Stein-Like Shrinkage Estimators for Coefficients of a Single-Equation in Simultaneous Equation Systems

Ali Mehrabani*

Department of Economics

University of Kansas

Abstract

Two stein-like shrinkage estimators are introduced to modify the 2SLS and the LIML estimators for coefficients of a single equation in a simultaneous system of equations. The proposed estimators are weighted averages of the 2SLS/LIML estimators and the OLS estimator. The shrinkage weight depends on the Wu-Hausman misspecification test statistic which evaluates the null of exogeneity against the alternative hypothesis of endogeneity. The approximate finite sample bias, mean squared errors, and density functions of the Stein-like shrinkage estimators are obtained using small-disturbance approximations. The dominance conditions of the Stein-like shrinkage estimators over the 2SLS/LIML estimator under the mean squared error and the concentration probability are obtained. The proposed method is further illustrated by simulation studies which demonstrate the good finite sample performance of the method, and is also applied to an empirical application of returns to education.

Key Words: Stein-like Estimator; Small-Disturbance Approximations; Simultaneous Equation Models; OLS; 2SLS; LIML.

JEL Classification: C13, C26, C52

*Correspondence to: ali.mehrabani@ku.edu.

1 Introduction

Simultaneous equations models which arise from economic theory in terms of operations of markets and the simultaneous determination of economic variables through an equilibrium model, are one of the many developments in econometrics. The study of estimating coefficients of a single equation in a complete system of simultaneous structural equations has provided many estimation methods, including ordinary least squares (OLS), two-stage least squares (2SLS) and limited information maximum likelihood (LIML) which are the most commonly used ones. Because of the presence of endogeneity in the model, the OLS estimator is biased and inconsistent, however, the 2SLS and the LIML estimators under appropriate general conditions are consistent (see e.g., Anderson and Rubin (1949)). Since these estimators are available, numerous articles have focused on the finite-sample properties of these estimators and their modifications.

One direction of modifying these estimators, in the hope that the modified estimation method may improve the existing estimators, has been made by linearly combining these estimators. Sawa (1973a) and Sawa (1973b) propose a combined estimator, to eliminate the bias of the 2SLS estimator. The combined estimator is a simple linear combination of the OLS and the 2SLS estimators. The coefficients of this combined estimator depends on the sample size and the number of included and excluded variables from the relevant equation. Besides, the estimator is unbiased to a certain order. Similarly, Morimune (1978) proposes a set of combined estimators which are convex linear combinations of the LIML estimator and the fixed $k-$class estimators of Theil (1961). The aim of this method is to eliminate the small-disturbance asymptotic bias of the LIML estimator to construct improved estimators that are unbiased up to a certain order. Morimune shows the inadmissibility of the LIML estimator in terms of asymptotic mean squared errors (see also, Morimune and Kunitomo (1980) for the same method in the problem of functional relationships). A comparison of the above modified estimators is given by Anderson et al. (1986).

Another direction of the modification considers a nonlinear function of the existing estimators. Stein (1956) is the pioneer of this method. Stein (1956) shows that the maximum likelihood estimator (MLE) for the mean of a multivariate normal distribution does not have the smallest risk, i.e., MLE is inadmissible. Later on this issue, James and Stein (1961) suggested a biased estimator which dominates the MLE estimator in the sense that its risk is smaller than that of the former, provided that at least three parameters are to be estimated. In the context of a single equation estimation in a linear simultaneous equations system, Zellner and Vandaele (1975) consider Stein-type estimators under a general quadratic loss function and obtain a minimum risk estimator by applying 2SLS method. However, the resulting estimator is unavailable in applications as it involves certain unknown parameters. On this regard, Ullah and Srivastava (1988) present a Stein-type estimator and analyze its properties and conditions under which the resulting estimator dominates the 2SLS estimator.

In reduced form estimation, Maasoumi (1978) constructs a Stein-like estimator which is the weighted average of the Least Squares (LS) estimator and the Three-Stage-Least-Squares (3SLS) estimator of the reduced form coefficients in a linear simultaneous equations system, where the weight depends on the inverse of an over-identification test statistic. Maasoumi shows that this estimator has a few advantages over the LS and 3SLS estimators as it has finite moments, thinner tails, and has the edge on the LS estimator as it is asymptotically equivalent to the 3SLS estimator. Following Maasoumi (1978), in the context of single equation instrumental variable models, Hansen (2017) constructs a Stein-like estimator which is a weighted average of the OLS estimator and the 2SLS estimator for estimating the structural coefficients of the model. The weight is defined similar to Maasoumi (1978), while the Wu-Hausman (1978) specification test statistic is used. Using the local asymptotic theory, Hansen (2017) shows that the asymptotic risk of the Stein-like estimator is strictly less than that of the 2SLS estimator when the number of included endogenous variables are more than two. See also Mehrabani and Ullah (2020) for a Stein-like shrinkage estimator in seemingly unrelated regression models.

There are several approaches to compare the estimators and their associated modified estimators in the literature. One approach is to derive the exact distributions of the estimators (see e.g., Anderson and Sawa (1979) and Phillips (1984)). However, the analytical expressions of the distributions are usually too complicated to permit meaningful general conclusions. An alternative approach is to approximate each distribution by one or more terms in an asymptotic expansion of the distribution. One term, most of the time, is not enough as it is the common term between several estimators, but three terms serve to distinguish between the estimators (see for example Rothenberg (1984)). In addition, the approximate distribution terms can be used to approximate the moments of an estimator, when these exist, or to produce pseudo-moments (of an approximate distribution) when they do not, see Phillips (1983) for more discussion.

The asymptotic expansions have been derived on the basis of limits as an index tends towards a value. In the large-sample approximation, the number of observations increase without bound. In this context, Nagar (1959) shows that $k-$class estimators in simultaneous equations models can be expanded in formal series where the successive terms are increasing powers of $T^{1/2}$, where T is the number of observation for each dependent variable. Nagar (1959) obtains the moments of the truncated series by keeping the first few terms in the expansion. These moments can be interpreted as the moments of a statistic which serves to approximate the estimator, while Sargan (1974) shows that under some conditions, these moments can be interpreted as approximations to the actual moments of the estimator. In the small-disturbance approximation, initiated by Kadane (1971), it is suggested that it might be more natural to consider a sequence indexed by the error variance. In this analysis, the reduced-form error-covariance matrix is written as $\sigma\Omega$, and while the sample size and the matrix Ω are held fixed, σ approaches zero. The large-sample and small-disturbance approximations can be related by the effect of them on the non-centrality parameter, which goes to infinity in both cases while in the small-disturbance approximation the sample size stays constant. However, in the large-sample approximation the sample size and the non-centrality parameter both tend to infinity at the same speed (Anderson (1977)).

In this paper, we propose two Stein-like shrinkage estimators for coefficients of a single equation in a complete system of simultaneous equations. The estimators are weighted averages of the OLS and the 2SLS (or the LIML) estimators where the weights are inspired by the weight in Hansen (2017). We obtain the analytical expression of the bias and mean squared errors (MSE) of the estimators using small-disturbance approximation and give the conditions under which the Stein-like shrinkage estimators dominate the 2SLS estimator or the LIML estimator.

There are two related papers in the literature that similar to this paper consider combining the 2SLS estimator and the OLS estimator. The first one is Sawa (1973a) that by giving fixed weights to the OLS and the 2SLS estimators creates an unbiased estimator. The weights are $w_{S,OLS} = -(K-N-1)/(T-K)$ and $w_{S,2SLS} = (T-N-1)/(T-K)$ where N and K are the number of equations and the number of excluded regressor, respectively. Sawa (1973a) shows that the combined estimator is dominated by the 2SLS estimator in terms of having smaller MSE when the condition $(T-K-2)(K-N-7) \leq 12$ holds. Under the local endogeneity assumption considered in this paper, it is easy to show that Sawa's combined estimator is always dominated by the 2SLS estimator (for more details, see Remark 1 in Appendix B). Hence, the MSE of the combined estimator proposed by Sawa (1973a) is strictly greater than that of the Stein-like shrinkage estimator in this paper. The other related paper in the literature is Hansen (2017) which considers a Stein-like estimator in instrumental variable regression models. Hansen (2017) derives the dominance conditions of the Stein-like estimator over the 2SLS estimator by minimizing the truncated asymptotic weighted risk of the estimator using asymptotic distributions of the estimators. There are several advantages to our approach compared to Hansen (2017). First, the method considered here derives the approximate moments, and distributions, however the analysis in Hansen (2017) is dealing with asymptotically minimizing a truncated risk. Second, Hansen (2017) minimizes a weighted risk where the weight matrix is set equal to the inverse of the difference of the asymptotic variances of the 2SLS and the OLS estimators which might not be practical in most of the empirical applications. However, we derive the MSE matrix which allows for deriving a weighted risk with any positive definite weight matrix. Third, Hansen (2017) only considers shrinking the 2SLS estimator toward the OLS estimator, while, we consider two estimators where one shrinks the 2SLS estimator and the other shrinks the LIML estimator. This is important as under weak instruments scenario the 2SLS estimator is biased in the direction of the OLS estimator, while the LIML estimator needs weaker conditions for the consistency (see for example Chao and Swanson (2005)).

Morimune (1978) considers combining the LIML with the OLS estimator. Morimune

(1978) uses fixed weights with the purpose of removing the higher order bias of the LIML estimator and shows that while Sawa (1973a)'s combined estimator is dominated by the 2SLS estimator, combining the LIML estimator with the OLS estimator dominates the LIML estimator when $K - N > 0$ and $T > K + 2$. Although, the main goal of Morimune (1978) is different from this paper, by comparing the MSEs of the combined estimator with our proposed estimator, under the local endogeneity assumption, it can be easily verified that the Stein-like shrinkage estimator significantly performs better than Morimune (1978)'s estimator when the sample size is large enough, e.g., when $T > 2(K+1)$ (for more details, see Remark 2 in Appendix B).

The rest of the paper is organized as follows. Sections 2 and 3 describe the model and introduce the estimators. The approximate distributions, bias, and mean squared errors of the Stein-like shrinkage estimators are given in section 4. Monte-Carlo simulations and the application results are provided in sections 5 and 6. Conclusions are given in section 7. Proofs and detailed calculations are listed in Appendices A – C.

2 The Model

Consider the following complete simultaneous equations model

$$Y_{T\times(N+1)}B_{(N+1)\times(N+1)} + X_{T\times K}\Gamma_{K\times(N+1)} = \sigma U_{T\times(N+1)}, \tag{2.1}$$

where in the system above, there are $N+1$ equations and $N+1$ endogenous variables, denoted by $Y = (y_1, y_2, \ldots, y_{(N+1)})$, there are K exogenous variables, $X = (x_1, x_2, \ldots, x_K)$, and $U = (u_1, u_2, \ldots, u_{(N+1)})$ are the structural disturbances. Each y_i, x_i, and u_i is a vector of $T \times 1$, where for example $y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$ and T is the number of observations in each equation. B is a nonsingular matrix of parameters with first column $(-1, \beta')'$, where β is a $N \times 1$ vector of unknown coefficients of interest in the first equation, and σ is a positive number.

The first equation of the above system, by assuming for simplicity that it includes no exogenous variables, may be written as

$$y_1 = Y_2\beta + \sigma u_1, \tag{2.2}$$

where y_1 is the first column of Y, and $Y_2 = (y_2, \ldots, y_{(N+1)})$ is $T \times N$, that contains the rest of the columns of Y and is the included endogenous variables. The model in (2.2) can be generalized to include exogenous variables in the first equation, but the results of the paper remain the same. Hence, for the sake of simplicity, we assume that the first equation of the system does not include exogenous variables.

We make the following assumptions about the distribution of the disturbances:

Assumption 1 *The rows of U are independently normally distributed with mean zero and variance-covariance matrix Σ. That is for all t and t' in $\{1, \ldots, T\}$ and i and $j \in \{1, \ldots, N\}$,*

$$\mathbb{E}(u_{it}) = 0,$$
$$Cov(u_{it}, u_{jt'}) = \begin{cases} \sigma_{ij} & \text{if } t = t' \\ 0 & \text{otherwise,} \end{cases}$$

and $\sigma_{11} = 1$. In matrix notation

$$\mathbb{E}(U) = \underset{T\times(N+1)}{\mathbf{0}},$$
$$\frac{1}{T}\,\mathbb{E}(U'U) = \underset{(N+1)\times(N+1)}{\Sigma} = \begin{bmatrix} 1 & \sigma_{12} & \ldots & \sigma_{1(N+1)} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2(N+1)} \\ & & \vdots & \\ \sigma_{(N+1)1} & \sigma_{(N+1)2} & \ldots & \sigma_{(N+1)(N+1)} \end{bmatrix} \equiv \begin{bmatrix} \sigma_1 & \sigma_2 & \ldots & \sigma_{N+1} \end{bmatrix}.$$

The reduced form of the structural equation (2.1) may be written as

$$Y = -X\Gamma B^{-1} + \sigma U B^{-1} \equiv X\Pi + \sigma V, \tag{2.3}$$

where $\Pi = -\Gamma B^{-1}$, $V = UB^{-1}$, $\Pi_{K\times(N+1)} = \begin{bmatrix} \underset{K\times 1}{\pi_1} & \underset{K\times N}{\Pi_2} \end{bmatrix}$, and $V_{T\times(N+1)} = \begin{bmatrix} \underset{T\times 1}{v_1} & \underset{T\times N}{V_2} \end{bmatrix}$.

Further, if we partition B^{-1} as

$$B^{-1} = \left[\dot{\beta}_{(N+1)\times 1} \quad \dot{B}_{(N+1)\times N}\right],$$

the reduced form system of equations can be written in partition as

$$y_1 = -X\Gamma\dot{\beta} + \sigma U\dot{\beta} = X\pi_1 + \sigma v_1, \tag{2.4}$$

and

$$Y_2 = -X\Gamma\dot{B} + \sigma U\dot{B} = X\Pi_2 + \sigma V_2 \equiv W + \sigma V_2, \tag{2.5}$$

where we define $W = X\Pi_2$.

Assumption 2 *Identification: Rank*$(\Pi_2) = N \leq K$.

Assumption 2 is the rank condition which ensures the identification of the system. Note also that for the case where $K - N > 2$, the first two moments of the 2SLS estimator exist, while it is well-known that the LIML estimator has no positive integer moments (see e.g., Mariano and Sawa (1972), Mariano and McDonald (1979), Phillips (1984), and Phillips (1985)). In this paper, we use small-disturbance expansions to approximate the distributions of the estimators of the parameters of the model, which then can be used to approximate the moments of these estimators where these exist, or to produce pseudo-moments of these estimators where they do not exist. For a discussion about the validity of the Nagar-type expansion of the k-class estimators see Sargan (1974).

