}. \tag{4.3}$$

[6]Corollary 3.1 shows that the submodel estimator of $\boldsymbol{\beta}_m$ has no asymptotic bias when $\boldsymbol{\Sigma}_i$ is a diagonal matrix for all i. The first term of (4.3) comes from the fact that $\mu(\boldsymbol{\beta}) - \mu(\boldsymbol{\beta}_m) = \mathbf{D}_{\boldsymbol{\beta}_2}'(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 + O\left(\Delta_{NT}^{-2}\right) = \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\Delta_{NT}^{-1}\boldsymbol{\delta} + O\left(\Delta_{NT}^{-2}\right)$.

Observe that $\text{AMSE}(\widehat{\mu}_m)$ is a function of $\mathbf{D}_{\boldsymbol{\beta}}$ in all cases. Since $\mathbf{D}_{\boldsymbol{\beta}}$ depends on the focus parameter μ, we can use (4.1)–(4.3) to select a proper submodel depending on the parameter of interest. This is the idea of the FIC proposed by Claeskens and Hjort (2003).

To use (4.1)–(4.3) 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. It turns out that the sample analog of $\text{AMSE}(\widehat{\mu}_m)$ is the same for both the general and degenerate cases. We first consider the variance part. For the covariance matrix $\boldsymbol{\Xi}_m$, we follow Pesaran (2006) and consider the nonparametric covariance matrix estimator

$$\widehat{\boldsymbol{\Xi}}_m = \frac{1}{N-1}\sum_{i=1}^{N}(\widehat{\boldsymbol{\beta}}_{mi} - \widehat{\boldsymbol{\beta}}_{\text{MG},m})(\widehat{\boldsymbol{\beta}}_{mi} - \widehat{\boldsymbol{\beta}}_{\text{MG},m})'. \tag{4.4}$$

The following lemma shows that $\widehat{\boldsymbol{\Xi}}_m$ is a consistent estimator for both $\boldsymbol{\Xi}_m$ and $\boldsymbol{\Xi}_{um}$. Thus, the estimator (4.4) is valid in both the general and degenerate cases.

Lemma 4.1. Suppose that Assumptions 3.1–3.5 hold. As $N, T \to \infty$ jointly, we have $\widehat{\boldsymbol{\Xi}}_m \xrightarrow{p} \boldsymbol{\Xi}_m$. Further, if Assumption 3.6 is satisfied, then $\widehat{\boldsymbol{\Xi}}_m \xrightarrow{p} \boldsymbol{\Xi}_{um}$.

We next 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 $\widehat{\mathbf{Q}}_{mi} = T^{-1}\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i$. 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 $\widehat{\mathbf{B}}_m = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\mathbf{P}}_{mi}\widehat{\mathbf{Q}}_{mi} - \mathbf{I}_k\right)\mathbf{S}_0$. Note that when $\boldsymbol{\Sigma}_i$ is a diagonal matrix for all i, the asymptotic bias only comes from the difference between $\mu(\boldsymbol{\beta})$ and $\mu(\boldsymbol{\beta}_m)$, as we discussed in equation (4.3). This implies that $\widehat{\mathbf{B}}_m \xrightarrow{p} \mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)$.

We now 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

$$\widehat{\boldsymbol{\delta}}_c = \sqrt{N}\widehat{\boldsymbol{\beta}}_{2,f} \xrightarrow{d} \mathbf{Z}_{\boldsymbol{\delta}} \sim \text{N}\left(\boldsymbol{\delta}_c, \mathbf{S}_0'\boldsymbol{\Xi}_f\mathbf{S}_0\right). \tag{4.5}$$

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

$$\widehat{\boldsymbol{\delta}_c\boldsymbol{\delta}_c'} = \widehat{\boldsymbol{\delta}}_c\widehat{\boldsymbol{\delta}}_c' - \mathbf{S}_0'\widehat{\boldsymbol{\Xi}}_f\mathbf{S}_0, \tag{4.6}$$

where $\widehat{\mathbf{\Xi}}_f = \frac{1}{N-1}\sum_{i=1}^{N}(\widehat{\boldsymbol{\beta}}_{fi} - \widehat{\boldsymbol{\beta}}_{\mathrm{MG},f})(\widehat{\boldsymbol{\beta}}_{fi} - \widehat{\boldsymbol{\beta}}_{\mathrm{MG},f})'$ is a consistent estimator of $\mathbf{\Xi}_f$ by Lemma 4.1.

We now follow Claeskens and Hjort (2003) and define the FIC for the large heterogeneous panel data model. The proposed FIC of the mth model is

$$\mathrm{FIC}_m = \widehat{\mathbf{D}}_{\boldsymbol{\beta}}' \left(\widehat{\mathbf{B}}_m (\widehat{\boldsymbol{\delta}}_c \widehat{\boldsymbol{\delta}}_c' - \mathbf{S}_0' \widehat{\mathbf{\Xi}}_f \mathbf{S}_0) \widehat{\mathbf{B}}_m' + \mathbf{S}_m \widehat{\mathbf{\Xi}}_m \mathbf{S}_m' \right) \widehat{\mathbf{D}}_{\boldsymbol{\beta}}, \tag{4.7}$$

which is an asymptotically unbiased estimator of $\mathrm{AMSE}(\widehat{\mu}_m)$ in both the general and degenerate cases. The proposed FIC aims to minimize the sample analog of AMSE and can be applied to all cases. In practice, we select the model with the lowest value of FIC_m.

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 theorem shows the asymptotic normality of the averaging estimator with fixed weights.

Theorem 5.1. Suppose that Assumptions 3.1–3.5 hold. As $N, T \to \infty$ jointly, we have

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

where

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

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

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

Theorem 5.1 shows the asymptotic normality of the averaging estimator with nonrandom weights for the general case. 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, the AMSE of $\widehat{\mu}(\mathbf{w})$ for the degenerate cases can be derived by the same approach. For example, when the rank condition is satisfied for all models, we have $\text{AMSE}(\widehat{\mu}(\mathbf{w})) = \mathbf{w}'\boldsymbol{\Psi}\mathbf{w}$ with

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

where $\boldsymbol{\Xi}_{u,m\ell} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \mathbf{R}_{mi}\mathbf{Q}_{mi}\boldsymbol{\Omega}_{\boldsymbol{\beta}}\mathbf{Q}'_{\ell i}\mathbf{R}'_{\ell i}$, or when $\boldsymbol{\Sigma}_i$ is a diagonal matrix for all i, we have $\text{AMSE}(\widehat{\mu}(\mathbf{w})) = \mathbf{w}'\boldsymbol{\Psi}\mathbf{w}$ with

$$\boldsymbol{\Psi}_{m\ell} = \mathbf{D}'_{\boldsymbol{\beta}} \left(\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\boldsymbol{\delta}_c\boldsymbol{\delta}'_c(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_\ell\boldsymbol{\Pi}_\ell)\mathbf{S}'_0 + \mathbf{S}_m\boldsymbol{\Xi}_{m\ell}\mathbf{S}'_m\right) \mathbf{D}_{\boldsymbol{\beta}}. \tag{5.5}$$

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 follow Liu (2015) and 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 $\boldsymbol{\Psi}$. 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 $\widehat{\boldsymbol{\Psi}}$ be a sample analogue of $\boldsymbol{\Psi}$ with the (m, ℓ)th element

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

where $\widehat{\boldsymbol{\delta}_c\boldsymbol{\delta}'_c}$ is defined in (4.6) and

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

The following lemma shows that $\widehat{\boldsymbol{\Xi}}_{m\ell}$ is a consistent estimator for both $\boldsymbol{\Xi}_{m\ell}$ and $\boldsymbol{\Xi}_{u,m\ell}$. Thus, the nonparametric covariance matrix estimator $\widehat{\boldsymbol{\Xi}}_{m\ell}$ is valid in both the general and degenerate cases.

