On Uniform Asymptotic Risk of Averaging GMM Estimators*

Xu Cheng† Zhipeng Liao‡ Ruoyao Shi§

This Version: August, 2018

Abstract

This paper studies the averaging GMM estimator that combines a conservative GMM estimator based on valid moment conditions and an aggressive GMM estimator based on both valid and possibly misspecified moment conditions, where the weight is the sample analog of an infeasible optimal weight. We establish asymptotic theory on uniform approximation of the upper and lower bounds of the finite-sample truncated risk difference between any two estimators, which is used to compare the averaging GMM estimator and the conservative GMM estimator. Under some sufficient conditions, we show that the asymptotic lower bound of the truncated risk difference between the averaging estimator and the conservative estimator is strictly less than zero, while the asymptotic upper bound is zero uniformly over any degree of misspecification. The results apply to quadratic loss functions. This uniform asymptotic dominance is established in non-Gaussian semiparametric nonlinear models.

Keywords: Asymptotic Risk, Finite-Sample Risk, Generalized Shrinkage Estimator, GMM, Misspecification, Model Averaging, Non-Standard Estimator, Uniform Approximation

*We thank Benedikt Pötscher for the insightful comments that greatly improved the quality of the paper. We appreciate useful suggestions from Karl Schmedders, the co-editor and three anonymous referees. The paper also benefited from discussions with Donald Andrews, Denis Chetverikov, Patrik Guggenberger, Jinyong Hahn, Jia Li, Rosa Matzkin, Hyunsik Roger Moon, Ulrich Mueller, Joris Pinkse, Frank Schorfheide, Shuyang Sheng, and participants at various seminars and conferences.

†Department of Economics, University of Pennsylvania, 133 S. 36th St, Philadelphia, PA 19104, USA. Email: xucheng@upenn.edu

‡Department of Economics, UCLA,8379 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095. Email: zhipeng.liao@econ.ucla.edu

§Department of Economics, UC Riverside, 3136 Sproul Hall, Riverside, CA, 92521. Email: ruoyao.shi@ucr.edu

1 Introduction

We are interested in estimating some finite dimensional parameter $\theta_F \in \mathbb{R}^{d_\theta}$ which is uniquely identified by the moment restrictions

$$\mathbb{E}_F[g_1(W, \theta_F)] = 0_{r_1 \times 1} \tag{1.1}$$

for some known vector function $g_1(\cdot) : \mathcal{W} \times \Theta \rightarrow \mathbb{R}^{r_1}$, where Θ is a compact subset of $\mathbb{R}^{d_\theta}$, W is a random vector with support $\mathcal{W}$ and joint distribution F, and $\mathbb{E}_F[\cdot]$ denotes the expectation operator under F. Suppose we have i.i.d. data $\{W_i\}_{i=1}^n$, where W_i has distribution F for $i = 1, \ldots, n$.[1] Let $\overline{g}_1(\theta) = n^{-1} \sum_{i=1}^n g_1(W_i, \theta)$. An efficient GMM estimator for θ_F is

$$\widehat{\theta}_1 = \arg\min_{\theta \in \Theta} \overline{g}_1(\theta)'(\overline{\Omega}_1)^{-1} \overline{g}_1(\theta), \tag{1.2}$$

where $\overline{\Omega}_1 = n^{-1} \sum_{i=1}^n g_1(W_i, \widetilde{\theta}_1) g_1(W_i, \widetilde{\theta}_1)' - \overline{g}_1(\widetilde{\theta}_1)\overline{g}_1(\widetilde{\theta}_1)'$ is the efficient weighting matrix with some preliminary consistent estimator $\widetilde{\theta}_1$.[2] In a linear instrumental variable (IV) example, $Y = X'\theta_F + U$ where the IV $Z_1 \in \mathbb{R}^{r_1}$ satisfies $\mathbb{E}_F[Z_1 U] = 0_{r_1 \times 1}$. The moments in (1.1) hold with $g_1(W, \theta) = Z_1(Y - X'\theta)$ and θ_F is uniquely identified if $\mathbb{E}_F[Z_1 X']$ has full column rank. Under certain regularity conditions, it is well-known that $\widehat{\theta}_1$ is consistent and achieves the lowest asymptotic variance among GMM estimators based on the moments in (1.1), see Hansen (1982).

If one has additional moments

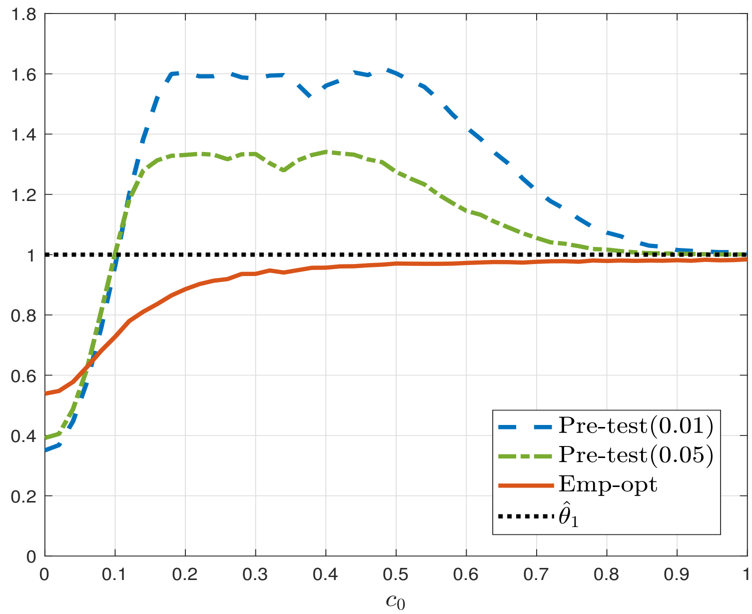
$$\mathbb{E}_F[g^*(W, \theta_F)] = 0_{r^* \times 1} \tag{1.3}$$

for some known function $g^*(\cdot) : \mathcal{W} \times \Theta \rightarrow \mathbb{R}^{r^*}$, imposing them together with (1.1) can further reduce the asymptotic variance of the GMM estimator. However, if these additional moments are misspecified in the sense that $\mathbb{E}_F[g^*(W, \theta_F)] \neq 0_{r^* \times 1}$, imposing (1.3) may result in inconsistent estimation. The choice of moment conditions is routinely faced by empirical researchers. Take the linear IV model for example. One typically starts with a large number of candidate IVs but only has confidence that a small number of them are valid, denoted by Z_1. The rest of them, denoted by Z^*, are valid only under certain economic hypothesis that yet to be tested. In this example, $g^*(W, \theta) = Z^*(Y - X'\theta)$. In contrast to the conservative estimator $\widehat{\theta}_1$, an aggressive estimator $\widehat{\theta}_2$ always imposes (1.3) regardless of its validity. Let $g_2(W_i, \theta) = (g_1(W_i, \theta)', g^*(W_i, \theta)')'$

[1] The main theory of the paper can be easily extended to time series models with dependent data, as long as the preliminary results in Lemma B.1 hold.

[2] For example, $\widetilde{\theta}_1$ could be the GMM estimator similar to $\widehat{\theta}_1$ but with an identity weighting matrix, see (B.5) in the Appendix.

Figure 1: Finite Sample (n = 500) MSEs of the Pre-test and the Averaging GMM Estimators



Note: "Pre-test(0.01)" and "Pre-test(0.05)" refer to the pre-test GMM estimator based on the J test statistic $n\overline{g}_2(\widehat{\theta}_2)'(\overline{\Omega}_2)^{-1}\overline{g}_2(\widehat{\theta}_2)$ with nominal size 0.01 and 0.05 respectively. "Emp-opt" refers to the averaging GMM estimator with weight defined in (4.3) of the paper. In this simulation, we set $\delta_F = c_0\omega$ where c_0 is in [0,1] and ω is a real vector. At each c_0, we consider 127 different values for ω and report the largest finite sample MSEs of the estimators. Details of the simulation design for this figure is provided in Subsection 6.1.

for $i = 1, \ldots, n$, and $\overline{g}_2(\theta) = n^{-1}\sum_{i=1}^{n} g_2(W_i, \theta)$. The aggressive estimator $\widehat{\theta}_2$ takes the form

$$\widehat{\theta}_2 = \arg\min_{\theta\in\Theta} \overline{g}_2(\theta)'(\overline{\Omega}_2)^{-1}\overline{g}_2(\theta), \tag{1.4}$$

where $\overline{\Omega}_2$ is constructed in the same way as $\overline{\Omega}_1$ except that $g_1(W_i, \theta)$ is replaced by $g_2(W_i, \theta)$.[3]

Because imposing $\mathbb{E}_F[g^*(W, \theta_F)] = 0_{r^*\times 1}$ is a double-edged sword, a data-dependent decision usually is made to choose between $\widehat{\theta}_1$ and $\widehat{\theta}_2$. To study such a decision and the subsequent estimator, let

$$\delta_F = \mathbb{E}_F[g^*(W, \theta_F)] \in \mathbb{R}^{r^*}. \tag{1.5}$$

The pre-testing approach tests the null hypothesis $H_0 : \delta_F = 0_{r^*\times 1}$ and constructs an estimator

$$\widehat{\theta}_{pre} = 1\{T_n > c_\alpha\}\widehat{\theta}_1 + 1\{T_n \leq c_\alpha\}\widehat{\theta}_2 \tag{1.6}$$

for some test statistic T_n with the critical value c_α at the significance level α. One popular test is the

[3]See the first line of equations (E.13) in the Supplemental Appendix for the definition of $\overline{\Omega}_2$. In particular, $\overline{\Omega}_2$ is constructed using $\widetilde{\theta}_1$, the preliminary consistent estimator based on the valid moment conditions in (1.1) only.

J-test, see Hansen (1982), and c_α is the $1-\alpha$ quantile of the chi-squared distribution with degree of freedom $r_2 - d_\theta$ where $r_2 = r_1 + r^*$. Because the power of this test against the fixed alternative is 1, $\widehat{\theta}_{pre}$ equals $\widehat{\theta}_1$ with probability 1 asymptotically ($n \to \infty$) for those fixed misspecified model where $\delta_F \neq 0_{r^* \times 1}$. Thus, it seems that $\widehat{\theta}_{pre}$ is immune to moment misspecification. However, we care about the finite-sample mean squared error (MSE) of $\widehat{\theta}_{pre}$ in practice and this standard pointwise asymptotic analysis (δ_F is fixed and $n \to \infty$) provides a poor approximation to the former.[4] To see the comparison between $\widehat{\theta}_{pre}$ and $\widehat{\theta}_1$, the dashed line and the dashed-dotted line in Figure 1 plot the maximum finite-sample ($n = 500$) MSEs of $\widehat{\theta}_{pre}$ with $\alpha = 0.01$ and 0.05 respectively, while the MSE of $\widehat{\theta}_1$ is normalized to 1.[5] For some values of δ_F, the MSE of $\widehat{\theta}_{pre}$ may be larger than that of $\widehat{\theta}_1$, sometimes by more than 50%. Note that the pre-test estimators exhibit multiple peaks because the simulation design allows for multiple potentially misspecified moments and considers two different ways of parametrizing δ_F. Given c_0, the norm of δ_F may be different in the two different parametrizations.

The goal of this paper is twofold. First, we propose a data-dependent averaging of $\widehat{\theta}_1$ and $\widehat{\theta}_2$ that takes the form

$$\widehat{\theta}_{eo} = (1 - \widetilde{\omega}_{eo})\widehat{\theta}_1 + \widetilde{\omega}_{eo}\widehat{\theta}_2 \tag{1.7}$$

where $\widetilde{\omega}_{eo} \in [0, 1]$ is a data-dependent weight specified in (4.7) below. The subscript in $\widetilde{\omega}_{eo}$ is short for empirical optimal because this weight is an empirical analog of an infeasible optimal weight ω_F^* defined in (4.3) below. We plot the finite-sample MSE of this averaging estimator as the solid line in Figure 1. This averaging estimator is robust to misspecification in the sense that the solid line is below 1 for all values of δ_F, in contrast to the bump in the dashed line that represents the pre-test estimator. Second, we develop a *uniform* asymptotic theory to justify the finite-sample robustness of this averaging estimator. We quantify the upper and lower bounds of the asymptotic risk differences between the averaging estimator and the conservative estimator, and show that this averaging estimator dominates the conservative estimator uniformly over a large class of models with different degrees of misspecification in certain asymptotic sense.[6] The uniform dominance is established under the truncated weighted loss function which is defined in (3.11) below.[7] Our uniform dominance result relies on the asymptotic properties of the GMM estimators, therefore it is weaker than the exact finite sample dominance result of the James-Stein estimator established

[4]The poor approximation of the pointwise asymptotics to the finite sample properties of the pre-test estimator has been noted in Shibata (1986), Pötscher (1991), Kabaila (1995, 1998) and Leeb and Pötscher (2005, 2008), among others.

[5]That is, the dashed line and the dashed-dotted line represent the ratios of the maximum MSEs of the two pre-test estimators divided by the MSE of $\widehat{\theta}_1$ respectively.

[6]The lower and upper bounds of asymptotic risk difference are defined in (3.12) below.

[7]Truncation at a large number is needed for the asymptotic analysis of the risk of general estimator without imposing stringent conditions such as uniform integrability.

in the Gaussian sampling models.

The rest of the paper is organized as follows. Section 2 discusses the literature related to our paper. Section 3 defines the parameter space over which the uniform result is established and defines uniform dominance. Section 4 introduces the averaging weight. Section 5 provides an analytical representation of the bounds of the asymptotic risk differences and applies it to show that the averaging GMM estimator uniformly dominates the conservative estimator. Section 6 investigates the finite sample performance of our averaging estimator using Monte Carlo simulations. Section 7 concludes. Proof of the main results of the paper and additional simulation results are given in the Appendix. Analysis of the pre-test estimator, extra simulation studies and proofs of some auxiliary results are included in the Supplemental Appendix of the paper.

Notation. For any real matrix A, we use $||A||$ to denote the Frobenius norm of A, that is $||A|| = (tr(A'A))^{1/2}$ where $tr(\cdot)$ denotes the trace operator of square matrices. If A is a real symmetric matrix, $\rho_{\min}(A)$ and $\rho_{\max}(A)$ denote the smallest and largest eigenvalues of A, respectively. For any positive integers d_1 and d_2, I_{d_1} and $0_{d_1 \times d_2}$ stand for the $d_1 \times d_1$ identity matrix and $d_1 \times d_2$ zero matrix, respectively. Let $\text{vec}(\cdot)$ denotes vectorization of a matrix and $\text{vech}(\cdot)$ denotes the half vectorization of a symmetric matrix. Let $\mathbb{R} = (-\infty, +\infty)$, $\mathbb{R}_+ = [0, +\infty)$, $\mathbb{R}_\infty = \mathbb{R} \cup \{\pm\infty\}$ and $\mathbb{R}_{+,\infty} = \mathbb{R}_+ \cup \{+\infty\}$. For any positive integer d and any set $\mathbb{S}$, $\mathbb{S}^d$ denotes the Cartesian product of d many sets: $\mathbb{S}_1 \times \cdots \times \mathbb{S}_d$ with $\mathbb{S}_j = \mathbb{S}$ for $j = 1, \ldots, d$. For any finite positive integer d and any set $\mathbb{S} \subset \mathbb{R}^d$, $int(\mathbb{S})$ denotes the interior of $\mathbb{S}$ under the Euclidean norm. We use $\mathbb{N}$ to denote the set of natural numbers and $\{p_n\} = \{p_n : n \in \mathbb{N}\}$ denote a subsequence of $\{n\}_{n \in \mathbb{N}}$. For any (possibly random) positive sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$, $a_n = O_p(b_n)$ means that $\sup_{n \in \mathbb{N}} \Pr(a_n/b_n > c) \to 0$ as $c \to \infty$; $a_n = o_p(b_n)$ means that for all $\varepsilon > 0$, $\lim_{n \to \infty} \Pr(a_n/b_n > \varepsilon) = 0$. Let "$\to_p$" and "$\to_D$" stand for convergence in probability and convergence in distribution, respectively. The notation $a \equiv b$ means a is defined as b.

