Efficient Shrinkage Estimation in Heterogeneous Panel Data Models with Application to Returns to Scale for U.S. Banks

Ali Mehrabani [†]

February 20, 2020

Abstract

This paper presents a Stein-like shrinkage method for estimating the slope coefficients in heterogeneous panel data models with cross-section dependence, when the cross-section dimension is fixed while the time dimension is allowed to increase without bounds. The shrinkage estimator is a weighted average of a feasible generalized least-squares (FGLS) estimator and a feasible restricted generalized least-squares estimator. The restricted estimator belongs to a set of restricted parameter space, where the restrictions represent possible model specifications. The shrinkage weight is inversely proportional to a Wald statistic that measures the importance of the restrictions. The asymptotic and higher-order approximations of bias, and mean squared error of the shrinkage estimator are given. It is shown that the shrinkage estimator is robust and uniformly superior in terms of asymptotic risks, relative to both the FGLS and the restricted estimators. Additionally, the shrinkage estimator achieves the lowest possible asymptotic risk in a high-dimensional large sample framework. A major advantage of this shrinkage method is that it is generalized to allow for the limitations of the existing model averaging techniques.

As an empirical illustration, the method is applied to estimate bank cost efficiency based on a cost system for U.S. commercial banks. This methodology has two major advantages over the existing studies: it considers the classification uncertainty about bank cost functions, and allows for general correlation patterns across the errors in the cross-section cost equations. The estimation results indicate that over the sample period, most U.S. banks faced increasing returns to scale, and small banks (with assets less than 100 million dollars) are more cost efficient than the other banks. Consequently, scale economies are a possible reason for bank mergers and that the tendency toward growth in average bank size is likely to continue.

Key Words: *Shrinkage estimator; Stein-like; High dimension; Panel data models; Higher order approximations.*

[†]Department of Economics, University of California, Riverside, CA 92521. E-mail: ali.mehrabani@email.ucr.edu

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 address 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 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, b), most importantly that the model selection and estimation are completely separated such that the uncertainty due to the initial model selection step is ignored. As shown by Magnus, 2002 and Leeb and Pötscher, 2003, 2006, among others, 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. Model averaging methods are distinct in two main strands based on whether the estimation of each candidate model and the choice of the associated weighting scheme are developed along frequentist or Bayesian paradigms. Shrinkage estimation methods, similar to model averaging, allow for uncertainty emerging from both model selection and estimation (see Hansen, 2014, 2016). In addition, as shown by Hansen, 2016, Stein-type shrinkage estimation methods, unlike recent model averaging techniques (such as focused information criterion of Claeskens and Hjort, 2003, the plug-in estimator of Liu, 2015, and the focused moment selection criterion of DiTraglia, 2016), have the minimax efficiency properties.

This paper investigates a Stein-like shrinkage estimation method in linear heterogeneous panel data models to deal with uncertainty issues about the slope parameters. We allow for cross-section dependence and to estimate the contemporaneous error covariances freely, it is assumed that the cross-section dimension is small and the time series dimension is large. The shrinkage estimator shrinks a feasible generalized least-squares (FGLS) estimator (the standard approach in this setup, see Zellner, 1962) 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 Wald statistic that measures the weighted distance of the FGLS estimator and the restricted estimator. The first and second moments, and risk of our proposed estimator are derived using both finite sample higher-order approximations and the asymptotic theory. Furthermore, we show the dominance properties of the Stein-like shrinkage estimator in terms of risk, which ensures that our proposed estimator is robust against arbitrary deviations from the restrictions. This is another advantage of our method relative to the “local asymptotic” argument that some previous studies rely on (see for example, Hansen, 2016; Claeskens and Hjort, 2003; Liu, 2015).

Another advantage of the shrinkage method considered in this paper is that, unlike most of the existing model averaging methods, it allows for cross-section dependence of errors. These cross correlations could be due to omitted common effects, spatial effects, or could arise as a result of interactions within socioeconomic networks. In addition, the presence of some forms of cross-sectional correlation of errors in panel data applications in economics is likely to be the rule rather than the exception. Ignoring the cross correlations can have serious consequences such that conventional panel estimators can result in misleading inference and even in inconsistent estimators, depending on the extent of the cross-sectional dependence and on whether the source generating the cross-sectional dependence (such as an unobserved common shock) is correlated with regressors (Phillips and Sul, 2003, 2007;Andrews, 2005; Sarafidis and Robertson, 2009, and see a survey by Chudick and Pesaran, 2015).

In Monte Carlo simulations, we compare the small sample performance of our shrinkage estimator with the FGLS estimator and a restricted estimator where the restrictions impose slope parameters homogeneity across cross-sections. The results show that the shrinkage estimator generally produces smaller risk than the restricted estimator and the FGLS estimator. As an empirical illustration, we apply our estimator to estimate cost efficiency of U.S. Commercial banks using a cost system method over period from 2000 to 2018. Since bank size is an important factor of production environment, a common approach in estimating banks cost functions is separating them into asset size classes. However, there is no industry standard to categorize banks, and existing studies classify banks in one, three, or more (twelve) groups. To address this model uncertainty issue, we apply our shrinkage method to estimate the cost efficiency of banks. Besides, our approach allows for general correlation patters across the errors in the different cross-section cost equations, which could arise as a result of interactions within socioeconomic network, and to our best knowledge, is not considered in the pervious studies. Due to bank mergers and changes in structure of bank technologies over time, we estimate cost efficiency for each asset size class independently across sample period. Hence, the time dimension in our panel model is considered as the observations for banks operating within each group. Our empirical results suggest that on average majority of banks have been operating under increasing returns to scale over the sample period. We find that on average all banks have increasing returns to scale. The results also suggest that small banks (banks with asset size less than $100 Million) are more cost efficient than the other banks. This is consistent with economic theory which suggests that scale economies tend to decline with increase in size. This finding is important for gauging the costs and benefits of policy interventions intended to control the size of banks for the purpose of ensuring competitive markets, reducing the number of “too big to fail” banks, or others.

The literature on shrinkage 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, its positive part (Baranchick, 1964). Judge and Bock, 1978 developed this method for most of econometric estimators. Recently, Maddala et al., 2001 and Hansen, 2016 use shrinkage estimation methods to deal with model uncertainty in the context of two models. The shrinkage estimator in this paper is similar to that of Hansen, 2016 and Maddala et al., 2001. The main difference is that the shrinkage weight in Hansen, 2016 is inversely related to a weighted quadratic loss function, hence is subject to rotations of the coefficient vector, unless investigators are interested in minimizing a mean squared forecast (prediction) error (MSFE). However, the one considered in this paper is proportional to a Wald statistic which is an excellent choice as it is invariant to these rotations. Also, Hansen, 2016 considers a homoscedastic likelihood framework, but this paper considers linear panel data models and allows for both heteroscedasticity and cross-section dependence. The difference between the method used here and the one in Maddala et al., 2001 is that they use small-disturbance approximations to study the performance of their estimator, which cannot be applied to a model with unknown cross-section dependence and variance-heteroscedasticity considered in this paper.

Penalized methods are alternatives to shrinkage estimations for dealing with the uncertainty of covariate selection in regression models, which is arguably the most pervasive situation in economics. Methods that simultaneously select variables and shrink coefficients by minimizing some penalized loss function include, among others, the least absolute shrinkage and selection operator (LASSO) of Tibshirani, 1996, the smoothly clipped absolute deviation (SCAD) penalty of Fan and Li, 2001, and the minimax concave penalty (MCP) of Zhang, 2010. LASSO-type methods have been shown to be particularly effective in high-dimensional settings with a true small-dimensional structure, or when the number of predictors exceeds the sample size (see, e.g., Fan and Lv, 2010; Chernozhukov et al., 2015; Belloni et al., 2017). However, shrinkage methods do not exploit sparsity, and can work well even when there are many non-zero but small parameters.

This paper is also related to a long-existing issue in the panel data analysis, referred by econometricians to as “to pool or not to pool ”, on which there is still no consensus. The issue is on how to model potential parameter heterogeneity across individual units. On one hand, parameter heterogeneity results in consistent estimators and violation of this assumption causes misleading estimates, see, for example, Robertson and Symons, 1992, Pesaran and Smith, 1995, Su and Chen, 2013, Durlauf et al., 2001, and Browning and Carro, 2007. On the other hand, parameter homogeneity causes higher variance efficiency, but at the cost of estimation bias and inconsistency of the associated estimators, which is supported by an increasing number of studies due to a better forecast performance of these estimators, see, for example, Maddala, 1991, Maddala and Hu, 1996, Baltagi and Griffin, 1984, Baltagi et al., 2000, and Hoogstrate et al., 2000. This shows the typical bias-variance trade-off that needs to be considered in choosing parameter specifications. In the literature, there are several ways to address this parameter heterogeneity such as the random coefficient model of Swamy, 1970, the pooled mean group estimator of Pesaran et al., 1999, and various group estimators, see for example Bonhomme and Manresa, 2015; Su et al.,

2016. These estimators are reasonable choices when investigators are interested in the average effect or know the true specification of the heterogeneity structure or the number of groups. However, researchers are often more interested in the individual parameters, and in most cases the true specification is unknown. As a result, a more useful approach could be model averaging and shrinkage estimation methods. Maddala et al., 2001 show the superior properties of shrinkage estimators among single-equation estimators and various averaging estimators in a heterogeneous panel data model under error homoscedasticity framework. Wang et al., 2019 propose a Mallow pooling averaging estimator for heterogeneous panel data models and conclude that the pooling estimator is preferred when the panel is heterogeneous and the signal-to-noise ratio is moderate or large. The Mallow model averaging estimator, however, is not asymptotically optimal in our framework since the condition (C.3) of Wang et al., 2019 does not hold here. The condition requires that there is no model for which the bias is zero, which does not hold in our framework since the FGLS estimator is unbiased.

Our paper is mainly concerned with point estimation and does not address the challenging issue of inference with shrinkage estimators. As a preliminary step in this direction, we study the mean squared errors of various estimators. However, since the distribution of shrinkage estimators are non-Gaussian distributions, it is still unclear how to use this knowledge to construct confidence intervals. We leave the full treatment of this nontrivial, interesting and important issue to a follow-up paper.

The paper is organized as follows. Section 2 describes the model and the estimators. In section 3, we study properties of the estimators. We give the bias, mean squared error (MSE) matrix, and the risk of the shrinkage estimator using Large-sample asymptotic theory in section 4. In section 5, the asymptotic bias, asymptotic MSE matrix and asymptotic risk of the shrinkage estimator are presented. Monte Carlo results are given is section 6. Results from our empirical example are given in section 7. Conclusions are given is section 8. Proofs and detailed calculations are listed in Appendix A.

2 The Model and Notation

Consider the following linear panel data model with heterogeneous slopes

$$y_i = X_i\beta_i + u_i, \qquad i = 1, 2, ...N, \tag{2.1}$$

where $y_i = (y_{i1}, y_{i2}, ..., y_{iT})'$ is a $T \times 1$ vector of observations on the dependent variable y_{it}, X_i is a $T \times k$ matrix of observations on the k vector of regressors including the intercept[1], β_i is a $k \times 1$ vector of unknown coefficients, and $u_i = (u_{i1}, u_{i2}, ..., u_{iT})'$ is a $T \times 1$ vector of disturbances for $i = 1, 2, ..., N$, where T is the time dimension, and N is the cross-section dimension. It is convenient

[1] That is the first column of X_i is a vector of ones for all $i = 1, 2, \dots, N$, which allows for fixed effects. Also, note that we do not assume that X_is are the same, nor do we assume they are different across equations. In other words, our model supports complete heterogeneity, partial heterogeneity, and complete homogeneity of regressors. Hence this setup includes the seemingly unrelated equations as well.

to stack the N equations above in the following form

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & & \\ \vdots & \ddots & & \\ 0 & \dots & 0 & X_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}, \tag{2.2}$$

or compactly as

$$\underset{NT\times 1}{y} = \underset{NT\times Nk}{\mathbb{X}} \underset{Nk\times 1}{\boldsymbol{\beta}} + \underset{NT\times 1}{u}. \tag{2.3}$$

We assume

Assumption 1: *The disturbances are normally distributed with mean zero and variance-covariance matrix* Ω.

Assumption 2: *The* $NT \times 1$ *vector of disturbances,* u *has zero conditional mean*

$$\mathbb{E}(u|X_1, X_2, \dots, X_N) = 0.$$

Assumption 3: *The disturbances are uncorrelated across observations but correlated across equations,*

$$\mathbb{E}(u_i u_j'|X_1, X_2, \dots, X_m) = \sigma_{ij} I_T,$$

or

$$\mathbb{E}(uu'|X_1, X_2, \dots, X_m) = \Omega = \begin{bmatrix} \sigma_{11}I_T & \sigma_{12}I_T & \dots & \sigma_{1N}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T & \dots & \sigma_{2N}I_T \\ & & \vdots & \\ \sigma_{N1}I_T & \sigma_{N2}I_T & \dots & \sigma_{NN}I_T \end{bmatrix} = \Sigma \otimes I_T,$$

where

$$\underset{N\times N}{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2N} \\ & & \vdots & \\ \sigma_{N1} & \sigma_{N2} & \dots & \sigma_{NN} \end{bmatrix},$$

where we assume Ω *is positive definite and* $\mathbb{E}(X_i'\Omega^{-1}X_i)$ *is nonsingular.*

The normality assumption of disturbances in assumption 1 is made for deriving finite sample properties of estimators, but we relax this assumption under asymptotic theory analysis of section 5. Assumption 2 states that the source generating cross-sectional dependence is not correlated with regressors. This assumption is required for the consistency of our proposed estimator.To estimate the variance-covariance matrix Ω freely, it is assumed the time series dimension is sufficiently large, and in particular $T > N$. Under assumptions 2 and 3 the model of equation (2.1) can also be viewed as a system of seemingly unrelated regressions.

We define some notations below, where

$$\underset{NT\times NT}{\mathbb{Q}} = \Omega^{-1}\Big(I_{NT} - \Psi\Big), \tag{2.4}$$

where

$$\underset{NT\times NT}{\Psi} = \mathbb{X}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}. \tag{2.5}$$

If we partition $\mathbb{Q}$ in the sub-matrices of $T \times T$ as below

$$\mathbb{Q} = \begin{bmatrix} Q_{11} & Q_{12} & \dots & Q_{1N} \\ Q_{21} & Q_{22} & \dots & Q_{2N} \\ \vdots & \vdots & & \vdots \\ Q_{N1} & Q_{N2} & \dots & Q_{NN} \end{bmatrix}, \tag{2.6}$$

then we define

$$\Pi = \begin{bmatrix} Q'_{11} & Q'_{12} & \dots & Q'_{1N} \\ Q'_{21} & Q'_{22} & \dots & Q'_{2N} \\ \vdots & \vdots & & \vdots \\ Q'_{N1} & Q'_{N2} & \dots & Q'_{NN} \end{bmatrix}. \tag{2.7}$$

Also, we define

$$\underset{T\times T}{\Phi} = \sum_{i=1}^{N} \Psi_{ii}, \tag{2.8}$$

where Ψ_{ii} is the diagonal $T \times T$ sub-matrix of Ψ which is partitioned as below

$$\Psi = \begin{bmatrix} \Psi_{11} & \Psi_{12} & \dots & \Psi_{1N} \\ \Psi_{21} & \Psi_{22} & \dots & \Psi_{2N} \\ \vdots & \vdots & & \vdots \\ \Psi_{N1} & \Psi_{N2} & \dots & \Psi_{NN} \end{bmatrix}. \tag{2.9}$$

We define $MSE(\hat{\boldsymbol{\beta}}) = \mathbb{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'$ and Risk $(\hat{\boldsymbol{\beta}}) = \mathbb{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'W(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ for any symmetric positive definite weight matrix W whose elements are of order $O(1)$, and any estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$.

3 Estimation

3.1 Unrestricted Estimator

The standard estimator of $\boldsymbol{\beta}$, is the feasible generalized least-squares (FGLS) estimator of Zellner, 1962. This estimator is defined as

$$\hat{\boldsymbol{\beta}} = (\mathbb{X}'\,\hat{\Omega}^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\hat{\Omega}^{-1}\,y = \boldsymbol{\beta} + (\mathbb{X}'\,\hat{\Omega}^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\hat{\Omega}^{-1}\,u, \tag{3.1}$$

where $\hat{\Omega} = \hat{\Sigma} \otimes I_T$, $\hat{\Sigma} = [s_{ij}]$, and s_{ij} is the estimate of σ_{ij}, using single-equation estimates of β_i, defined as $\breve{\beta}_i = (X_i'X_i)^{-1}X_i'y_i$, $i = 1, 2, ..., N$. Hence

$$s_{ij} = (y_i - X_i\breve{\beta}_i)'(y_j - X_j\breve{\beta}_j)/T = u_i'M_iM_ju_j/T, \tag{3.2}$$

where $M_i = I_T - X_i(X_i'X_i)^{-1}X_i$.

Theorem 1: *Under assumptions 1–3, the bias of the FGLS estimator up to order $O(T^{-1})$ is*

$$\textit{Bias}\ (\hat{\boldsymbol{\beta}}) = \mathbb{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = 0, \tag{3.3}$$

and the MSE matrix of the FGLS estimator up to order $O(T^{-2})$ is

$$\textit{MSE}\ (\hat{\boldsymbol{\beta}}) = (1 + \frac{N}{T})(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1} - \frac{1}{T}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}H(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}, \tag{3.4}$$

and the risk up to order $O(T^{-2})$ is

$$\textit{Risk}\ (\hat{\boldsymbol{\beta}}) = (1 + NT^{-1})\ tr\left[W(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\right] - T^{-1}\ tr\left[W(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}H(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\right], \tag{3.5}$$

where $H = \mathbb{X}'(\Sigma^{-1} \otimes \Phi)\,\mathbb{X} - \mathbb{X}'\,\Pi\,\mathbb{X}$, and W is a positive definite weight matrix of order $O(1)$.

Proof: Appendix A,(See page 32).

Two arbitrary choices of W are I_{Nk} and $T^{-1}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})$, where the latter one in risk, provides the mean squared forecast error (MSFE).

Corollary 1.1: *Under Assumptions 1–3, the MSFE of the FGLS estimator up to order $O(T^{-2})$ is*

$$\textit{MSFE}\ (\hat{\boldsymbol{\beta}}) = T^{-1}\left(1 + NT^{-1}\right)Nk - T^{-2}\ tr\left[H(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\right]. \tag{3.6}$$

3.2 Restricted Estimator

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

A common restricted parameter space, Θ_0, of particular interest in this setup is the homogeneity restriction of slope parameters across cross-sections known as pooling. In this case, we would form the restriction as

$$R\boldsymbol{\beta} = \begin{bmatrix} I_k & \mathbf{0} & \dots & \mathbf{0} & -I_k \\ \mathbf{0} & I_k & \dots & \mathbf{0} & -I_k \\ \vdots & \vdots & & \vdots & \\ \mathbf{0} & \mathbf{0} & \dots & I_k & -I_k \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} = \begin{bmatrix} \beta_1 - \beta_N \\ \beta_2 - \beta_N \\ \vdots \\ \beta_{(N-1)} - \beta_N \end{bmatrix} = \mathbf{0}, \tag{3.7}$$

this specifies a total of $d = (N-1)k$ restrictions on the $Nk \times 1$ parameter vector.

