Weighted Average Estimation in Panel Data

Ali Mehrabani[†] and Aman Ullah [‡]

April 8, 2022

Abstract

Mukhtar M. Ali has made many innovative and influential contributions in different areas of economics, finance, econometrics, and statistics. His contributions include developing econometric models to examine the determinants of the demand for casino gaming, investigating the approximate and exact distribution and moments of various econometric estimators and test statistics, and studying the statistical properties of time series based statistics under stationary and non-stationary processes (for example, see Ali and Thalheimer (1983, 2008), Ali (1977, 1979, 1984, 1989), Ali and Sharma (1993, 1996), Tsui and Ali (1992, 2002), Ali and Giaccotto (1982a, 1982b, 1984), Ali and Tiao (1971), and Ali and Silver (1985, 1989), among others). All of these have made significant impact on the profession and have been instrumental in advancing further research in statistics and econometrics. In this paper, we study the approximate first two moments of two weighted average estimators of the slope parameters in linear panel data models. The weighted average estimators shrink a generalized least squares estimator towards a restricted generalized least squares estimator, where the restrictions represent possible parameter specifications. The averaging weight is inversely proportional to a weighted quadratic loss function. The approximate bias and second moment matrix of the weighted average estimators using the large-sample approximations are provided. We give the conditions under which the weighted average estimators dominate the generalized least squares estimator on the basis of their mean squared errors.

Key Words: Asymptotic approximations; fixed-effects; panel data; random-effects; Stein-like shrinkage estimator.

AMS Subject Classification: 62J12, 62P20

1 Introduction

Estimation and forecasting under model uncertainty has been one of the fundamental issues in econometrics. In recent years, a large body of literature has been concerned with advancing a number of different approaches to overcome a variety of model uncertainty problems. The two most common approaches are model selection and model averaging. Model selection aims to find, among the set of models under consideration, the best approximate model for the unknown true

[†]College of Business and Analytics, Southern Illinois University, Carbondale, IL 62901. E-mail: ali.mehrabani@siu.edu

[‡]Department of Economics, University of California, Riverside, CA 92521. E-mail: aman.ullah@ucr.edu

data generating process. In this method, investigators typically first select the best performing model based on diagnostic tests (like Wald, F, t-ratios, R-squared, information criteria, etc.) and then carry out inference according to the selected model. This popular approach (also known as "pre-testing") is subject to many problems (Magnus (1999); Magnus and Durbin (1999); Danilov and Magnus (2004a, 2004b)). The most important problem is that the model selection and estimation are completely separated such that the uncertainty of the initial model selection step is ignored throughout the parameter estimation and inference, see for example Magnus (2002) and Leeb and Pötscher (2003, 2006), among others, who show the initial model selection step may have non-negligible effects on the statistical properties of the resulting estimators. Taking the above problems into consideration, model averaging is introduced as an alternative to model selection. In model averaging, the uncertainty is addressed by averaging (weighted) over the set of candidate models. However, one of the challenges of this method is how to assign weights to different candidates to minimize a specific loss function.

This paper investigates two weighted average estimation methods in linear panel data models to deal with uncertainty issues about the slope parameters. The weighted average estimators shrink a feasible generalized least-squares (FGLS) estimator towards a shrinkage direction, or equivalently a set of parameter restrictions. The restrictions are not necessarily believed to be true, but instead represent a belief about where the parameters of the model are likely to be close. Therefore, the proposed estimator is a weighted average of the FGLS estimator and a feasible restricted generalized least-squares estimator that belongs to the restricted parameter space. The shrinkage weight is inversely related to a weighted loss statistic that measures the weighted distance of the FGLS estimator and the restricted estimator. To evaluate our proposed estimators, we derive higher order approximations of the bias and mean squared error (MSE) of our proposed estimator using Nagar (1959) large sample approximations. Furthermore, we show the dominance properties of our weighted average estimators in terms of risk, which ensures that our proposed estimators are robust against arbitrary deviations from the restrictions.

The literature on weighted average estimation is substantial, which mainly was initiated by a seminal paper by Stein (1956). In that paper, Stein showed that the maximum likelihood estimator (MLE) for the mean of a multivariate normal distribution is inadmissible. This means that it is possible to construct an estimator with smaller risk than the MLE for the entire parameter space. James and Stein (1961) exhibited an estimator whose risk is uniformly smaller than that of the MLE. Paradoxically, the James-Stein estimator is itself inadmissible and can be dominated by another inadmissible estimate like its positive part (Baranchick (1964)). Judge and Bock (1978) and Ullah and Ullah (1978) developed this method for most of econometric estimators. Recently, Mehrabani and Ullah (2020), Hansen (2016) and Maddala et al. (2001) use weighted average estimation methods to deal with model uncertainty between two candidate models in seemingly unrelated regressions, cross-sectional models, and heterogenous panel data, respectively. See also Lee et al. (2021) who utilize weighted average estimation in structural breaks.

The paper is organized as follows. Sections 2 and 3 describe the model and the estimators. We give the analytical bias, mean squared error matrix and the risk of the weighted average estimators using the large-sample approximations in section 4. Monte Carlo experiments are presented in Section 5 to study the finite sample performance of our proposed estimators. Section 6 contains some concluding remarks, and proofs are given in Appendix A.

Notation: Throughout the paper, we adopt the following notation. For an $m \times n$ real matrix A we write the transpose A'. When A is symmetric, we use $\lambda_{max}(A)$ and $\lambda_{min}(A)$ to denote the largest and smallest eigenvalues, respectively. I_p and $\mathbf{0}_{p\times q}$ denote the $p \times p$ identity matrix and $p \times q$ matrix of zeros.

2 The Model and Assumptions

Consider the following linear panel data model

$$y_{it} = \alpha_i + x_{it}'\beta + u_{it}, \qquad i = 1, \dots, N, \text{ and } t = 1, \dots, T, \tag{2.1}$$

where y_{it} is the dependent variable, $x_{it} = (x_{it,1}, \dots, x_{it,k})'$ is a $k \times 1$ vector of exogenous regressors for unit i, and u_{it} is the unobserved error term, where T is the time dimension, and N is the cross-section dimension. β is a $k \times 1$ vector of common unknown slope coefficients of interest, and α_i is the individual effect (fixed effect or random effect).

Stacking the observations over t, we can express the model in (2.1) as

$$y_i = \alpha_i \iota_T + X_i\beta + u_i, \qquad i = 1, \dots, N, \tag{2.2}$$

where $y_i = (y_{i1}, \dots, y_{iT})'$ is a $T \times 1$ vector of observations on the dependent variable, $X_i = (x_{i1}, \dots, x_{iT})'$ is a $T \times k$ matrix of observations on the regressors, $u_i = (u_{i1}, \dots, u_{iT})'$ is a $T \times 1$ vector of disturbances for $i = 1, \dots, N$, and ι_T is a $T \times 1$ vector of ones. In a matrix form, we can write the model as

$$y = D\alpha + X\beta + u, \tag{2.3}$$

where $y = (y_1', \dots, y_N')'$, $u = (u_1', \dots, u_N')'$, $X = (X_1', \dots, X_N')'$, $\alpha = (\alpha_1, \dots, \alpha_N)'$, and $D = I_N \otimes \iota_T$ is a matrix of $NT \times N$.

We make the following assumptions.

Assumption 1: *The disturbances are normally distributed and for all $i, j = 1, \dots, N$,*

(i) $\mathbb{E}(u_i) = 0$.

(ii) $\mathbb{E}(u_i\, u_j') = \begin{cases} \sigma^2 I_T, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$

Assumption 2: *The matrix of regressors X, which is of order $NT \times k$ has full column rank and consists of non-stochastic elements.*

Assumption 3: *The individual effects, α_i, follow one of the followings* *(a) Fixed Effects Model: they are constant terms.*

(b) Random Effects Model: they are normally distributed and for all $i, j = 1, \ldots, N$, *and* $t = 1, \ldots, T$,

(i) $\mathbb{E}(\alpha_i) = 0$,

(ii) $\mathbb{E}(\alpha_i u_{it}) = 0$,

(iii) $\mathbb{E}(\alpha_i \, \alpha_j') = \begin{cases} \sigma_\alpha^2, & \textit{if } i = j, \\ 0, & \textit{otherwise.} \end{cases}$

3 Estimators

Our goal is to estimate the vector of slope parameters, β, in equation (2.1). We consider four estimators of the slope parameters: *i*) an unrestricted generalized least squares estimator, *ii*) a restricted generalized least squares estimator that shrinks the unrestrictive estimator towards a restricted parameters space, *iii*) a Stein-like weighted average estimator which is a weighted averages of the restricted and the unrestricted estimators where the weights are proportional to a weighted quadratic loss function, and *iv*) a weighted average minimal mean squared error (MMSE) estimator which is a weighted average of the restricted and the unrestricted estimators where the weights are derived by minimizing the risk.

We will examine the estimators for fixed effects and random effects models separately.

3.1 Fixed Effects Models

Since the individual effects, α_i's, are not our primary interest, we concentrate them out and obtain the following regression model from the model in (2.1)

$$\widetilde{y}_i = \widetilde{X}_i \beta + \widetilde{u}_i, \text{ for } i = 1, \ldots, N, \tag{3.1}$$

where for example $\widetilde{y}_i = M_{\iota_T} y_i = (\widetilde{y}_{i1}, \ldots, \widetilde{y}_{iT})'$, $M_{\iota_T} = I_T - \iota_T({\iota_T}'\iota_T)^{-1}{\iota_T}'$, therefore $\widetilde{y}_{it} = y_{it} - T^{-1}\sum_{t=1}^{T} y_{it}$. It is convenient to stack the N equations above in the following form

$$\widetilde{y} = \widetilde{X}\beta + \widetilde{u}, \tag{3.2}$$

where $\widetilde{y} = (\widetilde{y}_1', \ldots, \widetilde{y}_N')'$, $\widetilde{u} = (\widetilde{u}_1', \ldots, \widetilde{u}_N')'$, $\widetilde{X} = (\widetilde{X}_1', \ldots, \widetilde{X}_N')'$, and for example $\widetilde{y} = M_D y$, and $M_D = (I_N \otimes M_{\iota_T}) = I_{NT} - D(D'D)^{-1}D'$ which is an idempotent matrix.

Unrestricted Estimator

The typical estimator for the slope parameters in fixed effects models is a least squares estimator[1] defined as

$$\widehat{\beta} = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{y} = \beta + (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{u}. \tag{3.3}$$

Restricted Estimator

Because of a belief that the true parameter values may be close to a restricted parameter space $\Theta_0 = \{\beta \in \mathbb{R}^k : r(\beta) = 0\}$ where $r(\beta) = R\beta : \mathbb{R}^k \to \mathbb{R}^q$, we want to shrink $\hat{\beta}$ towards the restriction space Θ_0. The purpose of the restrictions can be a model specification, a structural model, a set of exclusion restrictions, or any other restrictions that are often tested by means of hypothesis testing to improve the estimation efficiency.

Hence, we can derive the restricted estimator from the following minimization

$$\underset{\beta}{\text{Minimize}}\ (\widetilde{y} - \widetilde{X}\beta)'(\widetilde{y} - \widetilde{X}\beta), \qquad \text{subject to } R\beta = 0.$$

The solution to the above minimization can be formulated as a restricted least squares estimator

$$\widetilde{\beta} = \widehat{\beta} - (\widetilde{X}'\widetilde{X})^{-1}R'\left[R(\widetilde{X}'\widetilde{X})^{-1}R'\right]^{-1}R\widehat{\beta} = (I_k - \widetilde{P})\widehat{\beta}, \tag{3.4}$$

where $\widetilde{P} = (\widetilde{X}'\widetilde{X})^{-1}R'\left[R(\widetilde{X}'\widetilde{X})^{-1}R'\right]^{-1}R$.

Remark 1: *A restricted parameter space, Θ_0, which is common in applied economics will take the form of an exclusion restriction. For example, if we partition*

$$\beta = \begin{bmatrix} \beta_c \\ \beta_a \end{bmatrix}, \tag{3.5}$$

where β_c, $(k-q) \times 1$, represents the slopes of the core regressors, and β_a, $q \times 1$, includes the slopes of included auxiliary regressors that are included in the model for robustness but may or may not be included in the model. Therefore, an exclusion restriction takes the form

$$R\beta = \begin{bmatrix} \mathbf{0}_{q\times(k-q)} & I_q \end{bmatrix} \beta = \beta_a = \mathbf{0}, \tag{3.6}$$

where the restriction sets the last q slope parameters equal to zero.

[1] We note that under Assumption 1 since the errors are homoscedastic and uncorrelated, the least squares estimator is identical to the generalized least squares estimator.