2 Related Literature

Our uniform dominance result is related to the Stein's phenomenon (Stein, 1956) in parametric models. The James-Stein (JS) estimator shrinks the maximum likelihood estimator (MLE) toward zero and has been shown to dominate the MLE in exact normal sampling, see James and Stein (1961). Green and Strawderman (1991) propose an averaging estimator in the Gaussian location model which shrinks an unbiased estimator toward a biased estimator with a JS-type of weight, and show that the averaging estimator dominates the unbiased estimator. Green and Strawderman (1991) assume that the unbiased estimator is independent of the biased estimator, which is relaxed in Kim and White (2001), Judge and Mittelhammer (2004) and Mittelhammer and Judge (2005).[8], [9] These papers propose averaging estimators which shrink asymptotically unbiased estimators toward biased estimators in the semiparametric setting, and show that the averaging estimators based on infeasible weights dominate the unbiased estimators in the Gaussian location models. Kim and White (2001) show that the infeasible weight can be consistently estimated when the asymptotic bias of the biased estimator is zero. Judge and Mittelhammer (2004) and Mittelhammer and Judge (2005) provide approximators of the infeasible optimal weights and show that these approximators can be consistently estimated. These estimators and our estimator are all linear combinations of the unbiased estimator and the biased estimator. However even in the Gaussian location models, the weights are different and their sufficient conditions and proofs for dominance are different, see Appendix A for details.[10]

Hansen (2016) considers the JS type averaging estimator in general parametric models and shows the Stein-dominance result in a pointwise local asymptotic sense.[11] Hansen (2017) proposes an averaging estimator that combines the ordinary least squares (OLS) estimator and the two-stage-least-squares (2SLS) estimator in linear IV models, and shows that the averaging estimator has smaller local asymptotic risk than the OLS estimator. DiTraglia (2016) also studies the averaging GMM estimator in the pointwise local asymptotic framework. The averaging weight of his estimator is based on the focused moment selection criterion with a targeted parameter. The simulation results in the paper show that this averaging estimator does not uniformly dominate the conservative estimator. Many other frequentist model averaging estimators are studied in the literature, including Buckland, Burnham, and Augustin (1997), Hjort and Claeskens (2003, 2006), Hansen (2007), Claeskens and Carroll (2007), Hansen and Racine (2012), Cheng and Hansen (2015), Lu and Su (2015), to name only a few.

Different from the aforementioned papers, our paper is the first to show global dominance based on uniform asymptotic approximation.[12] This uniform analysis is similar to those studied

[8] We thank an anonymous referee who referred Green and Strawderman (1991) and Judge and Mittelhammer (2004) to us.

[9] Judge and Mittelhammer (2007) propose an averaging estimator which combine different GMM estimators with weights determined by the empirical likelihood method. However, the properties of this averaging estimator are not fully investigated and no dominace results are established in this paper.

[10] In the Gaussian location model, our dominance results require $d_\theta \geq 4$; Green and Stawderman (1991) requires $d_\theta \geq 3$ by imposing independence between the unbiased and the biased estimators; and Kim and White (2001), Judge and Mittelhammer (2004) and Mittelhammer and Judge (2005) require $d_\theta \geq 5$.

[11] For a given real vector d, the pointwise local asymptotic analysis considers a sequence of local DGPs $\{F_n\}_n$ under which $\delta_{F_n} = dn^{-1/2}$, and derives the asymptotic (truncated) risk of the averaging estimator under $\{F_n\}_n$ for the given d. Such analysis will produce a pointwise risk function (on d) for the averaging estimator. Evaluation of the averaging estimator is then conducted using the pointwise local asymptotic risk function.

[12] In the uniform global asymptotic framework, one has to study the asymptotic behavior of the supermum and the infimum of the finite sample risk of the averaging estimator, where the supermum and the infimum are taken over a class of DGPs which include both the locally misspecified and many more severely misspecified DGPs. See Section

in Andrews and Guggenberger (2010) and Andrews, Cheng, and Guggenberger (2011) for uniform size control for inference in non-standard problems, but the present paper is for estimation rather than inference and focuses on a misspecification issue that is not studied in these papers.

The uniform dominance property of the averaging estimator does not contradict the risk properties of the post-model-selection estimators found in Yang (2005) and Leeb and Pötscher (2008). Measured by the MSE, the post-model-selection estimator usually does better than the unrestricted estimator in part of the parameter space and worse than the latter in other part of the parameter space. One standard example is the Hodge's estimator, whose scaled maximal MSE diverges to infinity with the growth of the sample size (see, e.g., Lehmann and Casella, 1998). Similar unbounded risk results are established in Yang (2005) and Leeb and Pötscher (2008) for post-model-selection estimator based on consistent model selection procedures. Such estimators have unbounded (scaled) maximal MSE because given the consistent model selection procedure: (i) there exist DGPs where the restrictions to be tested/selected are (locally) misspecified; (ii) the model selection procedures select these misspecified restrictions with high probabilities, converging to 1 asymptotically; (iii) the restricted estimator has unbounded (scaled) MSE under these DGPs.[13] In contrast, the empirical optimal weight of our averaging estimator is based on an infeasible optimal weight that satisfies: (i) when the aggressive/restricted GMM estimator has unbounded (scaled) MSE, the averaging weight on it is small, converging to 0 asymptotically. The resulting averaging estimator has the same asymptotic properties as the conservative GMM estimator; (ii) the Stein's dominance result applies in the asymptotic sense. Hence our averaging estimator is essentially different from the post-model-selection estimator.

There is a large literature studying the validity of GMM moment conditions. Many methods can be applied to detect the validity, including the over-identification tests (see, e.g., Sargan, 1958; Hansen, 1982; and Eichenbaum, Hansen and Singleton 1988), the information criteria (see, e.g., Andrews, 1999; Andrews and Lu, 2003; and Hong, Preston and Shum, 2003), and the penalized estimation methods (see, e.g., Liao, 2013; Cheng and Liao, 2014). Recently, misspecified moments and their consequences are considered by Ashley (2009), Berkowitz, Caner, and Fang (2012), Conley, Hansen, and Rossi (2012), Doko Tchatoka and Dufour (2008, 2014), Guggenberger (2012), Nevo and Rosen (2012), Kolesar, Chetty, Friedman, Glaeser, Imbens (2014), Small (2007), Small, Cai, Zhang, Kang (2015), among others. Moon and Schorfheide (2009) explore over-identifying moment

3 for more details.

[13]The post-model-selection estimator based on a conservative model selection procedure (e.g., hypothesis test with fixed critical value or Akaike information criterion) typically do not have unbounded (scaled) maximal MSE. However its asymptotic maximal MSE is not guaranteed to be less than or equal to the benchmark estimator (e.g., the conservative GMM estimator in the framework of this paper). The pre-test estimators in Figure 1 are good examples, since they are based on the J-test with nominal size 0.01 and 0.05.

inequalities to reduce the MSE. This paper contributes to this literature by providing new uniform results for potentially misspecified semiparametric models.

There is also a large literature studying adaptive estimation in nonparametric regression model using model averaging; see Yang (2000, 2003, 2004), Leung and Barron (2006), and the references therein. Since the unknown function can be written as a linear combination of (possibly infinitely but countably many) basis functions, the nonparametric model may be well approximated by parametric regression models in finite samples. These papers show that the averaging estimators which combine OLS estimators from different parametric models with data dependent weights may achieve the optimal convergence rate up to some logarithm factor. Our paper is different from these papers since the parameter of interest in our paper is a finite dimensional real value, not an unknown function, and the bias and variance trade-off of our averaging estimator is due to the possibly misspecified moment conditions. Moreover, there is a benchmark estimator in our paper, i.e., the conservative GMM estimator whose asymptotic properties are well-known. Our goal is to propose an averaging estimator with uniformly smaller risk than the conservative estimator.

3 Parameter Space and Uniform Dominance

Let $g_{2,j}(w,\theta)$ $(j=1,\ldots,r_2)$ denote the j-th component function of $g_2(w,\theta)$. We assume that $g_{2,j}(w,\theta)$ for $j=1,\ldots,r_2$ is twice continuously differentiable with respect to θ for any $w\in\mathcal{W}$. The first and second order derivatives of $g_2(w,\theta)$ with respect to θ are denoted by

$$g_{2,\theta}(w,\theta)\equiv\begin{pmatrix}\frac{\partial g_{2,1}(w,\theta)}{\partial\theta'}\\ \vdots\\ \frac{\partial g_{2,r_2}(w,\theta)}{\partial\theta'}\end{pmatrix}\text{ and } g_{2,\theta\theta}(w,\theta)\equiv\begin{pmatrix}\frac{\partial^2 g_{2,1}(w,\theta)}{\partial\theta\partial\theta'}\\ \vdots\\ \frac{\partial^2 g_{2,r_2}(w,\theta)}{\partial\theta\partial\theta'}\end{pmatrix},\tag{3.1}$$

respectively.[14] Let $\mathcal{F}$ be a set of distribution functions on $\mathcal{W}$. For $k=1$ and 2, define the expectation of the moment functions, the Jacobian matrix and the variance-covariance matrix as

$$\begin{aligned}M_{k,F}&\equiv\mathbb{E}_F\left[g_k(W,\theta_F)\right],\\ G_{k,F}&\equiv\mathbb{E}_F\left[g_{k,\theta}(W,\theta_F)\right],\text{ and}\\ \Omega_{k,F}&\equiv\mathbb{E}_F\left[g_k(W,\theta_F)g_k(W,\theta_F)'\right]-M_{k,F}M_{k,F}'\end{aligned}\tag{3.2}$$

for any $F\in\mathcal{F}$ respectively. The moments above exist by Assumption 3.2 below.

[14] By definition, $g_{1,\theta}(w,\theta)$ and $g_{1,\theta\theta}(w,\theta)$ are the leading $r_1\times d_\theta$ and $(r_1 d_\theta)\times d_\theta$ submatrices of $g_{2,\theta}(w,\theta)$ and $g_{2,\theta\theta}(w,\theta)$, respectively.

Let

$$Q_F(\theta) \equiv \mathbb{E}_F[g_2(W,\theta)]'\Omega_{2,F}^{-1}\mathbb{E}_F[g_2(W,\theta)] \tag{3.3}$$

for any $\theta \in \Theta$, which denotes the population criterion of the GMM estimation in (1.4). For any $\theta \in \Theta$, define $B_\varepsilon^c(\theta) = \{\theta^* \in \Theta : ||\theta^* - \theta|| \geq \varepsilon\}$. We consider the risk difference between two estimators uniformly over $F \in \mathcal{F}$, where $\mathcal{F}$ satisfies Assumptions 3.1-3.3 below.

Assumption 3.1 *The following conditions hold:*
(i) *for any* $F \in \mathcal{F}$, $\mathbb{E}_F[g_1(W,\theta_F)] = 0_{r_1\times 1}$ *for some* $\theta_F \in int(\Theta)$*;*
(ii) *for any* $\varepsilon > 0$, $\inf_{F\in\mathcal{F}} \inf_{\theta\in B_\varepsilon^c(\theta_F)} ||\mathbb{E}_F[g_1(W,\theta)]|| > 0$*;*
(iii) *for any* $F \in \mathcal{F}$, *there is* $\theta_F^* \in int(\Theta)$ *such that*

$$\inf_{F\in\mathcal{F}} \inf_{\theta\in B_\varepsilon^c(\theta_F^*)} [Q_F(\theta) - Q_F(\theta_F^*)] > 0 \text{ for any } \varepsilon > 0;$$

(iv) $\inf_{\{F\in\mathcal{F}:\ \|\delta_F\|>0\}} \frac{||G_{2,F}'\Omega_{2,F}^{-1}\delta_{2,F}||}{\|\delta_{2,F}\|^\tau} > 0$ *where* $\delta_{2,F}' = (0_{1\times r_1}, \delta_F')$ *and* $\tau > 0$ *is a fixed constant;*
(v) $0_{r^*\times 1} \in int(\Delta_\delta)$ *where* $\Delta_\delta = \{\delta_F : F \in \mathcal{F}\}$.

Assumptions 3.1.(i)-(ii) require that the true unknown parameter θ_F is uniquely identified by the moment conditions $\mathbb{E}_F[g_1(W,\theta_F)] = 0_{r_1\times 1}$ for any DGP $F \in \mathcal{F}$. Assumption 3.1.(iii) implies that for any $F \in \mathcal{F}$, a pseudo true value θ_F^* is identified by the unique minimizer of the population GMM criterion $Q_F(\theta)$ under possible misspecification. Assumption 3.1.(iv) requires that $\delta_{2,F}$ is not orthogonal to $\Omega_{2,F}^{-1}G_{2,F}$, which rules out the special case that θ_F may be consistently estimable even with severely misspecified moment conditions. Assumption 3.1.(v) implies that the set of distribution functions $\mathcal{F}$ is rich such that it includes the distributions under which the extra moment conditions are correctly specified, locally misspecified or severely misspecified. Uniform dominance can be easily established if we only allow for correctly specified models or severely misspecified models, because the desired dominance results hold trivially following a pointwise analysis. Assumption 3.1.(v) ensures that the extra moment conditions may have different degrees of misspecification in the parameter space.

Assumption 3.2 *The following conditions hold:*
(i) *For* $j = 1, \ldots, r_2$, $g_{2,j}(w,\theta)$ *is twice continuously differentiable with respect to* θ *for any* $w \in \mathcal{W}$*;*
(ii) $\sup_{F\in\mathcal{F}} \mathbb{E}_F\left[\sup_{\theta\in\Theta}\left(||g_2(W,\theta)||^{2+\gamma} + ||g_{2,\theta}(W,\theta)||^{2+\gamma} + ||g_{2,\theta\theta}(W,\theta)||^{2+\gamma}\right)\right] < \infty$ *for some* $\gamma > 0$*;*
(iii) $\inf_{F\in\mathcal{F}} \rho_{\min}(\Omega_{2,F}) > 0$*;*
(iv) $\inf_{F\in\mathcal{F}} \rho_{\min}(G_{1,F}'G_{1,F}) > 0$.

Assumption 3.2.(i) requires that the moment functions are smooth. Assumption 3.2.(ii) imposes $2+\gamma$ finite moment conditions on the GMM moment functions and their first and second derivatives. Assumptions 3.2.(iii) and 3.2.(iv) are important sufficient conditions for the local identification of the unknown parameter in GMM with valid moment conditions.

The next assumption is on the nuisance parameters of the DGP $F \in \mathcal{F}$. Write

$$v_F = \left(\text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})', \delta_F'\right) \tag{3.4}$$

for any $F \in \mathcal{F}$. It is clear that v_F includes the Jacobian matrix, the variance-covariance matrix, and the measure of misspecification of the moment conditions $\mathbb{E}_F[g^*(W, \theta_F)] = 0_{r^* \times 1}$. Let

$$\overline{v}_F = \left(\text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})'\right) \tag{3.5}$$

for any $F \in \mathcal{F}$.

Assumption 3.3 *The following conditions hold:*

(i) *For any $F \in \mathcal{F}$ with $\delta_F = 0_{r^* \times 1}$, there exists a constant $\varepsilon_F > 0$ such that for any $\widetilde{\delta} \in \mathbb{R}^{r^*}$ with $0 \leq ||\widetilde{\delta}|| < \varepsilon_F$, there is $\widetilde{F} \in \mathcal{F}$ with $\delta_{\widetilde{F}} = \widetilde{\delta}$ and $\left\|\overline{v}_{\widetilde{F}} - \overline{v}_F\right\| \leq C||\widetilde{\delta}||^{\kappa}$ for some $\kappa > 0$;*

(ii) *The set $\Lambda \equiv \{v_F \colon F \in \mathcal{F}\}$ is closed.*

Assumption 3.3.(i) requires that for any $F \in \mathcal{F}$ such that $\mathbb{E}_F[g_2(W, \theta_F)] = 0_{r^* \times 1}$ is valid, there are many DGPs $\widetilde{F} \in \mathcal{F}$ which are close to F. Here the closeness of any two DGPs F and $\widetilde{F}$ is measured by the distance between v_F and $v_{\widetilde{F}}$. Assumption 3.3 (i) and (ii) are useful to derive the exact expression of the asymptotic risk of the GMM estimator.