Another restricted parameter space, Θ_0, which is common in applied economics will take the form of an exclusion restriction for each cross-section equation. For example, if we partition

$$\beta_i = \begin{bmatrix} \beta_{i,1} \\ \beta_{i,2} \end{bmatrix}, i = 1, 2, \dots, N, \tag{3.8}$$

where $\beta_{i,1}$, $(k - d_i) \times 1$, represents the slopes of the core regressors, and $\beta_{i,2}$, $d_i \times 1$, 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

$$R\boldsymbol{\beta} = \begin{bmatrix} R_1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & R_2 & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & & \vdots & \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & R_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} = \begin{bmatrix} \beta_{1,2} \\ \beta_{2,2} \\ \vdots \\ \beta_{N,2} \end{bmatrix} = \mathbf{0}, \tag{3.9}$$

where R is a matrix of $d \times Nk$, with $R_i = (\mathbf{0}, I_{d_i})$, for $i = 1, 2, \dots, N$, and $d = \sum_{i=1}^{N} d_i$.

The restricted generalized least squares estimator is obtained as the solution to the following minimization

$$\underset{\text{s.t. } \boldsymbol{\beta}}{\text{Minimize}} \; (y - \mathbb{X}\boldsymbol{\beta})'\Omega(y - \mathbb{X}\boldsymbol{\beta}) \qquad \text{subject to } \boldsymbol{r}(\boldsymbol{\beta}) = 0, \tag{3.10}$$

and the solution can be formulated as a feasible restricted generalized least-squares estimator in below

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - (\mathbb{X}'\,\hat{\Omega}^{-1}\,\mathbb{X})^{-1}R'\left[R(\mathbb{X}'\,\hat{\Omega}^{-1}\,\mathbb{X})^{-1}R'\right]^{-1}R\hat{\boldsymbol{\beta}}. \tag{3.11}$$

Restrictions are often tested using hypothesis testing. The hypothesis to be tested is H_0 : $\boldsymbol{r}(\boldsymbol{\beta}) = 0$ against the alternative, $H_1 : \boldsymbol{r}(\boldsymbol{\beta}) \neq 0$. A conventional test static that has a limiting chi-squared distribution with d degrees of freedom when the null hypothesis is true is

$$F = \hat{\boldsymbol{\beta}}' R' \left[R(\mathbb{X}' \hat{\Omega}^{-1} \mathbb{X})^{-1} R' \right]^{-1} R \hat{\boldsymbol{\beta}} = \left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}} \right)' \mathbb{X}' \hat{\Omega}^{-1} \mathbb{X} \left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}} \right), \tag{3.12}$$

which can be recognized as a Wald statistic (Greene, 2008) and measures the weighted distance between $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ [2].

3.3 Shrinkage Estimator

We use the restrictions and the test statistic to construct a shrinkage estimator, and show that the proposed estimator improves estimation efficiency and makes an appropriate trade-off between bias due to possible incorrect restrictions and variance efficiency gains from imposing these restrictions.

Our proposed shrinkage estimator of $\boldsymbol{\beta}$ is a weighted average of the FGLS estimator and the restricted estimator

$$\hat{\boldsymbol{\beta}}_s = \omega \, \hat{\boldsymbol{\beta}} + (1 - \omega) \, \tilde{\boldsymbol{\beta}}, \tag{3.13}$$

where the weight takes the form

$$\omega = (1 - \frac{\tau}{F}). \tag{3.14}$$

In equation (3.14), $\tau \geq 0$ is a positive shrinkage parameter which controls the degree of shrinkage. We will defer describing our recommended optimal choice for this parameter in the following sections. Alternatively, ω can be replaced by its positive part, $(\omega)_+ = \omega \, \mathbb{1}(\omega \geq 0)$, as it can be easily verified that the risk of the estimator with the positive part is smaller. However, it will not affect the derivations of the approximations below, so for simplicity we do not impose it at this stage. Nevertheless, the Monte Carlo results and empirical results are reported using the positive-part weight.

The shrinkage estimator defined above, shrinks the FGLS estimator towards the restricted estimator by the ratio τ/F. When the difference between the restricted and FGLS estimators is small (Wald statistic is small), the shrinkage estimator gives a large weight to the restricted estimator, as it is most efficient estimator. However, when the difference between the restricted and unrestricted estimators is substantial or high ($F > \tau$), the bias of the restricted estimator could be more than its variance efficiency gain, so the shrinkage estimator is a weighted average of the restricted and FGLS estimators, with more weight on the FGLS estimator.

[2] The last equality in equation (3.12) holds because

$$F = \hat{\boldsymbol{\beta}}' R' \left[R(\mathbb{X}' \hat{\Omega}^{-1} \mathbb{X})^{-1} R' \right]^{-1} R(\mathbb{X}' \hat{\Omega}^{-1} \mathbb{X})^{-1} (\mathbb{X}' \hat{\Omega}^{-1} \mathbb{X}) (\mathbb{X}' \hat{\Omega}^{-1} \mathbb{X})^{-1} R' \left[R(\mathbb{X}' \hat{\Omega}^{-1} \mathbb{X})^{-1} R' \right]^{-1} R' \hat{\boldsymbol{\beta}}.$$

4 Finite Sample Approximation

In this section, we obtain the higher-order approximation bias, MSE matrix and risk of the shrinkage estimator using large-sample higher-order asymptotic of Nagar, 1959.

Theorem 2: *Under Assumptions 1–3, the bias of the shrinkage estimator up to order $O(T^{-1})$ is*

$$Bias\ (\hat{\boldsymbol{\beta}}_s) = \mathbb{E}(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}) = -\frac{\tau}{\phi} P\,\boldsymbol{\beta}, \tag{4.1}$$

where $P = (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}R'\Big[R(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}R'\Big]^{-1}R$, *and* $\phi = \boldsymbol{\beta}'\,P'(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}P\,\boldsymbol{\beta} = O(T)$, *and the MSE matrix of the shrinkage estimator up to order $O(T^{-2})$ is*

$$MSE\ (\hat{\boldsymbol{\beta}}_s) = MSE\ (\hat{\boldsymbol{\beta}}) + \frac{\tau}{\phi^2} P\,\boldsymbol{\beta}\,\boldsymbol{\beta}'\,P'\Big[\tau + 4\Big] - \frac{\tau}{\phi}\Big[(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}P' + P(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\Big], \tag{4.2}$$

and the risk of the shrinkage estimator up to order $O(T^{-2})$ is

$$Risk\ (\hat{\boldsymbol{\beta}}_s) = Risk\ (\hat{\boldsymbol{\beta}}) + \frac{\tau}{\phi^2}\,\boldsymbol{\beta}'\,P'W P\,\boldsymbol{\beta}\left[\tau - 2\Big[\frac{\phi}{\phi_w}\,tr\Big(W(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}P'\Big) - 2\Big]\right]. \tag{4.3}$$

where $\phi_w = \boldsymbol{\beta}'\,P'W P\,\boldsymbol{\beta}$.

The second term on the right hand side of equation (4.3) is negative when

$$0 < \tau < 2\Big[\frac{\phi}{\phi_w}\,tr\Big(W(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}P'\Big) - 2\Big], \tag{4.4}$$

that is when τ satisfies the condition above, the risk of the shrinkage estimator is less than that of the FGLS estimator. Also, the optimal value of the shrinkage parameter that minimizes the risk of the shrinkage estimator, when $tr\Big(W(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}P'\Big)\phi/\phi_w > 2$, *is*

$$\tau_{opt} = \frac{\phi}{\phi_w}\,tr\Big(W(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}P'\Big) - 2. \tag{4.5}$$

Proof: Appendix A,(See page 35).

As the optimal shrinkage parameter depends on unknown parameter values, one could replace them by their estimated values. The replacement still provides the dominance of the shrinkage estimator up to the order of interest. That is if $\mathrm{tr}\Big(W(\mathbb{X}'\,\hat{\Omega}^{-1}\,\mathbb{X})^{-1}\hat{P}'\Big)\hat{\phi}/\hat{\phi}_w \geq 2$ holds, we still have the dominance of the shrinkage estimator and the estimated optimal shrinkage parameter is

$$\hat{\tau}_{opt} = \mathrm{tr}\Big(W(\mathbb{X}'\,\hat{\Omega}^{-1}\,\mathbb{X})^{-1}\hat{P}'\Big)\hat{\phi}/\hat{\phi}_w - 2.$$

In the case of $W = T^{-1}\,\mathbb{X}'\,\hat{\Omega}^{-1}\,\mathbb{X}$, the shrinkage parameter does not depend on unknown parameter values and the MSFE of the shrinkage estimator will be more simplified, which is given in the following corollary.

Corollary 2.1: *Under Assumptions 1–3, the MSFE of the shrinkage estimator up to order $O_p(T^{-2})$ is*

$$MSFE\ (\hat{\boldsymbol{\beta}}_s) = MSFE\ (\hat{\boldsymbol{\beta}}) + \frac{\tau}{T\phi}\Big[\tau - 2(d-2)\Big]. \tag{4.6}$$

The above equation describes that when $0 < \tau < 2(d-2)$, the MSFE of the shrinkage estimator is less than that of the FGLS estimator. Also the optimal value of shrinkage parameter that minimizes the MSFE of the shrinkage estimator, provided $d > 2$, is

$$\tau_{opt,F} = d - 2. \tag{4.7}$$

5 Asymptotic Theory

In this section, we give the asymptotic bias, MSE matrix and risk of the shrinkage estimator defined in Section 3 under a general local asymptotic framework (Assumption 5), when the time horizon $T \to \infty$ while the cross-section dimension (N) is fixed. We make the following standard set of regulatory assumptions.

Assumption 4: *$V_T = T(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}$, and W_T have full ranks and tend to finite non-singular matrices, V and W respectively, as $T \to \infty$.*

Assumption 5: *We assume that*

$$\beta_i = \bar{\beta}_i + \alpha_i, \quad i = 1, 2, \ldots, N;$$

and

$$\alpha_i = T^{-\kappa}\delta_i, \quad \text{where } \kappa \geq 0\ , \ \text{and, } \delta_i \in \mathbb{R}^k,$$

where $\bar{\beta}_i$ is a centering value which belongs to the restricted parameter space Θ_0, $\delta_i \in \mathbb{R}^k$ is a localizing parameter which shows the difference between the unrestricted and restricted parameter space, and κ is the speed by which the localizing parameter converges to zero [3]. In a matrix form we can write the equations above as

$$\boldsymbol{\beta} = \bar{\boldsymbol{\beta}} + \boldsymbol{\alpha} = \bar{\boldsymbol{\beta}} + T^{-\kappa}\,\boldsymbol{\delta},$$

where $\underset{Nk\times 1}{\boldsymbol{\alpha}} = (\alpha_1', \alpha_2', \ldots, \alpha_N')'$ and $\underset{Nk\times 1}{\boldsymbol{\delta}} = (\delta_1', \delta_2', \ldots, \delta_N')'$.

[3] In the restricted parameter space where the restrictions impose parameter symmetry across cross-sections, $\bar{\beta}_i$ in assumption 5 can be viewed as a common mean, i.e. $\bar{\beta}_i = \bar{\beta}$ for $i = 1, 2, \ldots, N$. For the case of exclusion restrictions parameter space, the assumption can be explained as $\bar{\beta}_i = \beta_{i,c}$.

Theorem 3: *Under assumptions 2–4 , the asymptotic distribution of the FGLS estimator is*

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{p} Z \sim N(\mathbf{0}, V), \tag{5.1}$$

and together with assumption 5, the asymptotic distribution of the restricted estimator is

$$\sqrt{T}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{p} Z - VR'(RVR')^{-1}R(Z + \sqrt{T}\,\boldsymbol{\alpha}). \tag{5.2}$$

Further by using the above equations,

$$F(\,\hat{\boldsymbol{\beta}}\,,\,\tilde{\boldsymbol{\beta}}\,) \xrightarrow{p} \xi(Z) = (Z + \sqrt{T}\,\boldsymbol{\alpha})'R'(RVR')^{-1}R(Z + \sqrt{T}\,\boldsymbol{\alpha}) \sim \boldsymbol{\chi}^2_{Nk}(\boldsymbol{\theta}'\,A\,\boldsymbol{\theta}), \tag{5.3}$$

where [4] $\boldsymbol{\theta} = V^{-1/2}\sqrt{T}\,\boldsymbol{\alpha}$, $A = V^{1/2}R'(RVR')^{-1}RV^{1/2}$ *is an idempotent matrix, and for the shrinkage weight, we have*

$$(1 - \frac{\tau}{F}) \xrightarrow{p} \omega(Z) = (1 - \frac{\tau}{\xi(Z)}). \tag{5.4}$$

Therefore, the asymptotic distribution of the shrinkage estimator is

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}) \xrightarrow{p} Z_s = \omega(Z)Z + (1 - \omega(Z))(Z - VR'(RVR')^{-1}R'(Z + \sqrt{T}\,\boldsymbol{\alpha})). \tag{5.5}$$

Assumption 6: *We assume* $\sup_T \mathbb{E}(|Z_{s,T}|^{2+\epsilon}) < \infty, \quad \forall \epsilon > 0$, *where* $Z_{s,T} = \sqrt{T}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta})$.

Assumption 6 is sufficient for the uniform integrability of $T(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta})$ (seeBillingsley, 1986, pp. 32). Therefore, under assumption 6, we define the following terms. The asymptotic bias of the shrinkage estimator is,

$$\text{ABias }(\hat{\boldsymbol{\beta}}_s) = \lim_{T\to\infty} \mathbb{E}[\sqrt{T}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta})] = \mathbb{E}(Z_s),$$

the asymptotic MSE matrix of the shrinkage estimator is defined as,

$$\text{AMSE }(\hat{\boldsymbol{\beta}}_s) = \lim_{T\to\infty} \mathbb{E}[T(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta})'] = \mathbb{E}(Z_s Z_s'),$$

and the asymptotic risk of the shrinkage estimator for a weight matrix W_T, is

$$\text{ARisk }(\hat{\boldsymbol{\beta}}) = \lim_{T\to\infty} \mathbb{E}[T(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta})'W_T(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta})] = \mathbb{E}(Z_s'WZ_s).$$

Theorem 4: *Under assumptions 2–6 the asymptotic bias of the shrinkage estimator is*

$$ABias\ (\hat{\boldsymbol{\beta}}_s) = -T^{(1-2\kappa)/2}\ \frac{\tau}{d}\,VR'(RVR')^{-1}R\,\boldsymbol{\delta}\ e^{(-T^{1-2\kappa}\lambda)}\ {}_1F_1\Big(\frac{d}{2}\,,\,\frac{d}{2}+1;\,T^{1-2\kappa}\,\lambda\Big), \tag{5.6}$$

[4] $\chi^2_a(b)$ represents a chi-squared distribution with a degrees of freedom and a non-centrality parameter b.

and the asymptotic MSE matrix of the shrinkage estimator is

$$
\begin{aligned}
AMSE\ (\hat{\boldsymbol{\beta}}_s) = V &+ \frac{\tau}{d}\, VR'(RVR')^{-1}RV\ e^{(-T^{1-2\kappa}\lambda)} \\
&\left[\frac{\tau}{d-2}\ {}_1F_1\Big(\frac{d}{2}-1\,,\ \frac{d}{2}+1\,;\ T^{1-2\kappa}\lambda\Big) - 2\ {}_1F_1\Big(\frac{d}{2}\,,\ \frac{d}{2}+1\,;\ T^{1-2\kappa}\lambda\Big)\right] \\
&+ \tau\, T^{1-2\kappa}\ VR'(RVR')^{-1}R\,\boldsymbol{\delta}\,\boldsymbol{\delta}'\,R'(RVR')^{-1}RV\ e^{(-T^{1-2\kappa}\lambda)} \\
&\left[\frac{\tau}{d(d+2)}\ {}_1F_1\Big(\frac{d}{2}\,,\ \frac{d}{2}+2\,;\ T^{1-2\kappa}\lambda\Big) - 2\Big[\frac{1}{d+2}\ {}_1F_1\Big(\frac{d}{2}+1\,,\ \frac{d}{2}+2\,;\ T^{1-2\kappa}\lambda\Big)\right. \\
&\left. - \frac{1}{d}\ {}_1F_1\Big(\frac{d}{2}\,,\ \frac{d}{2}+1\,;\ T^{1-2\kappa}\lambda\Big)\Big]\right],
\end{aligned}
\tag{5.7}
$$

where $\lambda = \boldsymbol{\delta}'\,R'(RVR')^{-1}R\,\boldsymbol{\delta}\,/2$, *and* ${}_1F_1(.,.;.)$ *denotes the confluent hypergeometric function* [5].

Proof: Appendix A,(See page 37).

Corollary 4.1: *Under assumptions 2–6, and the results in Theorem 4 we have the followings*

(i) If $0 \leq \kappa < 1/2$ *, the shrinkage estimator is asymptotically unbiased and consistent and very close to the FGLS estimator*

$$
ABias\ (\hat{\boldsymbol{\beta}}_s) = -T^{-(1-2\kappa)/2}\ \frac{\tau}{2\lambda} VR'(RVR')^{-1}R\,\boldsymbol{\delta}\ \Big[1 + O(T^{-(1-2\kappa)})\Big], \tag{5.8}
$$

and the asymptotic MSE matrix is

$$
\begin{aligned}
AMSE\ (\hat{\boldsymbol{\beta}}_s) = V - T^{-(1-2\kappa)}\ \frac{\tau}{\lambda}\, VR'(RVR')^{-1}RV\ + T^{-(1-2\kappa)}\ \Big[\frac{\tau^2}{4\lambda^2} + \frac{\tau}{\lambda^2}\Big] \\
VR'(RVR')^{-1}R\,\boldsymbol{\delta}\,\boldsymbol{\delta}'\,R'(RVR')^{-1}RV + O(T^{-2(1-2\kappa)}),
\end{aligned}
\tag{5.9}
$$

(ii) Local Asymptotic: If $\alpha_i = \delta_i T^{-1/2}$, $i = 1, 2, \ldots, N$, *i.e.* $\kappa = 1/2$, *then the shrinkage estimator is asymptotically unbiased and efficient and we have*

$$
ABias\ (\hat{\boldsymbol{\beta}}_s) = -\frac{\tau}{d}\, e^{-\lambda}\, VR'(RVR')^{-1}R\,\boldsymbol{\delta}\ \ {}_1F_1\Big(\frac{d}{2}\,,\ \frac{d}{2}+1\,;\ \lambda\Big) + o(1), \tag{5.10}
$$

[5] The confluent hypergeometric function is given by

$$
{}_1F_1(a, c; x) = \sum_{n=0}^{\infty} \frac{(a)_n x^n}{(c)_n n!}
$$

where $(a)_n = a(a+1)\ldots(a+n-1)$, $(a)_0 = 1$. Also $(a)_n = \Gamma(a+n)/\Gamma(a)$ for positive a.

and

$$AMSE\,(\hat{\boldsymbol{\beta}}_s) = V + \frac{\tau}{d}\, e^{-\lambda} \left[\frac{\tau}{d-2}\, {}_1F_1\Big(\frac{d}{2}-1\,,\,\frac{d}{2}+1\,;\,\lambda\Big) - 2\,{}_1F_1\Big(\frac{d}{2}\,,\,\frac{d}{2}+1\,;\,\lambda\Big)\right] VR'(RVR')^{-1}RV$$

$$+ e^{\lambda}\,\tau \left[\frac{\tau}{d(d+2)}\, {}_1F_1\Big(\frac{d}{2}\,,\,\frac{d}{2}+2\,;\,\lambda\Big) - 2\Big[\frac{1}{d+2}\, {}_1F_1\Big(\frac{d}{2}+1\,,\,\frac{d}{2}+2\,;\,\lambda\Big) - \frac{1}{d}\, {}_1F_1\Big(\frac{d}{2}\,,\,\frac{d}{2}+1\,;\,\lambda\Big)\Big]\right]$$

$$VR'(RVR')^{-1}R\,\boldsymbol{\delta}\,\boldsymbol{\delta}'\,R'(RVR')^{-1}RV \;+ o(1). \tag{5.11}$$

Proof: Appendix A,(See page 40).