Under Assumption 1, the reduced form error is also normally distributed with

$$\mathbb{E}(V) = \mathbf{0},$$

$$\frac{1}{T}\,\mathbb{E}(V'V) = \frac{1}{T}\,\mathbb{E}(B^{'-1}U'UB^{-1}) = B^{'-1}\Sigma B^{-1} = \underset{(N+1)\times(N+1)}{\Omega} \equiv \begin{bmatrix} \underset{1\times 1}{\varpi_{11}} & \underset{1\times N}{\boldsymbol{\varpi_{12}}} \\ \underset{N\times 1}{\boldsymbol{\varpi_{21}}} & \underset{N\times N}{\Omega_{22}} \end{bmatrix}.$$

Following Nagar (1959), we define $\underset{N\times T}{\Psi'} = V_2' - qu_1'$, where the normally distributed matrix Ψ consists of residuals from the population regression of V_2 on u_1. Hence Ψ and u_1 are uncorrelated by construction. In addition, let

$$\underset{N\times 1}{q} = \frac{Cov(V_2, u_1)}{Var(u_1)} = \frac{\mathbb{E}(V_2' u_1)}{T} = \dot{B}'\sigma_1, \tag{2.6}$$

and define

$$\underset{N\times N}{C_1} = qq', \text{ and } \underset{N\times N}{C_2} = \frac{\mathbb{E}(\Psi'\Psi)}{T} = \dot{B}'\Sigma\dot{B} - qq', \text{ where it can be shown that } \Omega_{22} = C_1 + C_2. \tag{2.7}$$

We assume the following local endogeneity assumption.

Assumption 3 *Local Endogeneity:* $q = Cov(V_2, u_1)/T = \sigma\delta$, *where* $\delta_{N\times 1} \in R^N$.

We note that the local endogeneity assumption here is similar to the local asymptotic considered in Hansen (2017), where σ is replaced by $1/\sqrt{T}$. The local endogeneity assumption in Hansen (2017) is needed so that the estimator has asymptotically a non-degenerate asymptotic distribution. However, we need this assumption to simplify the derivations, see the discussion in Appendix A for more details.

3 Estimators

We consider three members of the $k-$class estimator of $\dot{\beta}$. The estimators are the OLS, the 2SLS, and the LIML estimators, which respectively correspond to $k = 0$, $k = 1$, and $k = \lambda$ where $\lambda = \tilde{\beta}'Y'Y\tilde{\beta}/\tilde{\beta}'Y'M_XY\tilde{\beta}$, $M_X = I_T - P_X$ is the projection onto the space orthogonal to the columns of X, with $P_X = X(X'X)^{-1}X'$, and I_T is the identity matrix. Moreover, we introduce two types of Stein-like shrinkage estimators which are, a weighted average of the 2SLS and the OLS estimators, and a weighted average of the LIML estimator and the OLS estimator.

3.1 $k-$Class Estimators

The $k-$class estimator is defined as

$$\hat{\beta}(k) = (Y_2'H_kY_2)^{-1}Y_2'H_ky_1 = \beta + \sigma(Y_2'H_kY_2)^{-1}Y_2'H_ku_1, \tag{3.1}$$

where $H_k = I_T - kM_X$.

3.2 Stein-Like Shrinkage Estimators

We define the Stein-like shrinkage estimators as the weighted averages of a first-order consistent $k-$class estimator (we consider the 2SLS estimator ($k = 1$), and the LIML ($k = \lambda$)) with the OLS estimator ($k = 0$), where the weights are inversely related to the Wu-Hausman (1978) misspecification test statistic. Hence, the Stein-like shrinkage estimators are defined as

$$\hat{\beta}_{c,k} = \omega_k\hat{\beta}(0) + (1 - \omega_k)\hat{\beta}(k), \quad \text{for } k = 1, \lambda \tag{3.2}$$

where $\omega_k = \tau/F_{k,\mathrm{WH}}$, and τ is a positive characterizing scalar which will be determined later. $F_{k,\mathrm{WH}}$ is the Wu-Hausman statistic test, defined as

$$F_{k,\mathrm{WH}} = \left(\hat{\beta}(k) - \hat{\beta}(0)\right)' \mathbb{R}_k \left(\hat{\beta}(k) - \hat{\beta}(0)\right), \tag{3.3}$$

and $\mathbb{R}_k$ is defined as

$$\mathbb{R}_k = \hat{\sigma}_{11,k}^{-1}\Big((Y_2'HY_2)^{-1} - (Y_2'Y_2)^{-1}\Big)^{-1}, \tag{3.4}$$

where $\hat{\sigma}_{11,k} = \hat{u}_1(k)'\hat{u}_1(k)/(T - N)$, in which $\hat{u}_1(k) = y_1 - Y_2\hat{\beta}(k)$.

4 The Approximate Distributions and MSE

Since the Stein-like shrinkage estimators are non-linear functions of the OLS and the 2SLS/LIML estimators, and consequently non-linear functions of the error terms, we first derive the approximate distributions of the Stein-like shrinkage estimators by following Kadane (1971) small-disturbance method. Then, we obtain the bias and mean squared error (MSE) matrices of the estimators up to orders σ^2 and σ^4, respectively.

The approximate density functions of the Stein-like shrinkage estimators are derived for the statistics

$$\hat{e}_{c,k} = \frac{1}{\sigma}(\hat{\beta}_{c,k} - \dot{\beta}), \quad \text{for } k = 1, \lambda, \tag{4.1}$$

as σ goes to zero.

Theorem 1 *Under assumptions 1–3, the asymptotic expansions of the density functions of $\hat{e}_{c,k}$ for $k = 1, \lambda$, as σ goes to zero are*

$$f_{c,1}(\xi) = f_1(\xi) + \phi_Q(\xi)\sigma^2 \frac{\tau}{T-N}\left[\alpha_1\delta'\xi + \frac{1}{2}\Big[\xi'C_2\xi - tr(C_2Q)\Big]\Big(\tau\alpha_2 - 2\alpha_1\Big)\right] + O(\sigma^3), \tag{4.2}$$

$$f_{c,\lambda}(\xi) = f_\lambda(\xi) + \phi_Q(\xi)\sigma^2 \frac{\tau}{T-N}\left[\alpha_1\delta'\xi + \frac{1}{2}\Big[\xi'C_2\xi - tr(C_2Q)\Big]\Big(\tau\alpha_3 - 2\alpha_1\Big)\right] + O(\sigma^3), \tag{4.3}$$

where ξ is an $N \times 1$ vector, and $f_1(\cdot)$ and $f_\lambda(\cdot)$ are the asymptotic expansions of the density functions of the 2SLS estimator and the LIML estimator, respectively. Also,

$$\alpha_1 = \frac{(T-K)(T-N)}{N}, \quad \alpha_2 = \frac{(T-K)(T-N-2)}{N(N-2)}, \quad \alpha_3 = \alpha_2 + c, \;\; \textit{and } c \leq 0.$$

Proof: Appendix B, (see page 37).

The expressions for $f_1(\cdot)$ and $f_\lambda(\cdot)$ are given in Lemma B7 in Appendix B. In the next theorem, the first and the second moments of the Stein-like shrinkage estimators based on the approximate expansions of their distributions are given.

Theorem 2 *Under assumptions 1–3, the asymptotic bias of the Stein-like shrinkage estimator* $\hat{\beta}_{c,k}$ *for* $k = 1, \lambda$, *as* σ *goes to zero is*

$$ABias(\hat{\beta}_{c,k}) = \mathbb{E}\left(\frac{1}{\sigma}(\hat{\beta}_{c,k} - \dot{\beta})\right) = 0 + O(\sigma^2), \tag{4.4}$$

and the asymptotic MSE matrices of the Stein-like shrinkage estimators are

$$AMSE(\hat{\beta}_{c,1}) = \mathbb{E}\left(\frac{1}{\sigma^2}(\hat{\beta}_{c,1} - \dot{\beta})(\hat{\beta}_{c,1} - \dot{\beta})'\right) = AMSE(\hat{\beta}(1)) + \frac{\tau}{T-N}\sigma^2\Big[\tau\alpha_2 - 2\alpha_1\Big]QC_2Q + O(\sigma^3), \tag{4.5}$$

$$AMSE(\hat{\beta}_{c,\lambda}) = \mathbb{E}\left(\frac{1}{\sigma^2}(\hat{\beta}_{c,\lambda} - \dot{\beta})(\hat{\beta}_{c,\lambda} - \dot{\beta})'\right) \leq AMSE(\hat{\beta}(\lambda)) + \frac{\tau}{T-N}\sigma^2\Big[\tau\alpha_2 - 2\alpha_1\Big]QC_2Q + O(\sigma^3), \tag{4.6}$$

where from Lemma B7 (equation (B.6)), we have

$$AMSE(\hat{\beta}(1)) = Q + \sigma^2\, tr(QC_2)Q + \sigma^2 QC_2Q\Big(2 - L_1\Big) + O(\sigma^3), \tag{4.7}$$

$$AMSE(\hat{\beta}(\lambda)) = Q + \sigma^2\, tr(QC_2)Q + \sigma^2 QC_2Q\frac{(L_1+2)(T-N-2)}{T-K-2} + O(\sigma^3). \tag{4.8}$$

Proof: Appendix B, (see page 40).

Corollary 2.1 *Under assumptions 1–3, we have*

$$AMSE(\hat{\beta}_{c,1}) - AMSE(\hat{\beta}(1)) = \frac{\tau}{T-N}\sigma^2 QC_2Q\Big[\tau\alpha_2 - 2\alpha_1\Big] + O(\sigma^3), \tag{4.9}$$

$$AMSE(\hat{\beta}_{c,\lambda}) - AMSE(\hat{\beta}(\lambda)) \leq \frac{\tau}{T-N}\sigma^2 QC_2Q\Big[\tau\alpha_2 - 2\alpha_1\Big] + O(\sigma^3), \tag{4.10}$$

where the right-hand side of the above equations are negative when $T - N > 2$ *and*

$$0 < \tau < \frac{2(T-N)(N-2)}{T-N-2}.$$

Therefore, The Stein-like shrinkage estimators dominate the 2SLS, and LIML estimators in terms of their MSEs when the number of included endogenous variables is more than 2. The *optimal value of the shrinkage parameter that minimizes the MSE of the Stein-like estimators is*

$$\tau_{opt} = \frac{(T-N)(N-2)}{T-N-2}.$$

As a comparison of the probability of concentration around the true $\dot{\beta}$, we compute

$$P(||Q^{-1/2}\hat{e}_{c,k}|| < z) - P(||Q^{-1/2}\hat{e}_k|| < z) = \int \cdots \int_{||Q^{-1/2}\xi|| < z} (f_{c,k}(\xi) - f_k(\xi))\, d\xi, \tag{4.11}$$

where $||\xi|| = \max\{|\xi_1|, \ldots, |\xi_N|\}$. Using Lemma B7 and Theorem 1 the next theorem follows.

Theorem 3 *Under assumptions 1–3,*

$$P(||Q^{-1/2}\hat{e}_{c,1}|| < z) - P(||Q^{-1/2}\hat{e}_1|| < z) = \sigma^2[\Phi(z) - \Phi(-z)]^N z\tilde{\phi}(z)\, tr(QC_2)d + O(\sigma^3), \tag{4.12}$$

$$P(||Q^{-1/2}\hat{e}_{c,\lambda}|| < z) - P(||Q^{-1/2}\hat{e}_\lambda|| < z) \geq \sigma^2[\Phi(z) - \Phi(-z)]^N z\tilde{\phi}(z)\, tr(QC_2)d + O(\sigma^3), \tag{4.13}$$

where $\tilde{\phi}(z) = \phi(z)/[\Phi(z) - \Phi(-z)]$, $d = \tau(2\alpha_1 - \tau\alpha_2)/(T-N)$, and $\Phi(\cdot)$ and $\phi(\cdot)$ are, respectively, the standard normal distribution and density functions.

Proof: Appendix B, (see page 41).

Corollary 3.1 *By Theorem 3, provided that $0 < \tau < 2(T-N)(N-2)/(T-N-2)$, and $T-N > 2$, we obtain*

$$P(||Q^{-1/2}\hat{e}_{c,k}|| < z) \geq P(||Q^{-1/2}\hat{e}_k|| < z) + O(\sigma^3), \quad k = 1, \lambda, \tag{4.14}$$

and the optimal value of τ that maximizes the concentration probability of the Stein-like shrinkage estimator is

$$\tau_{opt} = \frac{(T-N)(N-2)}{T-N-2}.$$

Table 1: Relative Median Squared Errors when $(T, N) = (100, 1)$

		$K=6$							$K=18$						
R^2	ρ	$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$
	0.01	0.130	0.674	0.136	0.054	0.665	0.057	2.365	0.260	0.747	0.288	0.036	0.668	0.044	6.403
	0.1	0.294	0.774	0.324	0.146	0.663	0.163	1.720	0.490	0.851	0.525	0.069	0.679	0.081	5.670
0.1	0.5	4.423	1.112	4.066	3.214	1.137	3.026	1.407	2.126	1.127	1.973	1.952	1.010	1.855	0.975
	0.9	7.431	1.043	1.061	16.935	0.986	1.066	0.415	2.143	1.018	1.002	12.190	1.004	1.175	0.173
	0.99	7.478	1.019	1.016	21.339	0.982	1.000	0.338	2.168	1.009	1.000	20.123	0.942	1.006	0.101
	0.01	0.445	0.854	0.485	0.396	0.833	0.433	1.096	0.509	0.890	0.568	0.377	0.865	0.430	1.311
	0.1	0.855	0.995	0.980	0.751	0.957	0.911	1.096	0.804	0.950	0.895	0.573	0.883	0.683	1.304
0.5	0.5	12.751	1.024	1.015	12.930	1.005	1.035	0.968	9.123	1.054	1.069	12.816	0.992	1.162	0.670
	0.9	43.458	1.022	1.000	42.806	0.976	1.000	0.970	11.000	1.020	1.000	45.395	1.011	1.000	0.240
	0.99	46.472	1.028	1.000	54.679	1.022	1.000	0.845	10.691	1.013	1.000	52.755	1.013	1.000	0.203
	0.01	0.882	0.977	0.927	0.864	0.983	0.882	1.026	0.872	0.982	0.903	0.787	0.908	0.787	1.024
	0.1	1.015	1.005	1.040	1.028	1.031	1.014	1.013	1.069	1.033	1.081	1.087	1.045	1.099	0.995
0.9	0.5	5.727	1.006	1.002	5.817	0.978	1.009	0.957	4.747	1.006	1.000	5.193	1.040	1.250	0.944
	0.9	16.956	1.007	1.000	16.511	1.012	1.000	1.032	14.032	0.995	1.000	17.258	1.021	1.000	0.834
	0.99	19.944	0.998	1.000	19.767	0.998	1.000	1.009	12.743	1.008	1.000	17.187	1.007	1.000	0.741

Note: This table reports the relative median squared errors of the OLS ($\hat{\beta}(0)$), the 2SLS ($\hat{\beta}(1)$), the LIML ($\hat{\beta}(\lambda)$) estimators, the Stein-like shrinkage estimator using the OLS and the 2SLS estimators ($\hat{\beta}_{c,1}$), the Stein-like shrinkage estimator using the OLS and the LIML estimators ($\hat{\beta}_{c,\lambda}$), and two pre-test estimators ($\hat{\beta}_{pre}$), i.e., $\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$ indicates the median squared errors of the OLS estimator divided by the median squared errors of the 2SLS estimator. The pre-test estimators use the Wu-Hausman test static under 5% critical value to choose between the OLS and the 2SLS/LIML estimators.