Example 3.1 (Linear IV Model) We study a linear IV model and provide a set of low-level conditions that imply Assumptions 3.1, 3.2 and 3.3. The parameters of interest θ_0 are the coefficients of the endogenous regressors X in

$$Y = X'\theta_0 + U, \tag{3.6}$$

with some valid IVs $Z_1 \in \mathbb{R}^{r_1}$ and some potentially misspecified IVs $Z^* \in \mathbb{R}^{r^*}$ such that

$$\mathbb{E}_{F^*}[U] = 0, \ \mathbb{E}_{F^*}[Z_1 U] = 0_{r_1 \times 1}, \text{ and} \tag{3.7}$$

$$Z^* = U\delta_0 + V, \text{ with } \mathbb{E}_{F^*}[V] = 0_{r^* \times 1} \text{ and } \mathbb{E}_{F^*}[VU] = 0_{r^* \times 1}, \tag{3.8}$$

where F^* denotes the joint distribution of $(X', Z_1', V', U)'$. In the reduced-form equation (3.8), δ_0 is a $r^* \times 1$ real vector which characterizes the degree of misspecification. Let $\mathcal{F}^*$ denote a class of distributions containing F^*, and let Θ and Δ_δ denote the parameter spaces of θ_0 and δ_0 respectively. The joint distribution of $W = (Y, Z_1', Z^{*\prime}, X')'$ is denoted as F which is determined by θ_0, δ_0 and F^* through the linear equations in (3.6) and (3.8).

For ease of discussion, we further assume that the random vector $(X', Z_1', V', U)'$ follows the normal distribution with mean ϕ and variance-covariance matrix Ψ. Under the normality assumption, each distribution F^* corresponds to a pair of ϕ and Ψ.

For notational simplicity, in Lemma 3.1 below, for any finite dimensional random vectors a_1 and a_2, let $\phi_{a_j} = \mathbb{E}_{F^*}[a_j]$ for $j = 1, 2$, $\Gamma_{a_1 a_2} = \mathbb{E}_{F^*}[a_1 a_2']$, and $\Omega_{a_1 a_2} = \mathbb{E}_{F^*}[a_1 a_2'] - \phi_{a_1}\phi_{a_2}'$.

Lemma 3.1 *Let $\mathcal{F}^*$ denote the set of normal distributions which satisfies:*
(i) $\phi_u = 0$, $\Gamma_{z_1 u} = 0_{r_1 \times 1}$ *and* $\Gamma_{vu} = 0_{r^* \times 1}$*;*
(ii) $\inf_{F^* \in \mathcal{F}^*} \rho_{\min}(\Gamma_{x z_1}\Gamma_{z_1 x}) > 0$, $\sup_{F^* \in \mathcal{F}^*} ||\phi||^2 < \infty$ *and*
$0 < \inf_{F^* \in \mathcal{F}^*} \rho_{\min}(\Psi) \leq \sup_{F^* \in \mathcal{F}^*} \rho_{\max}(\Psi) < \infty$*;*
(iii) $\inf_{F^* \in \mathcal{F}^*} \inf_{\{\|\delta\| \geq \varepsilon\}} \|\delta\|^{-1} ||(\Gamma_{x z_1}\Gamma_{z_1 z_1}^{-1}\Gamma_{z_1 v} - \Gamma_{xv})\delta - \Gamma_{xu}|| > 0$ *for some* $\varepsilon > 0$ *that is small enough;*[15]
(iv) $\theta_0 \in int(\Theta)$ *and* Θ *is compact and large enough such that the pseudo-true value* $\theta_F^* \in int(\Theta)$*;*[16]
(v) $\Delta_\delta = [c_{1,\Delta}, C_{1,\Delta}] \times \cdots \times [c_{r^*,\Delta}, C_{r^*,\Delta}]$ *where* $\{c_{j,\Delta}, C_{j,\Delta}\}_{j=1}^{r^*}$ *is a set of finite constants with* $c_{j,\Delta} < 0 < C_{j,\Delta}$ *for* $j = 1, \ldots, r^*$,
then Assumptions 3.1, 3.2 and 3.3 hold.

Condition (i) lists the moment conditions in (3.7) and (3.8). The inequality in Condition (ii) rules out DGPs under which $\rho_{\min}(\Gamma_{x z_1}\Gamma_{z_1 x})$ may be close to zero and (part of) the unknown parameter θ_0 is weakly identified. Condition (ii) also requires that the mean of the random vector $(X', Z_1', V', U)'$ is uniformly bounded and the eigenvalues of its variance-covariance matrix are uniformly finite and bounded away from 0. Condition (iii) requires that the projection residual of the vector Γ_{xu} on the subspace spanned by the matrix $\Gamma_{x z_1}\Gamma_{z_1 z_1}^{-1}\Gamma_{z_1 v} - \Gamma_{xv}$ is bounded away from zero. It is a sufficient condition for Assumption 3.1.(iv), which ensures that the aggressive estimator is inconsistent under severe misspecification. Condition (iv) is needed to derive the limit of the aggressive estimator under misspecification. The compactness assumption of Θ is not needed for the linear IV model. However, it is useful to verify Assumptions 3.1, 3.2 and 3.3 which do not assume any special structure on the model. Condition (v) specifies that the parameter space of δ_0 is a product space.

[15]The constant ε depends on the infimum and supremum in Condition (ii) and it is given in (B.3) in the Appendix.
[16]Specific restritions on Θ which ensures that $\theta_F^* \in int(\Theta)$ are given in (D.8) and Assumption D.1.(vi) in the Supplemental Appendix.

Lemma 3.1 provides simple conditions on θ_0, δ_0 and $\mathcal{F}^*$ on which uniformity results are subsequently established.[17] ∎

Now we get back to the general set up. For a generic estimator $\widehat{\theta}$ of θ, consider a weighted quadratic loss function

$$\ell(\widehat{\theta},\theta) = n(\widehat{\theta}-\theta)'\Upsilon(\widehat{\theta}-\theta), \tag{3.9}$$

where Υ is a $d_\theta \times d_\theta$ pre-determined positive semi-definite matrix. For example, if $\Upsilon = I_{d_\theta}$, $\mathbb{E}_F[\ell(\widehat{\theta},\theta_F)]$ is the MSE of $\widehat{\theta}$. If $\Upsilon = (\Sigma_{1,F}-\Sigma_{2,F})^{-1}$ where $\Sigma_{k,F}$ $(k=1,2)$ is defined in (4.4), the weighting matrix Υ rescales $\widehat{\theta}$ by the scale of variance reduction due to the additional moments. If $\Upsilon = \mathbb{E}_F[X_iX_i']$ for regressors X_i, $\mathbb{E}_F[\ell(\widehat{\theta},\theta_F)]$ is the MSE of $X_i'\widehat{\theta}$, an estimator of $X_i'\theta$.

Below we compare the averaging estimator $\widehat{\theta}_{eo}$ and the conservative estimator $\widehat{\theta}_1$. We are interested in the bounds of the truncated finite sample risk difference

$$\begin{aligned}\underline{RD}_n(\widehat{\theta}_{eo},\widehat{\theta}_1;\zeta) &\equiv \inf_{F\in\mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}_{eo},\theta_F)-\ell_\zeta(\widehat{\theta}_1,\theta_F)] \text{ and}\\ \overline{RD}_n(\widehat{\theta}_{eo},\widehat{\theta}_1;\zeta) &\equiv \sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}_{eo},\theta_F)-\ell_\zeta(\widehat{\theta}_1,\theta_F)],\end{aligned} \tag{3.10}$$

where

$$\ell_\zeta(\widehat{\theta},\theta_F) \equiv \min\{\ell(\widehat{\theta},\theta_F),\zeta\} \tag{3.11}$$

denotes the truncated loss function with an arbitrary trimming parameter ζ. The truncated loss function is employed to facilitate the asymptotic analysis of the bounds of the risk difference. The finite-sample bounds in (3.10) are approximated by

$$\begin{aligned}Asy\underline{RD}(\widehat{\theta}_{eo},\widehat{\theta}_1) &\equiv \liminf_{\zeta\to\infty}\ \liminf_{n\to\infty}\ \underline{RD}_n(\widehat{\theta}_{eo},\widehat{\theta}_1;\zeta) \text{ and}\\ Asy\overline{RD}(\widehat{\theta}_{eo},\widehat{\theta}_1) &\equiv \limsup_{\zeta\to\infty}\ \limsup_{n\to\infty}\overline{RD}_n(\widehat{\theta}_{eo},\widehat{\theta}_1;\zeta),\end{aligned} \tag{3.12}$$

which are called lower and upper bounds of the asymptotic risk difference respectively in this paper. The averaging estimator $\widehat{\theta}_{eo}$ asymptotically uniformly dominates the conservative estimator $\widehat{\theta}_1$ if

$$Asy\underline{RD}(\widehat{\theta}_{eo},\widehat{\theta}_1) < 0 \text{ and } Asy\overline{RD}(\widehat{\theta}_{eo},\widehat{\theta}_1) \le 0. \tag{3.13}$$

The bounds of the asymptotic risk difference build the uniformity over $F \in \mathcal{F}$ into the definition by taking $\inf_{F\in\mathcal{F}}$ and $\sup_{F\in\mathcal{F}}$ before $\liminf_{n\to\infty}$ and $\limsup_{n\to\infty}$ respectively. Uniformity is crucial

[17] Similar results have been established in Section D of the Supplemental Appendix for the linear IV model when the normality assumption on $(X', Z_1', V', U)'$ is relaxed. Section D of the Supplemental Appendix also provides proof for Lemma 3.1 with and without the normality assumption.

for the asymptotic results to give a good approximation to their finite-sample counterparts. These uniform bounds are different from pointwise results which are obtained under a fixed DGP. The sequence of DGPs $\{F_n\}$ along which the supremum or the infimum are approached often varies with the sample size.[18] Therefore, to determine the bounds of the asymptotic risk difference, one has to derive the asymptotic distributions of these estimators under various sequences $\{F_n\}$. Under $\{F_n\}$, the observations $\{W_{n,i}\}_{i=1}^n$ form a triangular array. For notational simplicity, $W_{n,i}$ is abbreviated to W_i throughout the paper.

To study the bounds of asymptotic risk difference, we consider sequences of DGPs $\{F_n\}$ such that δ_{F_n} satisfies[19]

$$\text{(i) } n^{1/2}\delta_{F_n} \to d \in \mathbb{R}^{r^*} \text{ or (ii) } ||n^{1/2}\delta_{F_n}|| \to \infty. \tag{3.14}$$

Case (i) models mild misspecification, where δ_{F_n} is a $n^{-1/2}$-local deviation from $0_{r^*\times 1}$. Case (ii) includes the severe misspecification where $\|\delta_{F_n}\|$ is bounded away from 0 as well as the intermediate case in which $\delta_{F_n} \to 0$ and $||n^{1/2}\delta_{F_n}|| \to \infty$. To obtain a uniform approximation, all of these sequences are necessary. Once we study the bounds of asymptotic risk difference along each of these sequences, we show that we can glue them together to obtain the bounds of asymptotic risk difference.

4 Averaging Weight

We start by deriving the joint asymptotic distribution of $\widehat{\theta}_1$ and $\widehat{\theta}_2$ under different degrees of misspecification. We consider sequences of DGPs $\{F_n\}$ in $\mathcal{F}$ such that (i) $n^{1/2}\delta_{F_n} \to d \in \mathbb{R}^{r^*}$ or $||n^{1/2}\delta_{F_n}|| \to \infty$; and (ii) G_{2,F_n}, Ω_{2,F_n} and M_{2,F_n} converges to $G_{2,F}$, $\Omega_{2,F}$ and $M_{2,F}$ for some $F \in \mathcal{F}$. [20]

For $k = 1, 2$ and any $F \in \mathcal{F}$, define

$$\Gamma_{k,F} = -\left(G'_{k,F}\Omega_{k,F}^{-1}G_{k,F}\right)^{-1} G'_{k,F}\Omega_{k,F}^{-1}. \tag{4.1}$$

Let $\mathcal{Z}_{2,F}$ denote a zero mean normal random vector with variance-covariance matrix $\Omega_{2,F}$ and $\mathcal{Z}_{1,F}$ denote its first r_1 components.

[18] In the rest of the paper, we use $\{F_n\}$ to denote $\{F_n \in \mathcal{F} : n = 1, 2, ...\}$.

[19] Since $F_n \in \mathcal{F}$, by Assumption 3.2.(ii), the sequence δ_{F_n} in (3.14) should satisfy $\|\delta_{F_n}\| \leq C$ for any n.

[20] The requirement on the convergence of G_{2,F_n}, Ω_{2,F_n} and M_{2,F_n} is not restrictive. Lemma B.7 in Appendix B.1 shows that the sequences G_{2,F_n}, Ω_{2,F_n} and M_{2,F_n} have subsequences that converge to $G_{2,F}$, $\Omega_{2,F}$ and $M_{2,F}$, respectively, for some $F \in \mathcal{F}$. The general result on the lower and upper bounds of the asymptotic risk difference, Lemma B.14 in Appendix B.2, only requires to consider the subsequence $\{F_{p_n}\}$ such that $G_{2,F_{p_n}}$, $\Omega_{2,F_{p_n}}$ and $M_{2,F_{p_n}}$ are convergent, where $\{p_n\}$ is a subsequence of $\{n\}$. The asymptotic properties of the GMM estimators established in this section under the full sequence of DGPs $\{F_n\}$ holds trivially for its subsequence.

Lemma 4.1 *Suppose Assumptions 3.1 and 3.2 hold. Consider any sequence of DGPs $\{F_n\}$ such that $v_{F_n} \to v_F = (\text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})', \delta_F')$ for some $F \in \mathcal{F}$, and $n^{1/2}\delta_{F_n} \to d$ for $d \in \mathbb{R}^{r^*}_{\infty}$.*

(a) *If $d \in \mathbb{R}^{r^*}$, then*

$$\begin{pmatrix} n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \\ n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) \end{pmatrix} \to_D \begin{pmatrix} \xi_{1,F} \\ \xi_{2,F} \end{pmatrix} \equiv \begin{pmatrix} \Gamma_{1,F}\mathcal{Z}_{1,F} \\ \Gamma_{2,F}\left(\mathcal{Z}_{2,F} + d_0\right) \end{pmatrix},$$

where $d_0 = (\mathbf{0}_{1\times r_1}, d')'$.

(b) *If $||d|| = \infty$, then $n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \to_D \xi_{1,F}$ and $||n^{1/2}(\widehat{\theta}_2 - \theta_{F_n})|| \to_p \infty$.*

Given the joint asymptotic distribution of $\widehat{\theta}_1$ and $\widehat{\theta}_2$, it is straightforward to study $\widehat{\theta}(\omega) = (1-\omega)\widehat{\theta}_1 + \omega\widehat{\theta}_2$ if ω is deterministic. Following Lemma 4.1.(a),

$$n^{1/2}(\widehat{\theta}(\omega) - \theta_{F_n}) \to_D \xi_F(\omega) \equiv (1-\omega)\xi_{1,F} + \omega\xi_{2,F} \tag{4.2}$$

for $n^{1/2}\delta_{F_n} \to d$, where $d \in \mathbb{R}^{r^*}$. In Section E of the Supplemental Appendix, a simple calculation shows that the asymptotic risk of $\widehat{\theta}(\omega)$ is minimized at the infeasible optimal weight

$$\omega_F^* \equiv \frac{\text{tr}(\Upsilon\left(\Sigma_{1,F} - \Sigma_{2,F}\right))}{d_0'\left(\Gamma_{2,F} - \Gamma_{1,F}^*\right)'\Upsilon\left(\Gamma_{2,F} - \Gamma_{1,F}^*\right)d_0 + \text{tr}(\Upsilon\left(\Sigma_{1,F} - \Sigma_{2,F}\right))}, \tag{4.3}$$

where Υ is the matrix specified in the loss function,

$$\Sigma_{k,F} \equiv \left(G_{k,F}'\Omega_{k,F}^{-1}G_{k,F}\right)^{-1} \text{ for } k = 1,2 \text{ and } \Gamma_{1,F}^* \equiv \left[\Gamma_{1,F}, \mathbf{0}_{d_\theta\times r^*}\right]. \tag{4.4}$$

To gain some intuition, consider the case where $\Upsilon = I_{d_\theta}$ such that the MSE of $\widehat{\theta}(\omega)$ is minimized at ω_F^*. In this case, the infeasible optimal weight ω_F^* yields the ideal bias and variance trade off. However, the bias depends on d, which cannot be consistently estimated. Hence, ω_F^* cannot be consistently estimated. Our solution to this problem follows the popular approach in the literature which replaces d by an estimator whose asymptotic distribution is centered at d, see Liu (2015) and Charkhi, Claeskens, and Hansen (2016) for similar estimators in the least square estimation and maximum likelihood estimation problems, respectively.