In the following corollary we give our recommended value of τ that minimizes the risk of the shrinkage estimator under the local asymptotic condition.

Corollary 4.2: *Under assumptions 2–6, when* $\kappa = 1/2$, $d \geq tr(C)/\varrho_{min}(V^{1/2}WV^{1/2})$, $tr(C)/\varrho_{max}(C) > 2$, *and* $0 < \tau \leq 2\Big[\,tr(C)/\varrho_{max}(C) - 2\Big]$, *then*

$$ARisk\,(\hat{\boldsymbol{\beta}}_s) < ARisk\,(\hat{\boldsymbol{\beta}}) - e^{-\lambda}\, \frac{1}{d-2}\, \frac{\lambda_W}{\lambda}\, {}_1F_1\Big(\frac{d}{2}-1\,,\,\frac{d}{2}\,;\,\lambda\Big) \left[2\tau\Big(\frac{tr(C)}{\varrho_{max}(C)} - 2\Big) - \tau^2 \right] \tag{5.12}$$

where $ARisk\,(\hat{\boldsymbol{\beta}}) = tr(W\,V)$, $C = V^{1/2}AVWVAV^{1/2}$, $\lambda_W = \boldsymbol{\delta}'\,AVWVA\,\boldsymbol{\delta}\,/2$, ϱ_{min} *and* ϱ_{max}, *respectively, denote the minimum and maximum eigenvalue. The above result shows the superiority of the shrinkage estimator relative to the FGLS estimator. The optimal shrinkage parameter that minimizes the risk is*

$$\tau_{opt} = tr(C)/\varrho_{max}(C) - 2, \tag{5.13}$$

which then results in

$$ARisk\,(\hat{\boldsymbol{\beta}}_{s,opt}) < ARisk\,(\hat{\boldsymbol{\beta}}) - e^{-\lambda}\, \frac{1}{d-2}\, \frac{\lambda_W}{\lambda}\, {}_1F_1\Big(\frac{d}{2}-1\,,\,\frac{d}{2}\,;\,\lambda\Big) \Big(\frac{tr(C)}{\varrho_{max}(C)} - 2\Big)^2 \tag{5.14}$$

Proof: Appendix A,(See page 40).

As the optimal shrinkage parameter depends on Ω, which is unknown, it can be estimated. That is one can replace Ω by its consistent estimator $\hat{\Omega}$, and use

$$\hat{\tau}_{opt} = \mathrm{tr}(\hat{C})/\varrho_{max}(\hat{C}) - 2. \tag{5.15}$$

In this case as $\hat{\tau}_{opt} \xrightarrow[T\to\infty]{p} \tau_{opt}$, the results of corollary 4.2 will still hold.

Corollary 4.3: *Under assumptions 2–6, when $\kappa = 1/2$, $d > 2$, and $0 < \tau \leq 2[d-2]$, the MSFE of the shrinkage estimator is*

$$MSFE\ (\hat{\boldsymbol{\beta}}_s) = MSFE\ (\hat{\boldsymbol{\beta}}) - e^{-\lambda}\ \frac{1}{d-2}\ {}_1F_1\Big(\frac{d}{2} - 1\,,\ \frac{d}{2}\,;\ \lambda\Big)\Big[2\tau(d-2) - \tau^2\Big], \tag{5.16}$$

where $MSFE\ (\hat{\boldsymbol{\beta}}) = I_{Nk}$. The value of τ that minimizes the MSFE of the shrinkage estimator is

$$\tau_{F,opt} = d - 2, \tag{5.17}$$

and the MSFE of the optimal shrinkage estimator is

$$MSFE\ (\hat{\boldsymbol{\beta}}_{s,opt}) = MSFE\ (\hat{\boldsymbol{\beta}}) -\ e^{-\lambda}\ (d-2)\ {}_1F_1\Big(\frac{d}{2} - 1\,,\ \frac{d}{2}\,;\ \lambda\Big). \tag{5.18}$$

Corollary 4.4: *Under assumptions 2–6, when $\kappa = 1/2$, and $d > 2$, if $\lambda \to \infty$* [6] *then*

$$ARisk\ (\hat{\boldsymbol{\beta}}_{s,opt}) = ARisk\ (\hat{\boldsymbol{\beta}}) + O(\frac{1}{\lambda}). \tag{5.19}$$

Proof: Appendix A,(See page 40).

The result in corollary 4.4 suggests that if the bias of the restricted estimator is very large (the restricted parameter space is too far from the true parameter space), the shrinkage estimator is asymptotically very close to the FGLS estimator, and asymptotically achieves the global minimax efficiency bond of van der Vaart, 1998. This condition assures that even for very large values of $\boldsymbol{\delta}$ in assumption 5, the shrinkage estimator remains asymptotically consistent and efficient by giving a weight one to the FGLS estimator.

5.1 High Dimensional Shrinkage

In this section, we study the performance of our estimator in a high dimensional case where the number of restrictions increase without bound. The asymptotic properties of our estimator is given in the following theorem using a sequential approximations by letting first the sample size, and then the number of instruments, tent to infinity.

Theorem 5: *Under assumptions 2–6, when $\kappa = 1/2$, if as $d \to \infty$, $\lim_{d\to\infty} \lambda/d \to 0$ then*

$$\lim_{d\to\infty} \frac{ARisk\ (\hat{\boldsymbol{\beta}}_{s,opt})}{ARisk\ (\hat{\boldsymbol{\beta}})} \leq 1 - \rho, \qquad \rho = \lim_{d\to\infty} \frac{tr(C)}{tr(WV)}, \tag{5.20}$$

where $0 \leq \rho \leq 1$, with $\rho = 1$ when all parameters of interest are restricted. If $W = V^{-1}$, then we

[6] Equivalently when $\boldsymbol{\delta} \to \infty$.

have

$$\lim_{d\to\infty} \frac{MSFE\ (\hat{\boldsymbol{\beta}}_{s,opt})}{MSFE\ (\hat{\boldsymbol{\beta}})} = 1 - \frac{d-2}{NK}. \tag{5.21}$$

Also, in the expressions above the optimal shrinkage estimator can be replaced with any shrinkage estimator in which the shrinkage parameter, τ, satisfies the condition below

$$\lim_{d\to\infty} \frac{\tau}{\tau_{opt}} \to 1. \tag{5.22}$$

Proof: Appendix A,(See page 41).

The right hand side of equation (5.20) is equal to the local minimax efficiency bound given in Theorem 5 of Hansen, 2016, which specifies that our proposed estimator asymptotically achieves this local minimax bound, while the FGLS estimator does not. Therefore, the shrinkage estimator proposed in this paper is locally efficient, thus, there is no need to consider alternative model averaging methods. A major advantage of our proposed shrinkage estimator relative to the Stein-type shrinkage estimator considered in Hansen, 2016 is that, the risk of our estimator does not depend on the bound size of localizing parameters, as a result the gain of our proposed estimator relative to the FGLS estimator can be quantified.

6 Monte Carlo Simulation

The results below are the simulation results of the model of section 2, where $x_{it,1=1}$ and the remaining regressors are independently generated from the standard normal distributions. The sample size varies from $T \in \{50, 100, 200\}$, $N \in \{3, 5\}$, $k \in \{4, 6\}$, leading to twelve combinations of m, T and k. u_1 is generated as $i.i.d\, N(0,1)$, while $u_i = cu_1 + v_i$, for $i = 2, \ldots, N$, where $v_i \sim i.i.d\, N(0,1)$ and $c = 0.25$. We consider two DGPs for generating β_i, the first one is under a completely heterogeneity in coefficients where we assume that

$$\text{DGP1:} \quad \beta_i = \bar{\beta} + (i \times \delta)/N,\ i = 1, 2, \ldots, N,$$

with $\bar{\beta} = (1, 1, \ldots, 1)'$, and the second DGP is under a weakly heterogeneity where we assume that

$$\text{DGP2:} \quad \beta_{i1}, \beta_{i2} = \begin{cases} 1 + (i \times \delta)/N, & \text{if } i = 1, \ldots, [N/2] \\ 1.2, & \text{if } i = [N/2]+1, \ldots, N \end{cases}, \beta_{il} = 2,\ l \in \{3, \ldots, k\},$$

where $[N/2]$ denotes the nearest integer value that is smaller than $m/2$, and δ takes values on a 10-point grid on $[0, 1]$.

The results of $1,000$ monte carlo simulations are given in Figures 1–8, where the vertical axis measure the relative mean squared error (RMSE) of the FGLS estimator, the restricted estimator, a pre-test estimator, and the optimal shrinkage estimator, to the FGLS estimator. The horizontal axis measures the degree of heterogeneity (δ) which is set between zero and one with 0.1 grid value.

The Monte Carlo results support our theoretical findings of the previous section. The figures show that the RMSE of the shrinkage estimator for the whole parameter heterogeneity is below that of the FGLS estimator. This shows the superiority of our proposed shrinkage estimator relative to the FGLS estimator.

The RMSE of the shrinkage estimator in DGP1 of a complete heterogeneous panel data model, is smaller than that of the restricted estimator except for very small values of parameter heterogeneity (δ). This is expected because as δ takes higher values, the bias of the restricted estimator increases, which then results in higher MSE. Also, when the sample size is higher the RMSE of the shrinkage estimator dominates that of the restricted estimator for most values of δ. In DGP2 where the model is characterized by some degree of homogeneity, the RMSE of the restricted estimator remains smaller than that of the FGLS estimator for even larger values of δ. In this case, the FGLS estimator can be inferior to the restricted estimator even with the presence of weak degrees of heterogeneity. This is because although the FGLS estimator is unbiased, it is inefficient, especially under small sample sizes, and high number of regressors. In contrast, the restricted estimator properly makes use of cross-section variation and thus provides a more accurate results.

In general, we find that the shrinkage estimator performs robustly well in heterogeneous panel data models with various degrees of heterogeneity. When there is a strong heterogeneity, the shrinkage estimator prevails. When there is a relatively weak heterogeneity, the shrinkage estimator tends to gain more from the efficiency of the restricted estimator by assigning a high weight to this estimator, and thus still remains one of the best choices.

7 Application: Returns to scale in US banking

In this section, we apply the shrinkage estimator studied in the previous sections to regressions of a cost system for U.S. commercial banks. We are interested in estimating the returns to scale and scale economies of these banks over the past recent years.

Over the past few years, the number of U.S. commercial banks fell by almost 70%, where in 1984 the total number of U.S. commercial banks was 14391 and dropped to 4773 in 2018. Over the same period of time, the average asset value of U.S. banks (adjusted for inflation), which is also a measure of bank size, increased about ten times, from $140 Million in 1984 to $1,400 Million in 2018 (See Figure 9). To support this bank size expansion, bank executives and analysts claim that due to the changes in regulation such as the permission of interstate branching and combination of banks, and because of technological and financial innovations such as communication technologies, the securitization and sale of bank loans over the past few years, the cost of production for larger banks has reduced and encouraged banks to grow larger and/or merge. On the other hand, critics contend that this decrease in the number of operating banks and having banks with large assets not only impact market competition, but result in agency problems, and disproportionate benefits of government policies in favor of large banks. In particular, the financial crisis of 2007 focused attention on large financial institutions and the role the "too-big-to-fail" doctrine played in driving their size. These together have brought attention of policy makers for regulatory limits on bank size. However, any policy intervention needs to consider the potential efficiency benefits of operating at a large scale. Therefore, estimation of scale economies is essential for analyzing the costs and benefits of any policy intervention to control the size of banks.

The estimation of scale economies for U.S. banking industry has stimulated a substantial body of studies. Older empirical studies that used data from the 1980s and 1990s did not find scale economies in banking industry except at very small banks. But recent research that used data from the 2000s, and more modern methods of estimating the banking models, has found considerably more evidence of scale economies in banking. These studies include Hughes et al., 1996, 2000, 2001, Berger and Mester, 1997, Hughes and Mester, 1998, 2013, Wheelock and Wilson, 2012, and Feng and Serletis, 2009. A common approach in analyzing banks' performance is to separate them into asset size classes (one group, three groups, or twelve groups, see Feng and Serletis, 2009, Hughes and Mester, 2013, and Mailkov et al., 2015), and then estimate the cost equation of each group independently from the other groups. However, not only it is very difficult to categorize banks based upon asset size and that there is not industry standard on asset ranges, but also it is hard to believe that these groups are not being affected by some unknown factors that caused correlations between groups. The uncertainty about the model classification exhibit two issues that need to be carefully considered by researchers. First, estimating cost efficiency by classifying banks in one group or three major groups has the advantage of producing estimators with smaller variance, but at the same time, the estimators could suffer from heterogeneity bias due to potential difference in technology of banks within these group. Second, partitioning banks in more groups could benefit from avoiding heterogeneity biases, but is subject to losing variance efficiency. Hence, there is a trade-off between bias and variance efficiency between these two estimators. As the shrinkage estimator that developed in the previous sections results in the optimal balance between bias and variance efficiency, we recommend using our estimator in estimation of the returns to scale for banking industry to reach minimum estimation loss.

7.1 The Model

We follow the framework broadly employed in the literature which is the so called "intermediation approach" of Sealey and Lindley, 1977. According to this approach, a bank's balance sheet is assumed to capture the essential structure of a bank's core business. Inputs are considered to be liabilities (core deposits and purchased funds), physical capital and labor. Inputs result in the bank's productions which are assets (other than the physical, includes loans and trading securities).

With regard to variable specification, we define five inputs and five outputs that are the ones used in the literature. We define the following output quantities: consumer loans (y_1), real estate loans (y_2), loans to business and other institutions (y_3), federal funds sold and securities purchased under agreements to resell (y_4), and other assets (y_5). The input variables are: labor quantities (x_1), premises and fixed assets (x_2), purchased funds (x_3), interest-bearing transaction accounts

(x_4), and non-transaction accounts (x_5). For each input x_j, its price w_j is obtained by dividing its total expenses by the corresponding input quantities. Further, z denotes the asset value.

For modeling the cost of banking industry, we consider a translog cost function and normalize it, so that the homogeneity (in input prices) property is automatically satisfied. We allow for individual (fixed) effects by adding intercepts in each regression, to control for specific group characteristics, heterogeneity in skills, and so on. Hence, the cost equation for each group $i = 1, 2, ..., m$, is considered as

$$\ln(C_i/w_{5,i}) = \beta_{0,i} + \sum_{j=1}^{4} \beta_{j,i} \ln(w_{j,i}/w_{5,i}) + \sum_{k=1}^{5} \gamma_{k,i} \ln(y_{k,i}) + \frac{1}{2} \sum_{k=1}^{5} \sum_{l=1}^{5} \gamma_{kl,i} \ln(y_{k,i}) \ln(y_{l,i})$$
$$+ \frac{1}{2} \sum_{j=1}^{4} \sum_{q=1}^{4} \eta_{jq,i} \ln(w_{j,i}/w_{5,i}) \ln(w_{q,i}/w_{5,i}) + \sum_{j=1}^{4} \sum_{k=1}^{5} \delta_{jk} \ln(w_{j,i}/w_{5,i}) \ln(y_{k,i}) + u_i, \tag{7.1}$$

where C_i is the total cost of banks in group i, defined as

$$C_i = w_{1,i}x_{1,i} + w_{2,i}x_{2,i} + w_{3,i}x_{3,i} + w_{4,i}x_{4,i} + w_{5,i}x_{5,i}, \qquad i = 1, \ldots, m. \tag{7.2}$$

The cost function is symmetric which requires the imposition of the following restriction on the parameters as below

$$\begin{aligned} \eta_{jq,i} &= \eta_{qj,i} \\ \gamma_{kl,i} &= \gamma_{lk,i}. \end{aligned} \tag{7.3}$$

The returns-to-scale (RTS) is defined as the inverse of the sum of cost elasticities. If we define the output elasticity of the model for output j as $Ecy_{j,i} = \frac{\partial \ln(C_i)}{\partial \ln(y_{j,i})}$, and the sum of cost elasticities as $Ecy_i = \sum_{j=1}^{5} \frac{\partial \ln(C_i)}{\partial \ln(y_{j,i})}$, then the RTS for group i is defined as

$$RTS_i = \frac{1}{Ecy_i} = \frac{1}{\sum_{j=1}^{5} \frac{\partial \ln(C_i)}{\partial \ln(y_{j,i})}}, \tag{7.4}$$

and scale economies are also defined as $(1 - Ecy_i)$. A bank with $RTS > 1$, has increasing returns to scale, that is for one percent increase in all outputs, cost is increased by less than one percent, and the bank is operating below its efficient scale size ($RTS = 1$) when $RTS < 1$.

7.2 The Data

The data we use is obtained from the Reports of Income and Condition (Call Reports)[7], over the period from 2000 to 2018. We omit observations for which negative values for assets, equity, outputs, and prices are reported. The summary of data is in Table 1.

[7]The data from 2000-2010 is downloaded from the Federal Reserve Bank of Chicago website, and the rest of the data from 2011-2018 is downloaded from the FFIEC Central Data Repository's Public Data Distribution website.

Following Feng and Serletis, 2009, we classify the banks into three major groups which is mainly based on the standard asset size categories that are used by the Federal Financial Institutions Examination Council (FFIEC), as specified in forms 031, 032, 033, and 034. Banks with over \$500 million in total assets are classified as large banks, banks with assets between \$100 million and \$500 million are classified as medium banks, and banks with under \$100 million in assets are classified as small banks. To avoid heterogeneity biases associated with asset size, our unrestricted parameter space classifies each of the three bank groups into several subgroups. Specifically, the further cutoffs are considered as \$20 million, \$40 million, \$60 million, and \$80 million within the small bank group, \$200 million, \$300 million, and \$400 million within the medium bank groups, and \$1 billion, and \$3 billion within the large bank group. Besides, in order to have a consistent partitions over time, the asset size caps in each year are justified upward by the growth in the CPI. Figure 10 and Figure 11 presents the bank partitions in twelve subgroups for years 2000 and 2018 with the number of banks and share of banks that place in each group.

7.3 Estimation

We estimate the model of equation (7.1) using the shrinkage estimation method developed in the previous sections and derive the RTS for each subgroup in each year separately. Basically, for each year, the cross-section dimension of our panel model is represented by the cost equations of (sub)groups, and the time dimension is considered as the observations of banks operating within each (sub)group. This is because, first, there is a substantial decrease in the number of operating banks due to bank mergers over the sample period. At the same time, the asset size of majority of banks has significantly increased. As a result, considering a same cost technology for banks over time may not be accurate. The second reason is to account for changes in production and cost structure of banks over time which could occur due to changes in banks management strategies. Hence our restricted heterogeneous panel model at each year consists of one cost equation representing all banks cost function, and the observations the operating banks data in that same year. Moreover, our unrestricted heterogeneous panel model at each year consists of twelve cost equations representing all of the subgroups, and the observations for each regression are the operating banks data under each subgroup in that specific year. Since the sample size for each group is different, we face a panel model with unequal number of observations and to estimate the variance-covariance matrix (Ω), we consider the following procedures recommended in the literature (See Schmidt, 1977, Baltagi et al., 1989):

1. Ignore the extra observations in estimating Ω.
2. Use the extra observations to estimate variances. This procedure has the disadvantage of giving estimates of Ω that are not positive definite.
3. Use the extra observations to estimate variances, and modifying the estimates of covariances using the method of Srivastava and Zaatar, 1973.
4. Use all observations in estimating, following the method of Hocking and Smith, 1968.

It is known in the literature that the results of the above procedures are much the same. Likewise, we find that all of the procedures above, generate similar results so we only report the results of method 3.