Weighted Average Estimators

We define the Stein-like weighted average estimator as

$$\widehat{\beta}_A = (1 - \frac{\tau}{\widehat{D}})\, \widehat{\beta} + \frac{\tau}{\widehat{D}}\, \widetilde{\beta}, \tag{3.7}$$

and the weighted average MMSE estimator [2] as below

$$\widehat{\beta}_M = (1 - \frac{\hat{\mu}}{\widehat{D}})\, \widehat{\beta} + \frac{\hat{\mu}}{\widehat{D}}\, \widetilde{\beta}, \tag{3.8}$$

where $\widehat{D}$ is a weighted quadratic loss function defined as

$$\widehat{D} = (\widehat{\beta} - \widetilde{\beta})' W (\widehat{\beta} - \widetilde{\beta}), \tag{3.9}$$

and W is an arbitrary symmetric positive definite weight matrix with elements of order $O(NT)$, $\hat{\mu} = \operatorname{tr}\left(\hat{\sigma}^2 \widetilde{P}(\widetilde{X}'\widetilde{X})^{-1} W\right)$, where $\hat{\sigma}^2 = \tilde{u}' M_{\widetilde{X}} \tilde{u}/(NT - N - k)$, $M_{\widetilde{X}} = I_{NT} - \widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'$, and τ is a positive characterizing parameter. We will defer describing the optimal choice of τ in the next section. It is worth mentioning that, the weighted average estimators are similar and the only difference is in the characterizing parameters. The Stein-like weighted average estimator leaves the characterizing parameter user-specified that can be determined by minimizing a statistic. However, the weighted average MMSE has a specific form for the characterizing parameter that is determined by minimizing the risk.

The idea behind the weighted average estimators defined above is that when the difference between the restricted and the unrestricted estimators is small ($\widehat{D}$ is small), the weighted average estimators give higher weights to the restricted estimator, as it is the most efficient estimator. However, when the difference between the restricted and the unrestricted estimators is substantial, the bias of the restricted estimator, which is caused by imposing the parameter restrictions, can be more than its variance efficiency gain, so the weighted average estimators assign higher weights to the unrestricted estimator.

[2] Consider the class of estimators $\widehat{\beta}_M = \omega\widehat{\beta} + (1-\omega)\widetilde{\beta}$, where ω is a scalar. Then the risk associated with this estimator is

$$\operatorname{Risk}(\widehat{\beta}_M) = \mathbb{E}(\widehat{\beta}_M - \beta)' W (\widehat{\beta}_M - \beta) = \omega^2 \operatorname{Risk}(\widehat{\beta}) + (1-\omega)^2 \operatorname{Risk}(\widetilde{\beta}) + 2\omega(1-\omega)\, \mathbb{E}(\widehat{\beta} - \beta)' W (\widetilde{\beta} - \beta).$$

The value of ω that minimizes the above risk, say ω^* is

$$\omega^* = \frac{\operatorname{tr}\left(\sigma^2 \widetilde{P}(\widetilde{X}'\widetilde{X})^{-1} W\right)}{\operatorname{Risk}(\widehat{\beta}_M)},$$

which can be approximated by $\hat{\omega} = \operatorname{tr}\left(\hat{\sigma}^2 \widetilde{P}(\widetilde{X}'\widetilde{X})^{-1} W\right)/\widehat{D}$.

3.2 Random Effects Models

In the random effects case, we can write the model in (2.1) as below

$$y_i = X_i\beta + \epsilon_i, \tag{3.10}$$

where the error term $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{it})'$ consists of the time-invariant random effects, α_i, and the random component u_i. Under Assumption 3(b), the variance-covariance matrix of ϵ is equal to $\bar{\Omega} = \sigma^2(M_D + \lambda^{-1}Z_D) \equiv \sigma^2\Omega$, where $\lambda = \sigma^2/\sigma_\eta^2$, $\sigma_\eta^2 = \sigma^2 + T\sigma_\alpha^2$, and $Z_D = D(D'D)^{-1}D'$. Ω can be estimated by replacing λ with $\widehat{\lambda} = \hat{\sigma}^2/\hat{\sigma}_\eta^2$, where

$$\hat{\sigma}^2 = \frac{u'(M_D - M_D X(X'M_D X)^{-1}X'M_D)u}{N(T-1)-k}, \tag{3.11}$$

$$\hat{\sigma}_\eta^2 = \frac{\epsilon'\left[Z_D - Z_D X\left(X'Z_D X\right)^{-1}X'Z_D\right]\epsilon}{N-k}. \tag{3.12}$$

Hence, $\widehat{\Omega} = M_D + \widehat{\lambda}^{-1}Z_D$ is an estimator of Ω.

Unrestricted Estimator

The typical estimator for the slope parameters in random effects models is a feasible generalized least squares (GLS) estimator defined as

$$\widehat{\beta} = (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}y = \beta + (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}\epsilon, \tag{3.13}$$

where $\widehat{\Omega}$ is the estimator of Ω.

Restricted Estimator

Because of a belief that the true parameter values may be close to a restricted parameter space $\Theta_0 = \{\beta \in \mathbb{R}^k : r(\beta) = 0\}$ where $r(\beta) = R\beta : \mathbb{R}^k \to \mathbb{R}^q$, we want to shrink $\hat{\beta}$ towards the restriction space Θ_0. The purpose of the restrictions can be a model specification, a structural model, a set of exclusion restrictions, or any other restrictions that are often tested by means of hypothesis testing to improve the estimation efficiency.

Hence, we can derive the restricted estimator from the following minimization

$$\underset{\beta}{\text{Minimize}} \;\; (y - X\beta)'\bar{\Omega}^{-1}(y - X\beta), \qquad \text{subject to } R\beta = 0.$$

The solution to the above minimization can be formulated as a feasible restricted GLS estimator

$$\widetilde{\beta} = \widehat{\beta} - (X'\widehat{\Omega}^{-1}X)^{-1}R'\left[R(X'\widehat{\Omega}^{-1}X)^{-1}R'\right]^{-1}R\widehat{\beta} = (I_k - \widehat{P})\widehat{\beta}, \tag{3.14}$$

where $\widehat{P} = (X'\widehat{\Omega}^{-1}X)^{-1}R'\left[R(X'\widehat{\Omega}^{-1}X)^{-1}R'\right]^{-1}R$.

Weighted Average Estimators

We define the Stein-like weighted average estimator as below

$$\widehat{\beta}_A = (1 - \frac{\tau}{\widehat{D}})\, \widehat{\beta} + \frac{\tau}{\widehat{D}}\, \widetilde{\beta}, \tag{3.15}$$

and the weighted average MMSE estimator as below

$$\widehat{\beta}_M = (1 - \frac{\hat{\mu}}{\widehat{D}})\, \widehat{\beta} + \frac{\hat{\mu}}{\widehat{D}}\, \widetilde{\beta}, \tag{3.16}$$

where $\widehat{D}$ is a weighted quadratic loss function defined as

$$\widehat{D} = (\widehat{\beta} - \widetilde{\beta})' W (\widehat{\beta} - \widetilde{\beta}), \tag{3.17}$$

and W is an arbitrary symmetric positive definite weight matrix with elements of order $O(NT)$, $\hat{\mu} = \operatorname{tr}\Big(\hat{\sigma}^2 \widehat{P} (X'\widehat{\Omega}^{-1}X)^{-1} W\Big)$ and τ is a positive characterizing parameter. We will defer describing the optimal choice for τ in the next section.

4 Large-Sample Approximate Bias and MSE

We employ the large-sample approximations method developed by Nagar (1959), to analyze the bias, mean squared error matrices (MSEM), and risks of the weighted average estimators for the fixed effects and the random effects models.

4.1 Fixed Effects Models

In the fixed effects case where the individual effects are constant terms, the unrestricted estimator is unbiased and we have

$$\text{Bias}(\widehat{\beta}) = \mathbb{E}(\widehat{\beta} - \beta) = 0, \tag{4.1}$$

$$\text{MSEM}(\widehat{\beta}) = \mathbb{E}(\widehat{\beta} - \beta)(\widehat{\beta} - \beta)' = \sigma^2 (\tilde{X}'\tilde{X})^{-1}, \tag{4.2}$$

$$\text{Risk}(\widehat{\beta}) = \mathbb{E}(\widehat{\beta} - \beta)' W (\widehat{\beta} - \beta) = \sigma^2 \operatorname{tr}\Big((\tilde{X}'\tilde{X})^{-1} W\Big), \tag{4.3}$$

Theorem 1: *Under Assumptions 1, 2, and 3(a), the bias of the Stein-like weighted average estimator up to order $O((NT)^{-1})$ is*

$$Bias(\widehat{\beta}_A) = \mathbb{E}(\widehat{\beta}_A - \beta) = -\frac{\tau}{\widetilde{\phi}} \widetilde{P} \beta, \tag{4.4}$$

and the MSEM of the Stein-like weighted average estimator up to order $O((NT)^{-2})$ is

$$MSEM(\widehat{\beta}_A) = \mathbb{E}\left[(\widehat{\beta}_A - \beta)(\widehat{\beta}_A - \beta)'\right] = MSEM(\widehat{\beta}) + \frac{\tau^2}{\widetilde{\phi}^2}\widetilde{P}\beta\beta'\widetilde{P}' - \frac{2\tau}{\widetilde{\phi}}\sigma^2\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}'$$
$$+ \frac{2\tau}{\widetilde{\phi}^2}\sigma^2\Big[\widetilde{P}\beta\beta'\widetilde{P}'W\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}' + \widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}'W\widetilde{P}\beta\beta'\widetilde{P}'\Big], \tag{4.5}$$

and for the symmetric positive definite weight matrix W of order $O(NT)$, the risk of the weighted average estimator up to order $O((NT)^{-1})$ is

$$Risk(\widehat{\beta}_A) = \mathbb{E}\left[(\widehat{\beta}_A - \beta)'W(\widehat{\beta}_A - \beta)\right] = Risk(\widehat{\beta}) + \frac{\tau}{\widetilde{\phi}}\left[\tau - 2\Big[\,tr(\widetilde{C}) - 2\frac{\widetilde{\phi}_c}{\widetilde{\phi}}\Big]\right], \tag{4.6}$$

where $\widetilde{C} = \sigma^2 W^{1/2}\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}'W^{1/2}$, $\widetilde{\phi} = \beta'\widetilde{P}'W\widetilde{P}\beta = O(NT)$, and $\widetilde{\phi}_c = \beta'\widetilde{P}'W^{1/2}\widetilde{C}W^{1/2}\widetilde{P}\beta = O(NT)$.

Proof: Appendix A (See page 21).

From Theorem 1, it follows that the Stein-like weighted average estimator dominates the unrestricted estimator in terms of having a smaller risk, when the second term on the right-hand side of equation (4.6) is negative, which holds when

$$0 < \tau < 2\Big[\,\mathrm{tr}(\widetilde{C}) - 2\frac{\widetilde{\phi}_c}{\widetilde{\phi}}\Big], \tag{4.7}$$

given the term in the square bracket is positive. Therefore, when τ satisfies the condition (4.7), the risk of the Stein-like weighted average estimator is less than the risk of the unrestricted estimator up to the order of interest. In addition, as the choice of the characteristic parameter is user-specified, its optimal value, τ_{opt}, that minimizes the risk of the Stein-like weighted average estimator up to order $O((NT)^{-1})$, is

$$\tau_{opt} = \mathrm{tr}(\widetilde{C}) - 2\frac{\widetilde{\phi}_c}{\widetilde{\phi}}, \tag{4.8}$$

since $\widetilde{\phi}_c/\widetilde{\phi}$ depends on the unknown slope coefficients, one can replace it with its supremum value which is equal to $\lambda_{max}(\widetilde{C})$ [3].