Table 2: Relative Median Squared Errors when $(T, N) = (100, 3)$

		$K=6$							$K=18$						
R^2	ρ	$\frac{\hat\beta(0)}{\hat\beta(1)}$	$\frac{\hat\beta_{c,1}}{\hat\beta(1)}$	$\frac{\hat\beta_{pre}}{\hat\beta(1)}$	$\frac{\hat\beta(0)}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{pre}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta_{c,1}}$	$\frac{\hat\beta(0)}{\hat\beta(1)}$	$\frac{\hat\beta_{c,1}}{\hat\beta(1)}$	$\frac{\hat\beta_{pre}}{\hat\beta(1)}$	$\frac{\hat\beta(0)}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{pre}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta_{c,1}}$
0.1	0.01	0.058	0.164	0.058	0.015	0.034	0.015	0.780	0.186	0.435	0.188	0.008	0.052	0.008	2.862
	0.1	0.075	0.169	0.075	0.016	0.031	0.016	0.832	0.253	0.494	0.254	0.012	0.056	0.012	2.401
	0.5	0.562	0.570	0.562	0.156	0.165	0.156	1.048	0.994	0.954	0.994	0.107	0.134	0.107	1.306
	0.9	1.685	1.078	1.680	0.540	0.478	0.540	1.384	1.258	1.067	1.251	0.375	0.355	0.374	1.117
	0.99	2.198	1.124	2.189	0.874	0.678	0.873	1.518	1.330	1.038	1.160	0.475	0.454	0.475	1.225
0.5	0.01	0.262	0.501	0.267	0.207	0.457	0.211	1.155	0.314	0.588	0.327	0.119	0.481	0.124	2.146
	0.1	0.320	0.540	0.323	0.257	0.498	0.260	1.147	0.441	0.672	0.463	0.177	0.467	0.187	1.733
	0.5	2.155	0.939	1.896	1.779	0.880	1.641	1.135	2.124	1.085	1.849	1.182	0.800	1.149	1.324
	0.9	7.078	1.001	1.000	6.017	0.915	1.000	1.075	4.064	1.100	1.000	4.302	0.894	1.001	0.768
	0.99	8.320	1.001	1.000	6.960	0.909	1.000	1.086	4.384	1.077	1.000	5.282	0.901	1.000	0.694
0.9	0.01	0.801	0.890	0.815	0.788	0.880	0.804	1.005	0.805	0.909	0.848	0.714	0.836	0.723	1.036
	0.1	0.796	0.898	0.830	0.779	0.878	0.808	0.998	0.889	0.913	0.894	0.781	0.832	0.784	1.038
	0.5	2.391	0.976	1.024	2.349	0.981	1.041	1.023	2.566	1.032	1.053	2.352	0.974	1.212	1.029
	0.9	6.307	0.981	1.000	6.271	1.002	1.000	1.028	6.104	1.020	1.000	6.721	0.975	1.000	0.868
	0.99	8.689	0.996	1.000	8.456	0.973	1.000	1.004	6.981	1.035	1.000	7.388	0.963	1.000	0.880

See the note to Table 1

Table 3: Relative Median Squared Errors when $(T, N) = (100, 6)$

R^2	ρ	$K=6$							$K=18$						
		$\frac{\hat\beta(0)}{\hat\beta(1)}$	$\frac{\hat\beta_{c,1}}{\hat\beta(1)}$	$\frac{\hat\beta_{pre}}{\hat\beta(1)}$	$\frac{\hat\beta(0)}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{pre}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta_{c,1}}$	$\frac{\hat\beta(0)}{\hat\beta(1)}$	$\frac{\hat\beta_{c,1}}{\hat\beta(1)}$	$\frac{\hat\beta_{pre}}{\hat\beta(1)}$	$\frac{\hat\beta(0)}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{pre}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta_{c,1}}$
	0.01	0.007	0.007	0.007	0.007	0.007	0.007	1.000	0.138	0.168	0.138	0.006	0.007	0.006	0.882
	0.1	0.007	0.007	0.007	0.007	0.007	0.007	1.000	0.175	0.203	0.175	0.007	0.008	0.007	0.926
0.1	0.5	0.039	0.039	0.039	0.039	0.039	0.039	1.000	0.535	0.539	0.535	0.037	0.037	0.037	0.997
	0.9	0.183	0.183	0.183	0.183	0.183	0.183	1.000	1.036	1.016	1.035	0.200	0.199	0.200	1.017
	0.99	0.271	0.270	0.271	0.271	0.270	0.271	1.000	1.126	1.068	1.126	0.263	0.263	0.263	1.053
	0.01	0.089	0.101	0.089	0.089	0.101	0.089	1.000	0.229	0.283	0.229	0.040	0.050	0.040	1.002
	0.1	0.104	0.112	0.104	0.104	0.112	0.104	1.000	0.260	0.320	0.261	0.042	0.052	0.042	1.004
0.5	0.5	0.446	0.387	0.445	0.446	0.387	0.445	1.000	0.931	0.811	0.933	0.177	0.168	0.177	1.089
	0.9	1.293	0.689	1.278	1.293	0.689	1.278	1.000	2.073	1.232	1.843	0.536	0.450	0.534	1.414
	0.99	1.534	0.689	1.513	1.534	0.689	1.513	1.000	2.341	1.204	1.077	0.678	0.531	0.676	1.524
	0.01	0.583	0.647	0.592	0.583	0.647	0.592	1.000	0.672	0.713	0.683	0.546	0.586	0.549	1.010
	0.1	0.594	0.660	0.624	0.594	0.660	0.624	1.000	0.711	0.757	0.731	0.580	0.632	0.591	1.022
0.9	0.5	1.575	0.887	1.085	1.575	0.887	1.085	1.000	1.766	1.014	1.211	1.420	0.830	1.227	1.019
	0.9	3.680	0.907	1.000	3.680	0.907	1.000	1.000	3.815	1.060	1.000	3.313	0.902	1.000	0.980
	0.99	4.326	0.910	1.000	4.326	0.910	1.000	1.000	4.473	1.037	1.000	3.911	0.897	1.000	0.989

See the note to Table 1.

Table 4: Relative Median Squared Errors when $(T, N) = (1000, 1)$

		$K=6$							$K=30$						
R^2	ρ	$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$
0.1	0.01	0.109	0.640	0.120	0.099	0.635	0.109	1.094	0.144	0.589	0.157	0.092	0.637	0.106	1.699
	0.1	2.029	1.098	2.211	1.880	1.076	2.075	1.058	2.340	1.185	2.340	1.412	1.040	1.625	1.454
	0.5	45.105	1.050	1.000	42.405	0.962	1.000	0.974	19.801	1.040	1.000	36.550	1.036	1.000	0.539
	0.9	134.530	1.001	1.000	156.554	0.996	1.000	0.855	19.361	1.009	1.000	166.104	0.998	1.000	0.115
	0.99	164.282	1.006	1.000	181.845	0.997	1.000	0.896	19.545	1.008	1.000	167.930	0.998	1.000	0.115
0.5	0.01	0.573	0.905	0.623	0.570	0.908	0.620	1.008	0.606	0.993	0.672	0.568	0.951	0.637	1.021
	0.1	4.887	1.037	1.949	4.832	1.004	1.921	0.980	6.630	1.071	2.356	6.428	1.036	2.414	0.998
	0.5	150.224	1.014	1.000	148.486	0.994	1.000	0.992	112.791	1.002	1.000	129.661	0.995	1.000	0.863
	0.9	457.784	1.002	1.000	437.146	1.012	1.000	1.058	254.561	1.008	1.000	432.566	1.006	1.000	0.587
	0.99	573.884	1.000	1.000	611.730	0.998	1.000	0.937	264.749	1.005	1.000	589.502	1.000	1.000	0.447
0.9	0.01	0.959	0.975	0.953	0.962	0.978	0.944	1.000	0.814	0.976	0.851	0.817	0.984	0.833	1.003
	0.1	2.181	0.982	1.078	2.170	0.986	1.067	1.009	2.544	0.991	1.102	2.478	0.994	1.146	1.029
	0.5	56.156	0.997	1.000	57.149	1.005	1.000	0.991	52.774	1.008	1.000	50.630	0.997	1.000	1.031
	0.9	145.938	1.003	1.000	151.439	1.004	1.000	0.965	148.529	1.000	1.000	151.989	1.000	1.000	0.977
	0.99	199.297	1.001	1.000	201.305	1.004	1.000	0.993	176.518	1.000	1.000	203.479	1.003	1.000	0.870

See the note to Table 1.

Table 5: Relative Median Squared Errors when $(T, N) = (1000, 3)$

R^2	ρ	$K=6$							$K=30$						
		$\frac{\hat\beta(0)}{\hat\beta(1)}$	$\frac{\hat\beta_{c,1}}{\hat\beta(1)}$	$\frac{\hat\beta_{pre}}{\hat\beta(1)}$	$\frac{\hat\beta(0)}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{pre}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta_{c,1}}$	$\frac{\hat\beta(0)}{\hat\beta(1)}$	$\frac{\hat\beta_{c,1}}{\hat\beta(1)}$	$\frac{\hat\beta_{pre}}{\hat\beta(1)}$	$\frac{\hat\beta(0)}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{pre}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta_{c,1}}$
0.1	0.010	0.037	0.325	0.037	0.030	0.356	0.031	1.319	0.061	0.341	0.063	0.018	0.516	0.019	5.236
	0.100	0.168	0.397	0.173	0.136	0.399	0.140	1.243	0.303	0.491	0.305	0.087	0.526	0.093	3.727
	0.500	3.817	1.015	2.032	3.219	0.931	2.035	1.087	3.109	1.175	2.632	1.871	0.890	1.629	1.259
	0.900	12.871	0.995	1.000	11.252	0.979	1.000	1.126	4.401	1.061	1.000	7.177	0.935	1.000	0.540
	0.990	15.196	1.000	1.000	12.799	0.930	1.000	1.104	4.575	1.037	1.000	8.282	0.947	1.000	0.505
0.5	0.010	0.263	0.570	0.270	0.260	0.565	0.267	1.002	0.265	0.524	0.275	0.223	0.508	0.234	1.150
	0.100	1.033	0.863	1.145	1.017	0.855	1.128	1.006	1.049	0.877	1.143	0.890	0.818	1.016	1.100
	0.500	20.138	0.983	1.000	19.731	0.984	1.000	1.021	18.582	1.052	1.000	18.887	0.970	1.000	0.907
	0.900	61.323	0.985	1.000	61.401	1.000	1.000	1.014	45.435	1.030	1.000	61.375	0.990	1.000	0.712
	0.990	74.706	0.999	1.000	72.588	0.979	1.000	1.009	52.269	1.035	1.000	73.052	0.993	1.000	0.686
0.9	0.01	0.767	0.898	0.799	0.764	0.895	0.790	1.001	0.732	0.846	0.760	0.721	0.836	0.734	1.004
	0.1	1.423	0.966	1.122	1.417	0.963	1.120	1.002	1.428	0.953	1.156	1.428	0.967	1.216	1.015
	0.5	19.660	1.011	1.000	19.791	1.006	1.000	0.989	19.662	1.002	1.000	19.361	1.000	1.000	1.013
	0.9	66.489	0.999	1.000	66.406	0.995	1.000	0.997	62.998	1.008	1.000	64.453	0.997	1.000	0.967
	0.99	79.751	1.002	1.000	77.864	0.989	1.000	1.010	74.223	1.006	1.000	81.686	1.010	1.000	0.912

See the note to Table 1.

Table 6: Relative Median Squared Errors when $(T, N) = (1000, 6)$

		$K=6$							$K=30$						
R^2	ρ	$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$
0.1	0.01	0.013	0.019	0.013	0.013	0.019	0.013	1.000	0.038	0.079	0.039	0.004	0.012	0.004	1.495
	0.1	0.037	0.043	0.037	0.037	0.043	0.037	1.000	0.109	0.159	0.110	0.010	0.020	0.010	1.351
	0.5	0.620	0.503	0.620	0.620	0.503	0.620	1.000	1.264	1.079	1.263	0.185	0.180	0.186	1.139
	0.9	2.220	0.733	2.163	2.220	0.733	2.163	1.000	2.287	1.174	1.005	0.700	0.482	0.693	1.341
	0.99	2.595	0.725	1.440	2.595	0.725	1.440	1.000	2.434	1.111	1.000	0.771	0.478	0.768	1.358
0.5	0.01	0.133	0.223	0.136	0.133	0.223	0.136	1.000	0.156	0.247	0.159	0.118	0.237	0.121	1.270
	0.1	0.356	0.422	0.375	0.356	0.422	0.375	1.000	0.391	0.453	0.420	0.291	0.386	0.324	1.143
	0.5	5.596	0.917	1.000	5.596	0.917	1.000	1.000	5.980	1.089	1.000	4.859	0.905	1.000	1.023
	0.9	18.385	0.954	1.000	18.385	0.954	1.000	1.000	16.255	1.104	1.000	16.442	0.964	1.000	0.863
	0.99	21.966	0.964	1.000	21.966	0.964	1.000	1.000	18.463	1.095	1.000	18.681	0.960	1.000	0.866
0.9	0.01	0.600	0.666	0.626	0.600	0.666	0.626	1.000	0.606	0.682	0.631	0.586	0.659	0.608	0.999
	0.1	1.071	0.869	1.066	1.071	0.869	1.066	1.000	0.981	0.840	1.023	0.948	0.820	1.012	1.010
	0.5	11.229	0.987	1.000	11.229	0.987	1.000	1.000	11.511	0.987	1.000	11.159	0.983	1.000	1.027
	0.9	36.470	1.001	1.000	36.470	1.001	1.000	1.000	34.401	1.009	1.000	34.509	0.995	1.000	0.983
	0.99	44.113	0.986	1.000	44.113	0.986	1.000	1.000	44.824	1.016	1.000	45.184	0.990	1.000	0.967

See the note to Table 1.