The RTS estimation results of years 2000 to 2018 are averaged over the subgroups within each asset class and are given in Figure 12–17. The Figure presents extreme deciles, quartiles and means for our estimated RTS. Comparing the banks over time shows that, on average the banks had increasing returns to scale over the whole sample, except for Large and Medium banks which is decreasing or constant returns to scale from 2006 to 2009. Specifically, we see signs of cost efficiency for more than 50% of banks within each of the asset size classes, with the exception of the Small banks that more than 75% of them show increasing returns to scale for years 2000, 2010, 2012, 2014, and 2016. Also, the results suggest higher magnitude of cost efficiency for Large and Medium banks at their 90$^{\text{th}}$ decile. The results are consistent with some recent studies (e.g. Feng and Serletis, 2009; Hughes and Mester, 2013; Wheelock and Wilson, 2012; Henderson et al., 2015; Mailkov et al., 2015), although we are not aware of any study from 2011 to 2018.

8 Conclusion

We introduce a new method of estimation and forecasting in heterogeneous panel data models under cross section-dependence to address the problem of model uncertainty. This method has four main advantages relative to the model averaging and shrinkage estimation methods. First, it allows for heteroscedasticity and cross-section dependence of error terms which is essential in most of the panel data model applications. Second, the dominance and optimality of the shrinkage estimator proposed here is not limited to MSFE and holds for any weighted quadratic loss function where the weight is positive definite and symmetric. Third, the shrinkage weight is proportional to a Wald statistics that controls for rotations of the coefficient vectors, hence provides a shrinkage estimator with a uniformly lowest MSE. Lastly, the framework considered here is not limited to a local misspecification, and the dominance properties of the shrinkage estimator is given against any arbitrary deviations from the restrictions.

This paper also contributes to the long-existing issue in the panel data analysis referred by econometricians to as “to pool or not to pool.” We compare the performance of our proposed estimator with the single-equation and pooling estimators, and show the reliability of the estimation results under our shrinkage estimator. Moreover, we apply our method to estimate cost efficiency of U.S. commercial banks which has gained popularity over the past recent years. Our method has two advantages over the method considered in the literature. First, it allows for correlation among banks with different asset sizes. This correlation could be due to omitted common effects, or could arise as a result of interactions within socioeconomic networks. Second, as there is a model specification uncertainty issue about the cost functions, our method, unlike previous studies, considers the uncertainty of model selection and estimation jointly. Therefore, the results are more robust and reliable than the single-equation or pooling estimators mostly considered in the literature.

References

Abadir, K. M., J. R. Magnus, 2005. *Matrix Algebra* New York: Cambridge University Press.

Andrews, D., 2005. Cross section regression with common shocks. *Econometrica* 73: 1551–85.

Berger, A. N., L. J. Mester, 1997. Inside the black box: What explains differences in the efficiencies of financial institutions? *Journal of Banking and Finance* 21(7): 895–947.

Baltagi, B. H., J. M. Griffin, 1984. Short and long run effects in pooled models. *International Economic Review* 25(3): 631–45.

Baltagi, B. H., S. Garvin, S. Kerman, 1989. Further Monte Carlo Evidence on Seemingly Unrelated Regressions with Unequal Number of Observations. *Annales D'Economie et de Statistique* 14: 103–15.

Baltagi, B. H., J. M. Griffin, W. Xiong, 2000. To pool or not to pool: Homogeneous versus heterogeneous estimations applied to cigarette demand. *The Review of Economics and Statistics* 82(1): 117–26.

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

Billingsley, P., 1986. Convergence of Probability Measures. *John Wiley, New York.*

Bonhomme, S., Manresa, E., 2015. Grouped patterns of heterogeneity in panel data. *Econometrica* 83: 1147–84.

Browning, M., J. Carro, 2007. Heterogeneity and microeconometrics modeling. In *Advances in Economics and Econometrics*. Edited by R. Blundell, W. Newey and T. Persson. Cambridge: Cambridge University Press, vol. 3, pp. 47–74.

Belloni, A., Chernozhukov, V., Fernandez-Val, I., Hansen, C., 2017. Program evaluation and causal inference with high-dimensional data. *Econometrica* 85: 233–298.

Chernozhukov, V., Hansen, C., Spindler, M., 2015. Post-selection and post-regularization inference in linear models with many controls and instruments. *American Economic Review* 105: 486–90.

Chudick, A., M. H. Pesaran, 2015. Large Panel Data Models With Cross-Sectional Dependence A Survey *The Oxford Handbook Of Panel Data, Edited By Badi H. Baltagi* New York: Oxford University Press.

Claeskens, G., N. L. Hjort, 2003. The focused information criterion. *Journal of American Statistical Association* 98: 900–45.

Danilov, D., J. R. Magnus, 2004a. On the harm that ignoring pretesting can cause. *Journal of Econometrics* 122:2746.

Danilov, D., J. R. Magnus, 2004b. Forecast accuracy after pretesting with an application to the stock market. *Journal of Forecasting* 23:251274.

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

Durlauf, S. N., A. Kourtellos, A. Minkin, 2001. The local Solow growth model. *European Economic Review* 45(4–6): 928–40.

Fan, J., R. Li, 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of American Statistical Association* 96: 1348–60.

Fan, J., J. Lv, 2010. A selective overview of variable selection in high dimensional feature space. *Statistical Sinica* 20: 102–148.

Feng, G., and A. Serletis, 2009. Efficiency and productivity of the US banking industry, 1998–2005: Evidence from the Fourier cost function satisfying global regularity conditions. *Journal of Applied Econometrics* 24(1): 105–38.

Greene, W. H. (2008). *Econometric Analysis* (7th Edition) Upper Saddle River, NJ: Prentice Hall.

Hansen, B. E. 2014. Model averaging, asymptotic risk, and regressor groups. *Quantitative Economics* 5: 495–530.

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

Henderson, Daniel J., Subal C. Kumbhakar, Qi Li, and Christopher F. Parameter. 2015. Smooth coefficient estimation of a seemingly unrelated regression. *Journal of Econometrics* 189(1): 148–62.

Hocking, R. R., W. B. Smith, 1968. Estimation of Parameters in the Multivariate Normal Distribution with Missing Observations. *Journal of the American Statistical Association* 63: 154–73.

Hoogstrate, A. J., F. C. Palm, G. A. Pfann, 2000. Pooling in dynamic panel-data models: An application to forecasting GDP growth rates. *Journal of business and Economic Studies* 18(3): 274–83.

Hughes, J. P., and L. J. Mester, 1998. Bank capitalization and cost: Evidence of scale economies in risk management and signaling. *The Review of Economics and Statistics* 80(2): 314–25.

Hughes, J. P., and L. J. Mester, 2013. Who said large banks dont experience scale economies? Evidence from a risk-return-driven cost function. *Journal of Financial Intermediation* 22(4): 559–85.

Hughes, J. P., W. Lang, L. J. Mester, C. Moon, 1996. Efficient banking under interstate branching. *Journal of Money, Credit, and Banking* 28(4): 1045-71.

Hughes, J. P., W. Lang, L. J. Mester, C. Moon, 2000. Recovering risky technologies using the almost ideal demand system: An application to US banking. *Journal of Financial Services Research* 18: 5–27.

Hughes, J. P., L. J. Mester, and C. Moon, 2001. Are scale economies in banking elusive or illusive?: Evidence obtained by incorporating capital structure and risk-taking into models of bank production. *Journal of Banking and Finance* 25(12): 2169–208.

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.*

Leeb, H. Pötscher, B. M., 2003. The finite-sample distribution of post-model-selection estimators and uniform versus nonuniform approximations. *Econometric Theory* 19:100142.

Leeb, H. Pötscher, B. M., 2006. Can one estimate the conditional distribution of post-model-selection estimators? *The Annals of Statistics* 34:25542591.

Liu, C., 2015. Distribution theory of the least squares averaging estimator. *Journal of Econometrics* 186: 142–59.

Maddala, G. S., 1991. To pool or not to pool: That is the question. *Journal of Quantitative Economics* 7: 255–64.

Maddala, G. S., W. Hu, 1996. The pooling problem. In *The Econometrics of Panel Data.* Edited by L. Matyas and P. Sevestre. Advanced Studies in Theoretical and Applied Econometrics, vol 33. Springer, Dordrecht, pp. 307–22.

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., 2002. Estimation of the mean of a univariate normal distribution with known variance. *Econometrics Journal* 5:225236.

Magnus, J. R., J. Durbin, 1999. Estimation of regression coefficients of interest when other regression coefficients are of no interest. *Econometrica* 67:639643.

Mailkov, Emir, Diego Restrepo-Tobon, and Subal C. Kumbhakar. 2015. Estimation of banking technology under credit uncertainty. *Empirical Economics* 49: 185–211.

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

Pesaran, M. H., Y. Shin, R. P. Smith, 1999. Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association.* 94: 621–34.

Pesaran, M. H., R. Smith, 1995. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics* 68(1): 79–113.

Phillips, P. C. B. and D. Sul, 2003. Dynamic panel estimation and homogeneity testing under cross section dependence. *The Econometrics Journal* 6: 217–59.

Phillips, P. C. B. and D. Sul, 2007. Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics* 137: 162–88.

Robertson, D., J. Symons, 1992. Some strange properties of panel data estimators. *Journal of Applied Econometrics* 7(2): 175–89.

Sarafidis, V. D. Robertson, 2009. On the impact of error cross-sectional dependence in short dynamic panel estimation. *The Econometrics Journal* 162: 62–81.

Sealey, C.W., and J. T. Lindley, 1977. Inputs, outputs, and a theory of production and cost at depository financial institutions. *Journal of Finance* 32(4): 1251–66.

Schmidt, P., 1977. Estimation of Seemingly Unrelated Regressions with Unequal Numbers of Observations. *Journal of Econometrics* 5(3): 365–77.

Slater, L.J., 1960. Confluent Hypergeometric Functions. *Cambridge University Press, London.*

Srivastava, J. N., and R. Tiwari, 1976. Evaluation of Expectations of Products of Stochastic Matrices. *Scandinavian Journal of Statistics* 3: 135–38.

Srivastava, J. N., M. K. Zaatar, 1973. A Monte Carlo Comparison of Four Estimators of Dispersion Matrix of a Bivariate Normal Population, Using Incomplete Data. *Journal of the American Statistical Association* 68(341): 180–83.

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.

Su, L., and Q. Chen, 2013. Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory* 29(6): 1079–135.

Su, L., Z. Shi, P. C. B. Phillips, 2016. Identifying latent structures in panel data. *Econometrica* 84: 2215–64.

Swamy, P. A. V. B., 1970. Efficient inference in a random coefficient regression model. *Econometrica* 38: 311–23.

Tibshirani, R., 1996. *Journal of the Royal Statistical Society* 58: 267–88.

Ullah, A., 1974. On the Sampling Distributions of Improved Estimators for Coefficients in Linear Regression. *Journal of Econometrics* 2: 143–50.

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

van der Vaart, A. W., 1998. Asymptotic Statistics. *Cambridge University Press, New York.*

Wang, W., X. Zhang, R. Paap, 2019. To Pool or Not to Pool: What is a Good Strategy? *Journal of Applied Econometrics* 34(5): 724–45.

Wheelock, D. C., P. W. Wilson, 2001. New evidence on returns to scale and product mix among US commercial banks. *Journal of Monetary Economics* 47(3): 653–74.

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

Zellner, Arnold. 1962. An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association* 57(298): 348–68.

A Appendix A

Lemma A1. It can be easily verified that for all $i, j, l, n, a, b = 1, 2, \ldots, N$, we have the followings

$$M_i X_i = 0, \tag{A.1}$$

$$\mathbb{Q}_{ij}\, X_j = 0, \tag{A.2}$$

$$\sum_{i=1}^{N} \sigma_{ji}\sigma^{il} = \begin{cases} 1, & \text{if } l = j \\ 0, & \text{otherwise,} \end{cases} \tag{A.3}$$

$$\mathbb{E}(u_j u_a' M_a M_i u_i u_l') = \sigma_{ja} M_a M_i \sigma_{il} + \sigma_{ai}\sigma_{jl}\,\mathrm{tr}(M_a M_i) I_T + \sigma_{al}\sigma_{ji} M_i M_a, \tag{A.4}$$

$$\begin{aligned} \mathbb{E}(u_a' M_a M_i u_i u_j' u_n u_l u_b') &= T\sigma_{al}\sigma_{ib}\sigma_{jn} M_a M_i + T\sigma_{ai}\sigma_{jn}\sigma_{lb}\,\mathrm{tr}(M_i M_a) + T\sigma_{ab}\sigma_{li}\sigma_{nj} M_i M_a \\ &\quad + T\sigma_{ai}\sigma_{lj}\sigma_{nb} + T\sigma_{ai}\sigma_{ln}\sigma_{jb} + T\sigma_{aj}\sigma_{lb}\sigma_{in} + T\sigma_{an}\sigma_{lb}\sigma_{ij} + o(1), \end{aligned} \tag{A.5}$$

where $\mathbb{Q}_{ij}$ is the (i, j)th submatrix of order $T \times T$ of $\mathbb{Q}$, and $\sigma^{il} = \sigma_{il}^{-1}$.

For the proof of the last two equality see Srivastava and Tiwari, 1976, and Ullah, 2004. ■

Lemma A2. Let us define

$$\Delta = \hat{\Omega} - \Omega, \tag{A.6}$$

where it can be easily verified that $\Delta = O_p(T^{-1/2})$. Employing this condition, and using the standard geometric expansion for the inverse of a matrix[8], for large T, we have the followings

$$\begin{aligned} \hat{\Omega}^{-1} &= (\Omega + \Delta)^{-1} = \Omega^{-1}[I_{NT} + \Delta\Omega^{-1}]^{-1} \\ &= \Omega^{-1}\Big[I_{NT} - \Delta\Omega^{-1} + \Delta\Omega^{-1}\Delta\Omega^{-1} - \Delta\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1} + \ldots\Big] \\ &= \Omega^{-1} - \underbrace{\Omega^{-1}\Delta\Omega^{-1}}_{O_p(T^{-1/2})} + \underbrace{\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}}_{O_p(T^{-1})} - \underbrace{\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}}_{O_p(T^{-3/2})} + O_p(T^{-2}), \end{aligned} \tag{A.7}$$

$$\begin{aligned} (\mathbb{X}'\,\hat{\Omega}^{-1}\,\mathbb{X})^{-1} &= \Big[\mathbb{X}'\,\Omega^{-1}\,\mathbb{X} - \mathbb{X}'\,\Omega^{-1}\Delta\Omega^{-1}\,\mathbb{X} + \mathbb{X}'\,\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}\,\mathbb{X} + \ldots\Big]^{-1} \\ &= (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\Big[I_{Nk} - \mathbb{X}'\,\Omega^{-1}\Delta\Omega^{-1}\,\mathbb{X}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1} + \ldots\Big]^{-1} \\ &= \underbrace{(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}}_{O_p(T^{-1})} + \underbrace{(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}\Delta\Omega^{-1}\,\mathbb{X}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}}_{O_p(T^{-3/2})} + O_p(T^{-2}), \end{aligned} \tag{A.8}$$

$$\mathbb{X}'\,\hat{\Omega}^{-1}\,u = \underbrace{\mathbb{X}'\,\Omega^{-1}u}_{O_p(T^{1/2})} - \underbrace{\mathbb{X}'\,\Omega^{-1}\Delta\Omega^{-1}u}_{O_p(1)} + \underbrace{\mathbb{X}'\,\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}u}_{O_p(T^{-1/2})} + O_p(T^{-1}), \tag{A.9}$$

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

$$\left[R(\mathbb{X}'\,\hat{\Omega}^{-1}\,\mathbb{X})^{-1}R'\right]^{-1} = S_1 + S_{1/2} + O_p(1), \tag{A.10}$$

where

$$S_1 = \left[R(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}R'\right]^{-1} = O_p(T), \tag{A.11}$$

and

$$S_{1/2} = -S_1 R(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}\Delta\Omega^{-1}\,\mathbb{X}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}R'S_1 = O_p(T^{1/2}), \tag{A.12}$$

$$(\,\mathbb{X}'\,\hat{\Omega}^{-1}\,\mathbb{X})^{-1}R'\left[R(\mathbb{X}'\,\hat{\Omega}^{-1}\,\mathbb{X})^{-1}R'\right]^{-1}R = P + P_{-1/2} + O_p(T^{-1}), \tag{A.13}$$

where

$$P = (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}R'\left[R(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}R'\right]^{-1}R = O_p(1), \tag{A.14}$$

$$P_{-1/2} = \left[I_{Nk} - P\right](\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}\Delta\Omega^{-1}\,\mathbb{X}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}P = O_p(T^{-1/2}) \tag{A.15}$$

■

Lemma A3. Let the $J \times 1$ vector ν is distributed normally with mean vector θ and covariance matrix I_J, and A is any $J \times J$ idempotent matrix. Also assume $\phi(.)$ is a Borel measurable function. Then

$$\mathbb{E}\left[\phi(\nu' A\nu)\nu\right] = \mathbb{E}\left[\phi(\chi^2_{r+2}(\theta' A\theta/2))\right]A\theta + \mathbb{E}\left[\phi(\chi^2_{r}(\theta' A\theta/2))\right]\Big(I_J - A\Big)\theta,$$

where $r = rank(A) = \mathrm{tr}(A)$.

Proof: Let P be an orthogonal matrix such that

$$PAP' = D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & & \\ & & \vdots & \\ 0 & \dots & 0 & d_J \end{bmatrix} = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{J-r} \end{bmatrix}; \quad d_i \in \{0,1\}.$$

Define the $J \times 1$ vector $\omega = \begin{bmatrix}\omega_1 & \dots & \omega_J\end{bmatrix}' = P\nu$, which has a $N(P\theta, I_J)$ distribution. Therefore

$$\mathbb{E}\left[\phi(\nu' A\nu)\nu\right] = \mathbb{E}\left[\phi(\omega' D\omega)P'\omega\right] = P'\,\mathbb{E}\left[\phi(\omega' D\omega)\omega\right],$$

Then

$$\mathbb{E}\left[\phi(\omega' D\omega)\omega\right] = \left[\mathbb{E}\left[\mathbb{E}[\phi\Big(d_1\omega_1^2 + \sum_{j=2}^{J} d_j\omega_j^2\Big)\omega_1|\omega_j,\, j\neq 1]\right], \ldots \mathbb{E}\left[\mathbb{E}[\phi\Big(d_J\omega_J^2 + \sum_{j=1}^{J-1} d_j\omega_j^2\Big)\omega_J|\omega_j,\, j\neq J]\right]\right]',$$

We now derive the expectation of one of the elements of the last matrix above,

$$\begin{aligned}\mathbb{E}\left[\phi(\omega' D\omega)\omega_i\right] &= p_i'\theta\,\mathbb{E}\left[\mathbb{E}\left[\phi\Big(d_i\chi_3^2((p_i'\theta)^2/2) + \sum_{j\neq i}\omega_j^2 d_j\Big)|\omega_j,\, j\neq i\right]\right] \\ &= \begin{cases} p_i'\theta\,\mathbb{E}\left[\phi(\chi_{r+2}^2(\theta' A\theta/2))\right], & \text{if } d_i = 1 \\ p_i'\theta\,\mathbb{E}\left[\phi(\chi_r^2(\theta' A\theta/2))\right], & \text{if } d_i = 0.\end{cases}\end{aligned}$$

where the first equality holds by Lemma 1 of Appendix B.1 Judge and Bock, 1978. Hence,

$$\begin{aligned}\mathbb{E}\left[\phi(\nu' A\nu)\nu\right] &= P'\,\mathbb{E}\left[\phi(\omega' D\omega)\omega\right] = P'DP\theta\,\mathbb{E}\left[\phi(\chi_{r+2}^2(\theta' A\theta/2))\right] + P'(I-D)P\theta\,\mathbb{E}\left[\phi(\chi_r^2(\theta' A\theta/2))\right] \\ &= A\theta\,\mathbb{E}\left[\phi(\chi_{r+2}^2(\theta' A\theta/2))\right] + (I-A)\theta\,\mathbb{E}\left[\phi(\chi_r^2(\theta' A\theta/2))\right].\end{aligned}$$

Therefore, the Lemma is proved. ■

Lemma A4. Let the $J\times 1$ vector ν is distributed normally with mean vector θ and covariance matrix I_J, and A is any $J\times J$ idempotent matrix. Also assume $\phi(.)$ is a Borel measurable function. Then

$$\begin{aligned}\mathbb{E}\left[\phi(\nu' A\nu)\nu\nu'\right] &= \mathbb{E}\left[\phi(\chi_{r+2}^2(\theta' A\theta/2))\right]A + \mathbb{E}\left[\phi(\chi_r^2(\theta' A\theta/2))\right]\Big(I_J - A\Big) \\ &+ \mathbb{E}\left[\phi(\chi_{r+4}^2(\theta' A\theta/2))\right]A\theta\theta' A + \mathbb{E}\left[\phi(\chi_r^2(\theta' A\theta/2))\right]\Big(I_J - A\Big)\theta\theta'\Big(I_J - A\Big) \\ &+ \mathbb{E}\left[\phi(\chi_{r+2}^2(\theta' A\theta/2))\right]\Big(\theta\theta' A + A\theta\theta' - 2A\theta\theta' A\Big),\end{aligned}$$

where $r = rank(A) = \mathrm{tr}(A)$.