The empirical analog of ω_F^* is constructed as follows. First, for $k = 1$ and 2, replace $\Sigma_{k,F}$ by its consistent estimator $\widehat{\Sigma}_k \equiv (\widehat{G}_k'\widehat{\Omega}_k^{-1}\widehat{G}_k)^{-1}$,[21] where

$$\widehat{G}_k \equiv n^{-1}\sum_{i=1}^{n} g_{k,\theta}(W_i, \widehat{\theta}_1) \text{ and } \widehat{\Omega}_k \equiv n^{-1}\sum_{i=1}^{n} g_k(W_i, \widehat{\theta}_1)g_k(W_i, \widehat{\theta}_1)' - \overline{g}_k(\widehat{\theta}_1)\overline{g}_k(\widehat{\theta}_1)'. \tag{4.5}$$

[21]The consistency of $\widehat{\Sigma}_k$ is proved in the proof of Lemma 4.2.

Note that $\widehat{G}_k$ and $\widehat{\Omega}_k$ are based on the conservative GMM estimator $\widehat{\theta}_1$. Hence they are consistent regardless of the degree of misspecification of the moment conditions in (1.3). Second, replace $(\Gamma_{2,F} - \Gamma^*_{1,F})d_0$ by its asymptotically unbiased estimator $n^{1/2}(\widehat{\theta}_2 - \widehat{\theta}_1)$ because

$$n^{1/2}(\widehat{\theta}_2 - \widehat{\theta}_1) \to_D (\Gamma_{2,F} - \Gamma^*_{1,F})(\mathcal{Z}_{2,F} + d_0), \tag{4.6}$$

for $d_0 = (\mathbf{0}_{1\times r_1}, d')'$ and $d \in \mathbb{R}^{r^*}$ following Lemma 4.1(a). Then the empirical optimal weight takes the form

$$\widetilde{\omega}_{eo} \equiv \frac{\mathrm{tr}(\Upsilon(\widehat{\Sigma}_1 - \widehat{\Sigma}_2))}{n(\widehat{\theta}_2 - \widehat{\theta}_1)'\Upsilon(\widehat{\theta}_2 - \widehat{\theta}_1) + \mathrm{tr}(\Upsilon(\widehat{\Sigma}_1 - \widehat{\Sigma}_2))}, \tag{4.7}$$

and the averaging GMM estimator takes the form

$$\widehat{\theta}_{eo} = (1 - \widetilde{\omega}_{eo})\widehat{\theta}_1 + \widetilde{\omega}_{eo}\widehat{\theta}_2. \tag{4.8}$$

Next we consider the asymptotic distribution of $\widehat{\theta}_{eo}$ under different degrees of misspecification.

Lemma 4.2 *Suppose that Assumptions 3.1-3.3 hold. Consider any sequence of DGPs* $\{F_n\}$ *such that* $v_{F_n} \to v_F = (\mathrm{vec}(G_{2,F})', \mathrm{vech}(\Omega_{2,F})', \delta_F')$ *for some* $F \in \mathcal{F}$*, and* $n^{1/2}\delta_{F_n} \to d$ *for* $d \in \mathbb{R}^{r^*}_{\infty}$.
(a) *If* $d \in \mathbb{R}^{r^*}$*, then*

$$\widetilde{\omega}_{eo} \to_D \overline{\omega}_F \equiv \frac{\mathrm{tr}(\Upsilon(\Sigma_{1,F} - \Sigma_{2,F}))}{(\mathcal{Z}_{2,F} + d_0)'(\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon(\Gamma_{2,F} - \Gamma^*_{1,F})(\mathcal{Z}_{2,F} + d_0) + \mathrm{tr}(\Upsilon(\Sigma_{1,F} - \Sigma_{2,F}))}$$

and

$$n^{1/2}(\widehat{\theta}_{eo} - \theta_{F_n}) \to_D \overline{\xi}_F \equiv (1 - \overline{\omega}_F)\xi_{1,F} + \overline{\omega}_F \xi_{2,F}.$$

(b) *If* $||d|| = \infty$*, then* $\widetilde{\omega}_{eo} \to_p 0$ *and* $n^{1/2}(\widehat{\theta}_{eo} - \theta_{F_n}) \to_D \xi_{1,F}$.

To study the bounds of the asymptotic risk difference between $\widehat{\theta}_{eo}$ and $\widehat{\theta}_1$, it is important to take into account the data-dependent nature of $\widetilde{\omega}_{eo}$. Unlike $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$, the randomness in $\widetilde{\omega}_{eo}$ is non-negligible in the mild misspecification case (a) of Lemma 4.2. In consequence, $\widehat{\theta}_{eo}$ does not achieve the same bounds of asymptotic risk difference as the ideal averaging estimator $(1 - \omega^*_F)\widehat{\theta}_1 + \omega^*_F\widehat{\theta}_2$ does. Nevertheless, below we show that $\widehat{\theta}_{eo}$ is insured against potentially misspecified moments because it uniformly dominates $\widehat{\theta}_1$.

5 Bounds of Asymptotic Risk Difference under Misspecification

In this section, we study the bounds of the asymptotic risk difference defined in (3.12). Note that the asymptotic distributions of $\widehat{\theta}_1$ and $\widehat{\theta}_{eo}$ in Lemma 4.1 and 4.2 only depend on d, $G_{2,F}$ and

$\Omega_{2,F}$. For notational convenience, define

$$h_{F,d} = (d', \text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})') \tag{5.1}$$

for any $F \in \mathcal{F}$ and any $d \in \mathbb{R}_{\infty}^{r^*}$. For the mild misspecification case, define the parameter space of $h_{F,d}$ as

$$H = \{h_{F,d} : d \in \mathbb{R}^{r^*} \text{ and } F \in \mathcal{F} \text{ with } \delta_F = 0_{r^* \times 1}\} \tag{5.2}$$

where δ_F is defined by (1.5) for a given F.

Theorem 5.1 *Suppose that Assumptions 3.1-3.3 hold. The bounds of the asymptotic risk difference satisfy*

$$\begin{aligned} Asy\overline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_1) &= \max\left\{\sup_{h \in H}[g(h)], 0\right\}, \\ Asy\underline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_1) &= \min\left\{\inf_{h \in H}[g(h)], 0\right\}, \end{aligned}$$

where $g(h) \equiv \mathbb{E}[\overline{\xi}_F' \Upsilon \overline{\xi}_F - \xi_{1,F}' \Upsilon \xi_{1,F}]$, $\xi_{1,F}$ *and* $\overline{\xi}_F$ *are given in Lemma 4.1 and Lemma 4.2 respectively, and the expectation is taken under the joint normal distribution with mean zero and variance-covariance matrix* $\Omega_{2,F}$.

The upper (or lower) bound of the asymptotic risk difference is determined by the maximum between $\sup_{h \in H}[g(h)]$ and zero (or the minimum between $\inf_{h \in H}[g(h)]$ and zero), where $\sup_{h \in H}[g(h)]$ (or $\inf_{h \in H}[g(h)]$) is related to the mildly misspecified DGPs and the zero component is associated with the severely misspecified DGPs. Since the GMM averaging estimator has the same asymptotic distribution as the conservative GMM estimator $\widehat{\theta}_1$ under the severely misspecified DGPs, their asymptotic risk difference is zero.

To show that $\widehat{\theta}_{eo}$ uniformly dominates $\widehat{\theta}_1$ following (3.13), Theorem 5.1 implies that it is sufficient to show that $\inf_{h \in H}[g(h)] < 0$ and $\sup_{h \in H}[g(h)] \leq 0$. We can investigate $\inf_{h \in H} g(h)$ and $\sup_{h \in H} g(h)$ by simulating $g(h)$. In practice, we replace $G_{2,F}$ and $\Omega_{2,F}$ by their consistent estimators and plot $g(h)$ as a function of d. Even if the uniform dominance condition does not hold, $\min\{\inf_{h \in H}[g(h)], 0\}$ and $\max\{\sup_{h \in H}[g(h)], 0\}$ quantify the most- and least-favorable scenarios for the averaging estimator.

Theorem 5.2 *Let* $A_F \equiv \Upsilon(\Sigma_{1,F} - \Sigma_{2,F})$ *for any* $F \in \mathcal{F}$. *Suppose that Assumptions 3.1-3.3 hold.*

If $\mathrm{tr}(A_F) > 0$ *and* $\mathrm{tr}(A_F) \geq 4\rho_{\max}(A_F)$ *for any* $F \in \mathcal{F}$ *with* $\delta_F = 0$, *we have*

$$Asy\underline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_1) < 0 \text{ and } Asy\overline{RD}(\widehat{\theta}_{eo}, \widehat{\theta}_1) = 0.$$

Thus, $\widehat{\theta}_{eo}$ *uniformly dominates* $\widehat{\theta}_1$.

Theorem 5.2 indicates that: (i) there exists $\varepsilon_1 < 0$ and some finite integer n_{ε_1} such that the minimum risk difference between $\widehat{\theta}_{eo}$ and $\widehat{\theta}_1$ is less than ε_1 for any n larger than n_{ε_1}; (ii) for any $\varepsilon_2 > 0$, there exists a finite integer n_{ε_2} such that the maximum risk difference between $\widehat{\theta}_{eo}$ and $\widehat{\theta}_1$ is less than ε_2 for any n larger than n_{ε_2}. Pre-test estimators fail to satisfy both properties (i) and (ii) above at the same time. Take the pre-test estimator based on the J-test for example[22] and consider three scenarios: (a) the critical value is fixed for any sample size; (b) the critical value diverges to infinity; and (c) the critical value converges to zero. In the pointwise asymptotic framework, the J-test based on the critical values in (a), (b) and (c) leads to inconsistent (but conservative) model selection, consistent model selection and no model selection results respectively. The pre-test estimator based on the J-test violates property (ii) in scenarios (a) and (b), and violates property (i) in scenario (c).

Different from the finite-sample results for the JS estimator established for the Gaussian location model, our comparison of the two estimators $\widehat{\theta}_{eo}$ and $\widehat{\theta}_1$ is based on the asymptotic bounds of the risk difference. For a given sample size n, we do not provide results on this asymptotic approximation error, and therefore our results do not state how the finite-sample upper bound $\overline{RD}_n(\widehat{\theta}_{eo}, \widehat{\theta}_1; \zeta)$ approaches to zero as $n \to \infty$ and then $\zeta \to \infty$ (e.g., from above or from below). For the Gaussian location model, the asymptotically uniform dominance here is weaker than the classical finite-sample results established for the JS estimator. However, the asymptotic results here apply to general nonlinear econometric models with non-normal random variables.[23]

To shed light on the sufficient conditions in Theorem 5.2, let us consider a scenario similar to the JS estimator: $\Sigma_{1,F} = \sigma_{1,F}^2 I_{d_\theta}$, $\Sigma_{2,F} = \sigma_{2,F}^2 I_{d_\theta}$, and $\Upsilon = I_{d_\theta}$. In this case, the sufficient conditions become $\sigma_{1,F} > \sigma_{2,F}$ and $d_\theta \geq 4$. The first condition $\mathrm{tr}(A_F) > 0$, which is reduced to $\sigma_{1,F} > \sigma_{2,F}$, requires that the additional moments $\mathbb{E}_F[g^*(W_i, \theta_F)] = 0$ are non-redundant in the sense that they lead to a more efficient estimator of θ_F. The second condition $\mathrm{tr}(A_F) \geq 4\rho_{\max}(A_F)$, which is reduced to $d_\theta \geq 4$, requires that we are interested in the total risk of several parameters rather than that of a single one. In a more general case where $\Sigma_{1,F}$ and $\Sigma_{2,F}$ are not proportional to the identity matrix, the sufficient conditions are reduced to $\Sigma_{1,F} > \Sigma_{2,F}$ and $d_\theta \geq 4$ under the choice

[22] See Section F in the Supplemental Appendix for definition and analysis of this estimator.

[23] In Section A of the Appendix, we show that the averaging GMM estimator has similar finite sample dominace results in the Gaussian location model.

$\Upsilon = (\Sigma_{1,F} - \Sigma_{2,F})^{-1}$, which rescales $\widehat{\theta}$ by the variance reduction $\Sigma_{1,F} - \Sigma_{2,F}$. In a simple linear IV model (Example 3.1) where Z_i^* is independent of $Z_{1,i}$ and the regression error U_i is homoskedastic conditional on the IVs, $\Sigma_{1,F} > \Sigma_{2,F}$ requires that $\mathbb{E}_{F^*}[Z_i^* X_i']$ and $\mathbb{E}_{F^*}[Z_i^* Z_i^{*\prime}]$ both have full rank.

6 Simulation Studies

In this section, we investigate the finite sample performance of our averaging GMM estimator in linear IV models. In addition to the empirical optimal weight $\widetilde{\omega}_{eo}$, we consider another averaging estimator based on the JS type of weight. Define the positive part of the JS weight[24]:

$$\omega_{JS} = 1 - \left(1 - \frac{\operatorname{tr}(\widehat{A}) - 2\rho_{\max}(\widehat{A})}{n(\widehat{\theta}_2 - \widehat{\theta}_1)'\Upsilon(\widehat{\theta}_2 - \widehat{\theta}_1)}\right)_+ \tag{6.1}$$

where $(x)_+ = \max\{0, x\}$ and $\widehat{A}$ is the estimator of A_F using $\widehat{\Sigma}_k$ for $k = 1, 2$. In the simulation study of this paper, we consider an alternative averaging estimator with the restricted JS weight

$$\omega_{R,JS} = (\omega_{JS})_+ . \tag{6.2}$$

By construction, $\omega_{JS} \leq 1$ and $0 \leq \omega_{R,JS} \leq 1$. We compare the finite-sample (truncated and untruncated) MSEs of our proposed averaging estimator with the empirical optimal weight, the JS-type of averaging estimator with the restricted weight in (6.2), the conservative GMM estimator $\widehat{\theta}_1$, and the pre-test GMM estimator based on the J-test. The finite-sample MSE of the conservative GMM estimator is normalized to 1. That is, we report the ratios of various MSEs to the MSE of the conservative GMM estimator and call these ratios as relative MSEs. Three different simulation designs are considered in this section.

6.1 Simulation Model 1

We consider a linear regression model with i.i.d. observed data

$$W_i = (Y_i, X_{1,i}, \ldots, X_{6,i}, Z_{1,i}, \ldots, Z_{12,i}, Z_{1,i}^*, \ldots, Z_{6,i}^*)' \text{ for } i = 1, ..., n, \tag{6.3}$$

where Y is the dependent variable, $(X_1, ..., X_6)$ are 6 endogenous regressors, $(Z_1, ..., Z_{12})$ are 12 valid IVs, and $(Z_1^*, ..., Z_6^*)$ are 6 potentially invalid IVs. The data are generated as follows. The

[24]This formula is a GMM analog of the generalized JS type shrinkage estimator in Hansen (2016) for parametric models. The shrinkage scalar τ is set to $\operatorname{tr}(\widehat{A}) - 2\rho_{\max}(\operatorname{tr}(\widehat{A}))$ in a fashion similar to the original JS estimator.

regression model is

$$Y = \sum_{j=1}^{6} \theta_j X_j + u, \tag{6.4}$$

where $(\theta_1, \ldots, \theta_6)$ is set to $2.5 \times \mathbf{1}_{1\times 6}$ and X_j is generated by

$$X_j = 2^{-1}(Z_j + Z_{j+6}) + Z_{j+12} + \varepsilon_j \text{ for } j = 1, \ldots, 6. \tag{6.5}$$

We first draw $(Z_1, ..., Z_{18}, \varepsilon_1, \ldots, \varepsilon_6, u^*)'$ from normal distribution with mean zero and variance-covariance matrix $\text{diag}(I_{18\times 18}, \Sigma_{7\times 7})$ where

$$\Sigma_{7\times 7} = \begin{pmatrix} I_{6\times 6} & 0.25 \times \mathbf{1}_{6\times 1} \\ 0.25 \times \mathbf{1}_{1\times 6} & 1 \end{pmatrix}. \tag{6.6}$$