5 Monte-Carlo Simulation

Our simulation experiment uses a design similar to that used by Hansen (2017), Kuersteiner and Okui (2010), Donald et al. (2009), and Donald and Newey (2001), where $T \in \{100, 1000\}$, $N = \{1, 3, 6\}$, and $K = \{6, 18, 30\}$. The observations are generated by the process

$y_1 = Y_2\beta + u_1,$

$Y_2 = X\Pi_2 + V_2,$

where u_1 has a standard normal distribution, V_2 and X have a multivariate normal distribution with mean zero, and variance-covariance matrix I_N and I_K, respectively. We set the correlation between u_1 and the rows of V_2 equal to $\rho/\sqrt{N}$, where ρ takes values on $\{0.01, 0.1, 0.5, 0.9, 0.99\}$. We set β to $0.1\iota_N$, where ι_q is a q-dimensional vector of unity. $\Pi_2 = c(I_N \otimes \iota_{K/N})$, where $\otimes$ denotes the Kronecker product, and $c = \sqrt{R^2/K(1-R^2)}$, hence R^2 is the reduced form population R^2 for each endogenous variable. This is important because R^2 measures the strength of the instruments. We consider three cases for the reduced form population R^2, which are $\{0.1, 0.5, 0.9\}$. The number of monte carlo simulations for each design is set to $1,000$. We set the value of $\tau = \tau_{opt}$ when $N = \{3, 6\}$ and set $\tau = 1/8$ when $N = 1$.

The simulation results are given in Table 1 – Table 6. We also consider the case where the error terms are generated from normalized chi-squared distribution with two degrees of freedom, and report the results in Table C.1– Table C.6 in Appendix C. Because of the concerns about the existence of moments of the estimators, we report the relative MEdian Squared Errors (MESE) of the OLS estimator, the 2SLS/LIML estimator, the Stein-like shrinkage estimator associated with the first two estimators, and a pre-test estimator. The pre-test estimator uses the Wu-Hausman test static under 5% critical value to choose between the OLS estimator and the 2SLS/LIML estimator. The relative MESE is calculated by dividing the MESE of an estimator by the MESE of the 2SLS/LIML estimator, hence the relative MESE of the 2SLS/LIML estimator is equal to one. In each table, we report the relative MESE of six estimators for different degrees of endogeneity (ρ), number of excluded exogenous variables (K), and R^2, where the first three columns report the relative MESE of the OLS estimator, the Stein-like shrinkage estimator using the OLS and the 2SLS estimators, and the pre-test estimator. The second three columns report the relative MESE of the OLS estimator, the Stein-like shrinkage estimator using the OLS and the LIML estimators, and the pre-test estimator. The last column reports the relative MESE of the Stein-like shrinkage estimator using the OLS and the LIML estimators to the Stein-like shrinkage estimator using the OLS and the 2SLS estimators.

We note that when the sample size, T, is short and R^2 is relatively small (weak instruments) the OLS estimator performs better than the 2SLS/LIML estimator up to mild degrees of endogeneity. This is because the 2SLS/LIML estimator has high dispersion, whereas the OLS estimator has smaller MSE. In this case, the Stein-like shrinkage estimator tends to gain from the efficiency of the OLS estimator by assigning a larger weight to this estimator, and prevails. However, when R^2 is relatively large, the 2SLS/LIML estimator performs better that the OLS estimator except for very small sizes of endogeneity. In this case, the Stein-like shrinkage estimator assigns a larger weight to the 2SLS/LIML estimator and dominates the OLS estimator. Moreover, when the number of endogenous variables increases, the OLS estimator gains from a higher efficiency and its MESE remains less than that of the 2SLS/LIML estimator even when the degree of endogeneity is moderate. Similarly, the Stein-like shrinkage estimator gains from the efficiency of the OLS estimator. We also report the results of the pre-test estimator which tests the null of endogeneity and assigns weight zero or one to the OLS or the 2SLS/LIML estimator based on the test results under 5% critical value. The relative MESE of the pre-test estimator is small when the degree of endogeneity is small, but is high for moderate degrees of endogeneity.

When the model is just-identified the 2SLS and the LIML estimators are identical. In this case, the relative MESEs of the Stein-like shrinkage estimators are smaller than those of the OLS estimators except for very small values of endogeneity and R^2. Further, the relative MESEs of the Stein-like shrinkage estimators for the whole parameter space are below that of the 2SLS/LIML estimator. When $K > N$, the 2SLS estimator performs better than the LIML estimator when R^2 is small. But, the LIML estimator performs better than the 2SLS estimator for moderate to large values of R^2 and ρ. In addition, the Stein-like shrinkage estimator that uses the OLS and the LIML estimators dominates the LIML estimator for all values of ρ and remains one of the best choices. The Stein-like shrinkage estimator that uses the OLS and the 2SLS estimators performs better than the 2SLS estimator when ρ is small to moderate.

In general, the monte carlo results support our theoretical findings of the previous sections. We find that the Stein-like shrinkage estimators perform robustly well in models with various degrees of endogeneity and instruments strength. When there is a strong degree of endogeneity or the sample size is large, the Stein-like shrinkage estimators prevail. When there is a relatively weak degree of endogeneity or weak instruments, the Stein-like shrinkage estimators tend to gain more from the efficiency of the OLS estimator by assigning a larger weight to this estimator, and thus still remain among the best choices.

6 Application To The Returns To Schooling

In this section, we present an empirical application that highlights the utility of the Stein-like shrinkage estimation in estimating returns to education.

In a seminal paper, Angrist and Krueger (1991) use quarters of birth as instruments to estimate the return to education, where they find that the 2SLS estimates of the return to education is different but close to the OLS estimates, suggesting the evidence of small bias in the OLS estimates of the return to education. The sample is drawn from the 1980 U.S. Census that consists of 329,509 men born between 1930-1939. Angrist and Krueger (1991) estimate an equation where the dependent variable is the log of the weekly wage, and the explanatory variable of interest is the number of years of schooling. The particular version of the model that we consider is the one with an intercept, 9 year-of-birth dummies, and 50 state-of-birth dummies included as explanatory variables. For this model, the OLS estimate of the coefficient on schooling in Angrist and Krueger (1991) is 0.0674 with a standard error of 0.0003, and the 2SLS estimate is 0.0928 with a standard error of 0.0093. As there is an evident sign of trade-off between the bias and variance efficiency of the OLS and the 2SLS estimators with a large number of instruments, we apply the Stein-like shrinkage estimation method which produces estimators with smaller MSE by gaining from the efficiency of the OLS estimator and the consistency of the 2SLS/LIML estimator.

Similar to Angrist and Krueger (1991), we consider a set of 180 excluded instruments which are: (*i*) 3 quarter-of-birth dummies; (*ii*) 27 dummy variables obtained by interacting quarter-of-birth with the 9 year-of-birth dummies; (*iii*) 150 dummy variables obtained by interacting the 3 quarter-of-birth with the 50 state-of-birth dummy variables. In addition, we consider 60 included exogenous variables that include the 9 year-of-birth dummies, the 50 state-of-birth dummies, and an intercept. Hence, the complete set of instruments contains 240 variables. Donald and Newey (2001) consider the same application to find the best subset of instruments for the 2SLS and the LIML estimators that can minimize the MSE of these two estimators. They find that the best subset of instruments for the 2SLS estimator includes only the 3 quarter-of-birth dummies, and for the LIML estimator includes the largest set of instruments. Therefore, we consider both cases here.

Table 7 contains the estimates of the return to education for different number of instruments for each of the 5 estimators: the OLS, the 2SLS, the LIML, and the two Stein-like shrinkage estimators. As expected the estimated coefficient of Years of education for OLS is smaller than the other estimates, the 2SLS estimates are smaller than the LIML estimates, and the Stein-like shrinkage estimations are smaller than the 2SLS/LIML estimates. The standard errors of the Stein-like shrinkage estimators are larger than the OLS estimator but smaller than the 2SLS/LIML estimator. The standard errors of the Stein-like shrinkage estimators are calculated using the bootstrap method. When the number of excluded instruments is three, the standard error of the Stein-like shrinkage estimator using the OLS estimator and the 2SLS estimator is around 14% smaller than the 2SLS estimator, and the standard error of the Stein-like shrinkage estimator using the OLS estimator and the LIML estimator is around 13% smaller than the LIML estimator. When the number of excluded instruments is 180, the standard error of the Stein-like shrinkage estimator using the OLS estimator and the 2SLS estimator is around 27% smaller, while the standard error of the Stein-like shrinkage estimator using the OLS estimator and the LIML estimator is around 6% smaller.

Table 7: Estimates of the returns to education for men born 1930-1939: 1980 Census

#IV	Independent variable	$\hat{\beta}(0)$	$\hat{\beta}(1)$	$\hat{\beta}(\lambda)$	$\hat{\beta}_{c,1}$	$\hat{\beta}_{c,\lambda}$
180	Years of eduction	0.0673	0.0928	0.1064	0.0920	0.1055
		(0.0003)	(0.0093)	(0.0116)	(0.0068)	(0.0109)
	9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes
	50 State-of-birth dummies	Yes	Yes	Yes	Yes	Yes
3	Years of eduction	0.0673	0.1077	0.1089	0.1053	0.1065
		(0.0003)	(0.0195)	(0.0198)	(0.0167)	(0.0173)
	9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes
	50 State-of-birth dummies	Yes	Yes	Yes	Yes	Yes

Note: This table reports the estimated coefficient on the education variable using OLS ($\hat{\beta}(0)$), 2SLS ($\hat{\beta}(1)$), LIML ($\hat{\beta}(\lambda)$), the Stein-like shrinkage estimation using the OLS and the 2SLS estimators ($\hat{\beta}_{c,1}$), and the Stein-like shrinkage estimation using the OLS and the LIML estimators ($\hat{\beta}_{c,\lambda}$). In each model an intercept, 9 dummy variables for the year of birth, and 50 dummy variables indicating the state of birth are also included. The top panel is the results when the excluded instrumental variables (IV) are the full 180 instruments. The bottom panel shows the results when the excluded instruments are only the 3 quarter-of-birth dummies. Standard errors are in parentheses below the corresponding point estimates.

7 Conclusion

In this paper, we introduce two Stein-like shrinkage estimators for estimating the structural parameters of a Simultaneous Equations Model. The estimators are weighed averages of the 2SLS/LIML and the OLS estimators where the weights are inversely related to a Wu-Hausman test statistic. The approximate distribution, bias, and MSE matrix of the Stein-like shrinkage estimators using Small-Disturbance approximations of Kadane (1971) are derived. The proposed method has several advantages relative to the existing methods. First, it allows us to study the performance of the weighted averages of any k-class estimators with the OLS estimator. This is important because under weak instruments the 2SLS estimator is biased towards the OLS estimator, and an alternative consistent estimator is required to allow balancing between the bias and variance efficiency of the OLS estimator. Second, the dominance and optimality of the Stein-like shrinkage estimators proposed here are not limited to a specific MSE and hold for any weighted quadratic loss function where the weight is positive definite and symmetric. Lastly, the framework considered here allows for studying the higher order terms, which is critical here because k-class estimators tend to have higher order bias.

Acknowledgments

The author gratefully thank the Editor, an Associate editor, and three anonymous referees for helpful and constructive comments that led to significant improvements in the contents and presentation of the paper. The author would like to acknowledge helpful comments from Aman Ullah, Tae-Hwy Lee, Gloria Gonzalez-Rivera, and Shahnaz Parsaeian.

References

Anderson, T. W. (1977). Asymptotic expansions of the distributions of estimates in simultaneous equations for alternative parameter sequence. *Econometrica,* 45, 506-518.

Anderson, T.W., N., Kunitomo and K., Morimune (1986). Comparing Single-Equation Estimators in a Simultaneous Equation System. *Econometric Theory*, 2 (1), 1-32.

Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a Single Equation in a Complete System of Stochastic Equations. *Annals of Mathematical Statistics,* 20, 1, 46-63.

Anderson, T. W. and T. Sawa (1979). Evaluation of the distribution function of the two-stage least squares estimate. *Econometrica,* 47: 163-182.

Angrist, J. and A. Krueger (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economic,* 106, 979-1014.

Chao, J. C., and N. R. Swanson (2005). Consistent Estimation with a Large Number of Weak Instruments. *Econometrica*, 73, 5, 1673-1692.

Donald, S. G., Newey W. K. (2001). Choosing the number of instruments. *Econometrica* 69, 1161–1191.

Donald, S. G., Imbens, G. W., Newey, W. K. (2009). Choosing instrumental variables in conditional moment restriction models. *Journal of Econometrics,* 152, 28–36.

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

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

Hansen, B. (2017). Stein-like 2SLS estimator. *Econometric Reviews*, 36, 840-852.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, 46, 1251-1271.

Kadane, J. B. (1970). Testing Overidentifying Restrictions When the Disturbances are Small. *Journal of the American Statistical Association*, 65, 182-185.

Kadane, J. B. (1971). Comparison of K-Class Estimators When the Disturbances Are Small. *Econometrica*, 39, 723-737.

Kuersteiner, G., R. Okui (2010). Constructing optimal instruments by first-stage prediction averaging. *Econometrica*, 78:697–718.

Maasoumi, E. (1978). A modified Stein-like estimator for the reduced form coefficients of simultaneous equations. *Econometrica*, 46, 695-703.

Mariano. R. S., J. B., McDonald (1979). A note on the distribution functions of LIML and 2SLS structural coefficient in the exactly identified case. *Journal of the American Statistical Association,* 74, 847-848.