Proof: Let P be an orthogonal matrix such that

$$PAP' = D = \begin{bmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & & \\ & & \vdots & \\ 0 & \ldots & 0 & d_J\end{bmatrix} = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{J-r}\end{bmatrix}; \quad d_i \in \{0,1\}.$$

Define the $J\times 1$ vector $\omega = \begin{bmatrix}\omega_1 & \ldots & \omega_J\end{bmatrix}' = P\nu$, which has a $N(P\theta, I_J)$ distribution. Therefore

$$\mathbb{E}\left[\phi(\nu' A\nu)\nu\nu'\right] = \mathbb{E}\left[\phi(\omega' D\omega)P'\omega\omega' P\right] = P'\,\mathbb{E}\left[\phi(\omega' D\omega)\omega\omega'\right]P.$$

Determine the diagonal and off-diagonal elements of $\mathbb{E}[\phi(\omega' D\omega)\omega\omega']$. The diagonal elements are of the form

$$\mathbb{E}\left[\phi\Big(\sum_{j=1}^{J} d_j\omega_j^2\Big)\omega_i^2\right] = \mathbb{E}\left[\mathbb{E}\left[\phi\Big(d_i\omega_i^2 + \sum_{j\neq i}\omega_j^2\Big)\omega_i^2|\omega_j^2, j\neq i\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\phi\Big(d_i\chi_3^2((P_i'\theta)^2/2) + \sum_{j\neq i}\omega_j^2\Big)|\omega_j^2, j\neq i\right]\right]$$

$$+ (P_i'\theta)^2\,\mathbb{E}\left[\mathbb{E}\left[\phi\Big(d_i\chi_5^2((P_i'\theta)^2/2) + \sum_{j\neq i}\omega_j^2\Big)|\omega_j^2, j\neq i\right]\right]$$

$$= \begin{cases} \mathbb{E}\left[\phi(\chi_{r+2}^2(\theta' A\theta/2)] + (P_i'\theta)^2\,\mathbb{E}[\phi(\chi_{r+4}^2(\theta' A\theta/2)\right], & \text{if } d_i = 1 \\ \mathbb{E}\left[\phi(\chi_{r}^2(\theta' A\theta/2)] + (P_i'\theta)^2\,\mathbb{E}[\phi(\chi_{r}^2(\theta' A\theta/2)\right], & \text{if } d_i = 0. \end{cases}$$

where the second equality holds by Lemma 1 of Appendix B.1 Judge and Bock, 1978.

The diagonal matrix may be written as

$$D\,\mathbb{E}\left[\phi(\chi_{r+2}^2(\theta' A\theta/2))\right] + \mathbb{E}\left[\phi(\chi_{r+4}^2(\theta' A\theta/2))\right] diag(DP\theta\theta' P' D)$$
$$+ (I_J - D)\,\mathbb{E}\left[\phi(\chi_{r}^2(\theta' A\theta/2))\right] + \mathbb{E}\left[\phi(\chi_{r}^2(\theta' A\theta/2))\right] diag((I_J - D)P\theta\theta' P'(I_J - D)).$$

The off-diagonal elements, for $i \neq j$, have the form

$$\mathbb{E}\left[\phi\Big(\sum_{k=1}^{J} d_k\omega_k^2\Big)\omega_i\omega_j\right] = \mathbb{E}\left[\omega_j\,\mathbb{E}\left[\phi\Big(d_i\omega_i^2 + \sum_{k\neq i} d_k\omega_k^2\Big)\omega_i|\omega_k, k\neq i\right]\right]$$

$$= \mathbb{E}\left[\omega_j P_i'\theta\,\mathbb{E}\left[\phi\Big(d_i\chi_3^2((P_i'\theta)^2/2) + \sum_{k\neq i} d_k\omega_k^2\Big)|\omega_k, k\neq i\right]\right]$$

$$= \mathbb{E}\left[\omega_j P_i'\theta_i\,\mathbb{E}\left[\phi\Big(d_i\chi_3^2((P_i'\theta)^2/2) + d_j\omega_j^2 + \sum_{k\neq i\& j} d_k\omega_k^2\Big)|\chi_3^2((P_i'\theta)^2/2), \omega_k, k\neq i\& j\right]\right]$$

$$= P_i'\theta P_j'\theta\,\mathbb{E}\left[\phi\Big(d_i\chi_3^2((P_i'\theta)^2/2) + d_j\chi_3^2((P_j'\theta)^2/2) + \sum_{k\neq i\& j} d_k\omega_k^2\Big)\right]$$

$$= P_i'\theta P_j'\theta \begin{cases} \mathbb{E}[\phi(\chi_{r+4}(\theta' A\theta/2))], & \text{if } d_i = d_j = 1 \\ \mathbb{E}[\phi(\chi_{r+2}(\theta' A\theta/2))], & \text{if } d_i = 1\& d_j = 0 \\ \mathbb{E}[\phi(\chi_{r}(\theta' A\theta/2))], & \text{if } d_i = d_j = 0 \end{cases}$$

where the second equality holds by lemma 2 of Appendix B.1 Judge and Bock, 1978. The off-diagonal matrix may be written as

$$\mathbb{E}\left[\phi(\chi_{r+4}^2(\theta' A\theta/2))\right](DP\theta\theta' P' D - diag(DP\theta\theta' P' D))$$
$$+\,\mathbb{E}\left[\phi(\chi_{r}^2(\theta' A\theta/2))\right]((I_J - D)P\theta\theta' P'(I_J - D) - diag((I_J - D)P\theta\theta' P'(I_J - D)))$$
$$+\,\mathbb{E}\left[\phi(\chi_{r+2}^2(\theta' A\theta/2))\right](P\theta\theta' P' - DP\theta\theta' P' D - (I_J - D)P\theta\theta' P'(I_J - D)).$$

Therefore, combining the diagonal and off-diagonal components, the Lemma is proved. ■

Lemma A5. Let $\chi^2_\alpha(\lambda)$ denote a non-central chi-square random variable with noncentrally parameter λ and α degree of freedom. Also let α denote a positive integer such that $\alpha > 2p$. Then

$$\mathbb{E}\left[\Big(\chi^2_\alpha(\lambda)\Big)^{-p}\right] = 2^{-p}\, e^{-\lambda}\, \frac{\Gamma(\frac{\alpha}{2}-p)}{\Gamma(\frac{\alpha}{2})}\, {}_1F_1\Big(\frac{\alpha}{2}-p\,,\,\frac{\alpha}{2}\,;\,\lambda\Big).$$

Proof: See Ullah, 1974. ■

Lemma A6. If x is bounded and suppose $a, c \to \infty$ such that $\lim_{a,c\to\infty} \frac{(c-a)x}{c} = 0$. Then

$${}_1F_1(a;c;x) = \exp(x)\Big[\sum_{j=0}^{p-1} \frac{(c-a)_j(-x)^j}{(c)_j j!} + O(|c|^{-p})\Big].$$

proof: See Slater, 1960, pp. 12, 65-66. ■

Proof of Theorem 1 :

Recall that from equation (3.1) and Lemma A2

$$\hat{\boldsymbol\beta} - \boldsymbol\beta = (\mathbb{X}'\,\hat\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\hat\Omega^{-1}\,u = A_{-1/2} + A_{-1} + A_{-3/2} + O_p(T^{-2}), \tag{A.16}$$

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

$$\begin{aligned}
A_{-1/2} &= (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}u = O_p(T^{-1/2}),\\
A_{-1} &= -(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}\Delta\,\mathbb{Q}\,u = O_p(T^{-1}),\\
A_{-3/2} &= (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}\Delta\,\mathbb{Q}\,\Delta\,\mathbb{Q}\,u = O_p(T^{-3/2}).
\end{aligned}$$

For the bias of $\hat{\boldsymbol\beta}$ up to order $O(T^{-1})$, we have

$$\mathbb{E}(\hat{\boldsymbol\beta} - \boldsymbol\beta) = \mathbb{E}(A_{-1/2}) + \mathbb{E}(A_{-1}) = 0, \tag{A.17}$$

the last equality holds because, both $A_{-1/2}$ and A_{-1} are odd functions of the error term which has a normal distribution. For the MSE up to order $O(T^{-2})$, we have

$$\begin{aligned}
\mathbb{E}\left[(\hat{\boldsymbol\beta} - \boldsymbol\beta)(\hat{\boldsymbol\beta} - \boldsymbol\beta)'\right] &= \mathbb{E}\left[A_{-1/2}A'_{-1/2} + A_{-1/2}A'_{-1} + A_{-1}A'_{-1/2} + A_{-1/2}A'_{-3/2} + A_{-3/2}A'_{-1/2} + A_{-1}A'_{-1}\right]\\
&= (1+NT^{-1})(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1} - T^{-1}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}H(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}, \qquad (A.18)
\end{aligned}$$

where the last equality holds by using the results in equations (A.19), (A.20), (A.22) and (A.24) .

$$\mathbb{E}(A_{-1/2}A'_{-1/2}) = (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}\,\mathbb{E}(uu')\Omega^{-1}\,\mathbb{X}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1} = (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}, \tag{A.19}$$

$$\begin{aligned}\mathbb{E}(A_{-1}A'_{-1/2}) &= -(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}\,\mathbb{E}(\Delta\,\mathbb{Q}\,uu')\Omega^{-1}\,\mathbb{X}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\\ &= -(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}\,\mathbb{E}\left(\hat{\Omega}\,\mathbb{Q}\,uu'\Omega^{-1}\,\mathbb{X}\right)(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1} = 0,\end{aligned} \tag{A.20}$$

where the second equality holds because $\mathbb{X}'\,\mathbb{Q} = 0$, and the last equality holds because, the (a, b)th submatrix, of order $T \times T$, of $\mathbb{E}(\hat{\Omega}\,\mathbb{Q}\,uu'\Omega^{-1}\,\mathbb{X})$ is zero, as it is shown below

$$\begin{aligned}\mathbb{E}(\sum_{i,j,l=1}^{N} s_{ai}\,\mathbb{Q}_{ij}\,u_j u_l'\sigma^{lb}\,\mathbb{X}_b) &= \frac{1}{T}\sum_{i,j,l=1}^{N}\sigma^{lb}u_a'M_aM_iu_i\,\mathbb{Q}_{ij}\,u_ju_l'\,\mathbb{X}_b\\ &= \frac{1}{T}\sum_{i,j,l=1}^{N}\sigma^{lb}\,\mathbb{Q}_{ij}\left[\sigma_{ja}\sigma_{il}M_aM_i + \sigma_{ai}\sigma_{jl}\,\mathrm{tr}(M_aM_i)I_T + \sigma_{al}\sigma_{ji}M_iM_a\right]\mathbb{X}_b\\ &= \frac{1}{T}\sum_{i,j=1}^{N}\sigma_{ja}\,\mathbb{Q}_{ij}\,M_a\underbrace{M_i\,\mathbb{X}_b}_{=0:\ b=i}\underbrace{\left(\sum_{l=1}^{N}\sigma^{lb}\sigma_{il}\right)}_{=0:\ b\neq i} + \frac{1}{T}\sum_{i,j=1}^{N}\mathrm{tr}(M_aM_i)\sigma_{ai}\underbrace{\mathbb{Q}_{ij}\,\mathbb{X}_b}_{=0:\ b=j}\underbrace{\left(\sum_{l=1}^{N}\sigma^{lb}\sigma_{jl}\right)}_{=0:\ b\neq j}\\ &+ \frac{1}{T}\sum_{i,j=1}^{N}\sigma_{ji}\,\mathbb{Q}_{ij}\,M_i\underbrace{M_a\,\mathbb{X}_b}_{=0:\ b=a}\underbrace{\left(\sum_{l=1}^{N}\sigma^{lb}\sigma_{al}\right)}_{=0:\ b\neq a} = 0,\end{aligned} \tag{A.21}$$

$$\begin{aligned}\mathbb{E}(A_{-3/2}A'_{-1/2}) &= (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}\,\mathbb{E}\left(\Delta\,\mathbb{Q}\,\Delta\,\mathbb{Q}\,uu'\right)\Omega^{-1}\,\mathbb{X}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\\ &= (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}\,\mathbb{E}(\hat{\Omega}\,\mathbb{Q}\,\hat{\Omega}\,\mathbb{Q}\,uu')\Omega^{-1}\,\mathbb{X}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\\ &- (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}\,\mathbb{E}(\hat{\Omega}\,\mathbb{Q}\,\Omega\,\mathbb{Q}\,uu')\Omega^{-1}\,\mathbb{X}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\\ &- (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}\,\mathbb{E}(\Omega\,\mathbb{Q}\,\hat{\Omega}\,\mathbb{Q}\,uu')\Omega^{-1}\,\mathbb{X}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\\ &+ (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}\Omega\,\mathbb{Q}\,\Omega\,\mathbb{Q}\,\mathbb{X}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\\ &= (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}\,\mathbb{E}(\hat{\Omega}\,\mathbb{Q}\,\hat{\Omega}\,\mathbb{Q}\,uu')\Omega^{-1}\,\mathbb{X}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1} = 0,\end{aligned} \tag{A.22}$$

where the second equality holds by substituting for Δ, the third equality holds by using $\mathbb{X}'\,\mathbb{Q} = 0$, and using equation (A.20), and the last equality holds because for the (a, b)th submatrix of $\mathbb{E}(\hat{\Omega}\,\mathbb{Q}\,\hat{\Omega}\,\mathbb{Q}\,uu')$, we have

$$\mathbb{E}(\sum_{i,j,m,l=1}^{N} s_{ai}\,\mathbb{Q}_{ij}\,s_{jm}\,\mathbb{Q}_{ml}\,u_lu_b') = \sum_{i,j,m,l=1}^{N}\mathbb{E}\left[\frac{1}{T^2}u_a'M_aM_iu_i\,\mathbb{Q}_{ij}\,u_j'M_jM_mu_m\,\mathbb{Q}_{ml}\,u_lu_b'\right]$$

$$= \sum_{i,j,m,l=1}^{N} \mathbb{E}\Big[\frac{1}{T^2} u_a' M_a M_i u_i \, \mathbb{Q}_{ij} \, u_j' u_m \, \mathbb{Q}_{ml} \, u_l u_b' + \frac{1}{T^2} u_a' u_i \, \mathbb{Q}_{ij} \, u_j' M_j M_m u_m \, \mathbb{Q}_{ml} \, u_l u_b'$$

$$- \frac{1}{T^2} u_a' u_i \, \mathbb{Q}_{ij} \, u_j' u_m \, \mathbb{Q}_{ml} \, u_l u_b' + O_p(T^{-2})\Big]$$

$$= \frac{1}{T} \sum_{i,j,m,l=1}^{N} \mathbb{Q}_{ij} \, \mathbb{Q}_{ml} \Bigg[\sigma_{jm}\Big(\sigma_{al}\sigma_{ib} M_a M_i + \sigma_{ab}\sigma_{li} M_i M_a\Big) + \sigma_{ai}\Big(\sigma_{lj}\sigma_{mb} M_j M_m + \sigma_{jb}\sigma_{ml} M_m M_j\Big)$$

$$+ \sigma_{lb}\sigma_{ai}\sigma_{jm}\Big(\operatorname{tr}(M_i M_a) + \operatorname{tr}(M_j M_m) - T\Big) I_T + \sigma_{lb}\Big(\sigma_{aj}\sigma_{im} + \sigma_{ij}\sigma_{am}\Big) I_T \Bigg] + O(T^{-2}), \quad \text{(A.23)}$$

where the third equality hold by using Lemma A1, further by using the above equation and Lemma A1 in equation (A.22), the result will hold.

$$\begin{aligned}
\mathbb{E}(A_{-1} A_{-1}') &= (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} \, \mathbb{X}' \, \Omega^{-1} \, \mathbb{E}\Big[\Delta \, \mathbb{Q} \, uu' \, \mathbb{Q} \, \Delta\Big] \Omega^{-1} \, \mathbb{X} (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} \\
&= (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} \, \mathbb{X}' \, \Omega^{-1} \, \mathbb{E}(\hat{\Omega} \, \mathbb{Q} \, uu' \, \mathbb{Q} \, \hat{\Omega}) \Omega^{-1} \, \mathbb{X} (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} \\
&+ (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} \, \mathbb{X}' \, \Omega^{-1} \, \mathbb{E}(\Omega \, \mathbb{Q} \, uu' \, \mathbb{Q} \, \Omega) \Omega^{-1} \, \mathbb{X} (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} \\
&- (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} \, \mathbb{X}' \, \Omega^{-1} \, \mathbb{E}(\Omega \, \mathbb{Q} \, uu' \, \mathbb{Q} \, \hat{\Omega}) \Omega^{-1} \, \mathbb{X} (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} \qquad \text{(A.24)} \\
&- (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} \, \mathbb{X}' \, \Omega^{-1} \, \mathbb{E}(\hat{\Omega} \, \mathbb{Q} \, uu' \, \mathbb{Q} \, \Omega) \Omega^{-1} \, \mathbb{X} (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} \\
&= (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} \, \mathbb{X}' \, \Omega^{-1} \, \mathbb{E}(\hat{\Omega} \, \mathbb{Q} \, uu' \, \mathbb{Q} \, \hat{\Omega}) \Omega^{-1} \, \mathbb{X} (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} \\
&= \frac{1}{T} \Big[N (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} - (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} H (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} \Big],
\end{aligned}$$

where the third equality holds because $\mathbb{X}' \, \mathbb{Q} = 0$, and the last equality holds, because for the (μ, η)th submatrix of $\mathbb{E}(\mathbb{X}' \, \Omega^{-1} \, \hat{\Omega} \, \mathbb{Q} \, uu' \, \mathbb{Q} \, \hat{\Omega} \, \Omega^{-1} \, \mathbb{X})$, we have

$$\mathbb{E}\Big(\sum_{a,i,l,b,j,m=1}^{N} \mathbb{X}_\mu' \, \sigma^{\mu a} s_{ai} \, \mathbb{Q}_{il} \, u_l u_b' \, \mathbb{Q}_{bj} \, s_{jm} \sigma^{m\eta} \, \mathbb{X}_\eta \Big)$$

$$= \frac{1}{T} \sum_{a,i,l,b,j,m=1}^{N} \sigma^{\mu a} \sigma^{m\eta} \, \mathbb{X}_\mu \, \mathbb{Q}_{il} \Bigg[\sigma_{jm}\Big(\sigma_{al}\sigma_{ib} M_a M_i + \sigma_{ab}\sigma_{li} M_i M_a\Big) + \sigma_{ai}\Big(\sigma_{lj}\sigma_{mb} M_j M_m + \sigma_{jb}\sigma_{ml} M_m M_j\Big)$$

$$+ \sigma_{lb}\sigma_{ai}\sigma_{jm}\Big(\operatorname{tr}(M_i M_a) + \operatorname{tr}(M_j M_m) - T\Big) I_T + \sigma_{lb}\Big(\sigma_{aj}\sigma_{im} + \sigma_{ij}\sigma_{am}\Big) I_T \Bigg] \mathbb{Q}_{bj} \, \mathbb{X}_\eta + O_p(T^{-2})$$