We consider two designs for generating the structural error u in (6.4). The first design (S1 hereafter) has non-Gaussian errors. Draw η from exponential distribution with mean 1 and η is independent of $(Z_1, \ldots, Z_{18}, \varepsilon_1, \ldots, \varepsilon_6, u^*)$. Generate the structural error

$$u = (u^* + \eta_0)/2, \tag{6.7}$$

where η_0 is the demeaned version of η to ensure that the mean of u is zero. The second design (S2 hereafter) has normal error

$$u = u^*. \tag{6.8}$$

The potentially invalid IVs are generated by

$$Z_j^* = (1 - c_j^2)^{1/2} Z_{j+12} + c_j(\varepsilon_j + u), \tag{6.9}$$

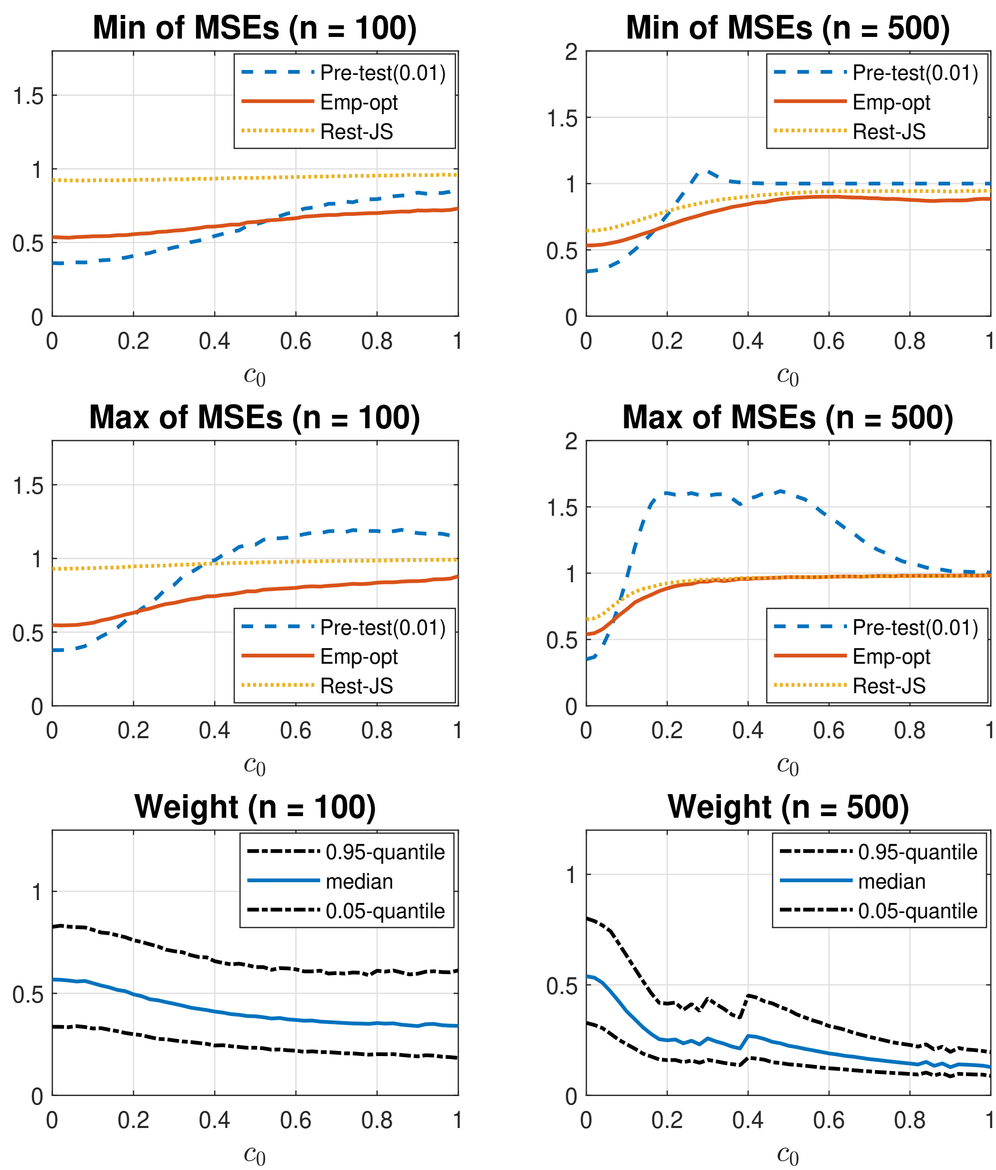
where $c_j \in [0, 1]$ for $j = 1, \ldots, 6$. In this simulation study, we consider different DGPs by choosing various values for $c = (c_1, ..., c_6)$ where $c_j \in [0, 1]$ for $j = 1, \ldots, 6$. Therefore,

$$\mathbb{E}\left[u Z_j^*\right] = \begin{cases} 5c_j/8 & \text{under (6.7)} \\ 5c_j/4 & \text{under (6.8)} \end{cases}. \tag{6.10}$$

From the above expression, we see that Z_j^* is a valid IV if c_j is zero while increasing c_j to 1 will enlarge the correlation coefficient between Z_j^* and u and hence the endogeneity of Z_j^*.

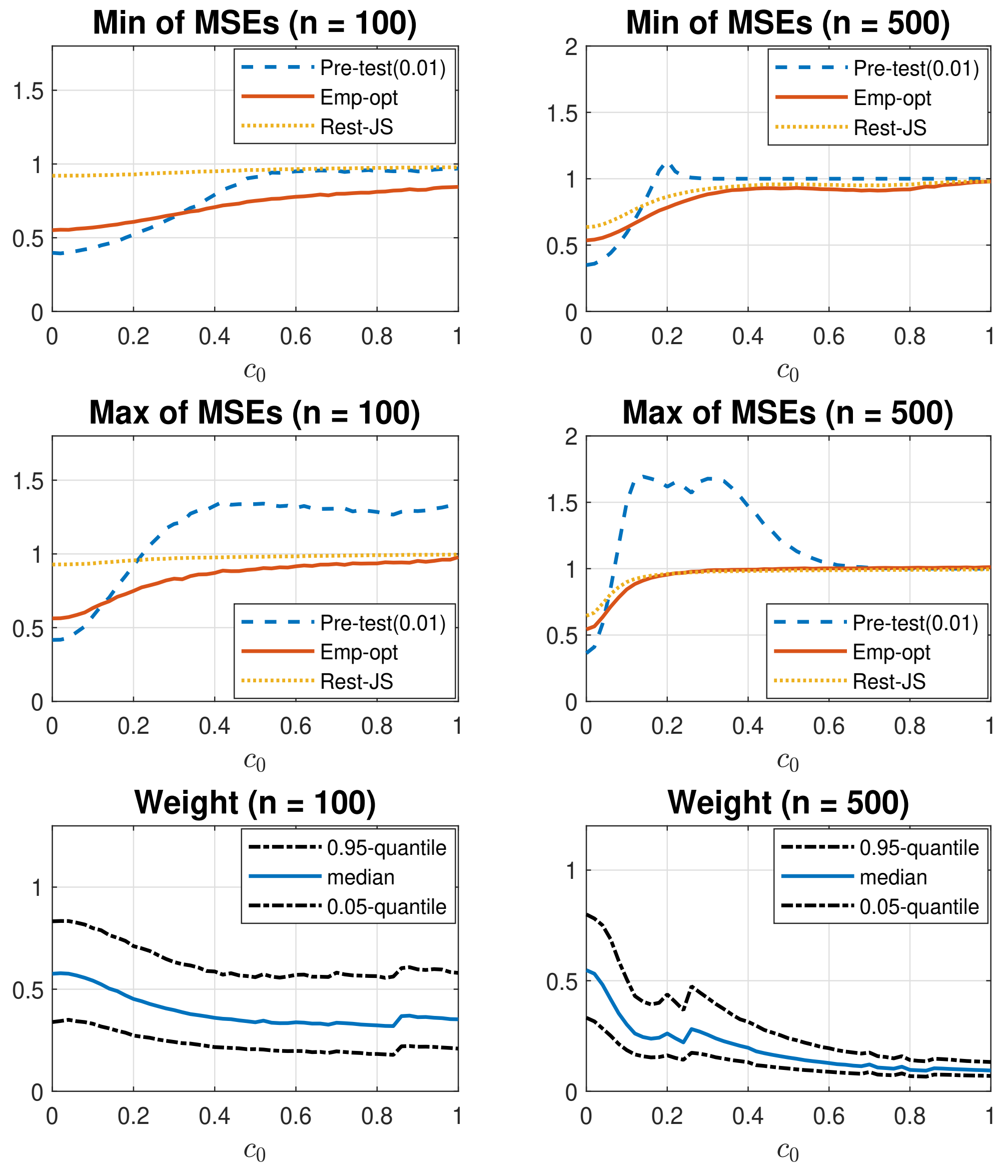
Given the sample size n, different DGPs of the simulated data $\{W_i : i = 1, ..., n\}$ are employed

Figure 2: Finite Sample MSEs of the Pre-test and Averaging GMM Estimators in S1



Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively.

Figure 3: Finite Sample MSEs of the Pre-test and Averaging GMM Estimators in S2



Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively.

in the simulation study by changing the values of $(c_1, ..., c_6)$. We consider

$$c_j = c_0\omega_j \text{ for } j = 1, \ldots, 6 \tag{6.11}$$

where c_0 is a scalar that takes values on the grid points between 0 and 1 with the grid length 0.02, and $(\omega_1, ..., \omega_6)$ is parametrized in two different ways. In the first one, we set $\omega_j = 0$ or 1 for $j = 1, \ldots, 6$ and rule out the case that $\omega_j = 0$ for all j (since this is the same as the case which sets $c_0 = 0$). In the second one, we consider the polar transformation and set

$$\begin{aligned} \omega_1 &= \sin(\alpha_1)\sin(\alpha_2)\sin(\alpha_3)\sin(\alpha_4)\sin(\alpha_5), \\ \omega_j &= \cos(\alpha_{j-1})\sin(\alpha_j) \times \cdots \times \sin(\alpha_5) \text{ for } j = 2, \ldots, 5, \\ \omega_6 &= \cos(\alpha_5), \end{aligned} \tag{6.12}$$

where $\alpha_1 \in \{\pi/4,\, 3\pi/4,\, 5\pi/4,\, 7\pi/4\}$ and $\alpha_j \in \{\pi/4,\, 3\pi/4\}$ for $j = 2, \ldots, 5$. Therefore, there are 127 different values for $(\omega_1, \ldots, \omega_6)$ for each of the 51 different values of c_0. For each DGP, we consider sample size $n = 50$, 100, 250, 500, 1000 and use 10000 simulation repetitions.

Given the sample size and the value of c_0, we report the minimum and the maximum of the 127 values of the finite sample relative MSEs for each estimator, and the weight $\widetilde{\omega}_{eo}$ in our averaging estimator in the DGP with the maximum relative MSE. Given each sample size, the maximum/minimum finite sample relative MSE and the weight are plotted as functions of c_0, see Figure 2 for S1 and Figure 3 for S2. In each figure, the left three panels and the right three panels include the results with sample size $n = 100$ and 500, respectively.[25] For each sample size, we also report the upper bound and the lower bound of the finite sample relative MSEs (among all 127×51 DGPs) of the averaging estimators and the pre-test estimator in Table 1.[26]

Our findings in the simulation designs S1 and S2 are summarized as follows. First, in both Figure 2 and Figure 3, we see that the minimum relative MSE of the averaging GMM estimator $\widehat{\theta}_{eo}$ is smaller than 1 (which is the normalized finite sample MSE of the conservative GMM estimator $\widehat{\theta}_1$) for all c_0 considered in both simulation designs. The maximum relative MSE of $\widehat{\theta}_{eo}$ is smaller than 1 when c_0 is small and approaches 1 when c_0 is close to 1. Table 1 provides detailed information

[25] We only report the untruncated MSEs with $n = 100$ and $n = 500$ here. The untruncated MSEs in S1 and S2 with $n = 50$, 250 and 1000 can be found in Figure C.1 and Figure C.2 in Section C of the Appendix, and the truncated MSEs ($\zeta = 1000$) in S1 and S2 with $n = 50$, 100, 250, 500 and 1000 can be found in Figure G.1, Figure G.2, Figure G.4 and Figure G.5 in Section G of the Supplemental Appendix. The simulation results on truncated MSEs are very similar to what we get without truncation. The maximum finite sample bias and finite sample variance for each c_0 are reported in Figure C.4 and Figure C.5 in Section C of the Appendix.

[26] The upper bound and the lower bound of the finite sample relative truncated MSEs are reported in Table G.1 in Section G of the Supplemental Appendix.

Table 1: The Lower and Upper Bounds of the Finite Sample Relative MSEs

		Design S1		Design S2		Design S3	
		Lower	Upper	Lower	Upper	Lower	Upper
	$\hat{\theta}_{oe}$	0.5732	0.7968	0.6113	0.8980	0.9302	1.0012
$n = 50$	$\hat{\theta}_{JS}$	0.9755	0.9959	0.9776	0.9978	0.9995	1.0003
	$\hat{\theta}_{pret}$	0.4424	0.9574	0.5057	1.0973	1.0324	1.4283
	$\hat{\theta}_{oe}$	0.5325	0.8789	0.5513	0.9781	0.9733	1.0040
$n = 100$	$\hat{\theta}_{JS}$	0.9208	0.9911	0.9202	0.9956	0.9996	1.0002
	$\hat{\theta}_{pret}$	0.3586	1.1940	0.3937	1.3539	0.9990	1.4709
	$\hat{\theta}_{oe}$	0.5316	0.9587	0.5384	1.0118	0.9720	1.0079
$n = 250$	$\hat{\theta}_{JS}$	0.7591	0.9787	0.7506	0.9923	0.9999	1.0000
	$\hat{\theta}_{pret}$	0.3360	1.5106	0.3598	1.6392	0.9753	1.4394
	$\hat{\theta}_{oe}$	0.5331	0.9846	0.5355	1.0112	0.9700	1.0096
$n = 500$	$\hat{\theta}_{JS}$	0.6443	0.9823	0.6359	0.9953	1.0000	1.0000
	$\hat{\theta}_{pret}$	0.3368	1.6196	0.3495	1.6937	0.9562	1.4236
	$\hat{\theta}_{oe}$	0.5335	0.9934	0.5341	1.0082	0.9681	1.0119
$n = 1,000$	$\hat{\theta}_{JS}$	0.5803	0.9890	0.5737	0.9978	1.0000	1.0000
	$\hat{\theta}_{pret}$	0.3395	1.6433	0.3451	1.6864	0.9473	1.3953

Note: 1. $\hat{\theta}_{JS}$ and $\hat{\theta}_{pret}$ denote the GMM averaging estimator based on the weight in (6.1) and the pre-testing GMM estimator based on J-test with nominal size 0.01 respectively; 2. the "Upper" and "Lower" refer to the upper bound and the lower bound of the finite sample relative MSEs among all DGPs considered in the simulation design given the sample size.

on the lower and upper bounds of the relative MSE of $\widehat{\theta}_{eo}$. In both simulation designs S1 and S2, the lower bound stays far below 1 while the upper bound approaches 1 with increasing sample size. These results are predicted by our theory because the key sufficient condition is satisfied in both S1 and S2.[27] Second, the pre-test GMM estimator has non-shrinking maximum relative MSE in S1 or S2, and therefore it fails to dominate the conservative GMM estimator $\widehat{\theta}_1$ in the asymptotic sense. For example, when $n = 500$, the pre-test GMM estimator in S1 has relative MSE above 1.5 when c_0 is between 0.2 and 0.4. From Table 1, we see that the upper bound of its relative MSE does not converge to 1 with increasing sample size. It stays around 1.61 and 1.69 in S1 and S2 respectively, when the sample size is large (e.g., $n = 500$ or 1000). Third, comparing the two averaging estimators, we find that the restricted JS estimator does not reduce the MSE as much as the averaging estimator based on $\widetilde{\omega}_{eo}$. Fourth, the weight $\widetilde{\omega}_{eo}$ becomes close to zero when c_0 is close to 1 for large n, which is clearly illustrated by the simulation with $n = 500$ in both S1 and S2. Last, the maximum relative MSE of the pre-test GMM estimator may show multiple peaks in

[27] It is easy to show that when $\delta_F = 0$ and Υ is identity matrix, we have $\text{tr}(A_F) = 4$ and $\text{tr}(A_F) - 4\rho_{\max}(A_F) = 4/3$ for S1 and $\text{tr}(A_F) = 8$ and $\text{tr}(A_F) - 4\rho_{\max}(A_F) = 8/3$ for S2.

Figure 2 and Figure 3, because given c_0 the Euclidean norm of $(c_1, \ldots, c_6)$ may be different under the two different parametrizations of $(\omega_1, \ldots, \omega_6)$. In the polar transformation, the Euclidean norm of $(c_1, \ldots, c_6)$ is c_0. However, the Euclidean norm of $(c_1, \ldots, c_6)$ is $c_0(\omega_1 + \cdots + \omega_6)^{1/2}$ in the other design and it may take 5 different values when we set $\omega_j = 0$ or 1 for $j = 1, \ldots, 6$ and rule out the case that $\omega_j = 0$ for all j. This also explains the kinks of the weight $\widetilde{\omega}_{eo}$ in the averaging estimator associated with the maximum MSE.