Theorem 2: *Under Assumptions 1, 2, and 3(a), the bias of the weighted average MMSE estimator*

[3]The inequality holds by noting that for any symmetric $n \times n$ matrix Q, we have

$$\lambda_{min}(Q) \leq \frac{\theta' Q\theta}{\theta'\theta} \leq \lambda_{max}(Q),$$

see Abadir and Magnus (2005), pages 181-182.

up to order $O((NT)^{-1})$ is

$$Bias(\widehat{\beta}_M) = \mathbb{E}(\widehat{\beta}_M - \beta) = -\frac{tr(\widetilde{C})}{\widetilde{\phi}}\widetilde{P}\beta, \tag{4.9}$$

and the MSEM of the estimator up to order $O((NT)^{-2})$ is

$$\begin{aligned} MSEM(\widehat{\beta}_M) = \mathbb{E}\left[(\widehat{\beta}_M - \beta)(\widehat{\beta}_M - \beta)'\right] &= MSEM(\widehat{\beta}) + \frac{tr(\widetilde{C})^2}{\widetilde{\phi}^2}\widetilde{P}\beta\beta'\widetilde{P}' - \frac{2\,tr(\widetilde{C})}{\widetilde{\phi}}\sigma^2\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}' \\ &+ \frac{2\,tr(\widetilde{C})}{\widetilde{\phi}^2}\sigma^2\left[\widetilde{P}\beta\beta'\widetilde{P}'W\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}' + \widetilde{P}(\widetilde{X}'X)^{-1}\widetilde{P}'W\widetilde{P}\beta\beta'\widetilde{P}'\right], \end{aligned} \tag{4.10}$$

and for the symmetric positive definite weight matrix W of order $O(NT)$, the risk of the estimator up to order $O((NT)^{-1})$ is

$$Risk(\widehat{\beta}_M) = \mathbb{E}\left[(\widehat{\beta}_M - \beta)'W(\widehat{\beta}_M - \beta)\right] = Risk(\widehat{\beta}) - \frac{tr(\widetilde{C})}{\widetilde{\phi}}\left[\,tr(\widetilde{C}) - 4\frac{\widetilde{\phi}_c}{\widetilde{\phi}}\right]. \tag{4.11}$$

Proof: Appendix A (See page 24).

From Theorem 2, it follows that the weighted average MMSE estimator dominates the unrestricted estimator in terms of having a smaller risk, when the second term on the right-hand side of equation (4.11) is negative, which holds when

$$\text{tr}(\widetilde{C}) > 4\frac{\widetilde{\phi}_c}{\widetilde{\phi}}.$$

Since the condition above depends on the slope parameters, it can be replaced with $\text{tr}(\widetilde{C}) > 4\lambda_{max}(\widetilde{C})$.

Furthermore, comparing the two weighted average estimators, it is clear that the dominance condition of the Stein-like weighted average estimator ($\text{tr}(\widetilde{C}) > 2\lambda_{max}(\widetilde{C})$) is weaker than the dominance condition for the weighted average MMSE estimator ($\text{tr}(\widetilde{C}) > 4\lambda_{max}(\widetilde{C})$.) Moreover, the risk of the Stein-like weighted average estimator using the optimal τ_{opt}, is smaller than the risk of the weighted average MMSE estimator.

4.2 Random Effects Models

In case the individual effects are random, the large-N (fixed T) approximate bias and MSEM of the unrestricted feasible GLS estimator up to order $O(N^{-1})$, and $O(N^{-2})$ respectively, are derived in Ullah and Huang (2006), which are equal to

$$\text{Bias}(\widehat{\beta}) = \mathbb{E}(\widehat{\beta} - \beta) = 0, \tag{4.12}$$

$$\text{MSEM}(\widehat{\beta}) = \mathbb{E}(\widehat{\beta} - \beta)(\widehat{\beta} - \beta)' = \sigma^2 (X'\Omega^{-1}X)^{-1} + \frac{2\lambda\sigma^2 T}{N(T-1)}\Lambda_1, \tag{4.13}$$

$$\text{Risk}(\widehat{\beta}) = \mathbb{E}(\widehat{\beta} - \beta)' W (\widehat{\beta} - \beta) = \sigma^2 \operatorname{tr}(W(X'\Omega^{-1}X)^{-1}) + \frac{2\lambda\sigma^2 T}{N(T-1)} \operatorname{tr}(W\Lambda_1), \tag{4.14}$$

where $A = X'\Omega^{-1}X$, $B = X'Z_D X$, $\Lambda_1 = A^{-1}(B - \lambda B A^{-1} B)A^{-1} = O(N^{-1})$.

Theorem 3: *Under Assumptions 1, 2, and 3(b), the bias of the Stein-like weighted average up to order $O(N^{-1})$ is*

$$Bias(\widehat{\beta}_A) = \mathbb{E}(\widehat{\beta}_A - \beta) = -\frac{\tau}{\phi} P\beta, \tag{4.15}$$

and the MSEM of the average estimator up to order $O(N^{-2})$ is

$$\begin{aligned} MSEM(\widehat{\beta}_A) = \mathbb{E}\left[(\widehat{\beta}_A - \beta)(\widehat{\beta}_A - \beta)'\right] &= MSEM(\widehat{\beta}) + \frac{\tau^2}{\phi^2} P\beta\beta'P' - \frac{2\tau}{\phi}\sigma^2 P(X'\Omega^{-1}X)^{-1}P' \\ &+ \frac{2\tau}{\phi^2}\sigma^2 \Big\{ P\beta\beta'P'WP(X'\Omega^{-1}X)^{-1}P' + P(X'\Omega^{-1}X)^{-1}P'WP\beta\beta'P' \Big\}, \end{aligned} \tag{4.16}$$

and for the symmetric positive definite weight matrix W of order $O(N)$, the risk of the Stein-like weighted average estimator up to order $O(N^{-1})$ is

$$Risk(\widehat{\beta}_A) = \mathbb{E}\left[(\widehat{\beta}_A - \beta)'W(\widehat{\beta}_A - \beta)\right] = Risk(\widehat{\beta}) + \frac{\tau}{\phi}\left[\tau - 2\Big[\, tr(C) - \frac{2}{\phi}\phi_c\Big]\right] \tag{4.17}$$

where $C = \sigma^2 W^{1/2} P (X'\Omega^{-1}X)^{-1} P' W^{1/2}$.

Proof: Appendix A (See page 26).

From Theorem 3, it follows that the Stein-like weighted average estimator dominates the unrestricted estimator in terms of having a smaller risk, when the second term on the right-hand side of equation (4.17) is negative, which holds when

$$0 < \tau < 2\Big[\operatorname{tr}(C) - \frac{2}{\phi}\phi_c\Big], \tag{4.18}$$

given the term in the bracket is positive. Therefore, when τ satisfies the condition (4.18), the risk of the Stein-like weighted average estimator is less than the risk of the unrestricted estimator up to the order of interest. In addition, as the choice of the characteristic parameter is user-specified, its optimal value, τ_{opt}, that minimizes the risk of the Stein-like weighted average estimator up to order $O(N^{-1})$, is

$$\tau_{opt} = \operatorname{tr}(C) - \frac{2}{\phi}\phi_c. \tag{4.19}$$

Further, since ϕ_c/ϕ depends on the unknown slope coefficients, one can replace it with its supremum value which is equal to $\lambda_{max}(C)$.

Theorem 4: *Under Assumptions 1, 2, and 3(b), the bias of the weighted average MMSE estimator up to order $O(N^{-1})$ is*

$$Bias(\widehat{\beta}_M) = \mathbb{E}(\widehat{\beta}_M - \beta) = -\frac{tr(C)}{\phi} P\beta, \tag{4.20}$$

and the MSEM of the estimator up to order $O(N^{-2})$ is

$$MSEM(\widehat{\beta}_M) = \mathbb{E}\left[(\widehat{\beta}_M - \beta)(\widehat{\beta}_M - \beta)'\right] = MSEM(\widehat{\beta}) + \frac{tr(C)^2}{\phi^2} P\beta\beta'P' - \frac{2\,tr(C)}{\phi}\sigma^2 P(X'\Omega^{-1}X)^{-1}P'$$
$$+ \frac{2\,tr(C)}{\phi^2}\sigma^2\left\{P\beta\beta'P'WP(X'\Omega^{-1}X)^{-1}P' + P(X'\Omega^{-1}X)^{-1}P'WP\beta\beta'P'\right\}, \tag{4.21}$$

and for the symmetric positive definite weight matrix W of order $O(N)$, the risk of the estimator up to order $O(N^{-1})$ is

$$Risk(\widehat{\beta}_M) = \mathbb{E}\left[(\widehat{\beta}_M - \beta)'W(\widehat{\beta}_M - \beta)\right] = Risk(\widehat{\beta}) - \frac{tr(C)}{\phi}\left[tr(C) - \frac{4}{\phi}\phi_c\right]. \tag{4.22}$$

Proof: Appendix A (See page 30).

From Theorem 4, it follows that the weighted average MMSE estimator dominates the unrestricted estimator in terms of having a smaller risk, when the second term on the right-hand side of equation (4.22) is negative, which holds when

$$\mathrm{tr}(C) > 4\frac{\phi_c}{\phi}.$$

Since the condition above depends on the slope parameters, it can be replaced with $\mathrm{tr}(C) > 4\lambda_{max}(C)$.

Furthermore, comparing the two weighted average estimators, it is clear that the dominance condition of the Stein-like weighted average estimator ($\mathrm{tr}(C) > 2\lambda_{max}(C)$) is weaker than the dominance condition for the weighted average MMSE estimator ($\mathrm{tr}(C) > 4\lambda_{max}(C)$.) Moreover, the risk of the Stein-like weighted average estimator using the optimal τ_{opt}, is smaller than the risk of the weighted average MMSE estimator.

5 Monte Carlo Simulation

In this section, we investigate the finite sample mean squared error of the Stein-like weighted average and the weighted average MMSE estimator via Monte Carlo experiments.

We consider the following data-generating process

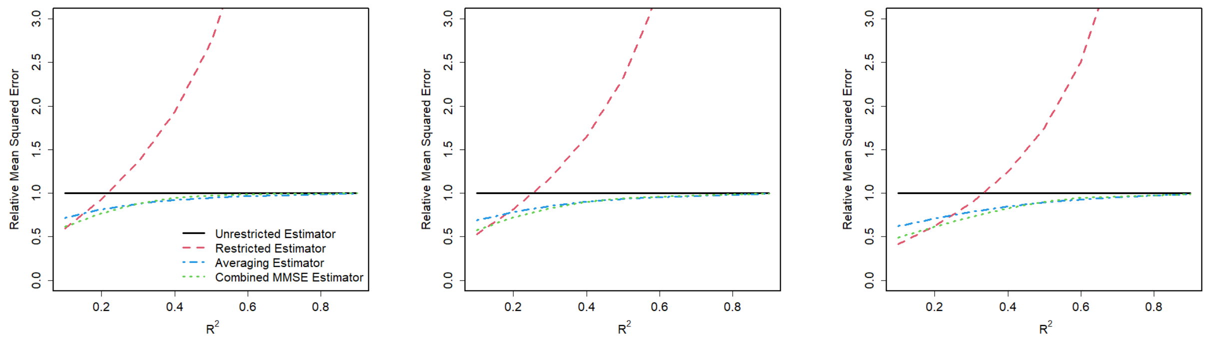
$$y_{it} = \alpha_i + x_{it}'\beta + u_{it},$$

where u_{it} is i.i.d. N(0,1), α_i is i.i.d. N(0,1). The regressors are generated for fixed effects as $x_{it} = \boldsymbol{v}_{it} + 0.2\alpha_i$, and for random effects case as $x_{it} = \boldsymbol{v}_{it}$, and $\boldsymbol{v}_{it} = (v_{1,it}, \ldots, v_{k,it})' \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, where the diagonal elements of $\boldsymbol{\Sigma}$ are σ_v^2 and off-diagonal elements are $\rho\sigma_v^2$, and we set $\sigma_v^2 = 0.1$. The number of regressors is $k = 6$ with two core regressors and four auxiliary regressors. The regression coefficients are determined by the rule

$$\beta = c\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{\sqrt{NT}}\left(1, \frac{q-1}{q}, \ldots, \frac{1}{q}\right)\right)',$$

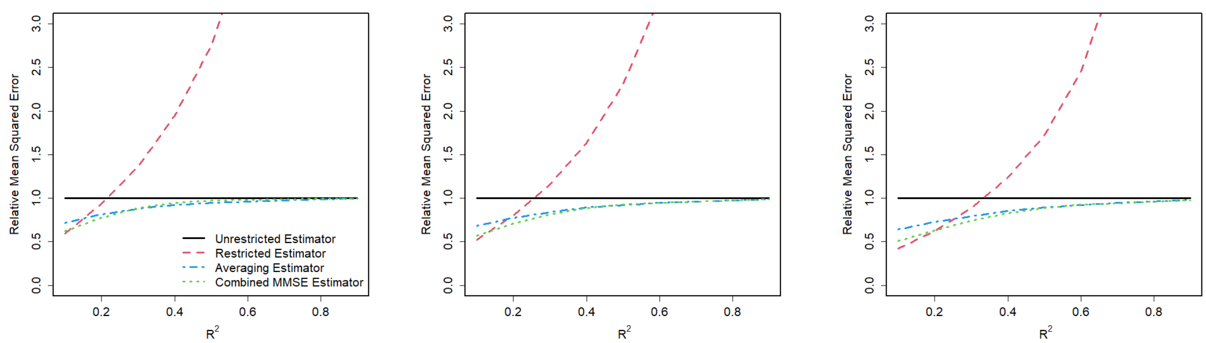
where q is the number of auxiliary regressors. The parameter c is selected to control the population R^2, and R^2 varies on a grid between 0.1 and 0.9.