(using $\mathbb{X}_\mu' \, \mathbb{Q}_{\mu.} = 0, then$)

$$= \frac{1}{T} \sum \mathbb{X}_\mu' \, \sigma_{ij} \sigma^{\mu\eta} \, \mathbb{Q}_{il} \, \sigma_{lb} \, \mathbb{Q}_{bj} \, \mathbb{X}_\eta + \frac{1}{T} \sum \mathbb{X}_\mu' \, \mathbb{Q}_{\eta l} \, \sigma_{lb} \, \mathbb{Q}_{b\mu} \, \mathbb{X}_\eta$$

(using $\mathbb{Q}_{.l} \, \sigma_{lb} \, \mathbb{Q}_{b.} = \mathbb{Q}_{..}$, then)

$$= \frac{1}{T} \sum \mathbb{X}_\mu' \, \sigma_{ij} \sigma^{\mu\eta} \, \mathbb{Q}_{ij} \, \mathbb{X}_\eta + \frac{1}{T} \, \mathbb{X}_\mu' \, \mathbb{Q}_{\eta\mu} \, \mathbb{X}_\eta$$

(substituting for $\mathbb{Q}_{ij}$, we obtain)

$$
\begin{aligned}
&= \frac{1}{T} \sum \mathbb{X}_\mu' \, \sigma_{ij} \sigma^{\mu\eta} \Big[\sigma^{ij} I_T - \sigma^{im} \, \mathbb{X}_m (\mathbb{X}' \, \Omega^{-1} \, \mathbb{X})^{-1} \, \mathbb{X}_l' \, \sigma^{lj} \Big] \, \mathbb{X}_\eta + \frac{1}{T} \, \mathbb{X}_\mu' \, \mathbb{Q}_{\mu\eta}' \, \mathbb{X}_\eta \\
&= \frac{N}{T} \, \mathbb{X}_\mu' \, \sigma^{\mu\eta} \, \mathbb{X}_\eta - \frac{1}{T} \, \mathbb{X}_\mu' \, \sigma^{\mu\eta} \underbrace{\sum_{i=1}^{N} \Psi_{ii}}_{\Phi} \, \mathbb{X}_\eta + \frac{1}{T} \, \mathbb{X}_\mu' \, \mathbb{Q}_{\mu\eta}' \, \mathbb{X}_\eta
\end{aligned} \tag{A.25}
$$

Using the above result, in equation (A.24), the last equality will hold.

■

Proof of Theorem 2 :

Using Lemma A1, in equation (3.12), we obtain

$$
\begin{aligned}
F &= \Big(\boldsymbol{\beta} + A_{-1/2} + A_{-1} + O_p(T^{-3/2}) \Big)' R' \Big(S_1 + S_{-1/2} + O_p(T^{-1}) \Big) R \Big(\boldsymbol{\beta} + A_{-1/2} + A_{-1} + O_p(T^{-3/2}) \Big) \\
&= \phi + F_{1/2} + O_p(1)
\end{aligned}
$$

where $\phi = \boldsymbol{\beta}' \, R' S_1 R \, \boldsymbol{\beta}$, and

$$
F_{1/2} = \boldsymbol{\beta}' \, R' S_{1/2} R \, \boldsymbol{\beta} + 2 \, \boldsymbol{\beta}' \, R' S_1 R A_{-1/2} = O_p(T^{1/2}). \tag{A.26}
$$

Therefore, we obtain that

$$
\frac{1}{F} = \frac{1}{\phi} \Big[1 + \frac{1}{\phi} F_{1/2} + O_p(T^{-1}) \Big]^{-1} = \underbrace{\frac{1}{\phi}}_{O_p(T^{-1})} - \underbrace{\frac{1}{\phi^2} F_{1/2}}_{O_p(T^{-3/2})} + O_p(T^{-2}). \tag{A.27}
$$

The shrinkage estimator defined in equation (TBD) may be written as

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta} &= (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \tau \Big(\frac{1}{\phi} - \frac{1}{\phi^2} F_{1/2} + O_p(T^{-2}) \Big) \Big(P + P_{-1/2} + O_p(T^{-1}) \Big) \hat{\boldsymbol{\beta}} \\
&= B_{-1/2} + B_{-1} + B_{-3/2} + O_p(T^{-2}),
\end{aligned} \tag{A.28}
$$

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

$$
\begin{aligned}
&B_{-1/2} = A_{-1/2} = O_p(T^{-1/2}), \\
&B_{-1} = A_{-1} - \frac{\tau}{\phi} P \, \boldsymbol{\beta} = O_p(T^{-1}), \\
&B_{-3/2} = A_{-3/2} - \frac{\tau}{\phi} P A_{-1/2} - \frac{\tau}{\phi} P_{-1/2} \, \boldsymbol{\beta} + \frac{\tau}{\phi^2} F_{1/2} P \, \boldsymbol{\beta} = O_p(T^{-3/2}).
\end{aligned}
$$

The bias of the shrinkage estimator to order $O(T^{-1})$ is

$$
\mathbb{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbb{E}(B_{-1/2} + B_{-1}) = \mathbb{E}(A_{-1/2} + A_{-1} - \frac{\tau}{\phi} P \, \boldsymbol{\beta}) = -\frac{\tau}{\phi} P \, \boldsymbol{\beta}, \tag{A.29}
$$

where the last equality holds by equation (A.17).

The MSE of the shrinkage estimator to order $O(T^{-2})$ is

$$\begin{aligned}
\text{MSE }(\hat{\boldsymbol{\beta}}_S) = \mathbb{E}(\Gamma_{-1} + \Gamma_{-3/2} + \Gamma_{-2}) &= (1 + NT^{-1})(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1} - \frac{1}{T}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}H(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1} \\
&- \frac{\tau}{\phi}\Big[(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}P' + P(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\Big] + \frac{\tau^2}{\phi^2}P\,\boldsymbol{\beta}\,\boldsymbol{\beta}'\,P' \\
&+ 2\frac{\tau}{\phi^2}\Big[(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}R'S_1R\,\boldsymbol{\beta}\,\boldsymbol{\beta}'\,P' + P\,\boldsymbol{\beta}\,\boldsymbol{\beta}'\,R'S_1R(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\Big],
\end{aligned} \tag{A.30}$$

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

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

below we give their expectations

$$\mathbb{E}(\Gamma_{-1}) = \mathbb{E}\left((\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\,\mathbb{X}'\,\Omega^{-1}uu'\Omega^{-1}\,\mathbb{X}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\right) = (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}, \tag{A.31}$$

$$\mathbb{E}(\Gamma_{-3/2}) = \mathbb{E}(B_{-1/2}B'_{-1}) + \mathbb{E}(B_{-1}B'_{-1/2}) = 0, \tag{A.32}$$

because

$$\mathbb{E}(B_{-1}B'_{-1/2}) = \mathbb{E}\left[\Big(A_{-1} - \frac{\tau}{\phi}P\,\boldsymbol{\beta}\Big)A'_{-1/2}\right] = \mathbb{E}(A_{-1}A'_{-1/2}) - \frac{\tau}{\phi}P\,\boldsymbol{\beta}\,\mathbb{E}(A_{-1/2}) = 0, \tag{A.33}$$

where the last equality holds by using equation (A.20), and we have

$$\begin{aligned}
\mathbb{E}(\Gamma_{-2}) &= \mathbb{E}(B_{-1/2}B'_{-3/2}) + \mathbb{E}(B_{-3/2}B'_{-1/2}) + \mathbb{E}(B_{-1}B'_{-1}) \\
&= -\frac{\tau}{\phi}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}P' + 2\frac{\tau}{\phi^2}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}R'S_1R\,\boldsymbol{\beta}\,\boldsymbol{\beta}'\,P' - \frac{\tau}{\phi}P(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1} + \frac{\tau^2}{\phi^2}P\,\boldsymbol{\beta}\,\boldsymbol{\beta}'\,P' \\
&+ 2\frac{\tau}{\phi^2}P\,\boldsymbol{\beta}\,\boldsymbol{\beta}'\,R'S_1R(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1} + \frac{1}{T}\Big[N(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1} - (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}H(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\Big],
\end{aligned} \tag{A.34}$$

the last equality above holds by using equations (A.35) and (A.36) below

$$\begin{aligned}
\mathbb{E}(B_{-1}B'_{-1}) &= \mathbb{E}(A_{-1}A'_{-1}) - \frac{\tau}{\phi}\,\mathbb{E}(A_{-1}\,\boldsymbol{\beta}'\,P') - \frac{\tau}{\phi}\,\mathbb{E}(P\,\boldsymbol{\beta}\,A'_{-1}) + \frac{\tau^2}{\phi^2}P\,\boldsymbol{\beta}\,\boldsymbol{\beta}'\,P' \\
&= \frac{1}{T}\Big(N(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1} - (\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}H(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\Big) + \frac{\tau^2}{\phi^2}P\,\boldsymbol{\beta}\,\boldsymbol{\beta}'\,P',
\end{aligned} \tag{A.35}$$

and

$$\mathbb{E}(B_{-3/2}B'_{-1/2}) = \mathbb{E}(A_{-3/2}A'_{-1/2}) - \frac{\tau}{\phi}P\,\mathbb{E}(A_{-1/2}A'_{-1/2}) - \frac{\tau}{\phi}\,\mathbb{E}(P_{-1/2}\,\boldsymbol{\beta}\,A'_{-1/2}) + \frac{\tau}{\phi^2}\,\mathbb{E}(F_{1/2}P\,\boldsymbol{\beta}\,A'_{-1/2})$$
$$= -\frac{\tau}{\phi}P(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1} + 2\frac{\tau}{\phi^2}P\,\boldsymbol{\beta}\,\boldsymbol{\beta}'\,R'S_1R(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}, \tag{A.36}$$

where the last equality holds by using equation (A.19), and

$$\mathbb{E}(A_{-3/2}A'_{-1/2}) = 0, \qquad \text{(by using equation (A.22))}, \tag{A.37}$$

$$\mathbb{E}(P_{-1/2}\,\boldsymbol{\beta}\,A'_{-1/2}) = 0,\ \text{(because it is product of odd numbers of the error term)}, \tag{A.38}$$

$$\mathbb{E}(F_{1/2}P\,\boldsymbol{\beta}\,A'_{-1/2}) = \mathbb{E}\left[P\,\boldsymbol{\beta}\,\boldsymbol{\beta}'\,R'S_{1/2}R\,\boldsymbol{\beta}\,u'\Omega^{-1}\,\mathbb{X}(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1}\right]$$
$$+ 2\,\mathbb{E}\left[P\,\boldsymbol{\beta}\,\boldsymbol{\beta}'\,R'S_1RA_{-1/2}A'_{-1/2}\right] = 2P\,\boldsymbol{\beta}\,\boldsymbol{\beta}'\,R'S_1R(\mathbb{X}'\,\Omega^{-1}\,\mathbb{X})^{-1} \tag{A.39}$$

where the last equality hold because the first term in the second equality is an odd number of the error term, and the second term is derived by using equation (A.19).

■

Proof of Theorem 4 :

Let $\zeta = \boldsymbol{\theta}'\,A\,\boldsymbol{\theta}\,/2$. Recall from equation (5.5) the asymptotic distribution of the shridage estimator is

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}) \xrightarrow{p} Z_s = \omega(Z)Z + (1-\omega(Z))(Z - VR(R'VR)^{-1}R'(Z + \sqrt{T}\,\boldsymbol{\alpha})),$$

employing Assumption 6, the asymptotic bias of $\sqrt{T}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta})$ is

$$\lim_{T\to\infty}\mathbb{E}\left[\sqrt{T}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta})\right] = \mathbb{E}(Z_s) = \mathbb{E}(Z) - \tau VR'(RVR')^{-1}R\,\mathbb{E}\left(\frac{Z + \sqrt{T}\,\boldsymbol{\alpha}}{\xi(Z)}\right)$$
$$= -\frac{\sqrt{T}\,\tau}{d}\,e^{-\zeta}\,VR'(RVR')^{-1}R\,\boldsymbol{\alpha}\ \ {}_1F_1\Big(\frac{d}{2},\frac{d}{2}+1;\zeta\Big)$$
$$-\frac{\sqrt{T}\,\tau}{d-2}\,e^{-\zeta}\,VR'(RVR')^{-1}R\Big[\,\boldsymbol{\alpha} - VR'(RVR')^{-1}R\,\boldsymbol{\alpha}\,\Big]\,{}_1F_1\Big(\frac{d}{2}-1,\frac{d}{2};\zeta\Big)$$
$$= -\frac{\sqrt{T}\,\tau}{d}\,e^{-\zeta}\,VR'(RVR')^{-1}R\,\boldsymbol{\alpha}\ \ {}_1F_1\Big(\frac{d}{2},\frac{d}{2}+1;\zeta\Big), \tag{A.40}$$

the last equality holds because second term of the third equality above is zero, and the second equality holds because

$$\mathbb{E}\left(\frac{Z+\sqrt{T}\,\boldsymbol{\alpha}}{\xi(Z)}\right) = V^{1/2}\,\mathbb{E}\left[V^{-1/2}\frac{Z+\sqrt{T}\,\boldsymbol{\alpha}}{(Z+\sqrt{T}\,\boldsymbol{\alpha})'R'(RVR')^{-1}R(Z+\sqrt{T}\,\boldsymbol{\alpha})}\right] = V^{1/2}\,\mathbb{E}\left(\frac{\nu}{\nu' A\nu}\right)$$

$$= V^{1/2}\left[\frac{1}{2}A\boldsymbol{\theta}\; e^{-\zeta}\,\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2}+1)}{}_1F_1(\frac{d}{2},\frac{d}{2}+1;\zeta) + \frac{1}{2}(I_d-A)\,\boldsymbol{\theta}\; e^{-\zeta}\,\frac{\Gamma(\frac{d}{2}-1)}{\Gamma(\frac{d}{2})}{}_1F_1(\frac{d}{2}-1,\frac{d}{2};\zeta)\right]$$

$$= \sqrt{T}\frac{1}{d}VR'(RVR')^{-1}R\,\boldsymbol{\alpha}\; e^{-\zeta}\,{}_1F_1(\frac{d}{2},\frac{d}{2}+1;\zeta) + \sqrt{T}\frac{1}{d-2}\,e^{-\zeta}\left[\boldsymbol{\alpha} - VR'(RVR')^{-1}R\,\boldsymbol{\alpha}\right]{}_1F_1(\frac{d}{2}-1,\frac{d}{2};\zeta), \tag{A.41}$$

where $\nu = V^{-1/2}(Z+\sqrt{T}\,\boldsymbol{\alpha}) \sim N(\boldsymbol{\theta}, I_d)$ and the third equality above holds by using Lemma A3 and Lemma A5.

Now we derive the expression for the asymptotic MSE of the shrinkage estimator.

$$\lim_{T\to\infty}\mathbb{E}\left[T(\hat{\boldsymbol{\beta}}_s-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_s-\boldsymbol{\beta})'\right] = \mathbb{E}(Z_sZ_s') = \mathbb{E}\left[\bar{\Pi}_1 - \bar{\Pi}_2 - \bar{\Pi}_2' + \bar{\Pi}_3\right], \tag{A.42}$$

where

$$\bar{\Pi}_1 = ZZ',$$
$$\bar{\Pi}_2 = \frac{\tau}{\xi(Z)}VR'(RVR')^{-1}R(Z+\sqrt{T}\,\boldsymbol{\alpha})Z',$$
$$\bar{\Pi}_3 = \frac{\tau^2}{\xi^2(Z)}VR'(RVR')^{-1}R(Z+\sqrt{T}\,\boldsymbol{\alpha})(Z+\sqrt{T}\,\boldsymbol{\alpha})'R'(RVR')^{-1}RV.$$

Now, in the following, we derive the expectations of $\bar{\Pi}_1 - \bar{\Pi}_3$,

$$\mathbb{E}(\bar{\Pi}_1) = \mathbb{E}(ZZ') = V, \tag{A.43}$$

$$\mathbb{E}(\bar{\Pi}_2) = \tau VR'(RVR')^{-1}R\,\mathbb{E}\left(\frac{(Z+\sqrt{T}\,\boldsymbol{\alpha})Z'}{\xi(Z)}\right)$$

$$= \tau VR'(RVR')^{-1}RV^{1/2}\,\mathbb{E}\left[\frac{V^{-1/2}(Z+\sqrt{T}\,\boldsymbol{\alpha})(Z+\sqrt{T}\,\boldsymbol{\alpha})'V^{-1/2}V^{1/2}}{\xi(Z)} - \frac{V^{-1/2}(Z+\sqrt{T}\,\boldsymbol{\alpha})\sqrt{T}\,\boldsymbol{\alpha}'}{\xi(Z)}\right]$$

$$= \tau VR(RVR')^{-1}R\left[V^{1/2}\,\mathbb{E}\left[(\nu' A\nu)^{-1}\nu\nu'\right]V^{1/2} - V^{1/2}\,\mathbb{E}\left[(\nu' A\nu)^{-1}\nu\right]\sqrt{T}\,\boldsymbol{\alpha}'\right]$$

$$= \frac{\tau}{2}VR'(RVR')^{-1}R\,e^{-\zeta}\left[V^{1/2}AV^{1/2}\,\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2}+1)}\,{}_1F_1\left(\frac{d}{2},\frac{d}{2}+1;\zeta\right)\right.$$

$$
+ (V - V^{1/2}AV^{1/2})\,\frac{\Gamma(\frac{d}{2}-1)}{\Gamma(\frac{d}{2})}\,{}_1F_1\Big(\frac{d}{2}-1,\frac{d}{2};\zeta\Big) + V^{1/2}A\,\boldsymbol{\theta}\boldsymbol{\theta}'\,AV^{1/2}\,\frac{\Gamma(\frac{d}{2}+1)}{\Gamma(\frac{d}{2}+2)}\,{}_1F_1\Big(\frac{d}{2}+1,\frac{d}{2}+2;\zeta\Big)
$$

$$
+ V^{1/2}(I-A)\,\boldsymbol{\theta}\boldsymbol{\theta}'(I-A)V^{1/2}\,\frac{\Gamma(\frac{d}{2}-1)}{\Gamma(\frac{d}{2})}\,{}_1F_1\Big(\frac{d}{2}-1,\frac{d}{2};\zeta\Big)
$$

$$
+ V^{1/2}(\boldsymbol{\theta}\boldsymbol{\theta}'\,A + A\,\boldsymbol{\theta}\boldsymbol{\theta}'\,-2A\,\boldsymbol{\theta}\boldsymbol{\theta}'\,A)V^{1/2}\,\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2}+1)}\,{}_1F_1\Big(\frac{d}{2},\frac{d}{2}+1;\zeta\Big)
$$

$$
- V^{1/2}A\,\boldsymbol{\theta}\sqrt{T}\,\boldsymbol{\alpha}'\;\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2}+1)}\,{}_1F_1\Big(\frac{d}{2},\frac{d}{2}+1;\zeta\Big) - V^{1/2}(I-A)\,\boldsymbol{\theta}\sqrt{T}\,\boldsymbol{\alpha}'\;\frac{\Gamma(\frac{d}{2}-1)}{\Gamma(\frac{d}{2})}\,{}_1F_1\Big(\frac{d}{2}-1;\frac{d}{2};\zeta\Big)\Big]
$$

$$
= \tau\,\frac{1}{d}\,e^{-\zeta}\,{}_1F_1\Big(\frac{d}{2},\frac{d}{2}+1;\zeta\Big)VR'(RVR')^{-1}RV + \tau T\,e^{-\zeta}\,VR'(RVR')^{-1}R\,\boldsymbol{\alpha}\boldsymbol{\alpha}'\,R'(RVR')^{-1}RV
$$

$$
\Big[\frac{1}{d+2}\,{}_1F_1\Big(\frac{d}{2}+1,\frac{d}{2}+2;\zeta\Big) - \frac{1}{d}\,{}_1F_1\Big(\frac{d}{2},\frac{d}{2}+1;\zeta\Big)\Big], \tag{A.44}
$$

where the fourth equality holds by using Lemma A3, Lemma A4 and Lemma A5.