6.2 Simulation Model 2

In this subsection, we investigate the finite sample properties of the pre-test GMM estimator and the averaging GMM estimators when the key uniform dominance condition in Theorem 5.2 does not hold. In this simulation design, the structural equation takes the same form as (6.4) with $(\theta_1, ..., \theta_6) = 2.5 \times \mathbf{1}_{1\times 6}$, but the regressors X_j $(j = 1, \ldots, 6)$ are generated in a different way. We draw i.i.d. random vectors $(Z_1, ..., Z_{13}, \varepsilon_1, ..., \varepsilon_5, u)$ from normal distribution with mean zero and variance-covariance matrix $diag(I_{13\times 13}, \Sigma_{6\times 6})$, where

$$\Sigma_{6\times 6} = \begin{pmatrix} I_{5\times 5} & 0.25 \times \mathbf{1}_{5\times 1} \\ 0.25 \times \mathbf{1}_{1\times 5} & 1 \end{pmatrix}. \tag{6.13}$$

The observed data are $W = (Y, X_1, ..., X_6, Z_6, Z_7, Z_8, Z_1^*, ..., Z_5^*)$, where $(X_1, ..., X_5)$ are exogenous regressors and X_6 is an endogenous regressor, $(X_1, ..., X_5, Z_6, Z_7, Z_8)$ are valid IVs and $(Z_1^*, ..., Z_5^*)$ are potentially invalid IVs. The exogenous variables are generated by

$$\begin{aligned} X_j &= 3^{-\frac{1}{2}}(Z_j + Z_{j+1} + Z_{j+8}), \text{ for } j = 1, ..., 4, \\ X_5 &= 3^{-\frac{1}{2}}(Z_5 + Z_1 + Z_{13}). \end{aligned} \tag{6.14}$$

The endogenous variable X_6 is generated by

$$X_6 = 2^{-1} \sum_{j=6}^{8} Z_j + 10^{-1/2} \sum_{j=1}^{5} (Z_{j+8} + \varepsilon_j). \tag{6.15}$$

The potentially invalid IVs are generated by

$$Z_j^* = (1 - c_j^2)^{1/2} Z_{j+8} + c_j(\varepsilon_j + u) \text{ for } j = 1, ..., 5. \tag{6.16}$$

Note that the key sufficient condition in Theorem 5.2 is not satisfied for this design.[28] We call the simulation design in this subsection as S3.

Given the sample size n, we consider different DGPs of the simulated data $\{W_i : i = 1, ..., n\}$ by changing the values of $(c_1, ..., c_5)$. We consider the following parametrization

$$c_j = c_0 \omega_j \text{ for } j = 1, ..., 5 \tag{6.17}$$

where c_0 is a scalar that takes values on the grid points between 0 and 1 with the grid length 0.02, $(\omega_1, ..., \omega_5)$ is parametrized in two different ways. In the first one, we set $\omega_j = 0$ or 1 for $j = 1, \ldots, 5$ and rule out the case that $\omega_j = 0$ for all j (since this is the same as the case which sets $c_0 = 0$). In the second one, we consider the polar transformation and set

$$\begin{aligned} \omega_1 &= \sin(\alpha_1)\sin(\alpha_2)\sin(\alpha_3)\sin(\alpha_4), \\ \omega_j &= \cos(\alpha_{j-1})\sin(\alpha_j) \times \cdots \times \sin(\alpha_4) \text{ for } j = 2, \ldots, 4, \\ \omega_5 &= \cos(\alpha_4), \end{aligned} \tag{6.18}$$

where $\alpha_1 \in \{\pi/4, 3\pi/4, 5\pi/4, 7\pi/4\}$ and $\alpha_j \in \{\pi/4, 3\pi/4\}$ for $j = 2, \ldots, 5$. Therefore, there are 63 different values for $(\omega_1, \ldots, \omega_5)$ for each of the 51 different values of c_0. For each DGP, we consider sample size $n = 50$, 100, 250, 500, 1000 and use 10000 simulation repetitions.

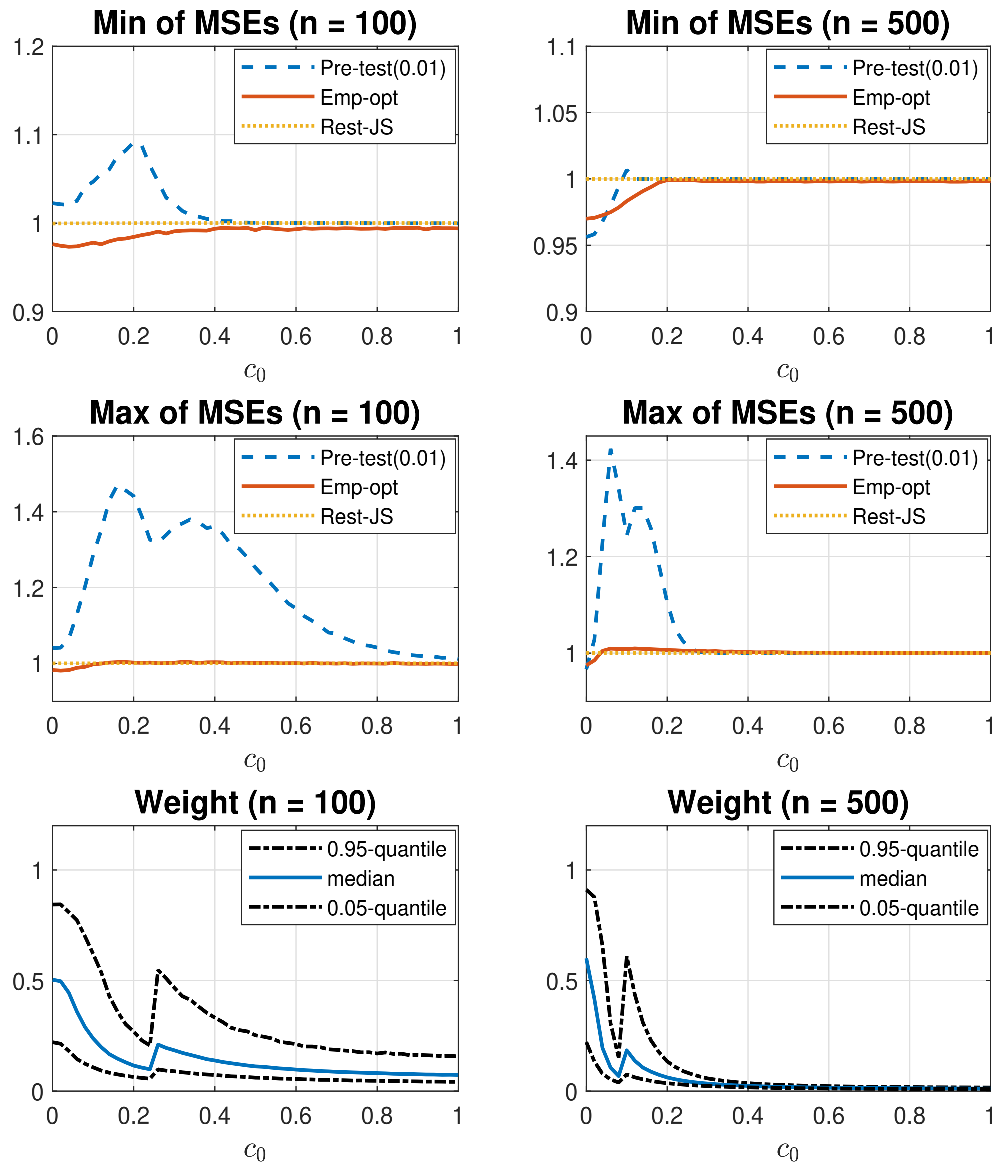
Given the sample size and the value of c_0, we report the minimum and maximum of the 63 values of the finite sample relative MSEs for each estimator, and the weight $\widetilde{\omega}_{eo}$ in our averaging estimator in the DGP with maximum relative MSE. Given each sample size, the maximum/minimum finite sample relative MSE and the weight are plotted as functions of c_0, see Figure 4.[29] For each sample size, the upper bound and the lower bound of the finite sample relative MSEs (among all 127×51 DGPs) of the averaging estimators and the pre-test estimator in this simulation design are also reported in Table 1.

Our findings in this simulation design are summarized as follows. First, compared to the conservative GMM estimator, the improvement of the pre-test GMM estimator or the averaging GMM estimator is small even when all the IVs Z_j^* ($j = 1, \ldots, 5$) are valid. This is because there is only one endogenous regressor and the improvement of using Z_j^* ($j = 1, \ldots, 5$) is mainly through the

[28] It is easy to show that, when $\delta_F = 0$, we have $\mathrm{tr}(A_F) = 0.4916$ and $\mathrm{tr}(A_F) - 4\rho_{\max}(A_F) = -1.4748 < 0$.

[29] We only report the untruncated MSEs with $n = 100$ and $n = 500$ here. The untruncated MSEs in S3 with $n = 50$, 250 and 1000 can be found in Figure C.3 in Section C of the Appendix, and the truncated MSEs ($\zeta = 1000$) in S3 with $n = 50$, 100, 250, 500 and 1000 can be found in Figure G.3 and Figure G.6 in Section G of the Supplemental Appendix. The simulation results on truncated MSEs are very similar to what we get without truncation. The maximum finite sample bias and finite sample variance for each c_0 are reported in Figure C.6 in Section C of the Appendix.

Figure 4: Finite Sample MSEs of the Pre-test and Averaging GMM Estimators in S3



Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively.

estimation of its coefficient. Second, both the pre-test GMM estimator and our averaging GMM estimator fail to dominate the conservative GMM estimator. However, the overall performance of the averaging GMM estimator is better than the pre-test GMM estimator. For example, when the sample size is 500, the maximum MSE of the pre-test GMM estimator is 1.4 times of that of the conservative GMM estimator. In contrast, the maximum MSE of our averaging GMM estimator is only slightly higher than (1.01 times of) that of the conservative GMM estimator. Third, the MSE of the JS-type averaging estimator is identical to the conservative GMM estimator even when all the IVs Z_j^* ($j = 1, \ldots, 5$) are valid. Therefore, this estimator performs the same as the conservative GMM estimator. Fourth, as c_0 goes to 1, the weight $\widetilde{\omega}_{eo}$ goes to zero for large sample size, which is well illustrated by the simulation with $n = 500$. Last, the maximum MSE of the pre-test GMM estimator and the weight $\widetilde{\omega}_{eo}$ in our averaging estimator may show multiple peaks for the same reason explained in the previous subsection.

7 Conclusion

This paper studies the averaging GMM estimator that combines the conservative estimator and the aggressive estimator with a data-dependent weight. The averaging weight is the sample analog of an optimal non-random weight. We provide a sufficient class of drifting DGPs under which the pointwise asymptotic results combine to yield uniform approximations to the finite-sample risk difference between two estimators. Using this asymptotic approximation, we show that the proposed averaging GMM estimator uniformly dominates the conservative GMM estimator for quadratic loss functions such as the mean square errors.

Inference based on the averaging estimator is an interesting and challenging problem. As pointed out in Pötscher (2006), the finite sample density of the averaging estimator can not be consistently estimated, which implies that directly applying an estimator of the finite-sample density may not yield uniformly valid inference. In addition to the uniform validity, a desirable confidence set should have smaller volume than that obtained from the conservative moments alone. We leave the inference issue to future investigation.

Appendix

A Illustration in Gaussian Location Model

This section shows that in a Gaussian location model, the averaging GMM estimator dominates the conservative GMM estimator in finite samples, i.e., it exhibits the JS phenomenon.

Suppose that we have one observation $(X', Y')'$ from the normal distribution

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \theta \\ \theta + d \end{pmatrix}, \sigma^2 I_{2k}\right) \tag{A.1}$$

where σ^2 is a known positive value, θ and d are $k \times 1$ vectors, and I_{2k} is a $2k \times 2k$ identity matrix ($k \geq 3$). It is clear that $d_\theta = k$ in model (A.1). We are interested in estimating θ.

Green and Strawderman (1991) consider the same model defined in (A.1). They propose the following JS-type of estimator

$$\widehat{\theta}_{GS} = X - \frac{\tau\sigma^2}{(X-Y)'(X-Y)}(X-Y) \tag{A.2}$$

where τ is a real constant in $(0, 2(k-2))$. Apparently, the above estimator $\widehat{\theta}_{GS}$ is an averaging estimator which combines an unbiased estimator X with a biased estimator Y with weight $\tau\sigma^2 \|X - Y\|^{-2}$ on the biased estimator. Green and Strawderman (1991) show that when $k \geq 3$, $\widehat{\theta}_{GS}$ has smaller MSE than the unbiased estimator X (which is the MLE of θ) for any $\theta \in \mathbb{R}^k$, any $d \in \mathbb{R}^k$ and any $\sigma^2 > 0$, and hence it uniformly dominates the MLE of θ.

Kim and White (2001), Judge and Mittelhammer (2004) and Mittelhammer and Judge (2005) propose averaging estimators which shrink the (asymptotic) unbiased estimator toward the biased estimator in semiparametric regression models. These papers show the dominance of the averaging estimator over the (asymptotic) unbiased estimator in the Gaussian location models using the joint normal distribution of the unbiased and biased estimators. In these papers, X and Y are the unbiased and biased estimators respectively with general variance-covariance matrix that allows for correlation between X and Y. The averaging estimator proposed in Kim and White (2001) is

$$\widehat{\theta}_{KW} = X - \left(c_1 + \frac{c_2}{(X-Y)'(X-Y)}\right)(X-Y) \tag{A.3}$$

where c_1 and c_2 constants. When $k \geq 5$, Kim and White (2001) show that there exist optimal values for c_1 and c_2 such that $\widehat{\theta}_{KW}$ dominates the unbiased estimator X. In the semiparametric setting, they show that these optimal values can be consistently estimated when $\mathbb{E}[Y] = \theta$. In Judge and Mittelhammer (2004) and Mittelhammer and Judge (2005), the averaging estimator takes the same form as $\widehat{\theta}_{GS}$ in (A.2) except $\tau\sigma^2$ is replaced by a constant. They show that when $k \geq 5$, there exists an optimal constant under which their averaging estimator dominates the unbiased estimator X. They provide an approximator of the infeasible constant and show that the approximator can be consistently estimated.

We next consider our averaging GMM estimator. Let Υ be the $k \times k$ identity matrix. The conservative GMM estimator $\widehat{\theta}_1 = X$ has risk $\sigma^2 \text{tr}(\Upsilon I_k) = \sigma^2 k$. On the other hand, the aggressive GMM estimator is $\widehat{\theta}_2 = (X + Y)/2$, which has risk $\sigma^2 k/2 + \|d\|^2/4$. The empirical optimal weight defined in (4.7) becomes

$$\widetilde{\omega}_{eo} = \frac{2k\sigma^2}{2k\sigma^2 + (X - Y)'(X - Y)}, \tag{A.4}$$

which together with the conservative and aggressive GMM estimators leads to the averaging GMM estimator

$$\widehat{\theta}_{eo} = X - \frac{k\sigma^2}{2k\sigma^2 + (X - Y)'(X - Y)}(X - Y). \tag{A.5}$$

From (A.2) and (A.5), we see that both $\widehat{\theta}_{GS}$ and $\widehat{\theta}_{eo}$ shrink the same unbiased estimator X to the same biased estimator Y but with different weights.

Lemma A.1 *When $k \geq 4$, the averaging estimator $\widehat{\theta}_{eo}$ defined in (A.5) satisfies*

$$\mathbb{E}\left[||\widehat{\theta}_{eo} - \theta||^2 - ||\widehat{\theta}_1 - \theta||^2\right] < 0 \tag{A.6}$$

for any $\theta \in \mathbb{R}^k$, any $d \in \mathbb{R}^k$ and any $\sigma^2 > 0$.