Mariano. R. S., T., Sawa (1972). The exact finite-sample distribution of the limited information maximum likelihood estimator in the case of two included endogenous variables. *Journal of the American Statistical Association,* 67, 159-163.

Morimune. K. (1978). Improving the Limited Information Maximum Likelihood Estimator When the Disturbances are Small. *Journal of the American Statistical Association,* 73, 867-871.

Morimune, K. and N. Kunitomo (1980). Improving the maximum likelihood estimate in linear functional relationships for alternative parameter sequences. *Journal of the American Statistical Association,* 75, 230-237.

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

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

Phillips, P.C.B. (1983). Exact small sample theory in the simultaneous equations model. *Handbook of Econometrics*, 1, 449-51.

Phillips, P.C.B. (1984). The exact distribution of LIML: I. *International Economic Review*, 25, 249-261.

Phillips, P.C.B. (1985). The exact distribution of LIML: II. *International Economic Review*, 26, 21-36.

Rothenberg, T. (1984). Approximating the Distribution of Econometric Estimators and Test Statistics. *Handbook of Econometrics,* Volume II, Edited by Z. Griliehes and M.D. lntriligator.

Sargan J. D. (1974). The Validity of Nagar's Expansion for the Moments of Econometric Estimators. *Econometrica,* 42, 1, 169-176.

Sawa, T. (1973a). Almost Unbiased Estimators in Simultaneous Equations Systems. *International Economic Review,* 14, 97-106.

Sawa, T. (1973b). The Mean Square Error of A Combined Estimator and Numerical Comparison with The TSLS Estimator. *Journal of Econometrics,* 1, 115-132.

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.

Theil, H. (1961). *Economic Forecast and Policy*, 2nd ed. Amsterdam: North Holland.

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.

Ullah, A., and V.K., Srivastava (1988). On the Improved Estimation of Structural Coefficients. *Sankhya: The Indian Journal of Statistics*, 50, 111-118.

Wu, D. M. (1973). Alternative Tests of Independence Between Stochastic Regressors and Disturbances. *Econometrica,* 41, 733-750.

Zellner, A., and Vandaele, W. (1975). Bayes-Stain estimators for k-means, regression and simultaneous equation models . In *Studies in Bayesian Econometrics and Statistics* (eds. S. E. Fienberg and A. Zellner), North-Holland Publishing Company, Amsterdam, 627-653.

A Appendix A

In this section, we analyze the proposed Stein-like shrinkage estimators without Assumption 3. We start by expanding $\hat{\beta}(k)$. Employing equation (2.5) in equation (3.1), we have

$$\begin{aligned}
\hat{\beta}(k) - \beta &= \Big[\Big(W + \sigma V_2\Big)' H_k \Big(W + \sigma V_2\Big)\Big]^{-1} \Big(W + \sigma V_2\Big)' H_k \sigma u_1 \\
&= \Big(W' H_k W + \sigma W' H_k V_2 + \sigma V_2' H_k W + \sigma^2 V_2' H_k V_2\Big)^{-1} \Big(\sigma W' H_k u_1 + \sigma^2 V_2' H_k u_1\Big) \\
&= \Big(I_N + \sigma Q W' H_k V_2 + \sigma Q V_2' H_k W + \sigma^2 Q V_2' H_k V_2\Big)^{-1} Q \Big(\sigma W' H_k u_1 + \sigma^2 V_2' H_k u_1\Big) \\
&= \Big(I_N + \sigma Q S + \sigma^2 Q V_2' H_k V_2\Big)^{-1} Q \Big(\sigma W' u_1 + \sigma^2 V_2' H_k u_1\Big),
\end{aligned} \tag{A.1}$$

where $Q_{N\times N} = (W'W)^{-1}$, $S = V_2'W + W'V_2$, and the use has been made of $W'H_k = W'$.

Using the standard geometric expansion for the inverse of a matrix (i.e., $(I + A)^{-1} = I - A + A^2 - A^3 + \dots$), the above equation can be written as

$$\begin{aligned}
\hat{\beta}(k) - \beta &= \sigma Q W' u_1 + \sigma^2 Q \Big(V_2' H_k u_1 - S Q W' u_1\Big) \\
&\quad + \sigma^3 Q \Big(S Q S Q W' u_1 - V_2' H_k V_2 Q W' u_1 - S Q V_2' H_k u_1\Big) + O_p(\sigma^4).
\end{aligned} \tag{A.2}$$

Now, we expand $\mathbb{R}_k$ defined in equation (3.4). Using equation (A.2), we have

$$\begin{aligned}
\hat{\sigma}_{11,k} &= \hat{u}_1(k)' \hat{u}_1(k)/(T - N) \\
&= \Big(y_1 - Y_2 \hat{\beta}(k)\Big)' \Big(y_1 - Y_2 \hat{\beta}(k)\Big)/(T - N) \\
&= \Big(\sigma u_1 - Y_2(\hat{\beta}(k) - \beta)\Big) \Big(\sigma u_1 - Y_2(\hat{\beta}(k) - \beta)\Big)/(T - N) \\
&= \frac{1}{T - N} \Big[\sigma M_W u_1 - \sigma^2 W Q V_2' H_k u_1 + \sigma^2 W Q S Q W' u_1 - \sigma^2 V_2 Q W' u_1 + O_p(\sigma^3)\Big]' \\
&\qquad \times \Big[\sigma M_W u_1 - \sigma^2 W Q V_2' H_k u_1 + \sigma^2 W Q S Q W' u_1 - \sigma^2 V_2 Q W' u_1 + O_p(\sigma^3)\Big] \\
&= \frac{1}{T - N} \Big[\sigma^2 u_1' M_W u_1 - 2\sigma^3 u_1' M_W V_2 Q W' u_1 + O_p(\sigma^4)\Big],
\end{aligned} \tag{A.3}$$

where $M_W = I_T - W(W'W)^{-1}W'$. Also, we have

$$(Y_2'H_kY_2)^{-1} = \Big(I_N - \sigma QS - \sigma^2 QV_2'H_kV_2 + \sigma^2 QSQS + \sigma^3 QSQV_2'H_kV_2 + \sigma^3 QV_2'H_kV_2QS\Big)Q + O_p(\sigma^4),$$

$$(Y_2'Y_2)^{-1} = \Big(I_N - \sigma QS - \sigma^2 QV_2'V_2 + \sigma^2 QSQS + \sigma^3 QSQV_2'V_2 + \sigma^3 QV_2'V_2QS\Big)Q + O_p(\sigma^4).$$

Hence the difference of the expressions above may be written as

$$(Y_2'H_kY_2)^{-1} - (Y_2'Y_2)^{-1} = \sigma^2 k\Big[QV_2'M_XV_2Q - \sigma QSQV_2'M_XV_2Q - \sigma QV_2'M_XV_2QSQ\Big] + O_p(\sigma^4). \tag{A.4}$$

Further, by using equation (A.4) in equation (3.4),

$$\begin{aligned}\mathbb{R}_k &= \hat{\sigma}_{11,k}^{-1}\Big((Y_2'HY_2)^{-1} - (Y_2'Y_2)^{-1}\Big)^{-1} \\ &= \frac{\hat{\sigma}_{11,k}^{-1}}{k\sigma^2}Q^{-1}\Big[I_N - \sigma(V_2'M_XV_2)^{-1}SQV_2'M_XV_2 - \sigma QS + O_p(\sigma^2)\Big]^{-1}(V_2'M_XV_2)^{-1}Q^{-1} \qquad (A.5) \\ &= \frac{\hat{\sigma}_{11,k}^{-1}}{k\sigma^2}Q^{-1}\Big[(V_2'M_XV_2)^{-1} + \sigma(V_2'M_XV_2)^{-1}SQ + \sigma QS(V_2'M_XV_2)^{-1} + O_p(\sigma^2)\Big]Q^{-1}.\end{aligned}$$

In addition from equation (A.2), we have

$$\hat{\beta}(0) - \hat{\beta}(k) = k\sigma^2\Big[QV_2'M_Xu_1 - \sigma QV_2'M_XV_2QW'u_1 - \sigma QSQV_2'M_Xu_1\Big] + O_p(\sigma^4). \tag{A.6}$$

Employing equations (A.5) and (A.6) in equation (3.3),

$$F_{k,\mathrm{WH}} = \frac{k\sigma^2}{\hat{\sigma}_{11,k}}\Big[u_1'M_XV_2(V_2'M_XV_2)^{-1}V_2'M_Xu_1 - 2\sigma u_1'M_XV_2QW'u_1 + O_p(\sigma^2)\Big]. \tag{A.7}$$

Therefore, we have the following expression

$$
\begin{aligned}
\frac{1}{F_{k,\mathrm{WH}}} &= \frac{\hat{\sigma}_{11,k}}{k\sigma^2 u_1' M_X V_2 (V_2' M_X V_2)^{-1} V_2' M_X u_1}\Big(1 + \frac{2\sigma u_1' M_X V_2 Q W' u_1}{u_1' M_X V_2 (V_2' M_X V_2)^{-1} V_2' M_X u_1} + O_p(\sigma^2)\Big) \\
&= \frac{1}{k(T-N)} \frac{1}{u_1' M_X V_2 (V_2' M_X V_2)^{-1} V_2' M_X u_1}\Big(u_1' M_W u_1 - 2\sigma u_1' M_W V_2 Q W' u_1 \\
&\quad + \frac{2\sigma u_1' M_W u_1}{u_1' M_X V_2 (V_2' M_X V_2)^{-1} V_2' M_X u_1} u_1' M_X V_2 Q W' u_1 + O_p(\sigma^2)\Big).
\end{aligned} \tag{A.8}
$$

Using equations (A.2), and (A.8) in equation (3.2), we can write the Stein-like shrinkage estimators as

$$
\begin{aligned}
\hat{\beta}_{c,k} - \beta &= (\hat{\beta}(k) - \beta) + \frac{\tau}{F_{k,WH}}\Big((\hat{\beta}(0) - \beta) - (\hat{\beta}(k) - \beta)\Big) \\
&= \sigma Q W' u_1 + \sigma^2 Q\Big(V_2' H u_1 - S Q W' u_1\Big) + \sigma^3 Q\Big(SQSQW'u_1 - V_2' H V_2 Q W' u_1 - SQV_2' H u_1\Big) \\
&+ \frac{\tau\sigma^2}{(T-N)} \frac{u_1' M_W u_1}{u_1' M V_2 (V_2' M_X V_2)^{-1} V_2' M_X u_1}\Big[Q V_2' M_X u_1 - \sigma Q V_2' M_X V_2 Q W' u_1 - \sigma Q S Q V_2' M_X u_1 \\
&+ \frac{2\sigma}{u_1' M_X V_2 (V_2' M_X V_2)^{-1} V_2' M_X u_1} u_1' M_X V_2 Q W' u_1 Q V_2' M_X u_1\Big] \\
&- \frac{2\tau\sigma^3}{(T-N)} \frac{u_1' M_W V_2 Q W' u_1}{u_1' M_X V_2 (V_2' M_X V_2)^{-1} V_2' M_X u_1} Q V_2' M_X u_1 + O_p(\sigma^4), \qquad \text{for } k = 1, \lambda.
\end{aligned} \tag{A.9}
$$

The above equation has the product of normally distributed and correlated terms in the denominator, which make the moments calculations complicated. However, under the local endogeneity assumption, Assumption 3, the random term in the denominator will be simplified and makes the derivation of the bias and MSE possible.

Under Assumption 3, (A.8) is equal to the following expression

$$
\begin{aligned}
\frac{1}{F_{k,WH}} &= \frac{1}{k(T-N)} \frac{u_1' M_W u_1}{u_1' M_X \Psi (\Psi' M_X \Psi)^{-1} \Psi' M_X u_1}\Big[1 + 2\sigma u_1' M_X \Psi (\Psi' M_X \Psi)^{-1} \delta \\
&+ \frac{2\sigma}{u_1' M_X \Psi (\Psi' M_X \Psi)^{-1} \Psi' M_X u_1}\Big(u_1' M_X \Psi Q W' u_1 - u_1' M_X \Psi (\Psi' M_X \Psi)^{-1} \delta u_1' M_X u_1\Big)\Big] \\
&- \frac{2\sigma}{k(T-N)} \frac{u_1' M_W \Psi Q W' u_1}{u_1' M_X \Psi (\Psi' M_X \Psi)^{-1} \Psi' M_X u_1} + O(\sigma^2).
\end{aligned} \tag{A.10}
$$

Using equation (A.10) in the Stein-like shrinkage estimator expression (equation (3.2)), we have

$$\hat{\beta}_{c,k} - \beta = (\hat{\beta}(k) - \beta) + \frac{\tau}{F_{k,WH}}\Big((\hat{\beta}(0) - \beta) - (\hat{\beta}(k) - \beta)\Big) \tag{A.11}$$

$$\begin{aligned}
&= \sigma QW'u_1 + \sigma^2 Q(V_2'Hu_1 - SQW'u_1) + \sigma^3 Q(SQSQW'u_1 - V_2'HV_2QW'u_1 - SQV_2'Hu_1) \\
&+ \frac{\tau\sigma^2}{T-N}\frac{u_1'M_W u_1}{u_1'M_X\Psi(\Psi'M_X\Psi)^{-1}\Psi'M_X u_1}\Bigg[\Big(1 + 2\sigma u_1'M_X\Psi(\Psi'M_X\Psi)^{-1}\delta\Big)QV_2'M_X u_1 \\
&- \sigma QV_2'M_X V_2 QW'u_1 - \sigma QSQV_2'M_X u_1 \\
&+ \frac{2\sigma}{u_1'M_X\Psi(\Psi'M_X\Psi)^{-1}\Psi'M_X u_1}\Big(u_1'M_X\Psi QW'u_1 - u_1'M_X\Psi(\Psi'M_X\Psi)^{-1}\delta u_1'M_X u_1\Big) \\
&QV_2'M_X u_1\Bigg] - \frac{2\sigma^3\tau}{T-N}\frac{u_1'M_W V_2 QW'u_1}{u_1'M_X\Psi(\Psi'M_X\Psi)^{-1}\Psi'M_X u_1}QV_2'M_X u_1 + O(\sigma^4).
\end{aligned} \tag{A.12}$$

Since the random term in the denominator includes Ψ and u_1 which are independent by construction, we can derive the bias and MSE of the Stein-like shrinkage estimators.