$$
\mathbb{E}(\bar{\Pi}_3) = \tau^2 VR'(RVR')^{-1}R\,\mathbb{E}\left[\frac{(Z+\sqrt{T}\,\boldsymbol{\alpha})(Z+\sqrt{T}\,\boldsymbol{\alpha})'}{\xi^2(Z)}\right]R'(RVR')^{-1}RV
$$

$$
= \tau^2 VR'(RVR')^{-1}RV^{1/2}\,\mathbb{E}\Big[(\nu' A\nu)^{-2}\nu\nu'\Big]V^{1/2}R'(RVR')^{-1}RV
$$

$$
= \tau^2\,\frac{1}{4}\,e^{-\zeta}\,VR'(RVR')^{-1}R\Bigg[V^{1/2}AV^{1/2}\,\frac{\Gamma(\frac{d}{2}-1)}{\Gamma(\frac{d}{2}+1)}\,{}_1F_1\Big(\frac{d}{2}-1,\frac{d}{2}+1;\zeta\Big)
$$

$$
+ (V - V^{1/2}AV^{1/2})\,\frac{\Gamma(\frac{d}{2}-2)}{\Gamma(\frac{d}{2})}\,{}_1F_1\Big(\frac{d}{2}-2,\frac{d}{2};\zeta\Big) + V^{1/2}A\,\boldsymbol{\theta}\boldsymbol{\theta}'\,AV^{1/2}\frac{1}{4}\,\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2}+2)}\,{}_1F_1\Big(\frac{d}{2},\frac{d}{2}+2;\zeta\Big)
$$

$$
+ V^{1/2}(I-A)\,\boldsymbol{\theta}\boldsymbol{\theta}'(I-A)V^{1/2}\,\frac{\Gamma(\frac{d}{2}-2)}{\Gamma(\frac{d}{2})}\,{}_1F_1\Big(\frac{d}{2}-2,\frac{d}{2};\zeta\Big)
$$

$$
+ V^{1/2}(\boldsymbol{\theta}\boldsymbol{\theta}'\,A + A\,\boldsymbol{\theta}\boldsymbol{\theta}'\,-2A\,\boldsymbol{\theta}\boldsymbol{\theta}'\,A)V^{1/2}\,\frac{\Gamma(\frac{d}{2}-1)}{\Gamma(\frac{d}{2}+1)}\,{}_1F_1\Big(\frac{d}{2}-1,\frac{d}{2}+1;\zeta\Big)\Bigg]R'(RVR')^{-1}RV
$$

$$
= \tau^2\,e^{-\zeta}\,\frac{1}{d(d-2)}\,VR'(RVR')^{-1}RV\,{}_1F_1\Big(\frac{d}{2}-1,\frac{d}{2}+1;\zeta\Big)
$$

$$
+ \tau^2\,e^{-\zeta}\,\frac{1}{d(d+2)}\,T\,VR'(RVR')^{-1}R\,\boldsymbol{\alpha}\boldsymbol{\alpha}'\,R'(RVR')^{-1}RV\,{}_1F_1\Big(\frac{d}{2},\frac{d}{2}+2;\zeta\Big), \tag{A.45}
$$

where the third equality holds by using Lemma A4 and Lemma A5.

Using the results in equations (A.43) - (A.45), the MSE of the shrinkage estimator is obtained.

■

Proof of Corollary 4.1 and 4.4 :

The results hold by noting that if $x > 0$ and $a, c > 0$, then as $x \to \infty$,

$$ {}_1F_1(a, c; x) = \frac{\Gamma(c)}{\Gamma(a)} e^{x} x^{-(c-a)} \left[\sum_{j=0}^{p-1} \frac{(c-a)_j (1-a)_j}{j!} x^{-j} + O(|x|^{-p}) \right]. $$

See Lebedev (1972), pp. 271.

■

Proof of Corollary 4.2 :

Note that we have

$$ \varrho_{min}(V^{1/2}AVWVAV^{1/2}) \leq \frac{\lambda_W}{\lambda} = \frac{\boldsymbol{\delta}'\, AVWVA\, \boldsymbol{\delta}}{\boldsymbol{\delta}'\, A\, \boldsymbol{\delta}} = \frac{\boldsymbol{\delta}'\, AV^{1/2}V^{1/2}AVWVAV^{1/2}V^{1/2}A\, \boldsymbol{\delta}}{\boldsymbol{\delta}'\, AV^{1/2}V^{1/2}A\, \boldsymbol{\delta}} \\ \leq \varrho_{max}(V^{1/2}AVWVAV^{1/2}), \tag{A.46} $$

and

$$ \varrho_{min}(V^{1/2}WV^{1/2}) \leq \frac{\lambda_W}{\lambda} = \frac{\boldsymbol{\delta}'\, AVWVA\, \boldsymbol{\delta}}{\boldsymbol{\delta}'\, A\, \boldsymbol{\delta}} = \frac{\boldsymbol{\delta}'\, AV^{1/2}V^{1/2}WV^{1/2}V^{1/2}A\, \boldsymbol{\delta}}{\boldsymbol{\delta}'\, AV^{1/2}V^{1/2}A\, \boldsymbol{\delta}} \leq \varrho_{max}(V^{1/2}WV^{1/2}), \tag{A.47} $$

by using $AVA = A$. [9]

Using equation (5.11), we have

$$ \begin{aligned} \text{Risk }(\hat{\boldsymbol{\beta}}_s) &= \text{tr}(W\,V) + e^{-\lambda}\, \frac{1}{d-2}\, {}_1F_1\Big(\frac{d}{2}-1\,,\, \frac{d}{2}\,;\, \lambda\Big) \Big[\frac{\lambda_W}{\lambda}\tau^2 - 2\tau\Big(\text{tr}(WVR'(RVR')^{-1}RV) - 2\frac{\lambda_W}{\lambda}\Big)\Big] \\ &\quad - e^{-\lambda}\, \frac{1}{d-2}\, {}_1F_1\Big(\frac{d}{2}-1\,,\, \frac{d}{2}+1\,;\, \lambda\Big) \Big(\frac{\lambda_W}{\lambda} - \frac{\text{tr}(WVR'(RVR')^{-1}RV)}{d}\Big) \Big[\tau^2 + 4\tau\Big] \\ &\leq \text{tr}(W\,V) + e^{-\lambda}\, \frac{1}{d-2}\, {}_1F_1\Big(\frac{d}{2}-1\,,\, \frac{d}{2}\,;\, \lambda\Big) \Big[\frac{\lambda_W}{\lambda}\tau^2 - 2\tau\Big(\text{tr}(WVR'(RVR')^{-1}RV) - 2\frac{\lambda_W}{\lambda}\Big)\Big], \end{aligned} \tag{A.48} $$

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

$$ \varrho_{min}(B) \leq \frac{\theta' B \theta}{\theta' \theta} \leq \varrho_{max}(B), $$

see Abadir and Magnus , 2005-Pages 181-182.

where $\lambda_W = \boldsymbol{\delta}' R'(RV_TR')^{-1}RV_TWV_TR'(RV_TR')^{-1}R\,\boldsymbol{\delta}\,/2$, and the last inequality holds because the third term on the right hand side of the first equality is non-positive when

$$\begin{aligned} & d \geq \operatorname{tr}(WVR'(RVR')^{-1}RV)\frac{\lambda}{\lambda_W} \Rightarrow \\ & d \geq \operatorname{tr}(WVR'(RVR')^{-1}RV)\max(\frac{\lambda}{\lambda_W}) = \operatorname{tr}(WVR'(RVR')^{-1}RV)/\varrho_{min}(V^{1/2}WV^{1/2}) \end{aligned} \tag{A.49}$$

where the last equality holds by equation (A.46). Note that when $W = V^{-1}$, the last line in equation (A.48) holds with equality, as the third term will be zero.

Also the second term of the same equality is negative when

$$\begin{aligned} \operatorname{tr}(WVR'(RVR')^{-1}RV)\frac{\lambda}{\lambda_W} &= \operatorname{tr}(V^{1/2}AVWVAV^{1/2})\frac{\lambda}{\lambda_W} > 2, \Rightarrow \\ & \operatorname{tr}(WVR'(RVR')^{-1}RV)\ \min(\frac{\lambda}{\lambda_W}) > 2 \Rightarrow \\ & \operatorname{tr}(WVR'(RVR')^{-1}RV)/\varrho_{max}(V^{1/2}AVWVAV^{1/2}) > 2 \end{aligned} \tag{A.50}$$

and

$$\begin{aligned} & 0 < \tau \leq 2\Big[\operatorname{tr}(WVR'(RVR')^{-1}RV)\frac{\lambda}{\lambda_W} - 2\Big] \text{ or } \Rightarrow \\ & 0 < \tau \leq 2\Big[\operatorname{tr}(WVR'(RVR')^{-1}RV)/\varrho_{max}(V^{1/2}AVWVAV^{1/2}) - 2\Big]. \end{aligned} \tag{A.51}$$

Therefore when $d \geq \operatorname{tr}(WVR'(RVR')^{-1}RV)/\varrho_{min}(V^{1/2}WV^{1/2})$,

$\operatorname{tr}(WVR'(RVR')^{-1}RV)\varrho_{max}(V^{1/2}AVWVAV^{1/2}) > 2$, and

$0 < \tau \leq 2\Big[\operatorname{tr}(WVR'(RVR')^{-1}RV)/\varrho_{max}(V^{1/2}AVWVAV^{1/2}) - 2\Big]$, we have

$$\text{Risk } (\hat{\boldsymbol{\beta}}_s) < \operatorname{tr}(W\,V) + e^{-\lambda}\,\frac{1}{d-2}\,{}_1F_1\Big(\frac{d}{2}-1\,,\,\frac{d}{2}\,;\,\lambda\Big)\Big[\frac{\lambda_W}{\lambda}\tau^2 - 2\tau\Big(\operatorname{tr}(WVR'(RVR')^{-1}RV) - 2\frac{\lambda_W}{\lambda}\Big)\Big]. \tag{A.52}$$

■

Proof of Theorem 5 :

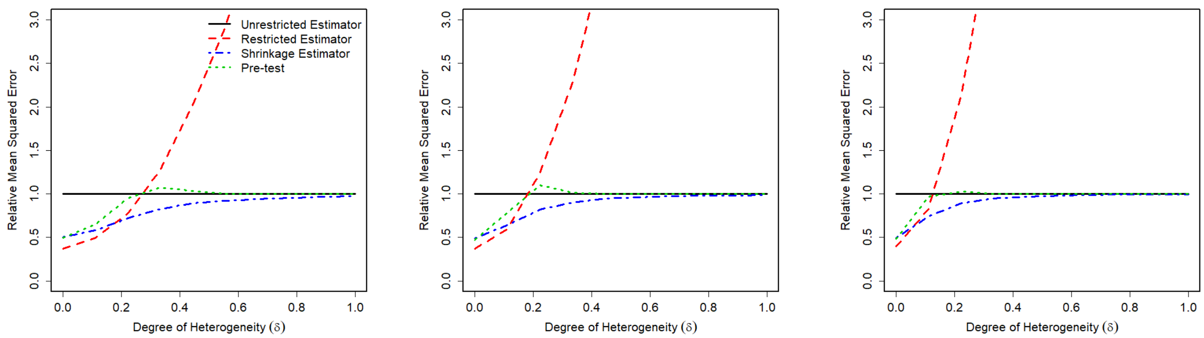
Using the results of Theorem 4 and Lemma A6, and noting that $\tau/d \to 1$ as $d \to \infty$, we have

$$\text{Risk } (\hat{\boldsymbol{\beta}}_s) < \operatorname{tr}(W\,V) - \tau[1 + O(\frac{1}{d})] \tag{A.53}$$

and by dividing both sides by the asymptotic risk of the FGLS estimator, and noting that and $\tau/tr(WV) \to \rho$ as $d \to \infty$, we have

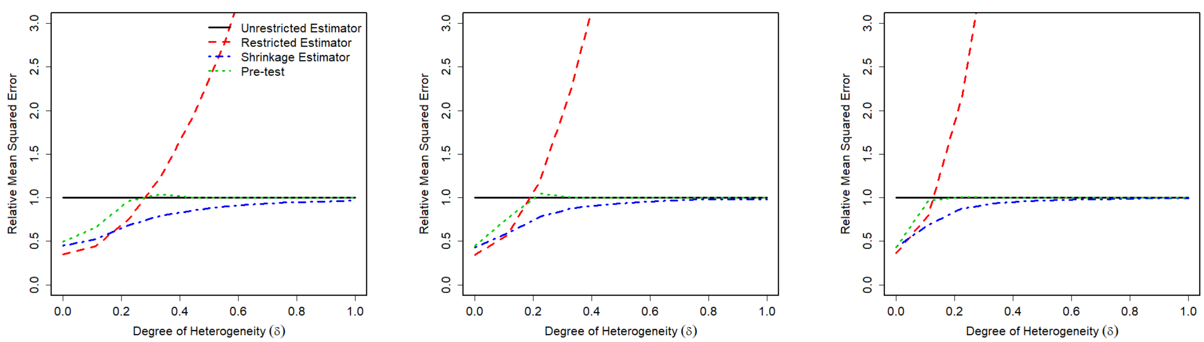
$$\lim_{d\to\infty} \frac{\text{ARisk }(\hat{\boldsymbol{\beta}}_s)}{\text{ARisk }(\hat{\boldsymbol{\beta}})} \leq 1 - \rho[1 + O(\frac{1}{d})], \qquad \rho = \lim_{d\to\infty} \frac{tr(C)}{tr(WV)}. \tag{A.54}$$

■



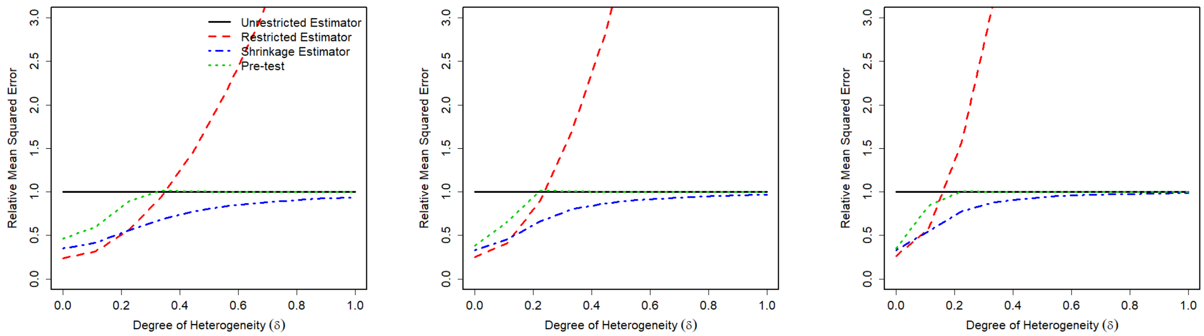
(a) T=50 (b) T=100 (c) T=200

Figure 1: Unrestricted, Restricted, Pre-test, and Shrinkage Estimators, DGP1, N=3, k=4



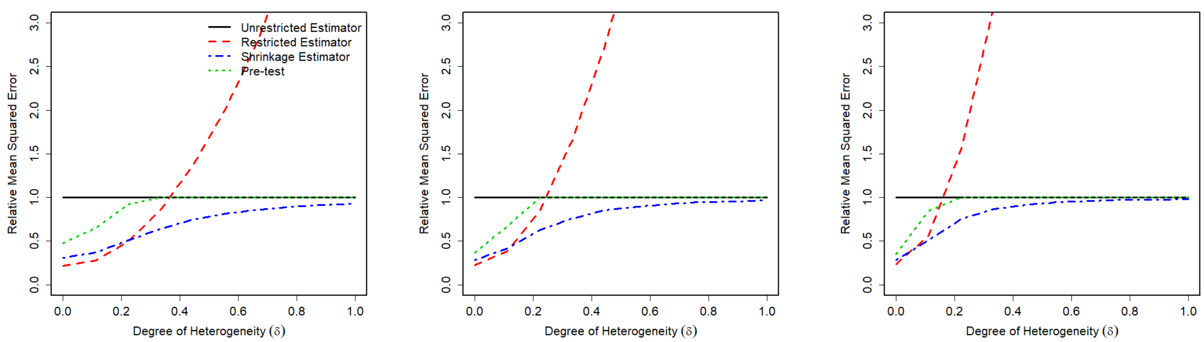
(a) T=50 (b) T=100 (c) T=200

Figure 2: Unrestricted, Restricted, Pre-test, and Shrinkage Estimators, DGP1, N=3, k=6



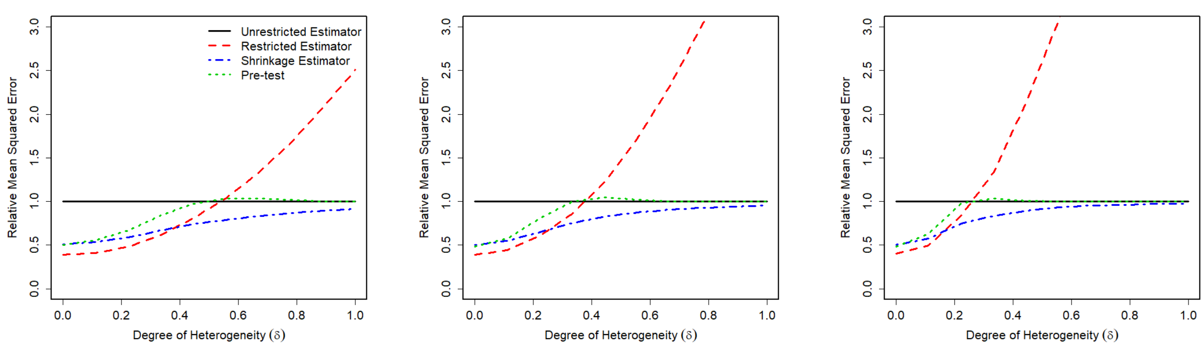
(a) T=50 (b) T=100 (c) T=200

Figure 3: Unrestricted, Restricted, Pre-test, and Shrinkage Estimators, DGP1, N=5, k=4