Lemma 5.1. Suppose that Assumptions 3.1–3.5 hold. As $N, T \to \infty$ jointly, we have $\widehat{\boldsymbol{\Xi}}_{m\ell} \xrightarrow{p} \boldsymbol{\Xi}_{m\ell}$. Further, if Assumption 3.6 is satisfied, then $\widehat{\boldsymbol{\Xi}}_{m\ell} \xrightarrow{p} \boldsymbol{\Xi}_{u,m\ell}$.

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

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

where $\mathbf{w}'\widehat{\boldsymbol{\Psi}}\mathbf{w}$ is an asymptotically unbiased estimator of $\mathbf{w}'\boldsymbol{\Psi}\mathbf{w}$ in both the general and degenerate cases. 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 $\boldsymbol{\Psi}$ by inserting $\widehat{\boldsymbol{\delta}}_c$ for $\boldsymbol{\delta}_c$ directly. Thus, the alternative estimator of $\boldsymbol{\Psi}_{m\ell}$ is

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

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

Theorem 5.2. Suppose that Assumptions 3.1–3.5 hold. As $N, T \to \infty$ jointly, we have

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

where Λ_m is defined in Theorem 4.1, and $\mathbf{w}^* = (w_1^*, ..., w_M^*)' = \underset{\mathbf{w} \in \mathcal{W}}{\operatorname{argmin}} \mathbf{w}'\boldsymbol{\Psi}^*\mathbf{w}$ and $\boldsymbol{\Psi}^*$ is an $M \times M$ matrix with the (m, ℓ)th element

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

Unlike the averaging estimator with fixed weights, Theorem 5.2 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) and Liu (2015). To conduct inference for the focus parameter μ, one might consider the simulation-based confidence intervals suggested by DiTraglia (2016) or follow Liu (2015) and construct a confidence interval. A rigorous demonstration of the validity of these inference methods is beyond the scope of the present paper and is left for future research.

6 Simulation Study

In this section, we study the finite sample mean squared error of the FIC and the plug-in averaging estimator via Monte Carlo experiments.

6.1 Simulation Setup

We consider the following data generating process:

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

where $f_{jt} \sim$ i.i.d.N$(0,1)$ for $j = 1, ..., r$, $\gamma_{ij} \sim$ i.i.d.N$(1, 0.25)$ for $j = 1, ..., r$, $\Gamma_{ij\ell} \sim$ i.i.d.N$(0.5, 2.25)$ for $j = 1, ..., r$ and $\ell = 1, ..., k$, $\varepsilon_{it} \sim$ i.i.d.N$(0, r)$, $\eta_{ij} \sim$ i.i.d.N$(0, 0.01)$ for $j = 1, ..., k$, and $\mathbf{v}_{it} = (v_{1it}, ..., v_{kit})' \sim$ N$(\mathbf{0}, \boldsymbol{\Sigma}_i)$ where the diagonal elements of $\boldsymbol{\Sigma}_i$ are $\sqrt{r}$ and off-diagonal elements are $\rho\sqrt{r}$. The parameter d is varied on a grid between 0.2 and 2. We set $\rho = 0$, 0.25, 0.5, and 0.75. The number of common factors $\mathbf{f}_t$ is varied between 3 and 15. 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 set the sample size $N = 25$, 50, 100, and 200, and the sample size $T = 25$, 50, 75, and 100.

We consider the following estimators: (1) CCEMG estimator for the full model estimator (Full), (2) averaging estimator with equal weights (Equal), (3) AIC model selection estimator (AIC),[7] (4) BIC model selection estimator (BIC),[8] (5) FIC model selection estimator (FIC), and (6) plug-in averaging estimator (Plug-In).

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 $5,000$ random samples. We normalize the risk by dividing by the risk of the infeasible optimal CCEMG estimator, that is, the risk of the best-fitting submodel m.

6.2 Simulation Results

The normalized risk functions are displayed in Figures 3–6. We first examine the finite sample performance of model selection and model averaging estimators in a general setting where the rank condition is not satisfied for all submodels. Figure 3 shows the normalized risk for $\rho = 0$, 0.25, 0.5, and 0.75 in four panels. It is clear that FIC achieves lower normalized risk than Full in all cases, and

[7] The AIC criterion for the mth model is $\text{AIC}_m = \sum_{i=1}^{N} T\log(\widehat{\sigma}_{mi}^2) + 2N(2k_1 + 2k_{2m} + 1)$, 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 from the submodel m.

[8] The BIC criterion for the mth model is $\text{BIC}_m = \sum_{i=1}^{N} T\log(\widehat{\sigma}_{mi}^2) + \log(TN)N(2k_1 + 2k_{2m} + 1)$.

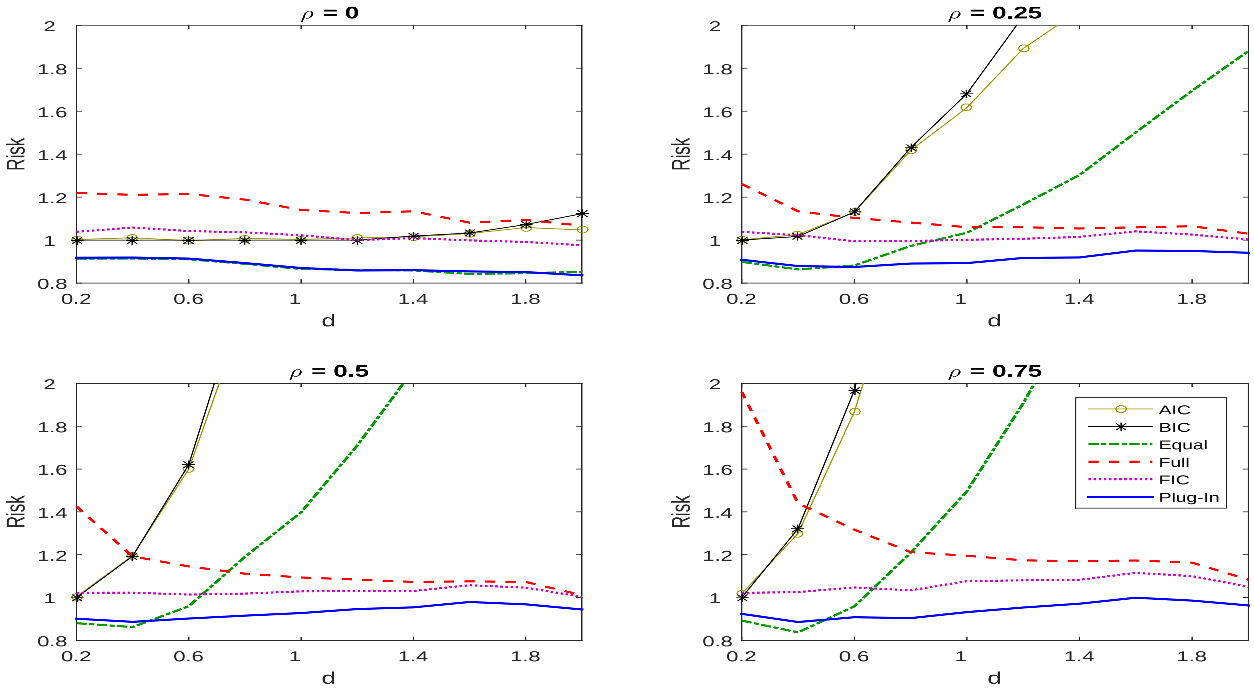

Figure 3: Normalized risk functions for $N = 100$, $T = 50$, and $r = 8$.