We consider four estimators for each fixed effects and random effects model: (1) the unrestricted estimator, (2) the restricted estimator where the restriction matrix follows the form of the restriction matrix in Remark 1 with $q = 4$, (3) the Stein-like weighted average estimator, (4) the weighted average MMSE estimator. Our parameters of interest are the slope parameters of the core regressors (the first two slope parameters). To evaluate the performance of our proposed estimators, we compute the risk based on the quadratic loss function. The risk (expected squared error) is calculated by averaging across 1000 random samples. We report the normalized risk by dividing the risk of each estimator by the risk of the unrestricted estimator in figures 1-5. The results show the normalized risk for $\rho = 0, 0.25, 0.5$ in three panels for different N and T. It is clear that both proposed estimators perform better than the unrestricted estimator over the whole range of R^2, which supports the theoretical findings of the previous section. The Stein-like weighted average estimator and the weighted average MMSE estimator have similar performance for all values of ρ. However, the weighted average MMSE estimator performs slightly better for small values of R^2 and the Stein-like weighted average estimator performs better for the rest. This is expected because for small values of R^2 the bias of the restricted estimator is very small, so it has a smaller risk than the unrestricted estimator, and as the weighted average MMSE estimator assigns a larger weight to the restricted estimator, initially it performs better and as the risk of the restricted estimator increases, the risk of the weighted average MMSE estimator becomes slightly larger than the Stein-like weighted average estimator. Furthermore, these figures show that the ranking of the estimators is quite similar across different sample sizes (T) and different number of cross-sections (N).



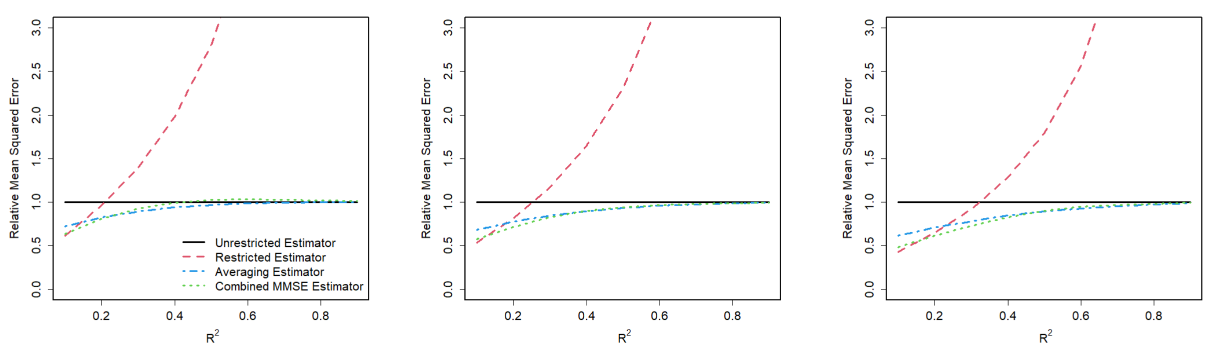
(a) $\rho = 0$ (b) $\rho = 0.25$ (c) $\rho = 0.5$

Figure 1: Relative MSE of Unrestricted, Restricted, Averaging, and Combined MMSE Estimators, for Fixed Effects model with $T = 100, N = 100, k = 6, q = 4$



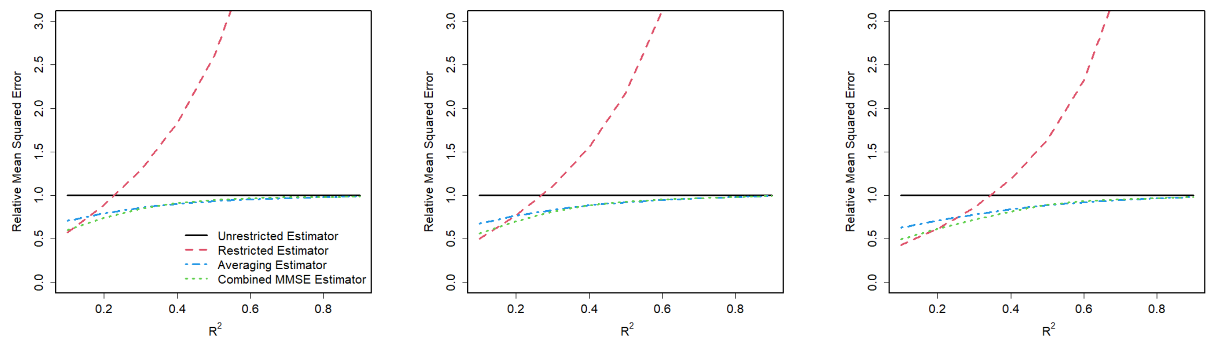
(a) $\rho = 0$ (b) $\rho = 0.25$ (c) $\rho = 0.5$

Figure 2: Relative MSE of Unrestricted, Restricted, Averaging, and Combined MMSE Estimators, for Fixed Effects model with $T = 100, N = 50, k = 6, q = 4$



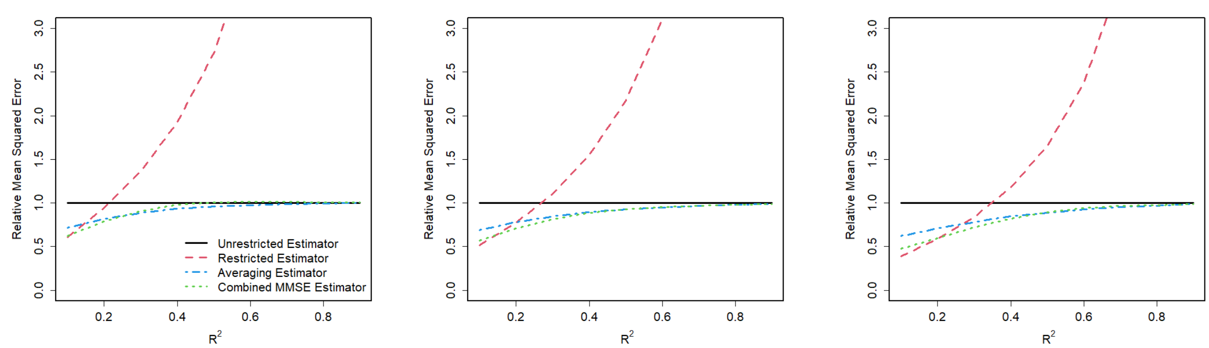
(a) $\rho = 0$ (b) $\rho = 0.25$ (c) $\rho = 0.5$

Figure 3: Relative MSE of Unrestricted, Restricted, Averaging, and Combined MMSE Estimators, for Fixed Effects model with $T = 50, N = 100, k = 6, q = 4$



(a) $\rho = 0$ (b) $\rho = 0.25$ (c) $\rho = 0.5$

Figure 4: Relative MSE of Unrestricted, Restricted, Averaging, and Combined MMSE Estimators, for Random Effects model with $T = 20, N = 100, k = 6, q = 4$



(a) $\rho = 0$ (b) $\rho = 0.25$ (c) $\rho = 0.5$

Figure 5: Relative MSE of Unrestricted, Restricted, Averaging, and Combined MMSE Estimators, for Random Effects model with $T = 20, N = 200, k = 6, q = 4$

6 Conclusions

In this paper, we introduce two weighted average estimators for estimating the slope parameters in linear panel data models. The introduced estimators are weighed averages of an unrestricted generalized least squares estimator, and a restricted generalized least squares estimator. The weights are inversely related to a weighted quadratic loss function which measures the weighted distance between the unrestricted and the restricted estimators. The analytical bias, MSE matrix, and risk of the weighted average estimators using large-sample approximations of Nagar (1959) are derived. The superiority conditions of the weighted average estimators in terms of the risk are given for any user-specific symmetric positive definite weight matrix.

References

Abadir, Karim M., and Jan R. Magnus. 2005. *Matrix Algebra*. New York: Cambridge University Press.

Ali, M. M. (1977). Analysis of ARMA Models - Estimation and Prediction. *Biometrika* Vol. 65, No. 3 , 677.

Ali, M. M. (1979). Analysis of Stationary Spatial-Temporal Processes: Estimation and Prediction. *Biometrika* Vol. 66, No. 3 , 513–518.

Ali, M. M. (1984). An Approximation to the Null Distribution and Power Properties of the Durbin-Watson Statistic. *Biometrika* Vol. 71, No. 2 , 253–261.

Ali, M. M. (1989). Tests for Autocorrelation and Randomness in Multiple Time Series. *Journal of American Statistical Association* Vol. 84, No. 406, 533–540.

Ali, M. M. and C. Giaccotto (1982a). Optimum Distribution-Free Tests and Further Evidence of Heteroskedasticity in the Market Model. *Journal of Finance* Vol. 27, No. 5, 1247–1257.

Ali, M. M. and C. Giaccotto (1982b). The Identical Distribution Hypothesis for Stock-Market Prices-Location and Scale Shift Alternatives. *Journal of American Statistical Assocication* Vol. 77, No. 377, 19–28.

Ali, M. M. and C. Giaccotto (1984). A Study of Several New and Existing Tests for Heteroscedasticity in the General Linear Model. *Journal of Econometrics* Vol. 26, 355–373.

Ali, M. M. and S. C. Sharma (1993). Robustness to Non-Normality of the Durbin-Watson Test for Autocorrelation. *Journal of Econometrics* Vol. 57, No. 1-3, 117–136.

Ali, M. M. and S. C. Sharma (1996). Moments of Regression F-Statistics Under Non-Normality. *Communication in Statistics: Theory and Methods* Vol. 25, No. 7, 1631–1658.

Ali, M. M. and J. L. Silver (1985). Tests for Equality between Sets of Coefficients in Two Linear Regressions under Heteroscedasticity. *Journal of American Statistical Association* Vol. 80, No. 391, 730–735.

Ali, M. M. and J. L. Silver (1989). Testing Slutsky Symmetry in Systems of Linear Demand Equations. *Journal of Econometrics* Vol. 41, 251–266.

Ali, M. M. and G. C. Tiao (1971). Analysis of Correlated Random Effects: Linear Model with Two Random Components. *Biometrika* Vol. 58, No. 1, 37–51.

Ali, M. M. and R. Thalheimer (1983). Stationarity Tests in Time Series Model Building. *Journal of Forecasting* Vol. 2, No. 3, 249–257.

Ali, M. M. and R. Thalheimer (2008). The Demand for Casino Gaming with Special Reference to a Smoking Ban. *Economic Inquiry* Vol. 46, No. 2 , 273–282.

Baranchick, A., 1964. Multiple Regression and Estimation of the Mean of A Multivariate Normal Distribution, Technical Report No. 51. *Department of Statistics, Stanford University.*

Danilov, D., J. R. Magnus, 2004a. On the harm that ignoring pretesting can cause. *Journal of Econometrics* 122:27–46.

Danilov, D., J. R. Magnus, 2004b. Forecast accuracy after pretesting with an application to the stock market. *Journal of Forecasting* 23:251–274.

James, W., C. Stein, 1961. Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1: 361–80.

Judge, G., M.E. Bock, 1978. The Statistical Implications of Pre-test and Stein- rule Estimators in Econometrics. *North-Holland.*

Hansen, B. E., 2016. Efficient Shrinkage in Parametric Models. *Journal of Econometrics* 190(1): 115–32.

Lee, T. , Parsaeian, Sh. and Ullah A. (2021). Optimal forecast under structural breaks. *Advances in Econometrics*, Vol. 43, ed. by Alexander Chudik, Cheng Hsiao and Allan Timmermann.

Leeb, H. Pötscher, B. M., 2003. The finite-sample distribution of post-model-selection estimators and uniform versus nonuniform approximations. *Econometric Theory* 19:100–142.

Leeb, H. Pötscher, B. M., 2006. Can one estimate the conditional distribution of post-model-selection estimators? *The Annals of Statistics* 34:2554–2591.

Maddala, G. S., Li H., V. K. Srivastava, 2001. A Comparative Study of Different Shrinkage Estimators for Panel Data Models. *Annals of Economics and Finance* 2: 1–30.

Magnus, J. R., 1999. The traditional pretest estimator. *Theory of Probability and Its Applications* 44:293–308.

Magnus, J. R., J. Durbin, 1999. Estimation of regression coefficients of interest when other regression coefficients are of no interest. *Econometrica* 67:639–643.

Magnus, J. R., 2002. Estimation of the mean of a univariate normal distribution with known variance. *Econometrics Journal* 5:225–236.

Mehrabani, A., Ullah, A. 2020. Improved Average Estimation in Seemingly Unrelated Regressions. *Econometrics* 8, 15.

Nagar, A. L. 1959. The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations. *Econometrica* 27(4): 575–95.