B Appendix B

Lemma B1 *Let A be a square constant matrix, and Ψ is $T \times N$ where its rows are independently normally distributes as $N(0, C_2)$. Then,*

(a) $\mathbb{E}(\Psi' A \Psi) = tr(A)C_2$

(b) $\mathbb{E}(\Psi A \Psi') = tr(C_2 B) I_T$

(c) $\mathbb{E}(\Psi A \Psi) = A'C_2$

Proof: See Kadane (1971), Lemmas B1-B3. ■

Lemma B2 *Let A and B be $T \times T$ symmetric, constant matrices, and the $T \times 1$ vector $u \sim N(0, I_T)$. Then*

$\mathbb{E}(uu'Auu') = tr(A)I_T + 2A,$

and

$$\mathbb{E}(uu'Auu'Buu') = \Big[tr(A)\, tr(B) + 2\, tr(AB)\Big] I_T + 2\, tr(A)B + 2\, tr(B)A + 4AB + 4BA,$$

Proof: See Ullah (2004). ■

Lemma B3 *Let χ^2_λ 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*

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

Proof: See Ullah (1974). ■

Lemma B4 *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 $\phi(\cdot)$ is a Borel measurable function. Then*

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

where $r = rank(A) = tr(A)$.

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 $\omega = (\omega_1, \ldots, \omega_J)' = P\nu$, which has a $N(P\theta, I_J)$ distribution. Therefore

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

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

$$\begin{aligned}
\mathbb{E}\left[\phi\Big(\sum_{j=1}^{J} d_j\omega_j^2\Big)\omega_i^2\right] &= \mathbb{E}\left[\mathbb{E}\left[\phi\Big(d_i\omega_i^2 + \sum_{j\neq i}\omega_j^2\Big)\omega_i^2|\omega_j^2, j\neq i\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\phi\Big(d_i\chi_3^2((P_i'\theta)^2/2) + \sum_{j\neq i}\omega_j^2\Big)|\omega_j^2, j\neq i\right]\right] \\
&\quad + (P_i'\theta)^2\,\mathbb{E}\left[\mathbb{E}\left[\phi\Big(d_i\chi_5^2((P_i'\theta)^2/2) + \sum_{j\neq i}\omega_j^2\Big)|\omega_j^2, j\neq i\right]\right] \\
&= \begin{cases} \mathbb{E}\left[\phi(\chi_{r+2}^2(\theta' A\theta/2)] + (P_i'\theta)^2\,\mathbb{E}[\phi(\chi_{r+4}^2(\theta' A\theta/2)\right], & \text{if } d_i = 1 \\ \mathbb{E}\left[\phi(\chi_{r}^2(\theta' A\theta/2)] + (P_i'\theta)^2\,\mathbb{E}[\phi(\chi_{r}^2(\theta' A\theta/2)\right], & \text{if } d_i = 0 \end{cases}
\end{aligned} \tag{B.1}$$

where the second equality holds by Lemma 1 of Appendix B.1 in Judge and Bock (1978). Hence, the matrix form of the diagonal elements can be written as

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

The off-diagonal elements, for any $i \neq j$, are

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

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

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

where the second equality holds by lemma 2 of Appendix B.1 in Judge and Bock (1978). Hence, the off-diagonal terms can be written as

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

Therefore the proof is complete by adding the diagonal and off-diagonal components.

■

Lemma B5 *Let M_1 and M_2 be two $T \times T$ idempotent matrices where $M_1 M_2 = \mathbf{0}$, and the $T \times 1$ vector $u \sim N(0, I_T)$. Then*

$$\begin{aligned}
\mathbb{E}\left(u\frac{u' M_1 u}{(u' M_2 u)^2}u'\right) &= \frac{tr(M_1) + 2}{(tr(M_2) - 2)(tr(M_2) - 4)}M_1 + \frac{tr(M_1)}{tr(M_2)(tr(M_2) - 2)}M_2 \\
&+ \frac{tr(M_1)}{(tr(M_2) - 2)(tr(M_2) - 4)}, (I_T - M_1 - M_2)
\end{aligned}$$

when $tr(M_2) > 4$.

Proof: Since M_1 and M_2 commute, let the orthogonal matrix Γ simultaneously diagonalize them such that

$$\Gamma' M_1 \Gamma = \begin{bmatrix} I_{N_1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = D_1, \text{ and } \Gamma' M_2 \Gamma = \begin{bmatrix} 0 & 0 & 0 \\ 0 & I_{N_2} & 0 \\ 0 & 0 & 0 \end{bmatrix} = D_1,$$

where $N_1 = \mathrm{tr}(M_1)$ and $N_2 = \mathrm{tr}(M_2)$. Further, define $\nu = \Gamma u$, and $\nu = (\nu_1', \nu_2', \nu_3')'$ be partitioned conformably with D_1 and D_2. Then

$$\begin{aligned}\mathbb{E}\Big(u\frac{u'M_1u}{(u'M_2u)^2}u'\Big) &= \Gamma\,\mathbb{E}\left(\nu\frac{\nu_1'\nu_1}{(\nu_2'\nu_2)\nu}\right)\Gamma' \\ &= \Gamma\Big[\frac{N_1+2}{(N_2-2)(N_2-4)}D_1 + \frac{N_1}{N_2(N_2-2)}D_2 + \frac{N_1}{(N_2-2)(N_2-4)}(I_T - D_1 - D_2)\Big]\Gamma',\end{aligned}$$

where the use has been made of Lemma B2, Lemma B3, and Lemma B4.

■

Lemma B6 *Under assumptions 1 and 2, the bias of the $k-$class estimators up to order σ^2, is*

$$\mathbb{E}(\hat{\beta}(k) - \beta) = \sigma^2 Qq(L_k - 1), \quad \textit{for fixed } k, \tag{B.2}$$

$$\mathbb{E}(\hat{\beta}(\lambda) - \beta) = -\sigma^2 Qq, \quad \textit{for LIML estimator.} \tag{B.3}$$

The mean squared error matrix up to order σ^4 is

$$\begin{aligned}\mathbb{E}(\hat{\beta}(k) - \beta)(\hat{\beta}(k) - \beta)' = \sigma^2 Q + \sigma^4\big\{&(3 - 2L_k)tr(C_1Q)Q + tr(QC_2)Q \\ &+ QC_1Q((L_k-2)^2 + 2 + 2s_k) + QC_2Q(2 + s_k - L_k)\big\},\end{aligned} \tag{B.4}$$

$$\begin{aligned}\mathbb{E}(\hat{\beta}(\lambda) - \beta)(\hat{\beta}(\lambda) - \beta)' = \sigma^2 Q + \sigma^4\big\{&3tr(C_1Q)Q + tr(QC_2)Q \\ &+ 6QC_1Q + QC_2Q[\frac{(L_1+2)(T-K+L_1-2)}{T-K-2}]\big\},\end{aligned} \tag{B.5}$$

where $L_k = (1-k)T + kK - N$ and $s_k = k(k-1)(T-K)$.

Proof: See Kadane (1971).

Lemma B7 *Under assumptions 1–3, the asymptotic expansions of the density functions of*

$\hat{e}_1 = \frac{1}{\sigma}(\hat{\beta}_1 - \beta)$, *and* $\hat{e}_\lambda = \frac{1}{\sigma}(\hat{\beta}_\lambda - \beta)$ *as* σ *goes to zero are given, respectively, by*

$$\begin{aligned} f_1(\xi) = \phi_Q(\xi)\Bigg[& 1 + \sigma^2\delta'\xi\Big(N + 1 + L_1 - \xi'Q^{-1}\xi\Big) \\ & + \frac{\sigma^2}{2}\Big[L_1\, tr(C_2Q) - \xi'C_2\xi\Big(N + 2 + L_1 - \xi'Q^{-1}\xi\Big)\Big]\Bigg] + O(\sigma^3), \end{aligned} \tag{B.6}$$

$$\begin{aligned} f_\lambda(\xi) = \phi_Q(\xi)\Bigg[& 1 + \sigma^2\delta'\xi\Big(N + 1 - \xi'Q^{-1}\xi\Big) + \frac{\sigma^2}{2}\Big[- \frac{L_1(T-N)}{T-K-2} tr(C_2Q) \\ & - \xi'C_2\xi\Big(N + 2 - \frac{L_1(T+N)}{T-K-2} - \xi'Q^{-1}\xi\Big)\Big]\Bigg] + O(\sigma^3), \end{aligned} \tag{B.7}$$

where ξ *is an* $N \times 1$ *vector and* $\phi_Q(\xi)$ *is the multivariate normal density function with mean* $\mathbf{0}$ *and covariance matrix* Q.

Proof: See Anderson et al. (1986).

Proof of Theorem 1: From equation (A.12), we have

$$\frac{1}{\sigma}(\hat{\beta}_{c,k} - \beta) = e_k^{(0)} + \sigma\Big(e_k^{(1)} + e_c^{(1)}\Big) + \sigma^2\Big(e_k^{(2)} + e_c^{(2)}\Big) + O(\sigma^3), \tag{B.8}$$

where $e_k^{(i)}, \quad i = 0, 1, 2$ are terms with order σ^i of $\frac{1}{\sigma}(\hat{\beta}(k) - \beta)$ and $e_c^{(i)}, \quad i = 0, 1, 2$ are the other terms with order σ^i which are defined below

$$\begin{aligned} e_k^{(0)} &= QW'u_1, \\ e_k^{(1)} &= Q(\Psi H_k^{(0)}u_1 - S_\Psi e_k^{(0)}), \\ e_k^{(2)} &= Q\Psi' H_k^{(1)}u_1 + Q\delta u_1' H_k^{(0)}u_1 - Q(\delta u_1' W + W'u_1\delta')e_k^{(0)} \\ &\quad + QS_\Psi QS_\Psi e_k^{(0)} - Q\Psi' H_k^{(0)}\Psi e_k^{(0)} - QS_\Psi Q\Psi' H_k^{(0)}u_1, \\ e_c^{(1)} &= \frac{\tau}{T-N}\frac{u_1'M_W u_1}{u_1'P_\Psi u_1}Q\Psi' M_x u_1, \end{aligned}$$

$$
\begin{aligned}
e_c^{(2)} &= \frac{\tau}{T-N}\frac{u_1' M_W u_1}{u_1' P_\Psi u_1}\Big[Q\delta u_1' M_x u_1 + 2u_1' M_x \Psi(\Psi' M_x \Psi)^{-1}\delta Q \Psi' M_x u_1 - Q\Psi' M_x \Psi e_k^{(0)} \\
&\quad - QS_\Psi Q\Psi' M_x u_1 + \frac{2}{u_1' P_\Psi u_1}\Big[u_1' M_x \Psi e_k^{(0)} Q\Psi' M_x u_1 - u_1' M_x \Psi(\Psi' M_x \Psi)^{-1}\delta u_1' M_x u_1 Q\Psi' M_x u_1\Big]\Big] \\
&\quad - \frac{\tau}{T-N}\frac{u_1' M_W \Psi e_k^{(0)}}{u_1' P_\Psi u_1} Q\Psi' M_X u_1,
\end{aligned}
$$

where $P_\Psi = M_x\Psi(\Psi' M_x \Psi)^{-1}\Psi' M_x$, and $S_\Psi = \Psi' W + W'\Psi$. When k is fixed $H_k = H_k^{(0)} = P_x$, so $H_k^{(1)} = 0$. When $k = \lambda$, since λ is random, we have $H_k^{(0)} = I_T - \lambda_0 M_x$, and $H_k^{(1)} = -\lambda_1 M_x$, because by Kadane (1970)

$$
\begin{aligned}
\lambda &= \frac{u_1' M_W u_1}{u_1' M_X u_1} + 2\sigma \frac{(u_1' W Q V_2' M_X u_1)(u_1' M_W u_1) - (u_1' W Q V_2' M_W u_1)(u_1' M_X u_1)}{(u_1' M_X u_1)^2} + O_p(\sigma^2) \\
&\equiv \lambda_0 + \sigma\lambda_1 + O_p(\sigma^2),
\end{aligned}
$$

where the definition of $\lambda_i, i = 0, 1$ should be apparent.

We derive the approximate expansions of the density function of $\hat{e}_{c,k}$ by inverting its characteristic function up to order σ^2. Using (B.8) the characteristic function of $\hat{e}_{c,k}$ can be expressed as

$$
\begin{aligned}
C_{c,k}(\theta) &= C_k(\theta) + \sigma\, \mathbb{E}(i\theta'\, \mathbb{E}(e_c^{(1)}|e_k^{(0)}) \exp(i\theta' e_k^{(0)})) + \sigma^2\, \mathbb{E}(i\theta'\, \mathbb{E}(e_c^{(2)}|e_k^{(0)}) \exp(i\theta' e_k^{(0)})) \\
&\quad + \frac{\sigma^2}{2}\, \mathbb{E}(i^2\theta'\, \mathbb{E}(e_c^{(1)} e_c^{(1)'}|e_k^{(0)})\theta \exp(i\theta' e_k^{(0)})) + \frac{\sigma^2}{2}\, \mathbb{E}(i^2\theta'\, \mathbb{E}(e_c^{(1)} e_k^{(1)'}|e_k^{(0)})\theta \exp(i\theta' e_k^{(0)})) \\
&\quad + \frac{\sigma^2}{2}\, \mathbb{E}(i^2\theta'\, \mathbb{E}(e_k^{(1)} e_c^{(1)'}|e_k^{(0)})\theta \exp(i\theta' e_k^{(0)})) + O(\sigma^3),
\end{aligned}
\tag{B.9}
$$

where θ is a $N \times 1$ vector, $C_k(\theta)$ is the characteristic function of the k-class estimator, and $\mathbb{E}(.|e_k^{(0)})$ denotes the conditional expectation given $e_k^{(0)}$. The conditional expectations given the first-order term, $e_k^{(0)}$, are calculated below.

$$
\mathbb{E}(e_c^{(1)}|e_k^{(0)}) = 0, \tag{B.10}
$$

as it is the product of odd numbers of normal distribution.

$$\mathbb{E}(e_c^{(2)}|e_k^{(0)}) = \frac{\tau}{T-N}\Big[Q\delta\frac{(T-N)(T-K)}{N}\Big] - QC_2e_k^{(0)}\Big[\frac{(T-N)(T-K)}{N}\Big]\Big], \tag{B.11}$$

$$\mathbb{E}(e_c^{(1)}e_c^{(1)'}|e_k^{(0)}) = \; = \frac{\tau^2}{(T-N)}\frac{(T-K)(T-N-2)}{N(N-2)}QC_2Q, \tag{B.12}$$

$$\mathbb{E}(e_c^{(1)}e_k^{(1)'}|e_k^{(0)}) = \begin{cases} 0, & \text{if } k=1 \\ \frac{\tau}{T-N}cQC_2Q, & \text{if } k=\lambda, \end{cases} \tag{B.13}$$

where c is a negative constant.