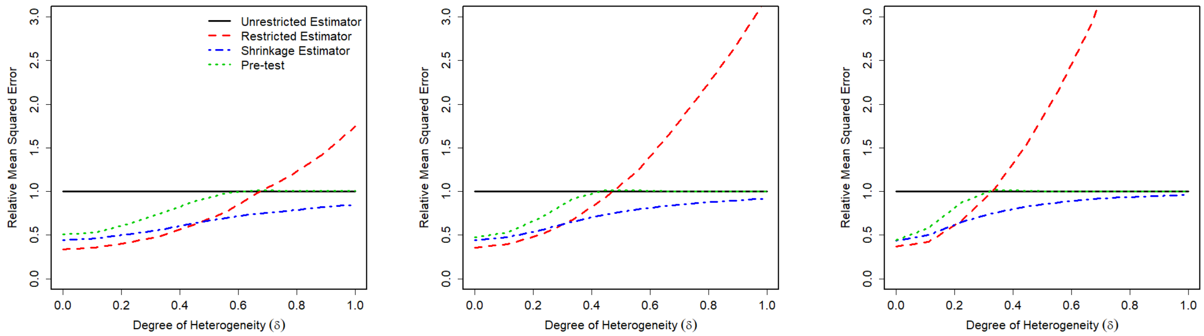
(a) T=50 (b) T=100 (c) T=200

Figure 4: Unrestricted, Restricted, Pre-test, and Shrinkage Estimators, DGP1, N=5, k=6



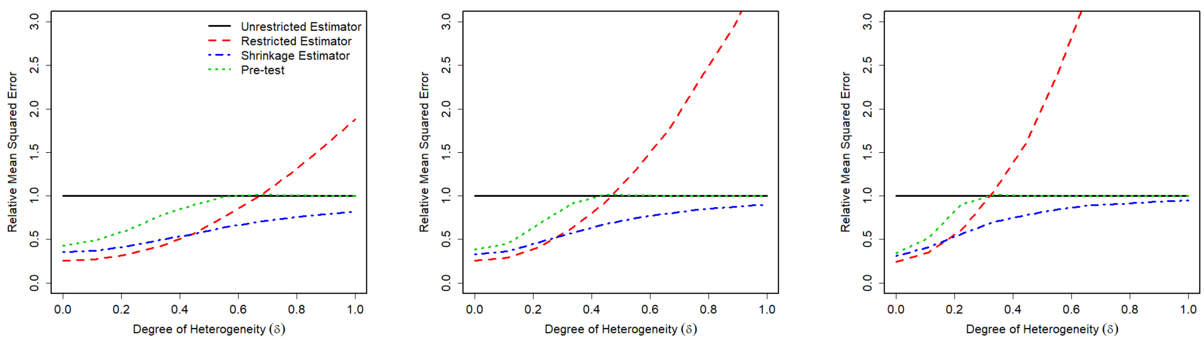
(a) T=50 (b) T=100 (c) T=200

Figure 5: Unrestricted, Restricted, Pre-test, and Shrinkage Estimators, DGP2, N=3, k=4



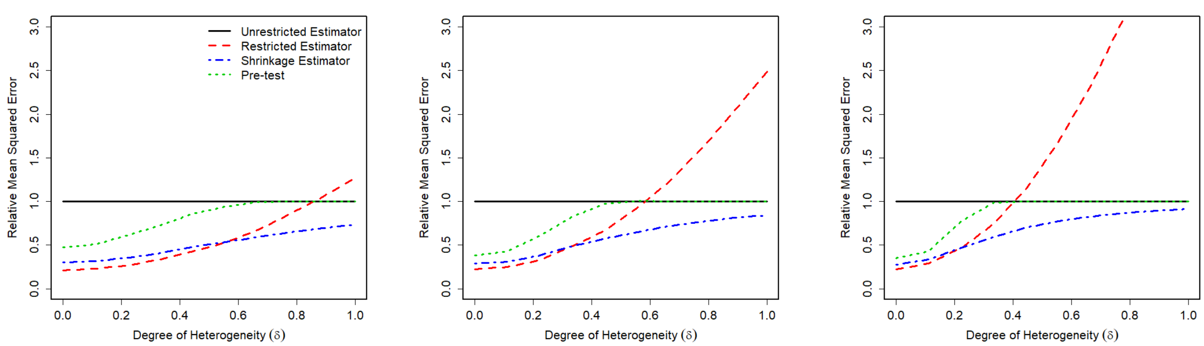
(a) T=50 (b) T=100 (c) T=200

Figure 6: Unrestricted, Restricted, Pre-test, and Shrinkage Estimators, DGP2, N=3, k=6



(a) T=50 (b) T=100 (c) T=200

Figure 7: Unrestricted, Restricted, Pre-test, and Shrinkage Estimators, DGP2, N=5, k=4



(a) T=50 (b) T=100 (c) T=200

Figure 8: Unrestricted, Restricted, Pre-test, and Shrinkage Estimators, DGP2, N=5, k=6

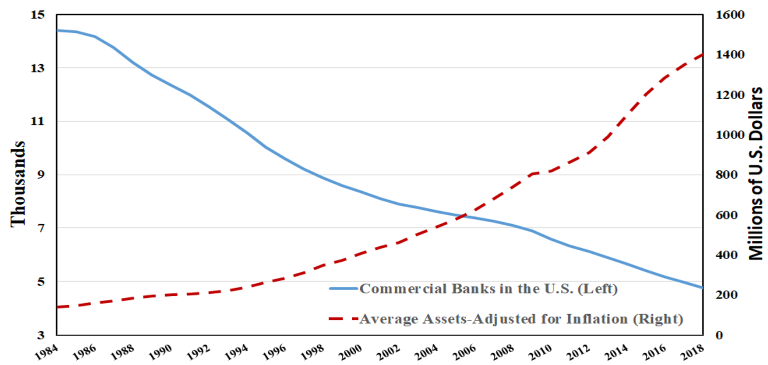


Figure 9: U.S. Commercial Banks, Number and Assets (Data from Federal Reserve Bank of St. Louis FRED database)

Table 1: Summary Statistics

Variable	**Min**		**Max**	
	2000	2018	2000	2018
C	136.49	151.78	18369359.04	26757643.39
w_1	7.01	15.32	163.36	261.24
w_2	0.01	0.01	12.38	43.47
w_3	0.00	0.00	0.21	0.05
w_4	0.00	0.00	0.26	0.15
w_5	0.00	0.00	0.05	0.02
y_1	1.25	1.00	42638250.00	173922000.00
y_2	0.25	170.50	168465250.00	686161250.00
y_3	45.75	125.50	178056500.00	457517750.00
y_4	89.50	0.25	144188250.00	703099250.00
y_5	39.00	42.50	86346000.00	704384250.00
z	4367.00	10157.00	584284000.00	2218960000.00
Variable	**Mean**		**STD**	
	2000	2018	2000	2018
C	24454.49	43803.47	331091.51	643872.50
w_1	26.18	49.39	7.08	15.33
w_2	0.23	0.27	0.35	0.92
w_3	0.03	0.00	0.01	0.00
w_4	0.01	0.00	0.01	0.00
w_5	0.02	0.00	0.00	0.00
y_1	57014.23	268717.11	754295.15	4722816.78
y_2	175843.57	881915.93	2961928.59	15632429.24
y_3	206836.22	929043.01	2423963.16	10526805.47
y_4	192820.92	862701.61	2985314.66	16392980.08
y_5	85019.93	842793.39	1554035.20	16868329.90
z	751262.62	3371943.77	10199407.63	53011852.59

Note: All variables are measured in thousands of dollars. Summary statistics for other years are available upon request.

Figure 10: Number of Banks in Each of the Twelve Subgroups from 2000 to 2018

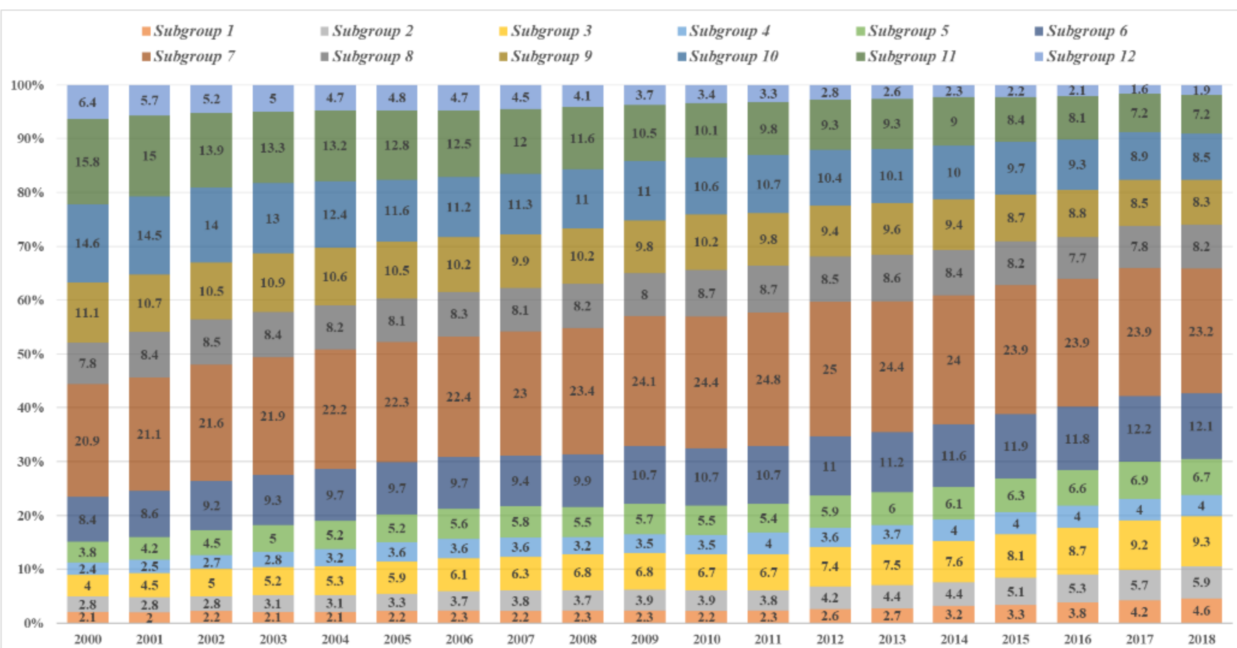


Figure 11: Share of Banks in Each of the Twelve Subgroups from 2000 to 2018

Figure 12: Mean of RTS for Each of Bank Asset Class from 2000 to 2018

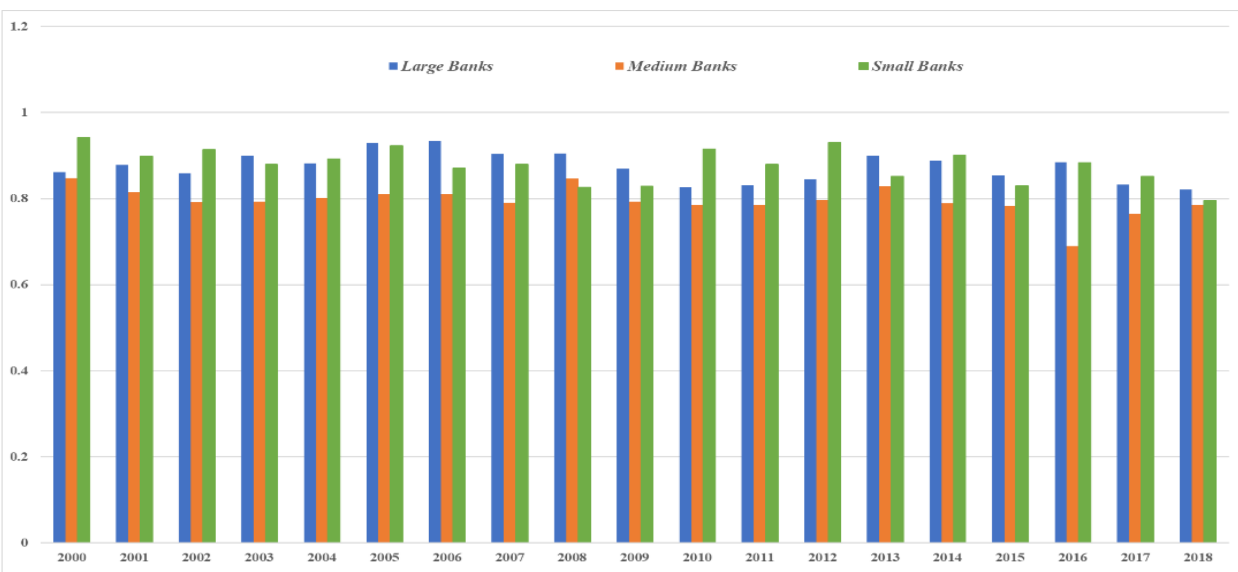


Figure 13: 10^{th} Decile RTS for Each of Bank Asset Class from 2000 to 2018

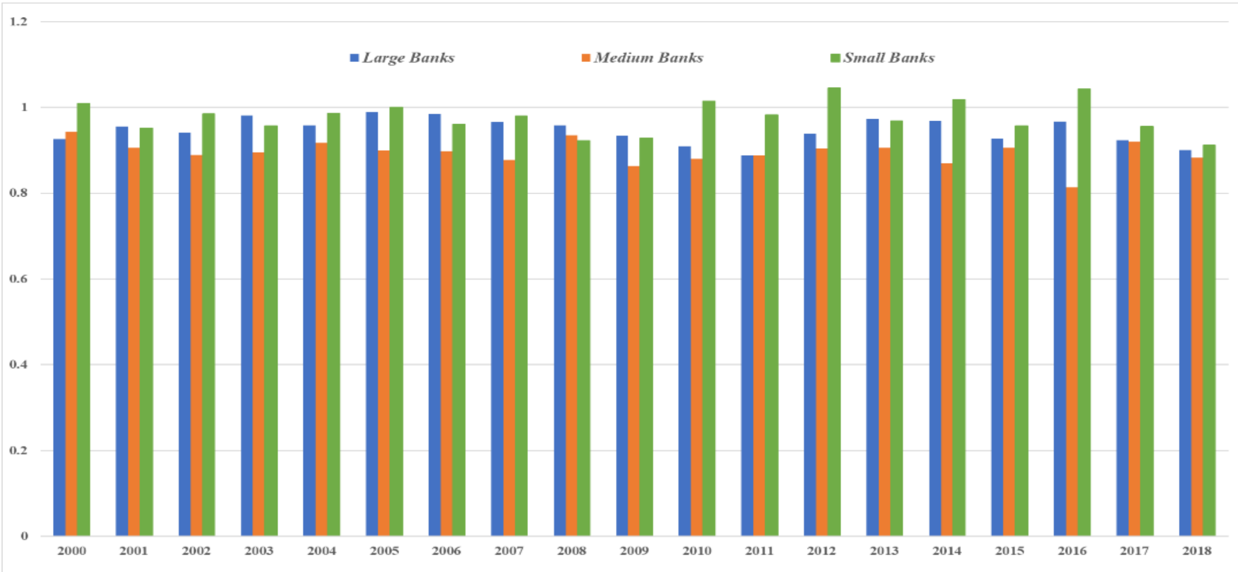


Figure 14: 25th Quartile of RTS for Each of Bank Asset Class from 2000 to 2018

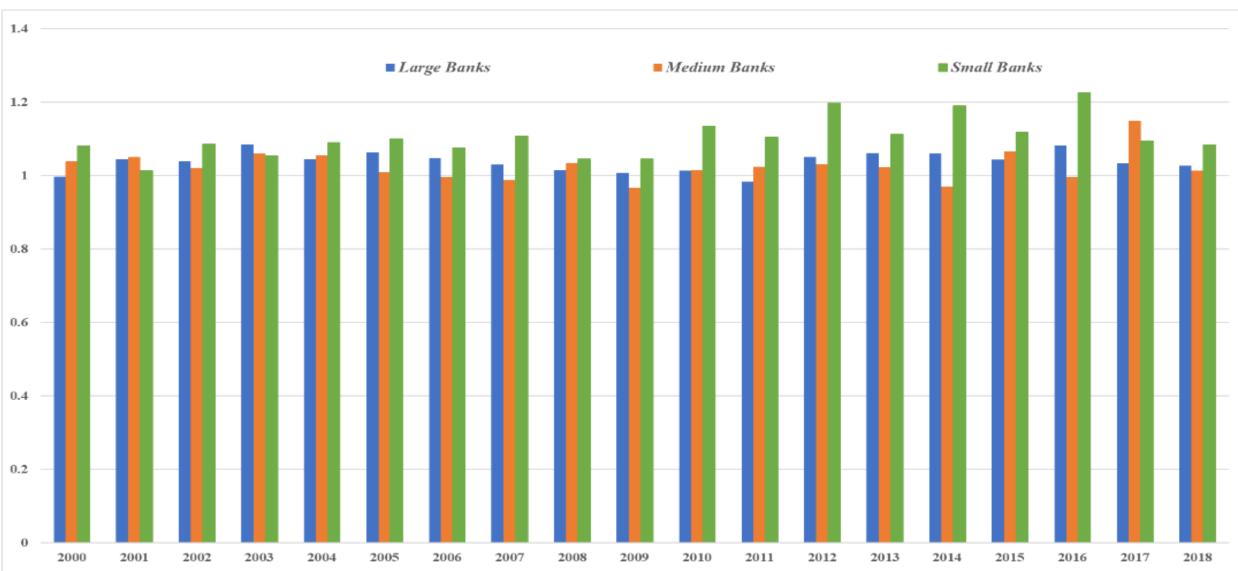


Figure 15: Median of RTS for Each of Bank Asset Class from 2000 to 2018

Figure 16: 75th Quartile of RTS for Each of Bank Asset Class from 2000 to 2018

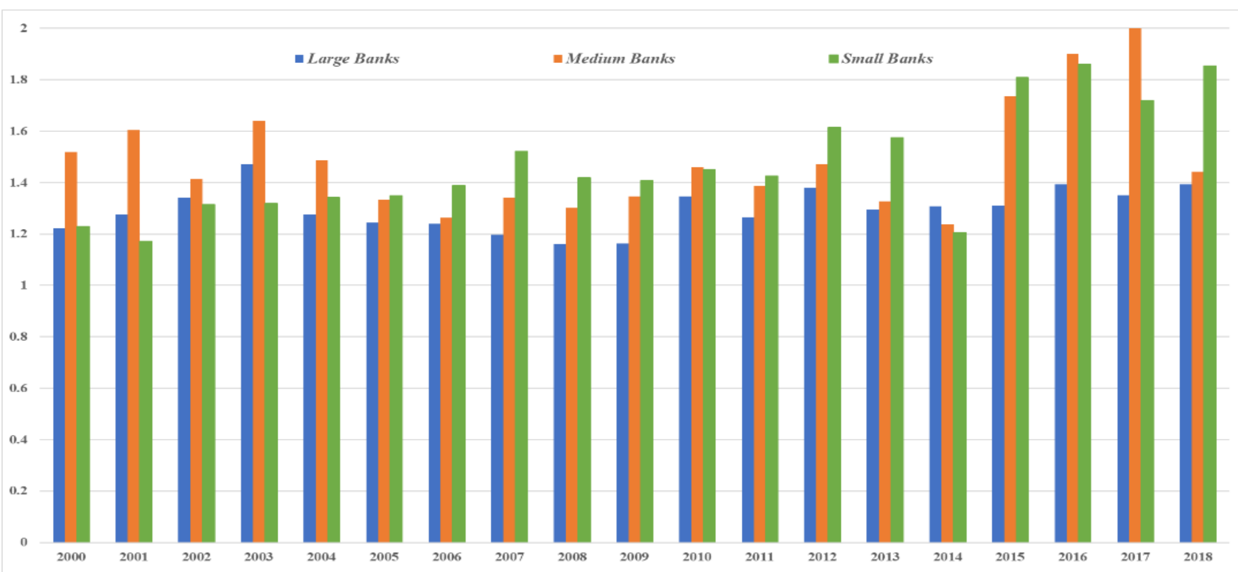


Figure 17: 90th Decile of RTS for Each of Bank Asset Class from 2000 to 2018