Stein C., 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of Berkeley Symposium on Mathematical Statistics and Probability* 1: 197–206.

Tsui A. K. and M. M. Ali (1992). Numerical Computation of Exact Moments of the Least Squares Estimator in a First Order Stationary Auto-Regressive Model. *Journal of Statiscial Simulation and Computation* Vol. 42, 65–78.

Tsui A. K. and M. M. Ali (2002). OLS Bias and MSE in the Random Walk Model with a Constant Initial Value. *Journal of Applied Statistical Science* Vol. 11, No. 1, 9–19.

Ullah, A., 2004. *Finite Sample Econometrics.* Oxford: Oxford University Press.

Ullah, A., and X. Huang. 2006. Finite Sample Properties of FGLS Estimator for Random-Effects Model under Non-Normality. *Contribution to Economics Analysis.* Edited by B. H. Baltagi. Elsevier, North Holland. Chp. 3.

Ullah, A., and S. Ullah (1978). Double k-Class Estimators of Coefficients in Linear Regression. *Econometrica* Vol. 46, No. 3, 705–722.

A Appendix A

Lemma A.1: *If $\Delta = \widehat{\Omega} - \Omega = O_p(n^{-1/2})$, then we have the followings*

$$\widehat{\Omega}^{-1} = \underbrace{\Omega^{-1}}_{O_p(1)} - \underbrace{\Omega^{-1}\Delta\Omega^{-1}}_{O_p(n^{-1/2})} + \underbrace{\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}}_{O_p(n^{-1})} - \underbrace{\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}}_{O_p(n^{-3/2})} + O_p(n^{-2}), \tag{A.1}$$

$$\begin{aligned}(X'\widehat{\Omega}^{-1}X)^{-1} = & \underbrace{(X'\Omega^{-1}X)^{-1}}_{O_p(n^{-1})} + \underbrace{(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}X(X'\Omega^{-1}X)^{-1}}_{O_p(n^{-3/2})} \\ & + \underbrace{(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}X(X'\Omega^{-1}X)^{-1}}_{O_p(n^{-2})} \\ & - \underbrace{(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}X(X'\Omega^{-1}X)^{-1}}_{O_p(n^{-2})} + O_p(n^{-5/2})\end{aligned} \tag{A.2}$$

$$X'\widehat{\Omega}^{-1}u = \underbrace{X'\Omega^{-1}u}_{O_p(n^{1/2})} - \underbrace{X'\Omega^{-1}\Delta\Omega^{-1}u}_{O_p(1)} + \underbrace{X'\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}u}_{O_p(n^{-1/2})} + O_p(n^{-1}), \tag{A.3}$$

let $\widehat{P} = (X'\widehat{\Omega}^{-1}X)^{-1}R'\Big[R(X'\widehat{\Omega}^{-1}X)^{-1}R'\Big]^{-1}R$, *then*

$$\widehat{P} = P + P_{-1/2} + O_p(n^{-1}), \tag{A.4}$$

where $P_{-1/2} = \Big[I_k - P\Big](X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}XP = O_p(n^{-1/2})$, *and the suffixes show the order of magnitude in probability.*

Proof:

Using the standard geometric expansion for the inverse of a matrix [4], *for large n, we have the followings*

$$\begin{aligned}\widehat{\Omega}^{-1} &= (\Omega + \Delta)^{-1} = \Omega^{-1}[I_n + \Delta\Omega^{-1}]^{-1} \\ &= \Omega^{-1}\Big[I_n - \Delta\Omega^{-1} + \Delta\Omega^{-1}\Delta\Omega^{-1} - \Delta\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1} + ...\Big] \\ &= \underbrace{\Omega^{-1}}_{O_p(1)} - \underbrace{\Omega^{-1}\Delta\Omega^{-1}}_{O_p(n^{-1/2})} + \underbrace{\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}}_{O_p(n^{-1})} - \underbrace{\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}}_{O_p(n^{-3/2})} + O_p(n^{-2}),\end{aligned}$$

which gives the results in equation (A.1). Now, by using equation (A.1), we have

$$\begin{aligned}(X'\widehat{\Omega}^{-1}X)^{-1} &= \Big[X'\Omega^{-1}X - X'\Omega^{-1}\Delta\Omega^{-1}X + X'\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}X + ...\Big]^{-1} \\ &= (X'\Omega^{-1}X)^{-1}\Big[I_k - X'\Omega^{-1}\Delta\Omega^{-1}X(X'\Omega^{-1}X)^{-1} + X'\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}X(X'\Omega^{-1}X)^{-1} + ...\Big]^{-1}\end{aligned}$$

[4] $(I + A)^{-1} = I - A + A^2 - A^3 + \dots$.

$$
\begin{aligned}
&= \underbrace{(X'\Omega^{-1}X)^{-1}}_{O_p(n^{-1})} + \underbrace{(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}X(X'\Omega^{-1}X)^{-1}}_{O_p(n^{-3/2})} \\
&+ \underbrace{(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}X(X'\Omega^{-1}X)^{-1}}_{O_p(n^{-2})} \\
&- \underbrace{(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}X(X'\Omega^{-1}X)^{-1}}_{O_p(n^{-2})} + O_p(n^{-5/2}),
\end{aligned}
$$

also we have

$$
X'\widehat{\Omega}^{-1}u = \underbrace{X'\Omega^{-1}u}_{O_p(n^{1/2})} - \underbrace{X'\Omega^{-1}\Delta\Omega^{-1}u}_{O_p(1)} + \underbrace{X'\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}u}_{O_p(n^{-1/2})} + O_p(n^{-1}).
$$

By using the above results, we have

$$
\Big[R(X'\widehat{\Omega}^{-1}X)^{-1}R'\Big]^{-1} = S_1 + S_{1/2} + O_p(1),
$$

where

$$
S_1 = \Big[R(X'\Omega^{-1}X)^{-1}R'\Big]^{-1} = O_p(n),
$$

and

$$
S_{1/2} = -S_1 R(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}X(X'\Omega^{-1}X)^{-1}R'S_1 = O_p(n^{1/2}),
$$

hence

$$
\widehat{P} = P + P_{-1/2} + O_p(n^{-1}),
$$

where

$$
\begin{aligned}
P &= (X'\Omega^{-1}X)^{-1}R'\Big[R(X'\Omega^{-1}X)^{-1}R'\Big]^{-1}R = O_p(1), \\
P_{-1/2} &= \Big[I_k - P\Big](X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}XP = O_p(n^{-1/2}).
\end{aligned}
$$

■

Proof. Theorem 1 :

From equation (3.3), we have

$$
\widehat{\beta} - \beta = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{u} = \xi_{-1/2}, \tag{A.5}
$$

where $\xi_{-1/2}$, is defined below, and the suffix shows the order of magnitude in probability,

$$\xi_{-1/2} = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{u} = O_p((NT)^{-1/2}).$$

Using equation (A.5) in equation (3.9), we have

$$\begin{aligned}
\frac{1}{\widehat{D}} &= \Big[(\widehat{\beta}-\widetilde{\beta})'W(\widehat{\beta}-\widetilde{\beta})\Big]^{-1} = \Big[\widehat{\beta}'\,\widetilde{P}'\,W\,\widetilde{P}\,\widehat{\beta}\Big]^{-1} = \Big[\Big(\beta+\xi_{-1/2}\Big)'\widetilde{P}'W\widetilde{P}\Big(\beta+\xi_{-1/2}\Big)\Big]^{-1} \\
&= \Big[\widetilde{\phi}+2\beta'\widetilde{P}'W\widetilde{P}\xi_{-1/2}\Big]^{-1} = \frac{1}{\widetilde{\phi}}\Big[1+\frac{2}{\widetilde{\phi}}\beta'\widetilde{P}'W\widetilde{P}\xi_{-1/2}\Big]^{-1} = \frac{1}{\widetilde{\phi}}\Big[1-\frac{2}{\widetilde{\phi}}\beta'\widetilde{P}'W\widetilde{P}\xi_{-1/2}\Big] + O_p((NT)^{-2}) \\
&\equiv \underbrace{\frac{1}{\widetilde{\phi}}}_{O_p((NT)^{-1})} - \underbrace{\frac{1}{\widetilde{\phi}^2}D_{1/2}}_{O_p((NT)^{-3/2})} + O_p((NT)^{-2}),
\end{aligned} \tag{A.6}$$

where $D_{1/2} = 2\beta'\widetilde{P}'W\widetilde{P}\xi_{-1/2} = O_p((NT)^{1/2})$, $\widetilde{\phi} = \beta'\widetilde{P}'W\widetilde{P}\beta = O(NT)$, and the last equality above holds by using the standard geometric expansion. The terms with order $O_p((NT)^{-2})$ and smaller are dropped, because they will not enter in the calculation of the bias and MSEM of the average estimator up to the orders of interest.

Employing equations (A.6) in equation (3.7), we obtain

$$\begin{aligned}
\widehat{\beta}_A - \beta &= (\widehat{\beta}-\beta) - \tau\Big[\frac{1}{\widetilde{\phi}} - \frac{1}{\widetilde{\phi}^2}D_{1/2} + O_p((NT)^{-2})\Big]\widetilde{P}\widehat{\beta} \\
&= \zeta_{-1/2} + \zeta_{-1} + \zeta_{-3/2} + O_p((NT)^{-2}),
\end{aligned} \tag{A.7}$$

where $\zeta_{-1/2}, \zeta_{-1}$ and $\zeta_{-3/2}$ are defined below

$$\begin{aligned}
\zeta_{-1/2} &= \xi_{-1/2} = O_p((NT)^{-1/2}), \\
\zeta_{-1} &= -\frac{\tau}{\widetilde{\phi}}\widetilde{P}\beta = O_p((NT)^{-1}), \\
\zeta_{-3/2} &= -\frac{\tau}{\widetilde{\phi}}\widetilde{P}\xi_{-1/2} + \frac{\tau}{\widetilde{\phi}^2}D_{1/2}\widetilde{P}\beta = O_p((NT)^{-3/2}).
\end{aligned}$$

The bias of the average estimator using the approximations in equation (A.7) up to order $O((NT)^{-1})$ is

$$\mathbb{E}(\widehat{\beta}_A - \beta) = \mathbb{E}(\zeta_{-1/2}+\zeta_{-1}) = \mathbb{E}(\xi_{-1/2}) - \frac{\tau}{\widetilde{\phi}}\widetilde{P}\beta = -\frac{\tau}{\widetilde{\phi}}\widetilde{P}\beta, \tag{A.8}$$

where the last equality holds because $\mathbb{E}(u) = 0$, hence

$$\mathbb{E}(\xi_{-1/2}) = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\,\mathbb{E}(\widetilde{u}) = 0. \tag{A.9}$$

The MSEM up to order $O((NT)^{-2})$ is

$$\mathbb{E}\left[(\widehat{\beta}_A-\beta)(\widehat{\beta}_A-\beta)'\right]=\mathbb{E}(\Gamma_{-1}+\Gamma_{-3/2}+\Gamma_{-2}), \tag{A.10}$$

where Γ_{-1}, $\Gamma_{-3/2}$ and Γ_{-2} are

$$\begin{aligned}
\Gamma_{-1}&=\zeta_{-1/2}\zeta_{-1/2}',\\
\Gamma_{-3/2}&=\zeta_{-1/2}\zeta_{-1}'+\zeta_{-1}\zeta_{-1/2}',\\
\Gamma_{-2}&=\zeta_{-1/2}\zeta_{-3/2}'+\zeta_{-3/2}\zeta_{-1/2}'+\zeta_{-1}\zeta_{-1}',
\end{aligned}$$

and we give their expectations below

$$\mathbb{E}(\Gamma_{-1})=\mathbb{E}\left((\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{u}\widetilde{u}'\widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}\right)=\sigma^2(\widetilde{X}'\widetilde{X})^{-1}, \tag{A.11}$$

$$\mathbb{E}(\Gamma_{-3/2})=\mathbb{E}(\zeta_{-1/2}\zeta_{-1}')+\mathbb{E}(\zeta_{-1}\zeta_{-1/2}')=0, \tag{A.12}$$

because

$$\mathbb{E}(\zeta_{-1}\zeta_{-1/2}')=\mathbb{E}\left[-\frac{\tau}{\widetilde{\phi}}\widetilde{P}\beta\xi_{-1/2}'\right]=-\frac{\tau}{\widetilde{\phi}}\widetilde{P}\beta\,\mathbb{E}(\xi_{-1/2}')=0, \tag{A.13}$$