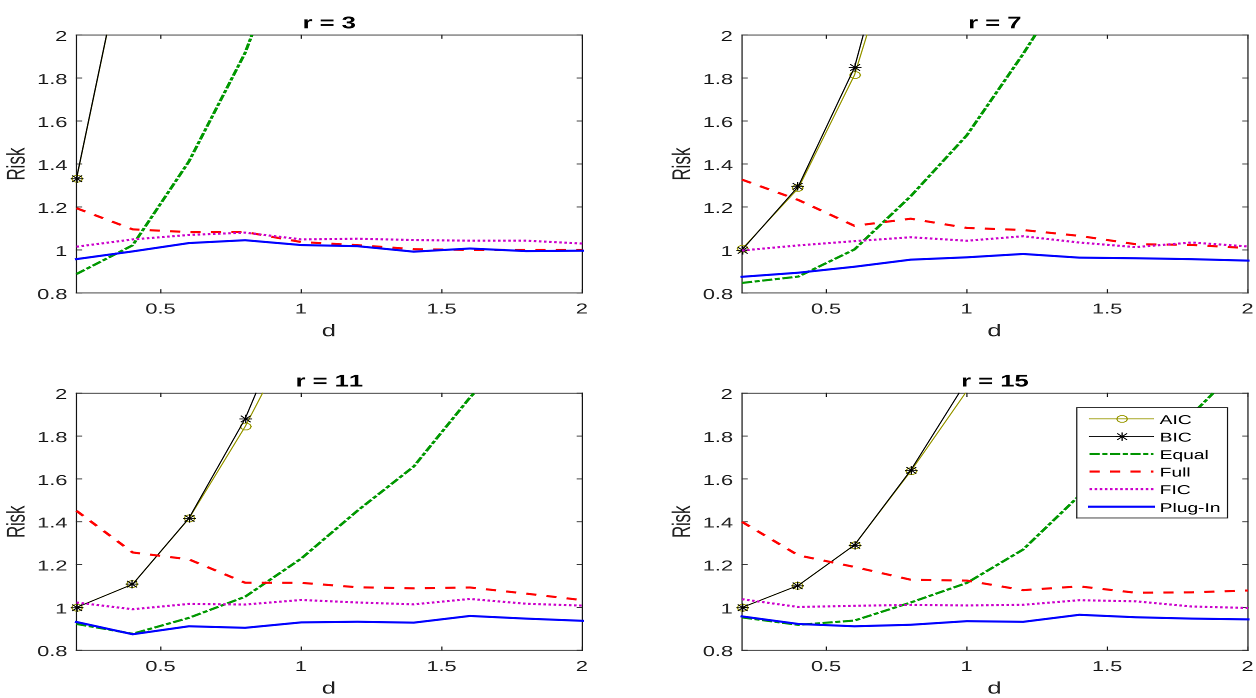

Figure 4: Normalized risk functions for $N = 100$, $T = 50$, and $\rho = 0.5$.

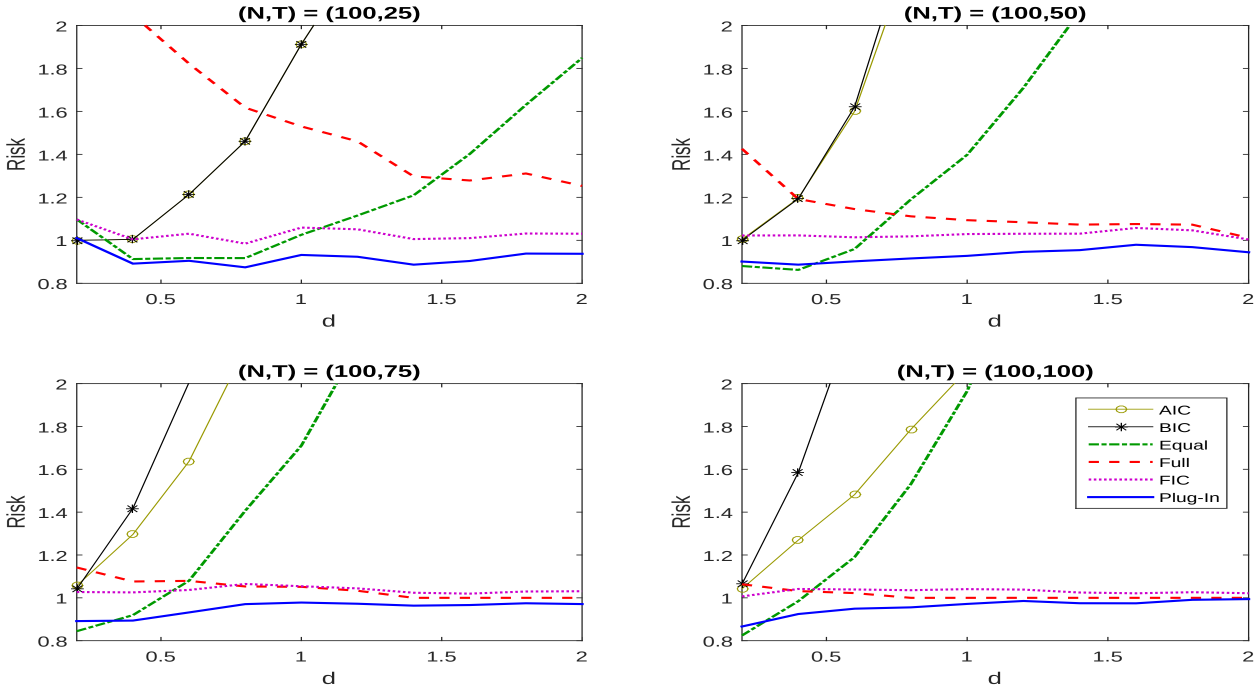

Figure 5: Normalized risk functions for $N = 100$, $r = 8$, and $\rho = 0.5$.

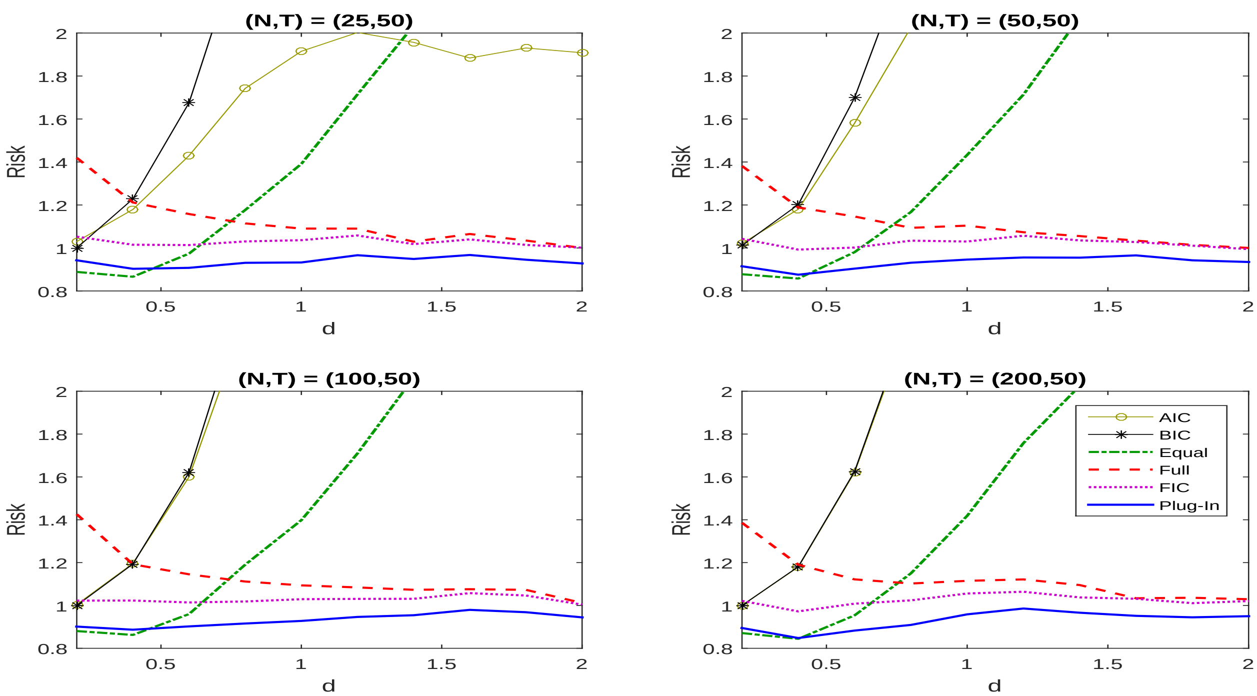

Figure 6: Normalized risk functions for $T = 50$, $r = 8$, and $\rho = 0.5$.