The inequality (A.6) shows that the risk of the averaging GMM estimator is strictly smaller than that of the conservative GMM estimator if $k \geq 4$, for any $\theta \in \mathbb{R}^k$, any $d \in \mathbb{R}^k$ and any $\sigma^2 > 0$, and hence it uniformly dominates the MLE of θ. The condition on k for the uniform dominance result of our averaging estimator is slightly stronger than the condition for Green and Strawderman (1991)'s estimator. The proof of Lemma A.1 is given in Section E of the Supplemental Appendix. It is different from the proof for that in Green and Strawderman (1991) and Judge and Mittelhammer (2004) because the two averaging estimators are different. But this proof is analogous to the proof of Theorem 5.2 for the general case. Thus we put it in the Supplemental Appendix.

B Proof of Results in Section 4 and Section 5

In Lemma 3.1, define

$$
\begin{aligned}
c_\rho &\equiv \min_{F^*\in\mathcal{F}^*}\left\{\rho_{\min}(\Gamma_{xz_1}\Gamma_{z_1x}),\rho_{\min}(\Psi)\right\}, \\
C_\rho &\equiv \max_{F^*\in\mathcal{F}^*}\left\{||\phi||^2,\rho_{\max}(\Psi)\right\}. \\
C_\Delta &\equiv \sup_{\delta_0\in\Delta_\delta}\|\delta_0\|^2 \qquad \text{(B.1)}
\end{aligned}
$$

Let

$$C_{*,W}\equiv 2(d_\theta+r_2+1)C_\rho,\ c_{*,\rho}\equiv\min\{1,c_\rho^2\}\text{ and }C_{*,\rho}\equiv C_{*,W}^2(2+C_\Delta^{1/2})^2. \qquad \text{(B.2)}$$

Then, in Lemma 3.1 (iii), the constant ε is given by

$$\varepsilon = c_{*,\rho}C_{*,\rho}^{-1}C_\Delta^{-1}, \qquad \text{(B.3)}$$

i.e., we require the condition to hold on a set bounded away from 0 by ε. The details of the proofs are given in Section D of the Supplemental Appendix.

B.1 Proof of the Results in Section 4

Let $\mu_n(g_2(W,\theta)) = n^{-1/2}\sum_{i=1}^n(g_2(W_i,\theta)-\mathbb{E}_{F_n}[g_2(W_i,\theta)])$. In the rest of the Appendix, we use C to denote a generic fixed positive finite constant which does not depend on any $F\in\mathcal{F}$ or n.

Lemma B.1 *Suppose that Assumption 3.2.(ii) holds and Θ is compact. Then we have*
(i) $\sup_{\theta\in\Theta}\|\overline{g}_2(\theta)-\mathbb{E}_{F_n}[g_2(W_i,\theta)]\| = o_p(1)$;
(ii) $\sup_{\theta\in\Theta}\left\|n^{-1}\sum_{i=1}^n g_2(W_i,\theta)g_2(W_i,\theta)'-\mathbb{E}_{F_n}[g_2(W_i,\theta)g_2(W_i,\theta)']\right\| = o_p(1)$;
(iii) $\sup_{\theta\in\Theta}\left\|n^{-1}\sum_{i=1}^n g_{2,\theta}(W_i,\theta)-\mathbb{E}_{F_n}[g_{2,\theta}(W_i,\theta)]\right\| = o_p(1)$;
(iv) $\mu_n(g_2(W,\theta))$ *is stochastic equicontinuous over* $\theta\in\Theta$*;*
(v) $\Omega_{2,F_n}^{-1/2}\mu_n(g_2(W,\theta_{F_n}))\to_D N(0_{r_2\times 1},I_{r_2})$.

Proof of Lemma B.1. See Lemma 11.3-11.5 of Andrews and Cheng (2013). ■

Define $M_{k,F}(\theta)=\mathbb{E}_F[g_k(W,\theta)]$, $G_{k,F}(\theta)=\mathbb{E}_F[g_{k,\theta}(W,\theta)]$ and $\Omega_{k,F}(\theta)=\mathrm{Var}_F[g_k(W,\theta)]$, for any $F\in\mathcal{F}$, for any $\theta\in\Theta$ and for $k=1,2$. The next lemma shows that $M_{2,F}(\cdot)$, $G_{2,F}(\cdot)$ and $\Omega_{2,F}(\cdot)$ are Lipschitz continuous uniformly over $F\in\mathcal{F}$.

Lemma B.2 *Under Assumptions 3.2.(i)-(ii), for any $F\in\mathcal{F}$ and any $\theta_1,\theta_2\in\Theta$, we have:*

(i) $\|M_{2,F}(\theta_1) - M_{2,F}(\theta_2)\| \leq C \|\theta_1 - \theta_2\|$;

(ii) $\|G_{2,F}(\theta_1) - G_{2,F}(\theta_2)\| \leq C \|\theta_1 - \theta_2\|$;

(iii) $\|\Omega_{2,F}(\theta_1) - \Omega_{2,F}(\theta_2)\| \leq C \|\theta_1 - \theta_2\|$.

Proof of Lemma B.2 is included in Section E of the Supplemental Appendix.

Lemma B.3 *Suppose that Assumptions 3.1.(i)-(ii) and 3.2.(i)-(ii) hold. Then for any sequence of DGPs* $\{F_n\}$*, we have*

$$\widetilde{\theta}_1 - \theta_{F_n} = o_p(1) \text{ and } \overline{\Omega}_2 = \Omega_{2,F_n} + o_p(1), \tag{B.4}$$

where $\widetilde{\theta}_1$ *is a preliminary estimator defined as*

$$\widetilde{\theta}_1 = \underset{\theta \in \Theta}{\arg\min}\ \overline{g}_1(\theta)'\overline{g}_1(\theta) \tag{B.5}$$

and $\overline{\Omega}_2$ *is defined in (E.13) of the Supplemental Appendix.*

Proof of Lemma B.3 is included in Section E of the Supplemental Appendix.

Lemma B.4 *Suppose that Assumptions 3.1.(i)-(ii) and 3.2 hold. Then for any sequence of DGPs* $\{F_n\}$*, we have*

$$n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) = \Gamma_{1,F_n}\mu_n(g_1(W, \theta_{F_n})) + o_p(1), \tag{B.6}$$

where $\Gamma_{1,F_n}\mu_n(g_1(W, \theta_{F_n})) \equiv -\left(G'_{1,F_n}\Omega^{-1}_{1,F_n}G_{1,F_n}\right)^{-1} G'_{1,F_n}\Omega^{-1}_{1,F_n} = O_p(1)$.

Proof of Lemma B.4 is included in Section E of the Supplemental Appendix.

Lemma B.5 *Suppose that Assumptions 3.1.(iii) and 3.2.(i)-(iii) hold. Then for any sequence of DGPs* $\{F_n\}$*, we have*

$$\widehat{\theta}_2 - \theta^*_{F_n} = o_p(1). \tag{B.7}$$

Proof of Lemma B.5 is included in Section E of the Supplemental Appendix.

Lemma B.6 *Suppose that Assumptions 3.1.(i)-(ii) and 3.2.(i)-(iii) hold. Consider any sequence of DGPs* $\{F_n\}$ *such that* $\delta_{F_n} = o(1)$*. Then we have*

$$\widehat{\theta}_2 - \theta_{F_n} = o_p(1). \tag{B.8}$$

If we further have Assumption 3.2.(iv), then

$$n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) = (\Gamma_{2,F_n} + o_p(1)) \left\{ \mu_n(g_2(W, \theta_{F_n})) + n^{1/2}\delta_{2,F_n} \right\} + o_p(1), \tag{B.9}$$

where $\Gamma_{2,F_n} = -\left(G'_{2,F_n}\Omega^{-1}_{2,F_n}G_{2,F_n}\right)^{-1} G'_{2,F_n}\Omega^{-1}_{2,F_n}$ *and* $\delta_{2,F_n} = (\mathbf{0}_{1\times r_1}, \delta'_{F_n})'$.

Proof of Lemma B.6 is included in Section E of the Supplemental Appendix.

Lemma B.7 *Under Assumptions 3.2.(ii) and 3.3.(ii), for any sequence of DGPs* $\{F_{p_n}\}$ *with* $F_{p_n} \in \mathcal{F}$ *where* $\{p_n\}$ *is a subsequence of* $\{n\}$*, there is a subsequence* $\{p_n^*\}$ *of* $\{p_n\}$ *such that* $v_{F_{p_n^*}}(\theta_{F_{p_n^*}}) \to v_F(\theta_F)$ *as* $p_n^* \to \infty$*, where* $F \in \mathcal{F}$*.*

Proof of Lemma B.7. Recall that $\Lambda = \{v_F : F \in \mathcal{F}\}$. By Assumptions 3.2.(ii) and 3.3.(ii), Λ is compact. Hence for any sequence $\{v_{F_{p_n}}(\theta_{F_{p_n}})\}$ in Λ, it has a convergent subsequence $\{v_{F_{p_n^*}}(\theta_{F_{p_n^*}})\}$ such that $v_{F_{p_n^*}}(\theta_{F_{p_n^*}}) \to v_F(\theta_F)$ as $p_n^* \to \infty$, where $F \in \mathcal{F}$. ■

Lemma B.8 *Suppose that Assumptions 3.1.(i)-(ii) and 3.2 hold. Consider any sequence of DGPs* $\{F_n\}$ *such that* $\overline{v}_{F_n} \to \overline{v}_F$ *for some* $F \in \mathcal{F}$*, and* $n^{1/2}\delta_{F_n} \to d$ *for* $d \in \mathbb{R}^{r^*}$*. Then*

$$\begin{pmatrix} n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \\ n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) \end{pmatrix} \to_D \begin{pmatrix} \xi_{1,F} \\ \xi_{2,F} \end{pmatrix} \equiv \begin{pmatrix} \Gamma_{1,F}\mathcal{Z}_{1,F} \\ \Gamma_{2,F}\left(\mathcal{Z}_{2,F} + d_0\right) \end{pmatrix},$$

where $d_0 = (\mathbf{0}_{1\times r_1}, d')'$.

Proof of Lemma B.8. In the proof, we use

$$G_{2,F_n} \to G_{2,F} \text{ and } \Omega_{2,F_n} \to \Omega_{2,F} \tag{B.10}$$

for some $F \in \mathcal{F}$, which is assumed in the lemma. Under Assumptions 3.1.(i)-(ii) and 3.2, for the sequence of DGPs $\{F_n\}$ considered in the lemma, we can apply Lemma B.4 and Lemma B.6 to deduce that

$$\begin{pmatrix} n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \\ n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) \end{pmatrix} = \begin{pmatrix} \Gamma_{1,F_n}\mu_n(g_1(W, \theta_{F_n})) \\ (\Gamma_{2,F_n} + o_p(1))\left\{\mu_n(g_2(W, \theta_{F_n})) + n^{1/2}\delta_{2,F_n}\right\} \end{pmatrix} + o_p(1), \tag{B.11}$$

where $\delta_{2,F_n} = (\mathbf{0}_{1\times r_1}, \delta'_{F_n})'$. By (B.10) and Assumption 3.2, we have

$$\Gamma_{1,F_n} = \Gamma_{1,F} + o(1) \text{ and } \Gamma_{2,F_n} = \Gamma_{2,F} + o(1) \tag{B.12}$$

where $\Gamma_{k,F} = -\left(G'_{k,F}\Omega^{-1}_{k,F}G_{k,F}\right)^{-1} G'_{k,F}\Omega^{-1}_{k,F}$ for $k = 1, 2$. Collecting the results in Lemma B.1.(v), (B.11) and (B.12), and then applying the continuous mapping theorem (CMT), we have

$$\begin{pmatrix} n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \\ n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) \end{pmatrix} \to_D \begin{pmatrix} \Gamma^*_{1,F} \\ \Gamma_{2,F} \end{pmatrix} (\mathcal{Z}_{2,F} + d_0), \tag{B.13}$$

where $\mathcal{Z}_{2,F} \sim N(\mathbf{0}_{r_2\times 1}, \Omega_{2,F})$, $\Gamma^*_{1,F} = (\Gamma_{1,F}, \mathbf{0}_{d_\theta\times r^*})$ and $d_0 = (\mathbf{0}_{1\times r_1}, d')'$. The claimed result follows from (B.13) and the definitions of $\Gamma^*_{1,F}$ and $\mathcal{Z}_{2,F}$. ■

Proof of Lemma 4.1. The claimed result in Part (a) has been proved in Lemma B.8.

We next consider the case that $n^{1/2}\delta_{F_n} \to d$ with $||d|| = \infty$. Note that the results in (B.6) and (B.12) do not depend on $||d|| < \infty$ or $||d|| = \infty$. Using (B.6), (B.12), Lemma B.1.(v) and the CMT, we have

$$n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \to_D \Gamma_{1,F}\mathcal{Z}_{1,F}. \tag{B.14}$$

To study the properties of $\widehat{\theta}_2$, we have to consider two separate scenarios: (1) $\delta_{F_n} = o(1)$; and (2) $\|\delta_{F_n}\| > c_\delta$ for some $c_\delta > 0$. In scenario (1), Assumption 3.2, Lemma B.1.(v) and Lemma B.6 imply that

$$n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) = (\Gamma_{2,F_n} + o_p(1))n^{1/2}\delta_{F_n} + O_p(1). \tag{B.15}$$

By Assumption 3.1.(iv) and $||n^{1/2}\delta_{F_n}|| \to \infty$,

$$n\delta'_{F_n}\Gamma'_{2,F_n}\Gamma_{2,F_n}\delta_{F_n} \geq C^{-2}n\delta'_{F_n}\delta_{F_n} \to \infty \tag{B.16}$$

which together with (B.15) implies that $||n^{1/2}(\widehat{\theta}_2 - \theta_{F_n})|| \to_p \infty$.

Finally, we consider the scenario (2) where $\|\delta_{F_n}\| > c_\delta$. By Assumption 3.1.(iv),

$$||G'_{2,F_n}\Omega^{-1}_{2,F_n}\delta_{F_n}|| > C^{-1}\|\delta_{F_n}\| > c_\delta C^{-1} \tag{B.17}$$

for any n. As $\theta^*_{F_n}$ is the minimizer of $Q_{F_n}(\theta)$, it has the following first order condition

$$0_{d_\theta\times 1} = G_{2,F_n}(\theta^*_{F_n})'\Omega^{-1}_{2,F_n}M_{2,F_n}(\theta^*_{F_n}), \tag{B.18}$$

which implies that

$$\begin{aligned} G'_{2,F_n}\Omega_{2,F_n}^{-1}\delta_{F_n} &= G_{2,F_n}(\theta_{F_n})'\Omega_{2,F_n}^{-1}M_{2,F_n}(\theta_{F_n}) - G_{2,F_n}(\theta^*_{F_n})'\Omega_{2,F_n}^{-1}M_{2,F_n}(\theta^*_{F_n}) \\ &= \left[G_{2,F_n}(\theta_{F_n}) - G_{2,F_n}(\theta^*_{F_n})\right]'\Omega_{2,F_n}^{-1}M_{2,F_n}(\theta_{F_n}) \\ &\quad + G_{2,F_n}(\theta^*_{F_n})'\Omega_{2,F_n}^{-1}\left[M_{2,F_n}(\theta_{F_n}) - M_{2,F_n}(\theta^*_{F_n})\right]. \end{aligned} \quad (B.19)$$

By Lemma B.2, the Cauchy-Schwarz inequality and Assumption 3.2.(ii)-(iii), we have

$$\begin{aligned} &\left\|\left[G_{2,F_n}(\theta_{F_n}) - G_{2,F_n}(\theta^*_{F_n})\right]'\Omega_{2,F_n}^{-1}M_{2,F_n}(\theta_{F_n})\right\| \\ \leq\ &\left\|G_{2,F_n}(\theta_{F_n}) - G_{2,F_n}(\theta^*_{F_n})\right\|\left\|\Omega_{2,F_n}^{-1}M_{2,F_n}(\theta_{F_n})\right\| \leq C\left\|\theta_{F_n} - \theta^*_{F_n}\right\|, \end{aligned} \quad (B.20)$$

where C is a fixed constant. Similarly, we have

$$\begin{aligned} &\left\|G_{2,F_n}(\theta^*_{F_n})'\Omega_{2,F_n}^{-1}\left[M_{2,F_n}(\theta_{F_n}) - M_{2,F_n}(\theta^*_{F_n})\right]\right\| \\ \leq\ &\left\|M_{2,F_n}(\theta_{F_n}) - M_{2,F_n}(\theta^*_{F_n})\right\|\left\|\Omega_{2,F_n}^{-1}G_{2,F_n}\theta^*_{F_n})\right\| \leq C\left\|\theta_{F_n} - \theta^*_{F_n}\right\|. \end{aligned} \quad (B.21)$$

Combining the results in (B.19), (B.20) and (B.21), and using the triangle inequality, we have

$$\left\|\theta_{F_n} - \theta^*_{F_n}\right\| \geq c_\delta C \quad (B.22)$$

for some fixed constant C. Using $\widehat{\theta}_2 = \theta^*_{F_n} + o_p(1)$ (which is proved in Lemma B.5) and the triangle inequality, we obtain

$$\left\|\widehat{\theta}_2 - \theta_{F_n}\right\| \geq \left|||\widehat{\theta}_2 - \theta^*_{F_n}|| - \left\|\theta^*_{F_n} - \theta_{F_n}\right\|\right| = \left\|\theta^*_{F_n} - \theta_{F_n}\right\|(1 + o_p(1)), \quad (B.23)$$

which together with (B.22) implies that $n^{1/2}||\widehat{\theta}_2 - \theta_{F_n}|| \to_p \infty$. This finishes the proof. ■

Lemma B.9 (a) $\Gamma^*_{1,F}d_0 = \mathbf{0}_{d_\theta\times 1}$; (b) $\Gamma^*_{1,F}\Omega_{2,F}\Gamma^{*\prime}_{1,F} = \Sigma_{1,F}$; (c) $\Gamma^*_{1,F}\Omega_{2,F}\Gamma'_{2,F} = \Sigma_{2,F}$; (d) $\Gamma_{2,F}\Omega_{2,F}\Gamma'_{2,F} = \Sigma_{2,F}$.