where the last equality holds by (A.9). Also, we have

$$\begin{aligned}
\mathbb{E}(\Gamma_{-2})&=\mathbb{E}(\zeta_{-1/2}\zeta_{-3/2}')+\mathbb{E}(\zeta_{-3/2}\zeta_{-1/2}')+\mathbb{E}(\zeta_{-1}\zeta_{-1}')\\
&=\frac{\tau^2}{\widetilde{\phi}^2}\widetilde{P}\beta\beta'\widetilde{P}'-\frac{\tau}{\widetilde{\phi}}\sigma^2\left[\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}+(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}'\right]\\
&\quad+\frac{2\tau}{\widetilde{\phi}^2}\sigma^2\left[\widetilde{P}\beta\beta'\widetilde{P}'W\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}+(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}'W\widetilde{P}\beta\beta'\widetilde{P}'\right],
\end{aligned} \tag{A.14}$$

where the last equality above holds by using equations (A.15) and (A.16) below

$$\mathbb{E}(\zeta_{-1}\zeta_{-1}')=\frac{\tau^2}{\widetilde{\phi}^2}\widetilde{P}\beta\beta'\widetilde{P}', \tag{A.15}$$

and

$$\begin{aligned}
\mathbb{E}(\zeta_{-3/2}\zeta_{-1/2}')&=-\frac{\tau}{\widetilde{\phi}}\widetilde{P}\,\mathbb{E}(\xi_{-1/2}\xi_{-1/2}')+\frac{\tau}{\widetilde{\phi}^2}\,\mathbb{E}(D_{1/2}\widetilde{P}\beta\xi_{-1/2}')\\
&=-\frac{\tau}{\widetilde{\phi}}\sigma^2\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}+\frac{2\tau}{\widetilde{\phi}^2}\sigma^2\widetilde{P}\beta\beta'\widetilde{P}'W\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1},
\end{aligned} \tag{A.16}$$

and the last equality above holds by using

$$\mathbb{E}(D_{1/2}\widetilde{P}\beta\xi_{-1/2}')=2\widetilde{P}\beta\beta'\widetilde{P}'W\widetilde{P}\,\mathbb{E}\left(\xi_{-1/2}\xi_{-1/2}'\right)=2\sigma^2\widetilde{P}\beta\beta'\widetilde{P}'W\widetilde{P}(\widetilde{X}\widetilde{X})^{-1}. \tag{A.17}$$

By employing the results of equations (A.11),(A.12) and (A.14), in equation (A.10), we obtain the MSEM of the average estimator up to order $O((NT)^{-2})$, as below

$$\begin{aligned} \mathrm{MSEM}(\widehat{\beta}_A) &= \mathrm{MSEM}(\widehat{\beta}) + \frac{\tau^2}{\widetilde{\phi}^2}\widetilde{P}\beta\beta'\widetilde{P}' - \frac{2\tau}{\widetilde{\phi}}\sigma^2\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}' \\ &\quad + \frac{2\tau}{\widetilde{\phi}^2}\sigma^2\Big[\widetilde{P}\beta\beta'\widetilde{P}'W\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}' + \widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}'W\widetilde{P}\beta\beta'\widetilde{P}'\Big], \end{aligned} \tag{A.18}$$

where the use has been made of

$$\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}' = \widetilde{P}(\widetilde{X}'\widetilde{X})^{-1} = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}'. \tag{A.19}$$

Further, the risk of the average estimator up to order $O((NT)^{-1})$, can be written as

$$\begin{aligned} \mathrm{Risk}(\widehat{\beta}_A) &= \mathbb{E}\left[(\widehat{\beta}_A-\beta)'W(\widehat{\beta}_A-\beta)\right] = \mathrm{tr}\left[W\,\mathbb{E}\left[(\widehat{\beta}_A-\beta)(\widehat{\beta}_A-\beta)'\right]\right] = \mathrm{tr}\left[W\,\mathrm{MSEM}(\widehat{\beta}_A)\right] \\ &= \mathrm{Risk}(\widehat{\beta}) + \frac{\tau^2}{\widetilde{\phi}} - \frac{2\tau}{\widetilde{\phi}}\sigma^2\,\mathrm{tr}\left[W\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}'\right] + \frac{4\tau}{\widetilde{\phi}^2}\sigma^2\Big[\beta'\widetilde{P}'W\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}'W\widetilde{P}\beta\Big]. \end{aligned} \tag{A.20}$$

■

Proof. Theorem 2 :

Note that, we have

$$\hat{\mu} = \hat{\sigma}^2\,\mathrm{tr}\left(\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}W\right) = \widetilde{\mu} + \mu_{-1/2} + O_p((NT)^{-1}), \tag{A.21}$$

where $\mu_{-1/2} = \mathrm{tr}\left(\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}W\right)(\hat{\sigma}^2-\sigma^2) = \frac{\widetilde{\mu}}{\sigma^2}(\hat{\sigma}^2-\sigma^2)$, and $\widetilde{\mu} = \sigma^2\,\mathrm{tr}\left(\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}W\right) = O(1)$.

Employing the results of (A.6) and (A.21) in equation (3.8), we obtain

$$\begin{aligned} \widehat{\beta}_M - \beta &= (\widehat{\beta}-\beta) - \Big[\widetilde{\mu} + \mu_{-1/2} + O_p((NT)^{-1})\Big]\Big[\frac{1}{\widetilde{\phi}} - \frac{1}{\widetilde{\phi}^2}D_{1/2} + O_p((NT)^{-2})\Big]\widetilde{P}\widehat{\beta} \\ &= \zeta_{-1/2} + \breve{\zeta}_{-1} + \breve{\zeta}_{-3/2} + O_p((NT)^{-2}), \end{aligned} \tag{A.22}$$

where $\zeta_{-1/2}, \breve{\zeta}_{-1}$ and $\breve{\zeta}_{-3/2}$ are defined below

$$\begin{aligned} \zeta_{-1/2} &= \xi_{-1/2} = O_p((NT)^{-1/2}), \\ \breve{\zeta}_{-1} &= -\frac{\widetilde{\mu}}{\widetilde{\phi}}\widetilde{P}\beta = O_p((NT)^{-1}), \\ \breve{\zeta}_{-3/2} &= -\frac{\widetilde{\mu}}{\widetilde{\phi}}\widetilde{P}\xi_{-1/2} + \frac{\widetilde{\mu}}{\widetilde{\phi}^2}D_{1/2}\widetilde{P}\beta - \frac{\mu_{-1/2}}{\widetilde{\phi}}\widetilde{P}\beta = O_p((NT)^{-3/2}), \end{aligned}$$

and $\xi_{-1/2}$ is defined in (A.5).

The bias of the weighted average MMSE estimator using equation (A.22) up to order $O((NT)^{-1})$

is

$$\mathbb{E}(\widehat{\beta}_M-\beta)=\mathbb{E}(\zeta_{-1/2}+\breve{\zeta}_{-1})=\mathbb{E}(\xi_{-1/2})-\frac{\widetilde{\mu}}{\widetilde{\phi}}\widetilde{P}\beta=-\frac{\widetilde{\mu}}{\widetilde{\phi}}\widetilde{P}\beta, \tag{A.23}$$

where the use has been made of equation (A.9).

The MSEM up to order $O((NT)^{-2})$ is

$$\mathbb{E}\left[(\widehat{\beta}_M-\beta)(\widehat{\beta}_M-\beta)'\right]=\mathbb{E}(\Gamma_{-1}+\breve{\Gamma}_{-3/2}+\breve{\Gamma}_{-2}), \tag{A.24}$$

where Γ_{-1}, $\breve{\Gamma}_{-3/2}$ and $\breve{\Gamma}_{-2}$ are

$$\begin{aligned}
\Gamma_{-1}&=\zeta_{-1/2}\zeta'_{-1/2},\\
\breve{\Gamma}_{-3/2}&=\zeta_{-1/2}\breve{\zeta}'_{-1}+\breve{\zeta}_{-1}\zeta'_{-1/2},\\
\breve{\Gamma}_{-2}&=\zeta_{-1/2}\breve{\zeta}'_{-3/2}+\breve{\zeta}_{-3/2}\zeta'_{-1/2}+\breve{\zeta}_{-1}\breve{\zeta}'_{-1},
\end{aligned}$$

and we give their expectations below

$$\mathbb{E}(\Gamma_{-1})=\mathbb{E}\left((\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{u}\widetilde{u}'\widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}\right)=\sigma^2(\widetilde{X}'\widetilde{X})^{-1}, \tag{A.25}$$

$$\mathbb{E}(\breve{\Gamma}_{-3/2})=\mathbb{E}(\zeta_{-1/2}\breve{\zeta}'_{-1})+\mathbb{E}(\breve{\zeta}_{-1}\zeta'_{-1/2})=0, \tag{A.26}$$

because

$$\mathbb{E}(\breve{\zeta}_{-1}\zeta'_{-1/2})=\mathbb{E}\left[-\frac{\widetilde{\mu}}{\widetilde{\phi}}\widetilde{P}\beta\xi'_{-1/2}\right]=-\frac{\widetilde{\mu}}{\widetilde{\phi}}\widetilde{P}\beta\,\mathbb{E}(\xi_{-1/2})=0, \tag{A.27}$$

where the last equality holds by (A.9). Also, we have

$$\begin{aligned}
\mathbb{E}(\breve{\Gamma}_{-2})&=\mathbb{E}(\zeta_{-1/2}\breve{\zeta}'_{-3/2})+\mathbb{E}(\breve{\zeta}_{-3/2}\zeta'_{-1/2})+\mathbb{E}(\breve{\zeta}_{-1}\breve{\zeta}'_{-1})\\
&=\frac{\widetilde{\mu}^2}{\widetilde{\phi}^2}\widetilde{P}\beta\beta'\widetilde{P}'-\frac{\widetilde{\mu}}{\widetilde{\phi}}\sigma^2\Big[\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}+(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}'\Big]\\
&\quad+\frac{2\widetilde{\mu}}{\widetilde{\phi}^2}\sigma^2\Big[\widetilde{P}\beta\beta'\widetilde{P}'W\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}+(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}'W\widetilde{P}\beta\beta'\widetilde{P}'\Big],
\end{aligned} \tag{A.28}$$

where the last equality above holds by using equations (A.29) and (A.30) below

$$\mathbb{E}(\breve{\zeta}_{-1}\breve{\zeta}'_{-1})=\frac{\widetilde{\mu}^2}{\widetilde{\phi}^2}\widetilde{P}\beta\beta'\widetilde{P}', \tag{A.29}$$

and

$$\mathbb{E}(\breve{\zeta}_{-3/2}\zeta'_{-1/2}) = -\frac{\widetilde{\mu}}{\widetilde{\phi}}\widetilde{P}\,\mathbb{E}(\xi_{-1/2}\xi'_{-1/2}) + \frac{\widetilde{\mu}}{\widetilde{\phi}^2}\,\mathbb{E}(D_{1/2}\widetilde{P}\beta\xi'_{-1/2}) - \mathbb{E}(\frac{\mu_{-1/2}}{\widetilde{\phi}}\widetilde{P}\beta\xi'_{-1/2})$$
$$= -\frac{\widetilde{\mu}}{\widetilde{\phi}}\sigma^2\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1} + \frac{2\widetilde{\mu}}{\widetilde{\phi}^2}\sigma^2\widetilde{P}\beta\beta'\widetilde{P}'W\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}, \tag{A.30}$$

where the last equality holds by using (A.17), and

$$\mathbb{E}(\mu_{-1/2}\xi_{-1/2}) = \frac{\widetilde{\mu}}{\sigma^2}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\,\mathbb{E}\left(\widetilde{u}\,(\hat{\sigma}^2 - \sigma^2)\right) = 0, \tag{A.31}$$

because of the normality of the errors.