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 $\rho = 0.25$, 0.5, and 0.75. The normalized risk of Plug-In and Equal is indistinguishable for $\rho = 0$, but Equal has much larger normalized risk for $\rho = 0.25$, 0.5, and 0.75. Overall, Plug-In performs well and dominates other estimators in most ranges of the parameter space. Figure 3 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.

We now examine the normalized risk when the rank condition is satisfied for some submodels. Figure 4 shows the normalized risk for $r = 3$, 7, 11, and 15 in four panels. 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 have better performance than AIC, BIC, and Equal. For $r = 7$, the full model is the only model that satisfies the rank condition. In this setting, 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 $r = 11$ and 15, 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 5 shows the normalized risk for a fixed $N = 100$ and for $T = 25$, 50, 75, and 100 in four 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 $T = 75$ and 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 6 shows the normalized risk for a fixed $T = 50$ and for $N = 25$, 50, 100, and 200 in four panels. Unlike the results shown in Figure 5, 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 methods to analyze the consumer response to changes in gasoline taxes. To directly examine the effect of gasoline taxes on gasoline consumption, several issues have been raised in the existing studies, for example, the frequency of the data used, the components of the gasoline price, the econometric methods, and the control variables. These unresolved issues could lead to opposite empirical results; see Espey (1998), Small and Van Dender (2007), Hughes, Knittel, Sperling et al. (2008), and Davis and Kilian (2011). Recently, Li, Linn, and Muehlegger (2014) decompose the retail gasoline price into tax and non-tax components, and provide robust evidence that gasoline taxes have more strongly negative effects on gasoline consumption than do non-tax components. We apply the focused information criterion and the plug-in averaging estimator to U.S. gasoline consumption and revisit the empirical exercise given in Table 2 of Li, Linn, and Muehlegger (2014).

7.1 Empirical Methodology and Data

We estimate the following model

$$\ln(q_{it}) = \alpha_{1i} \ln(p_{it}) + \alpha_{2i} \ln\left(1 + \frac{\tau_{it}}{p_{it}}\right) + \mathbf{x}_{it}'\boldsymbol{\beta}_i + \boldsymbol{\gamma}_i'\mathbf{f}_t + \varepsilon_{it} \tag{7.1}$$

where q_{it} is the gasoline consumption per adult by state i and year t, p_{it} is the tax-exclusive gasoline price, τ_{it} is the total gasoline tax, and $\mathbf{x}_{it}$ is a vector of state-level control variables.[9]

To compare the effects of a tax-inclusive gasoline price and gasoline taxes, we follow Li, Linn, and Muehlegger (2014) and consider two model specifications. In Model Setup I, the tax-inclusive gasoline price is the main regressor, while in Model Setup II, we decompose the tax-inclusive gasoline price into tax and tax-exclusive components. Model Setup I has one core regressor, the tax-inclusive gasoline price (INCPRICE), and Model Setup II has two core regressors, the tax-exclusive gasoline price (EXCPRICE) and the excise tax rate (TAX). Both model setups include the following seven auxiliary regressors: the average family size (FAMILY), the log road miles per adult (ROAD), the log real income per capita (INCOME), the log number of registered cars per capita (CAR), the log number of registered trucks per capita (TRUCK), the log number of licensed drivers per capita (LICENSE), and the fraction of the population living in metro areas (URBAN). For both model setups, we consider all possible subsets of auxiliary regressors and the number of submodels is 128.

The annual data are taken from Li, Linn, and Muehlegger (2014), which are available at the American Economic Journal: Economic Policy website. The total sample size is $2,064$ by state-year from 1966 to 2008; see Li, Linn, and Muehlegger (2014) for a detailed description of data and their source. Our empirical model (7.1) is more general than that used in Li, Linn, and Muehlegger (2014). First, we allow the heterogeneity of the coefficients, which means that we do not impose the assumption that all states have the same price elasticity of gasoline consumption. Second, we consider the presence of the unobserved common factors, $\mathbf{f}_t$, to characterize the global shocks, such as global supply and aggregate demand shocks and oil-specific market demand shocks.[10]

7.2 Empirical Results

We first examine the effect of the tax-inclusive gasoline price on gasoline consumption. In Model Setup I, the parameter of interest is the price elasticity of gasoline consumption, that is, the coefficient of INCPRICE. Table 1 presents the coefficient estimates and standard errors for Model

[9]Model (7.1) holds under the following two assumptions: (1) the retail gasoline price is orthogonal to the gasoline demand shocks in the U.S. market, and (2) the crude oil price is orthogonal to the gasoline demand shocks in the U.S. market. The first assumption can be justified by the fact that the gasoline prices are determined by a huge market, and the fluctuations of price caused by the local shocks are marginal; see Marion and Muehlegger (2011) and Rivers and Schaufele (2015). The second assumption is valid because the crude oil price is determined by the global demand and supply, and hence it is orthogonal to the gasoline demand shock.

[10]Kilian (2010) points that the fluctuation of gasoline consumption is contributed by gasoline demand shocks and global shocks. Stock and Watson (2016) and Caldara, Cavallo, and Iacoviello (2016) find that the fluctuation of the crude oil price is mainly affected by global shocks.

Table 1: Estimation results using the tax-inclusive gasoline price.

	Full	Equal	AIC	BIC	FIC	Plug-In
INCPRICE	-0.0583	-0.0559	-0.0336	-0.0596	-0.0453	-0.0520
	(0.0349)	(0.0273)	(0.0277)	(0.0360)	(0.0252)	(0.0244)
FAMILY	0.0163	0.0059			0.0174	0.0091
	(0.0089)	(0.0043)			(0.0094)	(0.0068)
ROAD	-0.0414	0.0337			0.1220	0.0811
	(0.1041)	(0.0356)			(0.0623)	(0.0437)
INCOME	0.2443	0.1381	0.2902		0.2688	0.2410
	(0.0433)	(0.0203)	(0.0467)		(0.0447)	(0.0367)
CAR	0.0223	0.0198	0.0316			0.0121
	(0.0266)	(0.0108)	(0.0251)			(0.0072)
TRUCK	0.0312	0.0243	0.0243		0.0304	0.0225
	(0.0158)	(0.0065)	(0.0167)		(0.0150)	(0.0106)
LICENSE	0.0743	0.0307			0.0596	0.0502
	(0.0400)	(0.0147)			(0.0298)	(0.0206)
URBAN	2.1189	0.5585		0.6783		0.0792
	(1.2528)	(0.5029)		(0.9057)		(0.2424)

Note: Standard errors are reported in parentheses.

Setup I.[11] The estimation results show that the coefficient estimates of INCPRICE are similar across different estimators, while AIC has a smaller coefficient estimate of INCPRICE. These results suggest that the price elasticity of gasoline consumption is negative, which is consistent with the conventional wisdom in the literature that the elasticity is salient. For auxiliary regressors, most coefficients have the same signs across different estimation methods except the estimated coefficient of ROAD by Full. One important finding from Table 1 is that Plug-In has the smallest standard error of INCPRICE as compared to other estimators.