Now, we invert the terms of the characteristic function of the Stein-like shrinkage estimator in (B.9) term by term. The inverse transformation of the first term in (B.9) is

$$\mathfrak{F}^{-1}[C_k(\theta)] = f_k(\xi), \tag{B.14}$$

where $f_k(\xi)$ is the approximate distribution of the k-class estimators given in Lemma B7. Note that, for any polynomial $g(.)$, $\mathfrak{F}^{-1}[h(-i\theta)\,\mathbb{E}(g(x)\exp(i\theta' x))] = h(\frac{\partial}{\partial\xi})g(\xi)\phi_Q(\xi)$, where $h(\cdot)$ is any polynomial, and $\partial/\partial\xi' = (\partial/\partial\xi_1, ..., \partial/\partial\xi_N)$. Using this fact, the inverse transformation of the rest of the terms in (B.9) are given below.

$$\mathfrak{F}^{-1}[(i\theta)'\,\mathbb{E}(\mathbb{E}(e_c^{(1)}|e_k^{(0)})\exp(i\theta' e_k^{(0)}))] = -\frac{\partial}{\partial\xi'}\{\mathbb{E}(e_c^{(1)}|e_k^{(0)}=\xi)\phi_Q(\xi)\} = 0, \tag{B.15}$$

$$\mathfrak{F}^{-1}[(i\theta)'\,\mathbb{E}(\mathbb{E}(e_c^{(2)}|e_k^{(0)})\exp(i\theta' e_k^{(0)}))] = -\frac{\partial}{\partial \xi'}\{\mathbb{E}(e_c^{(2)}|e_k^{(0)} = \xi)\phi_Q(\xi)\}$$
$$= \frac{\tau}{T-N}\phi_Q(\xi)\left[\alpha_1 \delta' \xi + \alpha_1 \Big[\operatorname{tr}(QC_2) - \xi' C_2 \xi\Big]\right], \tag{B.16}$$

where $\alpha_1 = (T-K)/N$, and

$$\mathfrak{F}^{-1}[i^2\theta'\,\mathbb{E}(\mathbb{E}(e_c^{(1)}e_c^{(1)'}|e_k^{(0)})\theta\exp(i\theta' e_k^{(0)}))] = \frac{\partial}{\partial \xi'}\{\mathbb{E}(e_c^{(1)}e_c^{(1)'}|e_k^{(0)} = \xi)\phi_Q(\xi)\}\frac{\partial}{\partial \xi}$$
$$= \frac{\tau^2}{T-N}\alpha_2\Big[\xi' C_2 \xi - \operatorname{tr}(QC_2)\Big]\phi_Q(\xi), \tag{B.17}$$

where $\alpha_2 = (T-K)(T-N-2)/N(N-2)$. Furthermore, the inverse transformation of the last term is

$$\mathfrak{F}^{-1}[i^2\theta'\,\mathbb{E}(\mathbb{E}(e_c^{(1)}e_k^{(1)'}|e_k^{(0)})\theta\exp(i\theta' e_k^{(0)}))] = \frac{\partial}{\partial \xi'}\{\mathbb{E}(e_c^{(1)}e_k^{(1)'}|e_k^{(0)} = \xi)\phi_Q(\xi)\}\frac{\partial}{\partial \xi}$$
$$= \begin{cases} 0, & \text{if } k = 1 \\ \frac{\tau}{T-N}c\Big[\xi' C_2 \xi - \operatorname{tr}(QC_2)\Big], & \text{if } k = \lambda. \end{cases} \tag{B.18}$$

Summation of the terms in equations (B.15)–(B.18) will provide the results in the theorem.

■

Proof of Theorem 2:

Using (4.2), the approximate bias of the Stein-like shrinkage estimator up to order of interest is equal to

$$\mathbb{E}(\frac{1}{\sigma}(\hat{\beta}_{c,k} - \beta)) = \mathbb{E}(\frac{1}{\sigma}(\hat{\beta}(k) - \beta)) + O(\sigma^2) = 0, \tag{B.19}$$

where the last equality holds by Lemma B6. The approximate MSE matrix of the Stein-like shrinkage estimator up to the order of interest is

$$\mathbb{E}\left(\frac{1}{\sigma^2}(\hat{\beta}_{c,1}-\beta)(\hat{\beta}_{c,1}-\beta)'\right)=\mathbb{E}\left(\frac{1}{\sigma^2}(\hat{\beta}(1)-\beta)(\hat{\beta}(1)-\beta)'\right)+\frac{\tau}{T-N}\sigma^2\alpha_1\int\xi\xi'\delta'\xi\phi_Q(\xi)d\xi$$
$$+\frac{1}{2}\frac{\tau}{T-N}\sigma^2\Big[\tau\alpha_2-2\alpha_1\Big]\int\Big(\xi\xi' C_2\xi\xi'-\mathrm{tr}(QC_2)\xi\xi'\Big)\phi_Q(\xi)d\xi$$
$$=\mathbb{E}\left(\frac{1}{\sigma^2}(\hat{\beta}(1)-\beta)(\hat{\beta}(1)-\beta)'\right)+\frac{\tau}{T-N}\sigma^2\Big[\tau\alpha_2-2\alpha_1\Big]QC_2Q. \tag{B.20}$$

Similarly,

$$\mathbb{E}\left(\frac{1}{\sigma^2}(\hat{\beta}_{c,\lambda}-\beta)(\hat{\beta}_{c,\lambda}-\beta)'\right)\ \leq\mathbb{E}\left(\frac{1}{\sigma^2}(\hat{\beta}(\lambda)-\beta)(\hat{\beta}(\lambda)-\beta)'\right)+\frac{\tau}{T-N}\sigma^2\Big[\tau\alpha_2-2\alpha_1\Big]QC_2Q. \tag{B.21}$$

■

Proof of Theorem 3:

To derive $\underset{||Q^{-1/2}\xi||<z}{\int\cdots\int}\left(f_{c,k}(\xi)-f_k(Q^{1/2}\xi)\right)d\xi$, we take the integral of each term of the difference of the approximate distributions below.

$$\underset{||\zeta||<z}{\int\cdots\int}\frac{\tau}{T-N}\alpha_1\delta' Q^{1/2}\zeta\phi_I(\zeta)d\zeta=0, \tag{B.22}$$

$$\underset{||\zeta||<z}{\int\cdots\int}\frac{1}{2}\frac{\tau}{T-N}\Big[\tau\alpha_2-2\alpha_1\Big]\,\mathrm{tr}(QC_2)\phi_I(\zeta)d\zeta=\frac{1}{2}\frac{\tau}{T-N}\Big[\tau\alpha_2-2\alpha_1\Big]\,\mathrm{tr}(QC_2)\Big[\Phi(z)-\Phi(-z)\Big]^N, \tag{B.23}$$

$$\underset{||\zeta||<z}{\int\cdots\int}\frac{1}{2}\frac{\tau}{T-N}\Big[\tau\alpha_2-2\alpha_1\Big]\zeta' Q^{1/2}C_2Q^{1/2}\zeta\phi_I(\zeta)d\zeta$$
$$=\frac{1}{2}\frac{\tau}{T-N}\Big[\tau\alpha_2-2\alpha_1\Big]\,\mathrm{tr}(QC_2)\Big\{-2z\phi(z)\Big[\Phi(z)-\Phi(-z)\Big]^{N-1}+\Big[\Phi(z)-\Phi(-z)\Big]^N\Big\},$$

$$(B.24)$$

where the last equality holds by using

$$\int_{|x|<z} x^2\phi(x)dx = -2z\phi(z) + \Phi(z) - \Phi(-z).$$

The results follow by adding the right-hand side of equations (B.22)–(B.24).

■

Remark 1 *Sawa (1973a) introduces a combined estimator using the OLS and the 2SLS estimators that takes the following form*

$$\hat{\beta}_S = \frac{T-N-1}{T-K}\hat{\beta}(1) - \frac{K-N-1}{T-K}\hat{\beta}(0). \tag{B.25}$$

Under assumptions 1–3, given $T > K$ and $K - N > 1$, the asymptotic MSE matrix of Sawa (1973a)'s combined estimator is equal to

$$\begin{aligned} AMSE(\hat{\beta}_S) &= \mathbb{E}\left(\frac{1}{\sigma^2}(\hat{\beta}_S - \dot{\beta})(\hat{\beta}_S - \dot{\beta})'\right) \\ &= AMSE(\hat{\beta}(1)) + \sigma^2\left(\frac{(K-N-1)^2}{T-K} + 2(K-N-1)\right)QC_2Q + O(\sigma^3) \geq 0 + O(\sigma^3), \end{aligned}$$

where the proof follows easily from Theorem 5.1 of Sawa (1973a). Since right hand side of the above equation is always nonnegative, it implies that Sawa (1973a)'s combined estimator is always dominated by the 2SLS estimator in terms of their MSEs.

■

Remark 2 *Morimune (1978) introduces a combined estimator using the OLS and the LIML estimators that takes the following form*

$$\hat{\beta}_M = \frac{T-N-1}{T-N}\hat{\beta}(\lambda) + \frac{1}{T-N}\hat{\beta}(0). \tag{B.26}$$

Under assumptions 1–3, given $T - K - 2 > 0$ and $K > N$, the asymptotic MSE matrix of Morimune (1978)'s combined estimator is equal to

$$\begin{aligned} AMSE(\hat{\beta}_M) &= \mathbb{E}\left(\frac{1}{\sigma^2}(\hat{\beta}_M - \dot{\beta})(\hat{\beta}_M - \dot{\beta})'\right) \\ &= AMSE(\hat{\beta}(\lambda)) - \sigma^2\left(1 + \frac{K-N}{T-K-2}\right)\left(2 - \frac{1}{T-N}\right) QC_2Q + O(\sigma^3), \end{aligned}$$

where the proof follows easily from Theorem 2 of Morimune (1978).

Furthermore, the MSE of the Stein-like shrinkage estimator using the OLS and the LIML estimators introduced in this paper, using the optimal choice of the tuning parameter, is equal to

$$\begin{aligned} AMSE(\hat{\beta}_{c,\lambda}) &= \mathbb{E}\left(\frac{1}{\sigma^2}(\hat{\beta}_{c,\lambda} - \dot{\beta})(\hat{\beta}_{c,\lambda} - \dot{\beta})'\right) \\ &= AMSE(\hat{\beta}(\lambda)) - \sigma^2\left(\frac{(T-K)(T-N)(N-2)}{(T-N-2)N}\right) QC_2Q + O(\sigma^3). \end{aligned}$$

Therefore, the Stein-like shrinkage estimator dominates Morimune (1978)'s combined estimator when the following condition holds,

$$\frac{(T-N-2)(2(T-N)-1)}{(T-K-2)(T-N)} \leq \frac{(T-K)(T-N)(N-2)}{(T-N-2)N}.$$

The condition depends on the sample size, the number of included endogenous variables, and the number of excluded exogenous variables. However, a necessary condition for the above inequality to hold is when $T \geq 2(K+1)$.

■

C Appendix C

This section provides further Monte Carlo simulation results considered in the main text when the error terms are generated from normalized chi-squared distribution with two degrees of freedom. The observations are generated by the process

$$y_1 = Y_2\beta + u_1,$$
$$Y_2 = X\Pi_2 + V_2,$$

where u_1 has a $(\chi^2 - 2)/2$ distribution. Similarly columns of V_2 are generated from independent normalized chi-squared distribution with two degrees of freedom, and X has a multivariate normal distribution with mean zero, and variance-covariance matrix I_K. We set the correlation between u_1 and the rows of V_2 equal to $\rho/\sqrt{N}$, where ρ takes values on $\{0.01, 0.1, 0.5, 0.9, 0.99\}$. We set β to $0.1\iota_N$, where ι_q is a q-dimensional vector of unity. $\Pi_2 = c(I_N \otimes \iota_{K/N})$, where $\otimes$ denotes the Kronecker product, and $c = \sqrt{R^2/K(1-R^2)}$. We consider three values for the reduced form population R^2, which are $\{0.1, 0.5, 0.9\}$. The number of monte carlo simulations for each design is set to $1,000$. We set the value of $\tau = \tau_{opt}$ when $N = \{3, 6\}$ and set $\tau = 1/8$ when $N = 1$. The results are reported for $T = \{100, 1000\}$, $N = \{1, 3, 6\}$, and $K = \{6, 18, 30\}$ in Table C.1– Table C.6.

Table C.1: Relative Median Squared Errors when $(T, N) = (100, 1)$

		$K = 6$							$K = 18$						
R^2	ρ	$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$
	0.01	0.120	0.603	0.126	0.058	0.593	0.063	2.028	0.237	0.810	0.257	0.034	0.663	0.039	5.708
	0.1	0.339	0.784	0.364	0.161	0.681	0.173	1.832	0.484	0.813	0.509	0.077	0.660	0.088	5.130
0.1	0.5	4.415	1.150	4.022	3.940	1.049	3.623	1.023	2.039	1.130	1.913	2.061	1.123	1.958	0.983
	0.9	7.981	1.043	1.054	17.363	1.006	1.077	0.443	2.179	1.013	1.002	12.653	0.962	1.213	0.164
	0.99	7.639	1.033	1.034	19.890	0.973	1.010	0.362	2.216	1.008	1.000	21.141	0.932	1.011	0.097
	0.01	0.537	0.934	0.659	0.480	0.864	0.586	1.036	0.519	0.827	0.579	0.364	0.813	0.406	1.402
	0.1	0.816	0.970	0.849	0.699	0.918	0.771	1.105	0.987	0.970	1.068	0.723	0.902	0.814	1.269
0.5	0.5	12.762	1.021	1.036	11.755	1.004	1.070	1.068	8.487	1.027	1.059	11.920	0.973	1.141	0.674
	0.9	39.751	1.018	1.000	46.086	0.969	1.000	0.820	11.603	1.025	1.000	44.403	0.986	1.000	0.251
	0.99	41.786	1.017	1.000	46.468	0.985	1.000	0.871	11.682	1.013	1.000	58.873	1.022	1.000	0.200
	0.01	0.880	0.942	0.906	0.877	0.928	0.884	0.989	0.959	0.989	0.992	0.971	1.025	0.992	1.024
	0.1	0.924	0.953	0.968	0.907	0.932	0.912	0.996	1.015	0.994	1.019	0.983	1.002	0.981	1.040
0.9	0.5	4.714	0.992	1.000	4.931	0.998	1.002	0.961	4.517	1.024	1.018	4.800	1.010	1.192	0.929
	0.9	15.640	0.990	1.000	16.512	1.013	1.000	0.969	11.677	0.995	1.000	14.066	1.031	1.000	0.861
	0.99	18.112	0.996	1.000	19.066	1.009	1.000	0.963	13.839	1.006	1.000	19.470	0.996	1.000	0.704

Note: This table reports the relative median squared errors of the OLS ($\hat{\beta}(0)$), the 2SLS ($\hat{\beta}(1)$), the LIML ($\hat{\beta}(\lambda)$) estimators, the Stein-like shrinkage estimator using the OLS and the 2SLS estimators ($\hat{\beta}_{c,1}$), the Stein-like shrinkage estimator using the OLS and the LIML estimators ($\hat{\beta}_{c,\lambda}$), and two pre-test estimators ($\hat{\beta}_{pre}$), i.e., $\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$ indicates the median squared errors of the OLS estimator divided by the median squared errors of the 2SLS estimator. The pre-test estimators use the Wu-Hausman test static under 5% critical value to choose between the OLS and the 2SLS/LIML estimators.