By employing the results of equations (A.25),(A.26) and (A.28), in equation (A.24), we obtain the MSEM of the estimator up to order $O((NT)^{-2})$, as below

$$\mathrm{MSEM}(\widehat{\beta}_M) = \mathrm{MSEM}(\widehat{\beta}) + \frac{\widetilde{\mu}^2}{\widetilde{\phi}^2}\widetilde{P}\beta\beta'\widetilde{P}' - \frac{2\widetilde{\mu}}{\widetilde{\phi}}\sigma^2\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}' + \frac{2\widetilde{\mu}}{\widetilde{\phi}^2}\sigma^2\Big[\widetilde{P}\beta\beta'\widetilde{P}'W\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}'$$
$$+ \widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}'W\widetilde{P}\beta\beta'\widetilde{P}'\Big], \tag{A.32}$$

where the use has been made of equation (A.19). Further, the risk of the estimator up to order $O((NT)^{-1})$, can be written as

$$\mathrm{Risk}(\widehat{\beta}_M) = \mathbb{E}\left[(\widehat{\beta}_M - \beta)'W\widehat{\beta}_M - \beta)\right] = \mathrm{tr}\left[W\,\mathbb{E}\left[(\widehat{\beta}_M - \beta)(\widehat{\beta}_M - \beta)'\right]\right] = \mathrm{tr}\left[W\,\mathrm{MSEM}(\widehat{\beta}_M)\right]$$
$$= \mathrm{Risk}(\widehat{\beta}) + \frac{\widetilde{\mu}^2}{\widetilde{\phi}} - \frac{2\widetilde{\mu}}{\widetilde{\phi}}\sigma^2\,\mathrm{tr}\left[W\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}'\right] + \frac{4\widetilde{\mu}}{\widetilde{\phi}^2}\sigma^2\beta'\widetilde{P}'W\widetilde{P}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{P}'W\widetilde{P}\beta. \tag{A.33}$$

■

Proof. Theorem 3 :

Note that

$$\hat{\lambda} = \lambda + \frac{\lambda f}{\sqrt{N}} + O_p(N^{-1}), \tag{A.34}$$

where

$$f = \frac{1}{\sqrt{N}\sigma^2}u'\Big(\frac{M_D}{T-1} - \lambda Z_D\Big)u - (1-\lambda)\Big(\frac{\alpha'\alpha}{\sqrt{N}\sigma_\alpha^2}\Big) - \frac{2}{\sqrt{N}\sigma_\eta^2}u'D\alpha, \tag{A.35}$$

for a proof see Ullah and Huang (2006). Hence, we have

$$\Delta = \widehat{\Omega} - \Omega = Z_D(\hat{\lambda}^{-1} - \lambda^{-1}) = -\frac{1}{\sqrt{N}\lambda} Z_D f + o_p(N^{-1/2}), \tag{A.36}$$

where the last equality holds by expansion of the inverse of $\hat{\lambda}^{-1}$.

Using the results of Lemma A.1, in equation (3.13), we have

$$\widehat{\beta} - \beta = (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}\epsilon = \xi_{-1/2} + \xi_{-1} + \xi_{-3/2} + O_p(N^{-2}), \tag{A.37}$$

where $\xi_{-1/2}$, $\xi_{-1/2}$, and $\xi_{-3/2}$ are defined below, and the suffixes show the order of magnitude in probability,

$$\begin{aligned}
\xi_{-1/2} &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon = O_p(N^{-1/2}),\\
\xi_{-1} &= -(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta Q\epsilon = O_p(N^{-1}),\\
\xi_{-3/2} &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta Q\Delta Q\epsilon = O_p(N^{-3/2})
\end{aligned}$$

and $Q = \Omega^{-1} - \Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$.

Using equation (A.37) in equation (3.17), we have

$$\begin{aligned}
\frac{1}{\widehat{D}} &= \left[(\widehat{\beta} - \widetilde{\beta})'W(\widehat{\beta} - \widetilde{\beta})\right]^{-1} = \left[\widehat{\beta}'\,\widehat{P}'\,W\,\widehat{P}\,\widehat{\beta}\right]^{-1}\\
&= \left[\Big(\beta + \xi_{-1/2} + O_p(N^{-1})\Big)'\left[P + P_{-1/2} + O_p(N^{-1})\right]'W\left[P + P_{-1/2} + O_p(N^{-1})\right]\Big(\beta + \xi_{-1/2} + O_p(N^{-1})\Big)\right]^{-1}\\
&= \left[\phi + 2\beta'P'WP\xi_{-1/2} + 2\beta'P'WP_{-1/2}\,\beta + O_p(1)\right]^{-1}\\
&= \frac{1}{\phi}\left[1 + \frac{2}{\phi}\beta'P'WP\xi_{-1/2} + \frac{2}{\phi}\beta'P'WP_{-1/2}\,\beta + O_p(N^{-1})\right]^{-1}\\
&= \frac{1}{\phi}\left[1 - \frac{2}{\phi}\beta'P'WP\xi_{-1/2} - \frac{2}{\phi}\beta'P'WP_{-1/2}\,\beta\right] + O_p(N^{-2})\\
&\equiv \underbrace{\frac{1}{\phi}}_{O_p(N^{-1})} - \underbrace{\frac{1}{\phi^2}D_{1/2}}_{O_p(N^{-3/2})} + O_p(N^{-2}),
\end{aligned} \tag{A.38}$$

where $D_{1/2} = 2\Big[\beta'P'WP\xi_{-1/2} + \beta'P'WP_{-1/2}\,\beta\Big] = O_p(N^{1/2})$, $\phi = \beta'P'WP\beta = O(N)$, and the last equality above holds by using the standard geometric expansion. Also, the use has been made of equations (A.1)–(A.4). The terms with order $O_p(N^{-2})$ and smaller are dropped, because they will not enter in the calculation of the bias and MSEM of the average estimator up to the orders of interest.

Employing equations (A.4)–(A.38) in equation (3.15), we obtain

$$\begin{aligned}\widehat{\beta}_A - \beta &= (\widehat{\beta} - \beta) - \tau\Big[\frac{1}{\phi} - \frac{1}{\phi^2}D_{1/2} + O_p(N^{-2})\Big]\Big[P + P_{-1/2} + O_p(N^{-1})\Big]\widehat{\beta} \\ &= \zeta_{-1/2} + \zeta_{-1} + \zeta_{-3/2} + O_p(N^{-2}),\end{aligned} \tag{A.39}$$

where $\zeta_{-1/2}, \zeta_{-1}$ and $\zeta_{-3/2}$ are defined below

$$\begin{aligned}&\zeta_{-1/2} = \xi_{-1/2} = O_p(N^{-1/2}),\\ &\zeta_{-1} = \xi_{-1} - \frac{\tau}{\phi}P\beta = O_p(N^{-1}),\\ &\zeta_{-3/2} = \xi_{-3/2} - \frac{\tau}{\phi}P\xi_{-1/2} - \frac{\tau}{\phi}P_{-1/2}\beta + \frac{\tau}{\phi^2}D_{1/2}P\beta = O_p(N^{-3/2}).\end{aligned}$$

The bias of the average estimator using the approximations in equation (A.39) up to order $O(N^{-1})$ is

$$\mathbb{E}(\widehat{\beta}_A - \beta) = \mathbb{E}(\zeta_{-1/2} + \zeta_{-1}) = \mathbb{E}(\xi_{-1/2}) + \mathbb{E}(\xi_{-1}) - \frac{\tau}{\phi}P\beta = -\frac{\tau}{\phi}P\beta, \tag{A.40}$$

where the last equality holds because $\mathbb{E}(\epsilon) = 0$, then

$$\mathbb{E}(\xi_{-1/2}) = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\,\mathbb{E}(\epsilon) = 0, \tag{A.41}$$

and by the normality of the errors

$$\mathbb{E}(\xi_{-1}) = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\,\mathbb{E}(\Delta Q\epsilon) = 0. \tag{A.42}$$

The MSEM up to order $O(N^{-2})$ is

$$\mathbb{E}\left[(\widehat{\beta}_A - \beta)(\widehat{\beta}_A - \beta)'\right] = \mathbb{E}(\Gamma_{-1} + \Gamma_{-3/2} + \Gamma_{-2}), \tag{A.43}$$

where Γ_{-1}, $\Gamma_{-3/2}$ and Γ_{-2} are

$$\begin{aligned}&\Gamma_{-1} = \zeta_{-1/2}\zeta'_{-1/2},\\ &\Gamma_{-3/2} = \zeta_{-1/2}\zeta'_{-1} + \zeta_{-1}\zeta'_{-1/2},\\ &\Gamma_{-2} = \zeta_{-1/2}\zeta'_{-3/2} + \zeta_{-3/2}\zeta'_{-1/2} + \zeta_{-1}\zeta'_{-1},\end{aligned}$$

and we give their expectations below

$$\mathbb{E}(\Gamma_{-1}) = \mathbb{E}\left((X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon\epsilon'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}\right) = \sigma^2(X'\Omega^{-1}X)^{-1}, \tag{A.44}$$

$$\mathbb{E}(\Gamma_{-3/2}) = \mathbb{E}(\zeta_{-1/2}\zeta'_{-1}) + \mathbb{E}(\zeta_{-1}\zeta'_{-1/2}) = \mathbb{E}(\xi_{-1}\xi'_{-1/2}) + \mathbb{E}(\xi_{-1/2}\xi'_{-1}), \tag{A.45}$$

because

$$\mathbb{E}(\zeta_{-1}\zeta'_{-1/2}) = \mathbb{E}\left[\Big(\xi_{-1} - \frac{\tau}{\phi}P\beta\Big)\xi'_{-1/2}\right] = \mathbb{E}(\xi_{-1}\xi'_{-1/2}) - \frac{\tau}{\phi}P\beta\,\mathbb{E}(\xi_{-1/2}) = \mathbb{E}(\xi_{-1}\xi'_{-1/2}), \tag{A.46}$$

where the last equality holds by (A.41). Also, we have

$$\begin{aligned}\mathbb{E}(\Gamma_{-2}) &= \mathbb{E}(\zeta_{-1/2}\zeta'_{-3/2}) + \mathbb{E}(\zeta_{-3/2}\zeta'_{-1/2}) + \mathbb{E}(\zeta_{-1}\zeta'_{-1})\\ &= \mathbb{E}(\xi_{-3/2}\xi'_{-1/2}) + \mathbb{E}(\xi_{-1/2}\xi'_{-3/2}) + \mathbb{E}(\xi_{-1}\xi'_{-1}) + \frac{\tau^2}{\phi^2}P\beta\beta'P'\\ &- \frac{\tau}{\phi}\sigma^2\Big[P(X'\Omega^{-1}X)^{-1} + (X'\Omega^{-1}X)^{-1}P'\Big]\\ &+ \frac{2\tau}{\phi^2}\sigma^2\Big[P\beta\beta'P'WP(X'\Omega^{-1}X)^{-1} + (X'\Omega^{-1}X)^{-1}P'WP\beta\beta'P'\Big],\end{aligned} \tag{A.47}$$

where the last equality above holds by using equations (A.48) and (A.49) below

$$\begin{aligned}\mathbb{E}(\zeta_{-1}\zeta'_{-1}) &= \mathbb{E}(\xi_{-1}\xi'_{-1}) - \frac{\tau}{\phi}\,\mathbb{E}(\xi_{-1}\beta'P') - \frac{\tau}{\phi}\,\mathbb{E}(P\beta\xi'_{-1}) + \frac{\tau^2}{\phi^2}P\beta\beta'P'\\ &= \mathbb{E}(\xi_{-1}\xi'_{-1}) + \frac{\tau^2}{\phi^2}P\beta\beta'P',\end{aligned} \tag{A.48}$$

and

$$\begin{aligned}\mathbb{E}(\zeta_{-3/2}\zeta'_{-1/2}) &= \mathbb{E}(\xi_{-3/2}\xi'_{-1/2}) - \frac{\tau}{\phi}P\,\mathbb{E}(\xi_{-1/2}\xi'_{-1/2}) - \frac{\tau}{\phi}\,\mathbb{E}(P_{-1/2}\beta\xi'_{-1/2}) + \frac{\tau}{\phi^2}\,\mathbb{E}(D_{1/2}P\beta\xi'_{-1/2})\\ &= \mathbb{E}(\xi_{-3/2}\xi'_{-1/2}) - \frac{\tau}{\phi}\sigma^2P(X'\Omega^{-1}X)^{-1} + \frac{2\tau}{\phi^2}\sigma^2P\beta\beta'P'WP(X'\Omega^{-1}X)^{-1},\end{aligned} \tag{A.49}$$

and the last equality holds by using

$$\begin{aligned}\mathbb{E}(D_{1/2}P\beta\xi'_{-1/2}) &= 2P\beta\beta'P'W\,\mathbb{E}\left[\Big(P\xi_{-1/2} + P_{-1/2}\,\beta\Big)\xi_{-1/2}\right]\\ &= 2\sigma^2P\beta\beta'P'WP(X'\Omega^{-1}X)^{-1} + 2P\beta\beta'P'W\,\mathbb{E}(P_{-1/2}\beta\xi'_{-1/2}),\end{aligned} \tag{A.50}$$

and

$$\begin{aligned}\mathbb{E}(P_{-1/2}\beta\xi'_{-1/2}) &= \mathbb{E}\left[(I_k - P)(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}XP\beta\epsilon'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}\right]\\ &= -\frac{1}{\sqrt{N}\lambda}\,\mathbb{E}\left[(I_k - P)(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Z_Df\Omega^{-1}XP\beta\epsilon'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}\right]\\ &= -\frac{1}{\sqrt{N}\lambda}(I_k - P)(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Z_D\Omega^{-1}XP\beta\,\mathbb{E}\left[f\epsilon'\right]\Omega^{-1}X(X'\Omega^{-1}X)^{-1}\\ &= 0,\end{aligned}$$

which holds by the normality of the errors.