We next examine the effect of gasoline taxes on gasoline consumption. In Model Setup II, we consider two different focus parameters. The first focus parameter is the price elasticity of demand, that is, the coefficient of EXCPRICE, while the second focus parameter is the tax elasticity of demand, that is, the coefficient of TAX. Table 2 presents the coefficient estimates and standard errors for Model Setup II for both focus parameters. It is clear that both gasoline taxes and the gasoline price have negative effects on gasoline consumption, and the effect of gasoline taxes is stronger than that of the gasoline price. Compared with Table 1, we find that the effect of the gasoline price on gasoline consumption shown in Model Setup II is stronger than that in Model Setup I, which is consistent with the results in Table 2 of Li, Linn, and Muehlegger (2014). Table 2 also shows that the standard errors of focus parameters obtained by Plug-In are smaller than those obtained by other estimators.

Tables 3–5 report the Plug-In weights placed on each submodel and regressor sets for each

[11] The standard error of the plug-in averaging estimator is calculated based on Theorem 5.1, i.e., $\widehat{\boldsymbol{\Xi}}(\mathbf{w}) = \sum_{m=1}^{M} \widehat{w}_m^2 \widehat{\mathbf{D}}_{\boldsymbol{\beta}}' \mathbf{S}_m \widehat{\boldsymbol{\Xi}}_m \mathbf{S}_m' \widehat{\mathbf{D}}_{\boldsymbol{\beta}} + 2\sum\sum_{m \neq \ell} \widehat{w}_m \widehat{w}_\ell \widehat{\mathbf{D}}_{\boldsymbol{\beta}}' \mathbf{S}_m \widehat{\boldsymbol{\Xi}}_{m\ell} \mathbf{S}_\ell' \widehat{\mathbf{D}}_{\boldsymbol{\beta}}$.

Table 2: Estimation results using the tax-exclusive gasoline price.

					μ : EXCPRICE		μ : TAX	
	Full	Equal	AIC	BIC	FIC	Plug-In	FIC	Plug-In
EXCPRICE	-0.0723	-0.0765	-0.0706	-0.1204	-0.0673	-0.0793	-0.0716	-0.0720
	(0.0371)	(0.0255)	(0.0275)	(0.0305)	(0.0238)	(0.0223)	(0.0289)	(0.0244)
TAX	-0.1198	-0.2137	-0.2128	-0.3460	-0.2243	-0.2315	-0.2017	-0.2197
	(0.0849)	(0.0546)	(0.0534)	(0.0751)	(0.0539)	(0.0505)	(0.0517)	(0.0461)
FAMILY	0.0040	0.0006				0.0003		1.94E-05
	(0.0085)	(0.0040)				(0.0015)		(2.10E-05)
ROAD	-0.0636	0.0383				0.0315		0.0226
	(0.1177)	(0.0351)				(0.0190)		(0.0099)
INCOME	0.2130	0.1224	0.2775		0.2859	0.1803	0.2522	0.1960
	(0.0470)	(0.0202)	(0.0474)		(0.0426)	(0.0272)	(0.0459)	(0.0327)
CAR	0.0217	0.0203	0.0574		0.0531	0.0217	0.0572	0.0340
	(0.0234)	(0.0095)	(0.0216)		(0.0205)	(0.0084)	(0.0218)	(0.0118)
TRUCK	0.0184	0.0202				0.0187		0.0182
	(0.0166)	(0.0058)				(0.0039)		(0.0047)
LICENSE	0.0581	0.0217			0.0491	0.0387		0.0214
	(0.0377)	(0.0140)			(0.0301)	(0.0241)		(0.0140)
URBAN	1.9469	0.3266		-0.1697		-0.0263	0.5667	0.0323
	(1.3752)	(0.4865)		(1.1090)		(0.4873)	(1.0438)	(0.6516)

Note: Standard errors are reported in parentheses.

submodel for Model Setup I and II. For Model Setup I, Plug-In puts most weights on the submodels with 4 or 5 auxiliary regressors, while for Model Setup II, Plug-In puts most weights on the submodels with 2–4 auxiliary regressors. One interesting observation is that for Model Setup II the submodels chosen by Plug-In for different focus parameters are almost completely different in Tables 4–5.

8 Conclusion

Many studies have revealed the importance of taking account of the model uncertainty by adopting a model averaging approach. In this paper, we extend the existing literature on frequentist model averaging to the 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 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 as compared to other methods.

Table 3: Submodel weights and regressor set for Model Setup I (μ : INCPRICE).

Weights	Regressor Set
0.1647	INCPRICE, FAMILY, ROAD, INCOME, TRUCK
0.1313	INCPRICE, FAMILY, ROAD, INCOME, CAR, TRUCK
0.0936	INCPRICE, LICENSE
0.0001	INCPRICE, FAMILY, ROAD, INCOME, LICENSE
0.0048	INCPRICE, CAR, LICENSE
0.0002	INCPRICE, INCOME, CAR, LICENSE
0.0003	INCPRICE, ROAD, INCOME, CAR, LICENSE
0.1248	INCPRICE, FAMILY, ROAD, INCOME, CAR, LICENSE
0.2663	INCPRICE, FAMILY, ROAD, INCOME, TRUCK, LICENSE
0.0356	INCPRICE, FAMILY, ROAD, INCOME, CAR, TRUCK , LICENSE
0.0018	INCPRICE, CAR, LICENSE, URBAN
0.1752	INCPRICE, INCOME, TRUCK, LICENSE, URBAN

Table 4: Submodel weights and regressor set for Model Setup II (μ : EXCPRICE).

Weights	Regressor Set
0.1438	EXCPRICE, TAX, FAMILY, ROAD, TRUCK
0.0175	EXCPRICE, TAX, FAMILY, ROAD, INCOME, TRUCK
0.0815	EXCPRICE, TAX, INCOME, LICENSE
0.4079	EXCPRICE, TAX, INCOME, CAR, LICENSE
0.0050	EXCPRICE, TAX, ROAD, TRUCK, URBAN
0.1590	EXCPRICE, TAX, INCOME, LICENSE, URBAN
0.0681	EXCPRICE, TAX, TRUCK, LICENSE, URBAN
0.1170	EXCPRICE, TAX, ROAD, TRUCK, LICENSE, URBAN

Table 5: Submodel weights and regressor set for Model Setup II (μ : TAX).

Weights	Regressor set
0.0925	EXCPRICE, TAX, INCOME, CAR
0.1105	EXCPRICE, TAX, ROAD, TRUCK
0.0001	EXCPRICE, TAX, ROAD, INCOME, TRUCK
0.0928	EXCPRICE, TAX, CAR, LICENSE
0.0602	EXCPRICE, TAX, INCOME, CAR, LICENSE
0.3275	EXCPRICE, TAX, INCOME, CAR, URBAN
0.0002	EXCPRICE, TAX, FAMILY, INCOME, CAR, URBAN
0.0249	EXCPRICE, TAX, ROAD, INCOME, TRUCK, URBAN
0.0022	EXCPRICE, TAX, FAMILY, INCOME, CAR, LICENSE, URBAN
0.2883	EXCPRICE, TAX, INCOME, TRUCK, LICENSE, URBAN

Appendix

A Proofs of Theorems

Proof of Theorem 3.1. 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

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

By Assumption 3.4, we have $\boldsymbol{\beta}_{mi} = \mathbf{S}'_m\boldsymbol{\beta}_i = \mathbf{S}'_m(\boldsymbol{\beta} + \boldsymbol{\eta}_i)$. Then we have

$$
\begin{aligned}
\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m} - \boldsymbol{\beta}_m) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\widehat{\boldsymbol{\beta}}_{mi} - \boldsymbol{\beta}_{mi} + \boldsymbol{\beta}_{mi} - \boldsymbol{\beta}_m) \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{S}'_m\boldsymbol{\eta}_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{X}'_{mi}\mathbf{M}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}'_{mi}\mathbf{M}_{hm}\mathbf{X}_i\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 \\
&\quad + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{X}'_{mi}\mathbf{M}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}'_{mi}\mathbf{M}_{hm}\mathbf{X}_i\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\boldsymbol{\eta}_{2i} \\
&\quad + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{X}'_{mi}\mathbf{M}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}'_{mi}\mathbf{M}_{hm}\mathbf{F}\boldsymbol{\gamma}_i \\
&\quad + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{X}'_{mi}\mathbf{M}_{hm}\mathbf{X}_{mi})^{-1}\mathbf{X}'_{mi}\mathbf{M}_{hm}\boldsymbol{\varepsilon}_i \\
&\equiv I_1 + I_2 + I_3 + I_4 + I_5. \qquad \text{(A.2)}
\end{aligned}
$$

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

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

Therefore, we have

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

where the last equality holds by Lemma C.1 (i). By Assumption 3.4, as $N, T \to \infty$ jointly, we have

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

where

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

and $\mathbf{R}_{mi} = (\mathbf{S}_m' \mathbf{Q}_{mi} \mathbf{S}_m)^{-1} \mathbf{S}_m'$.