Table C.2: Relative Median Squared Errors when $(T, N) = (100, 3)$

		$K=6$							$K=18$						
R^2	ρ	$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$
0.1	0.01	0.055	0.166	0.055	0.014	0.033	0.014	0.785	0.163	0.420	0.168	0.009	0.055	0.009	2.442
	0.1	0.073	0.190	0.073	0.018	0.037	0.018	0.811	0.225	0.461	0.227	0.012	0.071	0.012	2.830
	0.5	0.616	0.608	0.616	0.144	0.151	0.144	1.063	0.940	0.921	0.940	0.109	0.139	0.109	1.299
	0.9	1.765	1.104	1.752	0.601	0.508	0.601	1.350	1.319	1.087	1.309	0.510	0.462	0.510	1.099
	0.99	2.086	1.095	2.072	0.640	0.581	0.639	1.728	1.366	1.040	1.142	0.531	0.501	0.531	1.240
0.5	0.01	0.232	0.501	0.237	0.185	0.477	0.190	1.195	0.338	0.604	0.350	0.120	0.470	0.124	2.198
	0.1	0.319	0.551	0.320	0.270	0.547	0.271	1.170	0.443	0.659	0.456	0.170	0.526	0.176	2.078
	0.5	2.219	0.957	1.940	1.770	0.861	1.603	1.127	2.212	1.114	1.923	1.258	0.807	1.233	1.273
	0.9	7.121	0.994	1.004	6.047	0.923	1.000	1.094	4.027	1.091	1.000	4.297	0.895	1.005	0.768
	0.99	7.912	0.990	1.000	6.734	0.936	1.000	1.111	4.182	1.073	1.000	4.951	0.885	1.000	0.697
0.9	0.01	0.775	0.861	0.786	0.757	0.831	0.763	0.988	0.824	0.928	0.849	0.740	0.857	0.747	1.028
	0.1	0.802	0.874	0.857	0.790	0.865	0.826	1.005	0.815	0.888	0.832	0.747	0.839	0.753	1.031
	0.5	2.566	0.989	1.021	2.550	0.995	1.039	1.013	2.503	1.021	1.046	2.380	0.989	1.160	1.019
	0.9	7.333	0.988	1.000	7.092	0.986	1.000	1.032	5.941	1.025	1.000	6.199	0.991	1.000	0.926
	0.99	7.541	0.991	1.000	7.395	0.994	1.000	1.023	7.168	1.033	1.000	7.599	0.978	1.000	0.893

See the note to Table C.1.

Table C.3: Relative Median Squared Errors when $(T, N) = (100, 6)$

		$K = 6$							$K = 18$						
R^2	ρ	$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$
0.1	0.01	0.006	0.006	0.006	0.006	0.006	0.006	1.000	0.140	0.172	0.140	0.005	0.006	0.005	0.855
	0.1	0.008	0.008	0.008	0.008	0.008	0.008	1.000	0.149	0.177	0.149	0.006	0.007	0.006	0.865
	0.5	0.045	0.045	0.045	0.045	0.045	0.045	1.000	0.558	0.576	0.558	0.038	0.038	0.038	0.979
	0.9	0.168	0.168	0.168	0.168	0.168	0.168	1.000	1.055	1.038	1.055	0.183	0.182	0.183	1.012
	0.99	0.246	0.246	0.246	0.246	0.246	0.246	1.000	1.127	1.068	1.126	0.289	0.288	0.289	1.053
0.5	0.01	0.088	0.102	0.088	0.088	0.102	0.088	1.000	0.218	0.295	0.219	0.034	0.045	0.034	0.969
	0.1	0.117	0.127	0.117	0.117	0.127	0.117	1.000	0.242	0.295	0.242	0.036	0.043	0.036	0.993
	0.5	0.473	0.422	0.473	0.473	0.422	0.473	1.000	0.958	0.845	0.959	0.188	0.180	0.188	1.085
	0.9	1.285	0.678	1.267	1.285	0.678	1.267	1.000	2.039	1.230	1.757	0.554	0.453	0.551	1.354
	0.99	1.596	0.690	1.573	1.596	0.690	1.573	1.000	2.344	1.192	1.059	0.626	0.494	0.623	1.551
0.9	0.01	0.587	0.645	0.590	0.587	0.645	0.590	1.000	0.638	0.714	0.645	0.529	0.596	0.533	1.007
	0.1	0.625	0.676	0.644	0.625	0.676	0.644	1.000	0.668	0.707	0.688	0.547	0.592	0.556	1.024
	0.5	1.433	0.891	1.100	1.433	0.891	1.100	1.000	1.704	1.006	1.200	1.371	0.842	1.233	1.040
	0.9	3.733	0.918	1.000	3.733	0.918	1.000	1.000	3.857	1.046	1.000	3.407	0.895	1.000	0.969
	0.99	4.179	0.927	1.000	4.179	0.927	1.000	1.000	4.595	1.054	1.000	4.097	0.923	1.000	0.982

See the note to Table C.1.

Table C.4: Relative Median Squared Errors when $(T, N) = (1000, 1)$

		$K=6$							$K=30$						
R^2	ρ	$\frac{\hat\beta(0)}{\hat\beta(1)}$	$\frac{\hat\beta_{c,1}}{\hat\beta(1)}$	$\frac{\hat\beta_{pre}}{\hat\beta(1)}$	$\frac{\hat\beta(0)}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{pre}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta_{c,1}}$	$\frac{\hat\beta(0)}{\hat\beta(1)}$	$\frac{\hat\beta_{c,1}}{\hat\beta(1)}$	$\frac{\hat\beta_{pre}}{\hat\beta(1)}$	$\frac{\hat\beta(0)}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{pre}}{\hat\beta(\lambda)}$	$\frac{\hat\beta_{c,\lambda}}{\hat\beta_{c,1}}$
0.1	0.01	0.120	0.596	0.130	0.109	0.604	0.118	1.117	0.125	0.645	0.139	0.077	0.642	0.095	1.619
	0.1	2.132	1.149	2.251	1.992	1.139	2.140	1.061	2.239	1.187	2.302	1.332	1.022	1.486	1.448
	0.5	53.757	0.995	1.000	49.121	1.027	1.000	1.130	17.959	1.034	1.000	46.733	1.004	1.000	0.373
	0.9	140.893	1.003	1.000	186.025	1.033	1.000	0.780	18.323	1.009	1.000	147.932	0.995	1.000	0.122
	0.99	166.671	1.003	1.000	187.700	0.971	1.000	0.860	17.981	1.007	1.000	201.557	0.997	1.000	0.088
0.5	0.01	0.459	0.865	0.491	0.455	0.863	0.486	1.008	0.537	0.890	0.604	0.508	0.875	0.573	1.039
	0.1	5.509	1.018	2.080	5.624	1.047	2.163	1.007	5.497	1.042	2.099	5.193	1.062	2.193	1.079
	0.5	141.624	0.992	1.000	126.901	0.995	1.000	1.120	104.037	0.994	1.000	136.370	1.002	1.000	0.769
	0.9	436.847	1.001	1.000	423.353	1.005	1.000	1.036	222.012	1.008	1.000	459.761	0.998	1.000	0.478
	0.99	561.181	1.002	1.000	585.026	0.988	1.000	0.946	273.067	1.006	1.000	509.227	1.005	1.000	0.536
0.9	0.01	0.837	0.945	0.886	0.832	0.936	0.883	0.997	1.071	1.009	1.054	1.066	1.014	1.058	1.010
	0.1	2.136	0.995	1.090	2.143	1.002	1.093	1.003	2.520	1.027	1.038	2.595	1.020	1.081	0.965
	0.5	49.777	0.993	1.000	49.314	0.999	1.000	1.016	44.174	1.001	1.000	46.857	1.003	1.000	0.945
	0.9	166.949	0.997	1.000	162.073	0.997	1.000	1.030	169.999	0.998	1.000	176.045	1.000	1.000	0.968
	0.99	183.427	1.004	1.000	179.194	0.996	1.000	1.016	170.166	0.998	1.000	191.305	1.002	1.000	0.893

See the note to Table C.1.

Table C.5: Relative Median Squared Errors when $(T, N) = (1000, 3)$

R^2	ρ	$K=6$							$K=30$						
		$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$
0.1	0.01	0.035	0.328	0.036	0.029	0.358	0.029	1.330	0.065	0.345	0.068	0.018	0.528	0.020	5.631
	0.1	0.170	0.394	0.171	0.146	0.407	0.147	1.202	0.286	0.486	0.295	0.090	0.539	0.098	3.527
	0.5	3.733	0.957	2.033	3.104	0.909	1.906	1.142	2.911	1.168	2.496	1.887	0.878	1.660	1.160
	0.9	12.854	0.997	1.000	10.468	0.932	1.000	1.148	4.449	1.063	1.000	6.877	0.926	1.000	0.563
	0.99	15.729	1.016	1.000	13.351	0.939	1.000	1.090	4.757	1.042	1.000	8.873	0.923	1.000	0.475
0.5	0.01	0.258	0.544	0.267	0.255	0.546	0.265	1.012	0.279	0.546	0.291	0.235	0.523	0.250	1.136
	0.1	0.966	0.845	1.072	0.953	0.844	1.067	1.013	1.075	0.906	1.186	0.936	0.831	1.081	1.053
	0.5	20.098	1.006	1.000	19.896	0.991	1.000	0.995	18.476	1.035	1.000	18.087	0.998	1.000	0.985
	0.9	66.760	0.992	1.000	65.088	0.997	1.000	1.031	45.804	1.033	1.000	57.506	0.982	1.000	0.757
	0.99	79.496	1.003	1.000	77.485	0.992	1.000	1.015	48.778	1.029	1.000	71.379	0.989	1.000	0.657
0.9	0.01	0.793	0.871	0.838	0.788	0.867	0.833	1.001	0.748	0.878	0.765	0.741	0.868	0.754	0.997
	0.1	1.418	0.983	1.178	1.407	0.978	1.172	1.003	1.498	0.988	1.186	1.470	0.978	1.219	1.009
	0.5	20.074	1.006	1.000	19.898	0.992	1.000	0.995	21.057	0.989	1.000	20.755	1.011	1.000	1.037
	0.9	67.314	0.999	1.000	67.734	0.999	1.000	0.994	58.662	0.996	1.000	61.731	0.999	1.000	0.953
	0.99	77.247	1.002	1.000	77.033	1.001	1.000	1.002	74.963	1.000	1.000	77.572	0.994	1.000	0.961

See the note to Table C.1.

Table C.6: Relative Median Squared Errors when $(T, N) = (1000, 6)$

		$K = 6$							$K = 30$						
R^2	ρ	$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{c,1}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(1)}$	$\frac{\hat{\beta}(0)}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{pre}}{\hat{\beta}(\lambda)}$	$\frac{\hat{\beta}_{c,\lambda}}{\hat{\beta}_{c,1}}$
0.1	0.01	0.013	0.018	0.013	0.013	0.018	0.013	1.000	0.040	0.077	0.040	0.004	0.010	0.004	1.329
	0.1	0.035	0.043	0.035	0.035	0.043	0.035	1.000	0.109	0.163	0.109	0.011	0.023	0.011	1.400
	0.5	0.618	0.496	0.619	0.618	0.496	0.619	1.000	1.265	1.067	1.264	0.207	0.201	0.208	1.152
	0.9	2.011	0.703	1.957	2.011	0.703	1.957	1.000	2.273	1.189	1.005	0.706	0.490	0.698	1.327
	0.99	2.742	0.735	1.563	2.742	0.735	1.563	1.000	2.422	1.110	1.000	0.825	0.485	0.822	1.282
0.5	0.01	0.139	0.217	0.141	0.139	0.217	0.141	1.000	0.159	0.251	0.166	0.118	0.237	0.126	1.274
	0.1	0.339	0.393	0.362	0.339	0.393	0.362	1.000	0.390	0.446	0.410	0.288	0.380	0.316	1.154
	0.5	5.587	0.914	1.000	5.587	0.914	1.000	1.000	6.070	1.083	1.000	4.895	0.907	1.000	1.039
	0.9	17.486	0.951	1.000	17.486	0.951	1.000	1.000	16.425	1.117	1.000	15.779	0.953	1.000	0.888
	0.99	21.467	0.923	1.000	21.467	0.923	1.000	1.000	18.801	1.105	1.000	19.390	0.960	1.000	0.842
0.9	0.01	0.616	0.696	0.635	0.616	0.696	0.635	1.000	0.586	0.675	0.624	0.571	0.661	0.605	1.004
	0.1	0.997	0.864	1.065	0.997	0.864	1.065	1.000	0.997	0.901	1.030	0.966	0.875	1.014	1.002
	0.5	11.985	0.985	1.000	11.985	0.985	1.000	1.000	11.498	1.002	1.000	11.263	0.986	1.000	1.005
	0.9	36.093	0.990	1.000	36.093	0.990	1.000	1.000	37.182	1.027	1.000	35.950	0.982	1.000	0.989
	0.99	43.981	0.998	1.000	43.981	0.998	1.000	1.000	43.738	1.008	1.000	43.534	0.992	1.000	0.989

See the note to Table C.1.