By employing the results of equations (A.44),(A.45) and (A.47), in equation (A.43), we obtain the MSEM of the average estimator up to order $O(N^{-2})$, as below

$$\begin{aligned}\mathrm{MSEM}(\widehat{\beta}_A) &= \mathrm{MSEM}(\widehat{\beta}) + \frac{\tau^2}{\phi^2}P\beta\beta'P' - \frac{\tau}{\phi}2\sigma^2P(X'\Omega^{-1}X)^{-1}P' \\ &\quad + \frac{2\tau}{\phi^2}\sigma^2\Big[P\beta\beta'P'WP(X'\Omega^{-1}X)^{-1}P' + P(X'\Omega^{-1}X)^{-1}P'WP\beta\beta'P'\Big],\end{aligned} \tag{A.51}$$

where the use has been made of

$$P(X'\Omega^{-1}X)^{-1}P' = P(X'\Omega^{-1}X)^{-1} = (X'\Omega^{-1}X)^{-1}P'. \tag{A.52}$$

Further, the risk of the average estimator up to order $O(N^{-1})$, can be written as

$$\begin{aligned}\mathrm{Risk}(\widehat{\beta}_A) &= \mathbb{E}\Big[(\widehat{\beta}_A-\beta)'W(\widehat{\beta}_A-\beta)\Big] = \mathrm{tr}\Big[W\,\mathbb{E}\Big[(\widehat{\beta}_A-\beta)(\widehat{\beta}_A-\beta)'\Big]\Big] = \mathrm{tr}\Big[W\,\mathrm{MSEM}(\widehat{\beta}_A)\Big] \\ &= \mathrm{Risk}(\widehat{\beta}) + \frac{\tau^2}{\phi} - \frac{2\tau\sigma^2}{\phi}\,\mathrm{tr}\Big[WP(X'\Omega^{-1}X)^{-1}P'\Big] + \frac{4\tau\sigma^2}{\phi^2}\Big[\beta'P'WP(X'\Omega^{-1}X)^{-1}P'WP\beta\Big].\end{aligned} \tag{A.53}$$

■

Proof. Theorem 4 :

Using the results of Lemma A.1, we have

$$\hat{\mu} = \hat{\sigma}^2\,\mathrm{tr}\Big(\widehat{P}(X'\widehat{\Omega}^{-1}X)^{-1}W\Big) = \mu + \mu_{-1/2} + O_p(N^{-1}), \tag{A.54}$$

where $\mu_{-1/2} = \sigma^2\,\mathrm{tr}\Big(P(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}X(X'\Omega^{-1}X)^{-1}W + P_{-1/2}(X'\Omega^{-1}X)^{-1}W\Big) + \sigma^2\nu_\epsilon = O_p(N^{-1/2})$, $\mu = \sigma^2\,\mathrm{tr}\Big(P(X'\Omega^{-1}X)^{-1}P'W\Big) = \mathrm{tr}\Big(P(X'\bar{\Omega}^{-1}X)^{-1}P'W\Big) = O(1)$, and $\nu_\epsilon = \Big(\epsilon'M_D\epsilon/\sigma^2N(T-1)\Big) - 1$.

Employing the results of (A.4), (A.38) and (A.54) in equation (3.16), we obtain

$$\begin{aligned}\widehat{\beta}_M - \beta &= (\widehat{\beta}-\beta) - \Big[\mu + \mu_{-1/2} + O_p(N^{-1})\Big]\Big[\frac{1}{\phi} - \frac{1}{\phi^2}D_{1/2} + O_p(N^{-2})\Big]\Big[P + P_{-1/2} + O_p(N^{-1})\Big]\widehat{\beta} \\ &= \zeta_{-1/2} + \breve{\zeta}_{-1} + \breve{\zeta}_{-3/2} + O_p(N^{-2}),\end{aligned} \tag{A.55}$$

where $\zeta_{-1/2}, \breve{\zeta}_{-1}$ and $\breve{\zeta}_{-3/2}$ are defined below

$$\begin{aligned}\zeta_{-1/2} &= \xi_{-1/2} = O_p(N^{-1/2}), \\ \breve{\zeta}_{-1} &= \xi_{-1} - \frac{\mu}{\phi}P\beta = O_p(N^{-1}), \\ \breve{\zeta}_{-3/2} &= \xi_{-3/2} - \frac{\mu}{\phi}P\xi_{-1/2} - \frac{\mu}{\phi}P_{-1/2}\beta + \frac{\mu}{\phi^2}D_{1/2}P\beta - \frac{\mu_{-1/2}}{\phi}P\beta = O_p(N^{-3/2}),\end{aligned}$$

and $\xi_{-1/2}$ is given in equation (A.37).

The bias of the weighted average MMSE estimator using equation (A.55) up to order $O(N^{-1})$ is

$$\mathbb{E}(\widehat{\beta}_M - \beta) = \mathbb{E}(\zeta_{-1/2} + \breve{\zeta}_{-1}) = \mathbb{E}(\xi_{-1/2}) + \mathbb{E}(\xi_{-1}) - \frac{\mu}{\phi} P\beta = -\frac{\mu}{\phi} P\beta, \tag{A.56}$$

where the use has been made of equation (A.41) and (A.42).

The MSEM up to order $O(N^{-2})$ is

$$\mathbb{E}\left[(\widehat{\beta}_M - \beta)(\widehat{\beta}_M - \beta)'\right] = \mathbb{E}(\Gamma_{-1} + \breve{\Gamma}_{-3/2} + \breve{\Gamma}_{-2}), \tag{A.57}$$

where Γ_{-1}, $\breve{\Gamma}_{-3/2}$ and $\breve{\Gamma}_{-2}$ are

$$\begin{aligned}
\Gamma_{-1} &= \zeta_{-1/2}\zeta'_{-1/2},\\
\breve{\Gamma}_{-3/2} &= \zeta_{-1/2}\breve{\zeta}'_{-1} + \breve{\zeta}_{-1}\zeta'_{-1/2},\\
\breve{\Gamma}_{-2} &= \zeta_{-1/2}\breve{\zeta}'_{-3/2} + \breve{\zeta}_{-3/2}\zeta'_{-1/2} + \breve{\zeta}_{-1}\breve{\zeta}'_{-1},
\end{aligned}$$

and we give their expectations below

$$\mathbb{E}(\Gamma_{-1}) = \mathbb{E}\left((X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon\epsilon'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}\right) = \sigma^2(X'\Omega^{-1}X)^{-1}, \tag{A.58}$$

$$\mathbb{E}(\breve{\Gamma}_{-3/2}) = \mathbb{E}(\zeta_{-1/2}\breve{\zeta}'_{-1}) + \mathbb{E}(\breve{\zeta}_{-1}\zeta'_{-1/2}) = \mathbb{E}(\xi_{-1}\xi'_{-1/2}) + \mathbb{E}(\xi_{-1/2}\xi'_{-1}), \tag{A.59}$$

because

$$\mathbb{E}(\breve{\zeta}_{-1}\zeta'_{-1/2}) = \mathbb{E}\left[\left(\xi_{-1} - \frac{\mu}{\phi}P\beta\right)\xi'_{-1/2}\right] = \mathbb{E}(\xi_{-1}\xi'_{-1/2}) - \frac{\mu}{\phi}P\beta\,\mathbb{E}(\xi_{-1/2}) = \mathbb{E}(\xi_{-1}\xi'_{-1/2}), \tag{A.60}$$

where the last equality holds by (A.41). Also, we have

$$\begin{aligned}
\mathbb{E}(\breve{\Gamma}_{-2}) &= \mathbb{E}(\zeta_{-1/2}\breve{\zeta}'_{-3/2}) + \mathbb{E}(\breve{\zeta}_{-3/2}\zeta'_{-1/2}) + \mathbb{E}(\breve{\zeta}_{-1}\breve{\zeta}'_{-1})\\
&\qquad = \mathbb{E}(\xi_{-3/2}\xi'_{-1/2}) + \mathbb{E}(\xi_{-1/2}\xi'_{-3/2}) + \mathbb{E}(\xi_{-1}\xi'_{-1}) + \frac{\mu^2}{\phi^2}P\beta\beta'P'\\
&\qquad - \frac{\mu}{\phi}\sigma^2\Big[P(X'\Omega^{-1}X)^{-1} + (X'\Omega^{-1}X)^{-1}P'\Big]\\
&\qquad + \frac{2\mu}{\phi^2}\sigma^2\Big[P\beta\beta'P'WP(X'\Omega^{-1}X)^{-1} + (X'\Omega^{-1}X)^{-1}P'WP\beta\beta'P'\Big],
\end{aligned} \tag{A.61}$$

where the last equality above holds by using equations (A.62) and (A.63) below

$$\begin{aligned}
\mathbb{E}(\breve{\zeta}_{-1}\breve{\zeta}'_{-1}) &= \mathbb{E}(\xi_{-1}\xi'_{-1}) - \frac{\mu}{\phi}\mathbb{E}(\xi_{-1})\beta'P' - \frac{\mu}{\phi}P\beta\,\mathbb{E}(\xi'_{-1}) + \frac{\mu^2}{\phi^2}P\beta\beta'P'\\
&\qquad = \mathbb{E}(\xi_{-1}\xi'_{-1}) + \frac{\mu^2}{\phi^2}P\beta\beta'P',
\end{aligned} \tag{A.62}$$

and

$$\mathbb{E}(\breve{\zeta}_{-3/2}\zeta'_{-1/2}) = \mathbb{E}(\xi_{-3/2}\xi'_{-1/2}) - \frac{\mu}{\phi}P\,\mathbb{E}(\xi_{-1/2}\xi'_{-1/2}) - \frac{\mu}{\phi}\,\mathbb{E}(P_{-1/2}\beta\xi'_{-1/2}) + \frac{\mu}{\phi^2}\,\mathbb{E}(D_{1/2}P\beta\xi'_{-1/2})$$
$$- \mathbb{E}(\frac{\mu_{-1/2}}{\phi}P\beta\xi'_{-1/2}) = \mathbb{E}(\xi_{-3/2}\xi'_{-1/2}) - \frac{\mu}{\phi}\sigma^2 P(X'\Omega^{-1}X)^{-1}$$
$$+ \frac{2\mu}{\phi^2}\sigma^2 P\beta\beta'P'WP(X'\Omega^{-1}X)^{-1}, \tag{A.63}$$

where the last equality holds by using (A.50), and

$$\mathbb{E}(\mu_{-1/2}\xi_{-1/2}) = -\frac{\sigma^2}{\sqrt{N}\lambda}\breve{\mu}(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\,\mathbb{E}(\epsilon f) + \sigma^2(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\,\mathbb{E}(\epsilon\nu_\epsilon) = 0,$$

and

$$\breve{\mu} = \mathrm{tr}\Big(P(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Z_D\Omega^{-1}X(X'\Omega^{-1}X)^{-1}W$$
$$+ (I_k - P)(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Z_D\Omega^{-1}XP(X'\Omega^{-1}X)^{-1}W\Big).$$

By employing the results of equations (A.58),(A.59) and (A.61), in equation (A.57), we obtain the MSEM of the estimator up to order $O(N^{-2})$, as below

$$\mathrm{MSEM}(\widehat{\beta}_M) = \mathrm{MSEM}(\widehat{\beta}) + \frac{\mu^2}{\phi^2}P\beta\beta'P' - 2\frac{\mu}{\phi}\sigma^2 P(X'\Omega^{-1}X)^{-1}P'$$
$$+ \frac{2\mu}{\phi^2}\sigma^2\Big[P\beta\beta'P'WP(X'\Omega^{-1}X)^{-1}P' + P(X'\Omega^{-1}X)^{-1}P'WP\beta\beta'P'\Big], \tag{A.64}$$

where the use has been made of equation (A.52). Further, the risk of the estimator up to order $O(N^{-1})$, can be written as

$$\mathrm{Risk}(\widehat{\beta}_M) = \mathbb{E}\Big[(\widehat{\beta}_M - \beta)'W\widehat{\beta}_M - \beta)\Big] = \mathrm{tr}\Big[W\,\mathbb{E}\Big[(\widehat{\beta}_M - \beta)(\widehat{\beta}_M - \beta)'\Big]\Big] = \mathrm{tr}\Big[W\,\mathrm{MSEM}(\widehat{\beta}_M)\Big]$$
$$= \mathrm{Risk}(\widehat{\beta}) + \frac{\mu^2}{\phi} - \frac{2\mu}{\phi}\sigma^2\,\mathrm{tr}\Big[WP(X'\Omega^{-1}X)^{-1}P'\Big]$$
$$+ \frac{4\mu}{\phi^2}\sigma^2\beta'P'WP(X'\Omega^{-1}X)^{-1}P'WP\beta. \tag{A.65}$$

■