We next consider the second term of (A.2). By Assumption 3.5 and Lemma C.1 (i), we have

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

where $\mathbf{A}_m = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbf{R}_{mi} \mathbf{Q}_{mi} \mathbf{S}_0 (\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m' \boldsymbol{\Pi}_m)$ 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 3.3, we have

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

where the last equality holds by Lemma C.1 (i), (iii), and (iv). Therefore, by Assumption 3.3, as $N, T \to \infty$ jointly, we have

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

where

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

Note that $\boldsymbol{\eta}_i$ are distributed independently of $\boldsymbol{\iota}_i$ across i by Assumption 3.4. 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 3.1 and 3.2. It follows that $\mathrm{E}(I_5) = 0$ and

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

where the second equality holds by Lemma C.1 (i) and (ii), and the last equality holds by the fact that $T^{-1}\mathbf{X}_i'\mathbf{M}_{gm}\boldsymbol{\varepsilon}_i = O_p\left(T^{-1/2}\right)$. 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 4.1. 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

$$\begin{aligned}\mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \boldsymbol{\beta}_{2m}^c) &= \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \mathbf{0}) + \frac{\partial \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \mathbf{0})'}{\partial \boldsymbol{\beta}_{2m}^c} \boldsymbol{\beta}_{2m}^c + O\left(\frac{1}{\Delta_{NT}^2}\right) \\ &= \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \mathbf{0}) + \frac{\partial \mu(\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2m}, \mathbf{0})'}{\partial \boldsymbol{\beta}_2} (\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 + O\left(\frac{1}{\Delta_{NT}^2}\right) \\ &= \mu(\boldsymbol{\beta}_m) + \mathbf{D}_{\boldsymbol{\beta}_2}'(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\beta}_2 + O\left(\frac{1}{\Delta_{NT}^2}\right). \end{aligned} \tag{A.8}$$

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 Assumptions 3.1–3.5, the result from Theorem 3.1, and the delta method, we have

$$\sqrt{N}(\mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}) - \mu(\boldsymbol{\beta}_m)) \xrightarrow{d} \mathbf{D}_{\boldsymbol{\beta}_m}'\left(\mathbf{A}_m\boldsymbol{\delta}_c + \mathbf{U}_m + \mathbf{V}_m\right). \tag{A.9}$$

Combining (A.8) and (A.9), it follows that

$$\begin{aligned}\sqrt{N}(\mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}) - \mu(\boldsymbol{\beta})) &= \sqrt{N}(\mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}) - \mu(\boldsymbol{\beta}_m)) - \sqrt{N}(\mu(\boldsymbol{\beta}) - \mu(\boldsymbol{\beta}_m)) \\ &\xrightarrow{d} \mathbf{D}_{\boldsymbol{\beta}_m}'(\mathbf{A}_m\boldsymbol{\delta}_c + \mathbf{U}_m + \mathbf{V}_m) - \mathbf{D}_{\boldsymbol{\beta}_2}'(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\delta}_c \\ &= \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m(\mathbf{U}_m + \mathbf{V}_m) + \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{B}_m\boldsymbol{\delta}_c \\ &\sim \mathrm{N}(\mathbf{D}_{\boldsymbol{\beta}}'\mathbf{B}_m\boldsymbol{\delta}_c, \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_m\boldsymbol{\Xi}_m\mathbf{S}_m'\mathbf{D}_{\boldsymbol{\beta}}),\end{aligned}$$

where $\mathbf{B}_m = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N ((\mathbf{P}_{mi}\mathbf{Q}_{mi} - \mathbf{I}_k)\mathbf{S}_0)$, $\mathbf{P}_{mi} = \mathbf{S}_m\left(\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'$, and the second equality holds by the facts that

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

where the first equality holds by the fact that $\mathbf{S}_m\mathbf{A}_m = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \mathbf{S}_m\left(\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_m\right)^{-1}\mathbf{S}_m'\mathbf{Q}_{mi}\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m) = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \mathbf{P}_{mi}\mathbf{Q}_{mi}\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)$ and the third equality holds by the fact that $\mathbf{S}_0\boldsymbol{\Pi}_m' = \mathbf{S}_m(\mathbf{0}_{k_1\times k_{2m}}', \mathbf{I}_{k_{2m}})'$.

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 3.1. Therefore, the bias term becomes $\mathbf{D}'_{\boldsymbol{\beta}_2}(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\boldsymbol{\delta}_c = \mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_0(\mathbf{I}_{k_2} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\boldsymbol{\delta}_c$. When Assumption 3.6 holds, $\mathbf{V}_m = o_p(1)$ as $N, T \to \infty$ and $\sqrt{N}/T \to 0$ as we discussed in Corollary 3.2. This completes the proof. □

Proof of Theorem 5.1. By Assumptions 3.1–3.5 and the result from Theorem 3.1, we have

$$\begin{aligned}\sqrt{N}(\mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}) - \mu(\boldsymbol{\beta}_m)) &= \widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\mathbf{S}_m\sqrt{N}(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m} - \boldsymbol{\beta}_m) + o_p(1) \\ &= \widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big(\mathbf{S}_m\widehat{\mathbf{R}}_{mi}\widehat{\mathbf{Q}}_{mi}\boldsymbol{\eta}_i + \mathbf{S}_m\widehat{\mathbf{R}}_{mi}\widehat{\boldsymbol{\Sigma}}_{mi}(\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}}) \\ &\quad + \widehat{\mathbf{P}}_{mi}\widehat{\mathbf{Q}}_{mi}\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\Delta_{NT}^{-1}\boldsymbol{\delta}\Big) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).\end{aligned}$$

Therefore, by (A.8) and the fact that the weights are non-random, it follows that

$$\begin{aligned}\sqrt{N}\left(\widehat{\mu}(\mathbf{w}) - \mu(\boldsymbol{\beta})\right) &= \sum_{m=1}^{M} w_m\Big(\mu(\widehat{\boldsymbol{\beta}}_{\mathrm{MG},m}) - \mu(\boldsymbol{\beta}_m) + \mu(\boldsymbol{\beta}_m) - \mu(\boldsymbol{\beta})\Big) \\ &= \sum_{m=1}^{M} w_m\widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{S}_m\widehat{\mathbf{R}}_{mi}\widehat{\mathbf{Q}}_{mi}\boldsymbol{\eta}_i + \sum_{m=1}^{M} w_m\widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{S}_m\widehat{\mathbf{R}}_{mi}\widehat{\boldsymbol{\Sigma}}_{mi}(\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}}) \\ &\quad + \sum_{m=1}^{M} w_m\widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{P}}_{mi}\widehat{\mathbf{Q}}_{mi}\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\sqrt{N}\Delta_{NT}^{-1}\boldsymbol{\delta} \\ &\quad - \sum_{m=1}^{M} w_m\mathbf{D}'_{\boldsymbol{\beta}}\mathbf{S}_0(\mathbf{I} - \boldsymbol{\Pi}'_m\boldsymbol{\Pi}_m)\sqrt{N}\Delta_{NT}^{-1}\boldsymbol{\delta} + O\left(\frac{1}{\Delta_{NT}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \\ &\equiv L_1 + L_2 + L_3 + L_4 + O\left(\frac{1}{\Delta_{NT}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).\end{aligned}$$

For L_1, by Assumption 3.4, as $N, T \to \infty$ jointly, we have

$$L_1 = \widehat{\mathbf{D}}'_{\boldsymbol{\beta}}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\sum_{m=1}^{M} w_m\mathbf{S}_m\widehat{\mathbf{R}}_{mi}\widehat{\mathbf{Q}}_{mi}\right)\boldsymbol{\eta}_i \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Xi}_u(\mathbf{w})) \equiv \mathbf{U}(\mathbf{w}), \tag{A.10}$$

where

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

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

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

Similarly, by Assumption 3.3, as $N, T \to \infty$ jointly, we have

$$L_2 = \widehat{\mathbf{D}}_{\boldsymbol{\beta}}'\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\sum_{m=1}^{M}w_m\mathbf{S}_m\widehat{\mathbf{R}}_{mi}\widehat{\boldsymbol{\Sigma}}_{mi}\right)(\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Xi}_v(\mathbf{w})) \equiv \mathbf{V}(\mathbf{w}), \tag{A.11}$$

where

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

For L_3 and L_4, by Assumption 3.5, we have

$$\begin{aligned}
L_3 + L_4 &= \sum_{m=1}^{M}w_m\left(\widehat{\mathbf{D}}_{\boldsymbol{\beta}}'\frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{P}}_{mi}\widehat{\mathbf{Q}}_{mi}\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m) - \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\right)\sqrt{N}\Delta_{NT}^{-1}\boldsymbol{\delta} \\
&\xrightarrow{p} \sum_{m=1}^{M}w_m\left(\mathbf{D}_{\boldsymbol{\beta}}'\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbf{P}_{mi}\mathbf{Q}_{mi}\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m) - \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\right)\boldsymbol{\delta}_c \\
&= \sum_{m=1}^{M}w_m\mathbf{D}_{\boldsymbol{\beta}}'\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(\mathbf{P}_{mi}\mathbf{Q}_{mi}-\mathbf{I}_k)\mathbf{S}_0(\mathbf{I}-\boldsymbol{\Pi}_m'\boldsymbol{\Pi}_m)\boldsymbol{\delta}_c \\
&= \mathbf{D}_{\boldsymbol{\beta}}'\sum_{m=1}^{M}w_m\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(\mathbf{P}_{mi}\mathbf{Q}_{mi}-\mathbf{I}_k)\mathbf{S}_0\boldsymbol{\delta}_c \\
&= \mathbf{D}_{\boldsymbol{\beta}}'\mathbf{B}(\mathbf{w})\boldsymbol{\delta}_c.
\end{aligned} \tag{A.12}$$

where $\mathbf{B}(\mathbf{w}) = \sum_{m=1}^{M}w_m\mathbf{B}_m$, $\mathbf{B}_m = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}((\mathbf{P}_{mi}\mathbf{Q}_{mi}-\mathbf{I}_k)\mathbf{S}_0)$, and the third equality holds by the fact that $\mathbf{S}_0\boldsymbol{\Pi}_m' = \mathbf{S}_m(\mathbf{0}_{k_1\times k_{2m}}', \mathbf{I}_{k_{2m}})'$.

Since $\boldsymbol{\eta}_i$ and $\boldsymbol{\iota}_i$ are mutually independent by Assumption 3.4, $\mathbf{U}(\mathbf{w})$ and $\mathbf{V}(\mathbf{w})$ are two stochastically independent normal random vectors. Combining (A.10)–(A.12), 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 Theorem 5.2. We first show the limiting distribution of $\widehat{\mathbf{\Psi}}_{m\ell}$. 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 C.1 (i) and the continuous mapping theorem, we have $\widehat{\mathbf{B}}_m \xrightarrow{p} \mathbf{B}_m$, Also, by Lemma 5.1, we have $\widehat{\mathbf{\Xi}}_{m\ell} \xrightarrow{p} \mathbf{\Xi}_{m\ell}$. Recall that $\widehat{\boldsymbol{\delta}}_c \xrightarrow{d} \mathbf{Z}_{\boldsymbol{\delta}} \sim \mathrm{N}\left(\boldsymbol{\delta}_c, \mathbf{S}_0'\mathbf{\Xi}_f\mathbf{S}_0\right)$. Then by the application of Slutsky's theorem, it follows that

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

We next show the limiting distribution of $\mathbf{w}'\widehat{\mathbf{\Psi}}\mathbf{w}$. Using the results from the proof of Theorem 5.1 and the Cramer-Wold Theorem, we can show that

$$\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{R}_{1i}\mathbf{Q}_{1i}\boldsymbol{\eta}_i, ..., \frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{R}_{Mi}\mathbf{Q}_{Mi}\boldsymbol{\eta}_i\right)' \xrightarrow{d} (\mathbf{U}_1, ..., \mathbf{U}_M)'. \tag{A.13}$$

Similarly, we have

$$\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{R}_{1i}\mathbf{\Sigma}_{1i}(\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}}), ..., \frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{R}_{Mi}\mathbf{\Sigma}_{Mi}(\boldsymbol{\iota}_i - \bar{\boldsymbol{\iota}})\right)' \xrightarrow{d} (\mathbf{V}_1, ..., \mathbf{V}_M)'. \tag{A.14}$$

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 (4.5). Since all of $\mathbf{\Psi}_{m\ell}^*$ 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 $\widehat{\mathbf{\Psi}}_{m\ell}$ to $\mathbf{\Psi}_{m\ell}^*$. Hence, it follows that $\mathbf{w}'\widehat{\mathbf{\Psi}}\mathbf{w} \xrightarrow{d} \mathbf{w}'\mathbf{\Psi}^*\mathbf{w}$. 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 $\mathbf{w}'\mathbf{\Psi}^*\mathbf{w}$, 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 Proofs of Lemmas and Corollaries

Proof of Lemma 4.1. The argument is similar to the proof of Lemma 5.1 and we omit it for brevity. □

Proof of Lemma 5.1. From (A.1), we have

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

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

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

By Assumptions 3.4 and 3.5, we have

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

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

For K_3, we have

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

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

For K_4, we have

$$\begin{aligned}
K_4 &= \left(\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\mathbf{M}_{hm}\mathbf{X}_i)\mathbf{S}_m\right)^{-1}\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\mathbf{M}_{hm}\boldsymbol{\varepsilon}_i) \\
&= \left(\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i)\mathbf{S}_m\right)^{-1}\mathbf{S}_m'(T^{-1}\mathbf{X}_i'\mathbf{M}_{gm}\boldsymbol{\varepsilon}_i) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) \\
&= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),
\end{aligned} \tag{B.3}$$

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

Combining (B.1), (B.2), and (B.3), it follows that

$$\widehat{\boldsymbol{\beta}}_{mi} - \boldsymbol{\beta}_m = \boldsymbol{\Phi}_{1imT} + \boldsymbol{\Phi}_{2imT} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O\left(\frac{1}{\Delta_{NT}}\right) \tag{B.4}$$

and

$$\frac{1}{N}\sum_{i=1}^{N}(\widehat{\boldsymbol{\beta}}_{mi}-\boldsymbol{\beta}_{m})=\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Phi}_{1imT}+\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Phi}_{2imT}+O_p\left(\frac{1}{N}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)+O\left(\frac{1}{\Delta_{NT}}\right). \quad \text{(B.5)}$$

By subtracting equation (B.4) from equation (B.5), we obtain

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

By Assumptions 3.3 and 3.4, $\boldsymbol{\Phi}_{1imT}$ and $\boldsymbol{\Phi}_{2imT}$ are mutually independent. Thus, we have

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

This completes the proof. □

Proof of Corollary 3.1. From the proof of Theorem 3.1, it is easy to see that the limits of I_1, I_3, I_4, and I_5 of (A.2) remain the same 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 3.3 and the fact that $\mathbf{X}_i\mathbf{S}_0=\mathbf{X}_{2i}=\mathbf{F}\boldsymbol{\Gamma}_{2i}+\mathbf{V}_{2i}$, we have

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

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

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

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 3.2. We first consider the case that $k_m+1>r$ when Assumption 3.6 holds. We follow the same strategy used in Karabiyik, Reese, and Westerlund (2016). 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

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

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 (2016)

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

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 a similar argument to Pesaran (2006), we have

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

Therefore, Lemma C.1 (ii) becomes

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

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

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

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 \leq 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|| \leq ||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|| \leq ||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|| \leq ||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 \leq ||\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

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

Therefore, we can rewrite (i) and (iii) of Lemma C.1 as

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

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

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

and

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

We now consider the case that $k_m + 1 = r$ when Assumption 3.6 holds. In this case, we have $\mathbf{M}_{gm} = \mathbf{I}_T - \bar{\mathbf{G}}_m(\bar{\mathbf{G}}_m'\bar{\mathbf{G}}_m)^{-}\bar{\mathbf{G}}_m' = \mathbf{I}_T - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-}\mathbf{F}' = \mathbf{M}_f$. Thus, it follows that $\mathbf{M}_{gm}\mathbf{F} = \mathbf{M}_f\mathbf{F} = \mathbf{0}$. Therefore, Lemma C.1 (iii) becomes

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

Similarly, the fourth term of (A.2) becomes

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

Also, the remaining terms of (A.2) have the same limits as shown in the proof of Theorem 3.1. Furthermore, when Assumption 3.6 holds, we can show that $\mathbf{Q}_{mi} = p\lim_{T\to\infty}(\mathbf{X}_i'\mathbf{M}_{gm}\mathbf{X}_i) = \boldsymbol{\Sigma}_i$ since

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

where the last equality holds by the facts that $T^{-1}\mathbf{V}_i'\mathbf{F} = O_p(T^{-1/2})$ and $T^{-1}\mathbf{F}'\mathbf{F} = O_p(1)$. This completes the proof. $\square$

Proof of Corollary 3.3. 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 element 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. $\square$

Proof of Corollary 3.4. From the proof of Theorem 3.1, 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 3.5′ hold. 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. $\square$

C Supplementary Lemmas and Their Proofs

Lemma C.1. Suppose that Assumptions 3.1–3.5 hold. Then we have

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

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

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

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

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

Proof of Lemma C.1. The proofs of (i)-(iii) follow the proof structure adopted in Pesaran (2006). We highlight 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

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

After taking the cross-sectional averages under the equal 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

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

We show (i)-(iii) by establishing that $\mathrm{E}||\bar{\mathbf{u}}_{mt}||^2 = O(N^{-1})$ with Assumption 3.5. Note that the local to zero assumption is imposed on $\boldsymbol{\beta}_2$ only. Thus, by Lemma 1 of Pesaran (2006), we have $\mathrm{Var}(N^{-1}\sum_{i=1}^N \varepsilon_{it}) = O(N^{-1})$, $\mathrm{Var}(N^{-1}\sum_{i=1}^N \mathbf{v}_{1it}) = O(N^{-1})$, $\mathrm{Var}(N^{-1}\sum_{i=1}^N \boldsymbol{\Pi}_m\mathbf{v}_{2it}) = O(N^{-1})$, 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

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

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

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

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 $\text{Cov}\left(N^{-1}\sum_{i=1}^{N}\varepsilon_{it}+N^{-1}\sum_{i=1}^{N}\boldsymbol{\beta}_{1i}'\mathbf{v}_{1it}+N^{-1}\sum_{i=1}^{N}\boldsymbol{\beta}_{2i}'\mathbf{v}_{2it},N^{-1}\sum_{i=1}^{N}\mathbf{v}_{2it}\right)=O\left(N^{-1}\right)+O\left((N\Delta_{NT})^{-1}\right)$. Combining these results, we have $\text{Var}(\bar{\mathbf{u}}_{mt})=O(N^{-1})$ no matter Δ_{NT}^{-1} is $O(1)$ or Δ_{NT}^{-1} converges to zero at any rate. Therefore, we have $\text{E}||\bar{\mathbf{u}}_{mt}||^2=O(N^{-1})$. 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 $\mathbf{M}_{gm}=\mathbf{I}_T-\bar{\mathbf{G}}_m(\bar{\mathbf{G}}_m'\bar{\mathbf{G}}_m)^{-}\bar{\mathbf{G}}_m'$ and $\mathbf{M}_{gm}\bar{\mathbf{G}}_m=\mathbf{0}$. Then, we have $\mathbf{M}_{gm}\mathbf{F}(\bar{\gamma}+\overline{\boldsymbol{\Gamma}_2\boldsymbol{\beta}_2}+\overline{\boldsymbol{\Gamma}_1\boldsymbol{\beta}_1})=\mathbf{0}$, $\quad\mathbf{M}_{gm}\mathbf{F}\bar{\boldsymbol{\Gamma}}_1=\mathbf{0}$, and $\mathbf{M}_{gm}\mathbf{F}\bar{\boldsymbol{\Gamma}}_2\boldsymbol{\Pi}_m'=\mathbf{0}$. Also,

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

By Assumption 3.5 and the above results, we have

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

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

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

By a similar argument, we have

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

This completes the proof. □

References

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

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

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

Caldara, D., M. Cavallo, and M. Iacoviello (2016): "Oil Price Elasticities and Oil Price Fluctuations," Working paper, Federal Reserve Board.

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

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

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

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

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

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

Davis, L. W. and L. Kilian (2011): "Estimating the Effect of a Gasoline Tax on Carbon Emissions," *Journal of Applied Econometrics*, 26, 1187–1214.

DiTraglia, F. (2016): "Using Invalid Instruments on Purpose: Focused Moment Selection and Averaging for GMM," *Journal of Econometrics*, 195, 187–208.

Espey, M. (1998): "Gasoline Demand Revisited: An International Meta-Analysis of Elasticities," *Energy Economics*, 20, 273–295.

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

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

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

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

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

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

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

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

Hughes, J., C. Knittel, D. Sperling, et al. (2008): "Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand," *The Energy Journal*, 29, 113–134.

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

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

Karabiyik, H., S. Reese, and J. Westerlund (2016): "On the Role of the Rank Condition in CCE Estimation of Factor-Augmented Panel Regressions," Forthcoming. *Journal of Econometrics.*

Kilian, L. (2010): "Explaining Fluctuations in Gasoline Prices: A Joint Model of the Global Crude Oil Market and the US Retail Gasoline Market," *The Energy Journal*, 87–112.

Kim, J. and D. Pollard (1990): "Cube Root Asymptotics," *The Annals of Statistics*, 18, 191–219.

Kitagawa, T. and C. Muris (2016): "Model Averaging in Semiparametric Estimation of Treatment Effects," *Journal of Econometrics*, 193, 271–289.

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

Li, S., J. Linn, and E. Muehlegger (2014): "Gasoline Taxes and Consumer Behavior," *American Economic Journal: Economic Policy*, 6, 302–342.

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, 40–58.

Marion, J. and E. Muehlegger (2011): "Fuel Tax Incidence and Supply Conditions," *Journal of Public Economics*, 95, 1202–1212.

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

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

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, 967–1012.

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

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

Rivers, N. and B. Schaufele (2015): "Salience of Carbon Taxes in the Gasoline Market," *Journal of Environmental Economics and Management*, 74, 23–36.

Small, K. A. and K. Van Dender (2007): "Fuel Efficiency and Motor Vehicle Travel: the Declining Rebound Effect," *The Energy Journal*, 25–51.

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

——— (2016): "Factor Models and Structural Vector Autoregressions in Macroeconomics," in Forthcoming. *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, North Holland.

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

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

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

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

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

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

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

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

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

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

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