Proof of Lemma B.9 is included in Section E of the Supplemental Appendix.

B.2 Proof of the Results in Section 5

We first present some generic results on the bounds of asymptotic risk difference between two estimators under some high-level conditions. Then we apply these generic results to the two specific estimators we consider in this paper: $\widehat{\theta}_{eo}$ and $\widehat{\theta}_1$. The proof uses the subsequence techniques used to show the asymptotic size of a test in Andrews, Cheng, and Guggenberger (2011) but we adapt the proof and notations to the current setup and extend results from test to estimators.

Recall that $h_{F,d} = (d', \text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})')$ and $\overline{v}_F = (\text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})')$ for any $F \in \mathcal{F}$ and any $d \in \mathbb{R}^{r^*}_{\infty}$. We have defined

$$H = \{h_{F,d} : d \in \mathbb{R}^{r^*} \text{ and } F \in \mathcal{F} \text{ with } \delta_F = 0_{r^* \times 1}\} \tag{B.24}$$

where δ_F is defined by (1.5) for a given F. Define

$$H^*_{\infty} = \{h_{F,d} : d \in \mathbb{R}^{r^*}_{\infty} \text{ with } ||d|| = \infty \text{ and } F \in \mathcal{F}\}. \tag{B.25}$$

Let $d_h = r^* + d_\theta r_2 + (r_2 + 1)r_2/2$. It is clear that $h_{F,d}$ is a d_h-dimensional vector.

Condition B.1 (i) *For any sequence of DGPs $\{F_{p_n}\}$ with $F_{p_n} \in \mathcal{F}$ where $\{p_n\}$ is a subsequence of $\{n\}$, there exists a subsequence $\{p^*_n\}$ of $\{p_n\}$ and some $F \in \mathcal{F}$ such that $v_{F_{p^*_n}} \to v_F$ as $p^*_n \to \infty$;*
(ii) *$M_{1,F}(\theta) = 0_{r_1 \times 1}$ has a unique solution at $\theta_F \in \Theta$ for any $F \in \mathcal{F}$;*
(iii) *$M_{2,F}(\cdot)$ is uniform equicontinuous over $F \in \mathcal{F}$;*
(iv) *for any subsequence $\{p_n\}$ of $\{n\}_{n\in\mathbb{N}}$, if $(p_n)^{1/2}\delta_{F_{p_n}} \to d$ for $d \in \mathbb{R}^{r^*}_{\infty}$ and $v_{F_{p_n}} \to v_F$, then*

$$\lim_{n\to\infty} \mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta}, \theta_{F_{p_n}})] = R_\zeta(h_{F,d}) \text{ and } \lim_{n\to\infty} \mathbb{E}_{F_{p_n}}[\ell_\zeta(\widetilde{\theta}, \theta_{F_{p_n}})] = \widetilde{R}_\zeta(h_{F,d})$$

where $R_\zeta(h_{F,d})$ and $\widetilde{R}_\zeta(h_{F,d})$ are some non-negative functions that are bounded from above by ζ for any $F \in \mathcal{F}$ and any $d \in \mathbb{R}^{r^*}_{\infty}$;
(v) *for any $F \in \mathcal{F}$ with $\delta_F = 0_{r^* \times 1}$, there exists a constant $\varepsilon_F > 0$ such that for any $\widetilde{\delta} \in \mathbb{R}^{r^*}$ with $0 \leq ||\widetilde{\delta}|| < \varepsilon_F$, there is $\widetilde{F} \in \mathcal{F}$ with $\delta_{\widetilde{F}} = \widetilde{\delta}$ and $||\overline{v}_F - \overline{v}_{\widetilde{F}}|| \leq C||\widetilde{\delta}||^\kappa$ for some $\kappa > 0$;*
(vi) *for any $h_{F,d} \in H^*_{\infty}$ and $h_{F,\widetilde{d}} \in H^*_{\infty}$, we have*

$$R_\zeta(h_{F,d}) = R_\zeta(h_{F,\widetilde{d}}) \text{ and } \widetilde{R}_\zeta(h_{F,d}) = \widetilde{R}_\zeta(h_{F,\widetilde{d}})$$

for any $\zeta > 0$.

Condition B.1.(i) requires that for any sequence of $\{v_{F_{p_n}}\}$, it has a convergent subsequence $\{v_{F_{p^*_n}}\}$ with limit being v_F for some $F \in \mathcal{F}$. This condition is verified under Assumptions 3.2.(ii) and 3.3.(ii) in Lemma B.7. Condition B.1.(ii) is the unique identification condition of θ_F which holds under Assumptions 3.1.(i)-(ii). Condition B.1.(iii) holds under Assumption 3.2.(ii) by Lemma

B.2. Condition B.1.(iv) is a key assumption to derive an explicit upper bound of asymptotic risk. This condition can be verified by using Lemma 4.1 as we shall show in the proof of Theorem 5.1. Condition B.1.(v) enables us to show that the upper bound we derived for the asymptotic risk is also a lower bound. This condition is assumed in Assumption 3.3.(i). Condition B.1.(vi), in our context, requires that the asymptotic (truncated) risk of $\widehat{\theta}$ (or $\widetilde{\theta}$) under the subsequences of DGPs $\{F_{p_n}\}$ satisfying the restrictions in Condition B.1.(iv) are identical whenever $(p_n)^{1/2}\delta_{F_{p_n}} \to d$ with $||d|| = \infty$. Condition B.1 is verified in the proof of Theorem 5.1 below.

Lemma B.10 *Under Conditions B.1.(i) - B.1.(iv), we have*

$$AsyR_\zeta(\widehat{\theta}) \leq \max\left\{\sup_{h\in H} R_\zeta(h), \sup_{h\in H^*_\infty} R_\zeta(h)\right\}, \tag{B.26}$$

where $AsyR_\zeta(\widehat{\theta}) \equiv \limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta},\theta_F)]$.

Proof of Lemma B.10. Let $\{F_n\}$ be a sequence such that

$$\limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell_\zeta(\widehat{\theta},\theta_{F_n})] = \limsup_{n\to\infty}\left(\sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta},\theta_F)]\right) \equiv AsyR_\zeta(\widehat{\theta}). \tag{B.27}$$

Such a sequence always exists by the definition of supremum. The sequence $\{\mathbb{E}_{F_n}[\ell_\zeta(\widehat{\theta},\theta_{F_n})]\colon n \geq 1\}$ may not converge. However, by the definition of limsup, there exists a subsequence of $\{n\}_{n\in\mathbb{N}}$, say $\{p_n\}$, such that $\{\mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta},\theta_{F_{p_n}})]\colon n \geq 1\}$ converges and

$$\lim_{n\to\infty} \mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta},\theta_{F_{p_n}})] = AsyR_\zeta(\widehat{\theta}). \tag{B.28}$$

Below we show that for any subsequence $\{p_n\}$ of $\{n\}_{n\in\mathbb{N}}$ such that $\{\mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta},\theta_{F_{p_n}})]\colon n \geq 1\}$ is convergent, there exists a subsequence $\{p^*_n\}$ of $\{p_n\}$ such that

$$\lim_{n\to\infty} \mathbb{E}_{F_{p^*_n}}[\ell_\zeta(\widehat{\theta},\theta_{F_{p^*_n}})] = R_\zeta(h) \text{ for some } h \in H \text{ or } H^*_\infty. \tag{B.29}$$

Because $\lim_{n\to\infty} \mathbb{E}_{F_{p^*_n}}[\ell_\zeta(\widehat{\theta},\theta_{F_{p^*_n}})] = \lim_{n\to\infty} \mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta},\theta_{F_{p_n}})]$, which combined with (B.28) and (B.29) implies that

$$AsyR_\zeta(\widehat{\theta}) = R_\zeta(h) \text{ for some } h \in H \text{ or } H^*_\infty. \tag{B.30}$$

The desired result in (B.26) follows immediately by (B.30).

To show that there exists a subsequence $\{p^*_n\}$ of $\{p_n\}$ such that (B.29) holds, it suffices to show that for any sequence $\{F_n\}$ and any subsequence $\{p_n\}$ of $\{n\}_{n\in\mathbb{N}}$, there exists a subsequence $\{p^*_n\}$

of $\{p_n\}$ for which we have

$$(p_n^*)^{1/2}\delta_{F_{p_n^*}} \to d \text{ for } d \in \mathbb{R}_\infty^{r^*} \text{ and } v_{F_{p_n^*}} \to v_F \tag{B.31}$$

for some $F \in \mathcal{F}$. If (B.31) holds, then we can use Condition B.1.(iv) to deduce that

$$\lim_{n\to\infty} \mathbb{E}_{F_{p_n^*}}[\ell_\zeta(\widehat{\theta}, \theta_{F_{p_n^*}})] = R_\zeta(h_{F,d}) \tag{B.32}$$

for the sequence of DGPs $\{F_{p_n^*}\}$ that satisfies (B.31). As $d \in \mathbb{R}_\infty^{r^*}$, we have either $\|d\| < \infty$ or $\|d\| = \infty$. In the first case, $\|d\| < \infty$ together with $(p_n^*)^{1/2}\delta_{F_{p_n^*}} \to d$ and $\delta_{F_{p_n^*}} \to \delta_F$ (which is implied by $v_{F_{p_n^*}} \to v_F$) implies that $\delta_F = 0_{r^*\times 1}$, which implies that $h_{F,d} \in H$ by the definition of H. In the second case, $h_{F,d} \in H_\infty^*$ by the definition of H_∞^*. We have proved that $h_{F,d}$ in (B.32) belongs either to H or H_∞^* which together with (B.32) proves (B.29).

Finally, we show that for any sequence $\{F_n\}$ and any subsequence $\{p_n\}$ of $\{n\}_{n\in\mathbb{N}}$, there exists a subsequence $\{p_n^*\}$ of $\{p_n\}$ for which (B.31) holds. Let $\delta_{p_n,j}$ denote the j-th component of δ_{p_n} and $p_{1,n} = p_n$ for any $n \geq 1$. For $j = 1$, either (a) $\limsup_{n\to\infty} |p_{j,n}^{1/2}\delta_{p_{j,n},j}| < \infty$; or (b) $\limsup_{n\to\infty} |p_{j,n}^{1/2}\delta_{p_{j,n},j}| = \infty$. If (a) holds, then for some subsequence $\{p_{j+1,n}\}$ of $\{p_{j,n}\}$, $p_{j+1,n}^{1/2}\delta_{p_{j+1,n},j} \to d_j$ for some $d_j \in \mathbb{R}$. If (b) holds, then for some subsequence $\{p_{j+1,n}\}$ of $\{p_{j,n}\}$, $p_{j+1,n}^{1/2}\delta_{p_{j+1,n},j} \to \infty$ or $-\infty$. As r^* is a fixed positive integer, we can apply the same arguments successively for $j = 1, ..., r^*$ to obtain a subsequence $\{p_{r^*,n}\}$ of $\{p_n\}$ such that $(p_{r^*,n})^{1/2}\delta_{p_{r^*,n}} \to d \in \mathbb{R}_\infty^{r^*}$. By Condition B.1.(i), we know that there exists a subsequence $\{p_n^*\}$ of $\{p_{r^*,n}\}$ such that $v_{p_n^*} \to v_F$ for some $F \in \mathcal{F}$, which finishes the proof of (B.31). ■

Lemma B.11 *Suppose that Condition B.1.(v) holds. Then (i) for any $h_{F,d} \in H$, there exists a sequence of DGPs $\{F_n\}$ with $F_n \in \mathcal{F}$ such that*

$$n^{1/2}\delta_{F_n} \to d,\ G_{2,F_n} \to G_{2,F} \text{ and } \Omega_{2,F_n} \to \Omega_{2,F}; \tag{B.33}$$

(ii) for any $h_{F,d} \in H_\infty^*$, there exists a sequence of DGPs $\{F_n\}$ with $F_n \in \mathcal{F}$ such that

$$||n^{1/2}\delta_{F_n}|| \to \infty,\ G_{2,F_n} \to G_{2,F},\ \Omega_{2,F_n} \to \Omega_{2,F} \text{ and } \delta_{F_n} \to \delta_F. \tag{B.34}$$

Proof of Lemma B.11. (i) By the definition of H, we have $\delta_F = 0_{r^*\times 1}$ for any F such that $h_{F,d} \in H$. Let N_{ε_F} be the smallest n such that $\|d\|\, n^{-1/2} < \varepsilon_F$. By Condition B.1.(v), for any

$n \geq N_{\varepsilon_F}$ we can find a DGP F_n such that

$$\delta_{F_n} = n^{-1/2}d \text{ and } \|\overline{v}_{F_n} - \overline{v}_F\| \leq n^{-\kappa/2}C||d||^{\kappa}. \tag{B.35}$$

For any $n < N_{\varepsilon_F}$ such that $\|d\|\, n^{-1/2} \geq \varepsilon_F$, we let $F_n = F$. The desired properties in (B.33) holds under the constructed sequence of DGPs $\{F_n\}$ by (B.35), because C is a fixed constant and $\kappa > 0$.

(ii) For any $h_{F,d} \in H^*_\infty$, we have either $\delta_F = 0_{r^*\times 1}$ or $||\delta_F|| > 0$. We first consider the case that $\delta_F = 0_{r^*\times 1}$. Let $1_{r^*\times 1}$ denote the $r^* \times 1$ vector of ones. Let N_{ε_F} be the smallest n such that $n^{-1/4}(r^*)^{1/2} < \varepsilon_F$. By Condition B.1.(v), for any $n \geq N_{\varepsilon_F}$ we can find a DGP F_n such that

$$\delta_{F_n} = n^{-1/4}1_{r^*\times 1} \text{ and } \left\|\overline{v}_{F_{pn}} - \overline{v}_F\right\| \leq Cn^{-\kappa/4}(r^*)^{\kappa/2}. \tag{B.36}$$

For any $n < N_{\varepsilon_F}$ such that $n^{-1/4}(r^*)^{1/2} \geq \varepsilon_F$, we let $F_n = F$. The desired properties in (B.34) holds under the constructed sequence of DGPs $\{F_n\}$ by (B.36), because C is a fixed constant and $\kappa > 0$. When $||\delta_F|| > 0$, we define a trivial sequence of DGPs $\{F_n\}$ as $F_n = F$ for any n. It is clear that (B.34) holds trivially in this case. ∎

Lemma B.12 *Under Condition B.1, we have*

$$AsyR_\zeta(\widehat{\theta}) = \max\left\{\sup_{h\in H} R_\zeta(h),\ \sup_{h\in H^*_\infty} R_\zeta(h)\right\}. \tag{B.37}$$

Proof of Lemma B.12. In view of the upper bound in (B.26) in Lemma B.10, it is sufficient to show that

$$AsyR_\zeta(\widehat{\theta}) \geq \max\left\{\sup_{h\in H} R_\zeta(h),\ \sup_{h\in H^*_\infty} R_\zeta(h)\right\}. \tag{B.38}$$

First, we note that for any $h_{d,F} = (d', \text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})') \in H$, there exists a sequence $\{F_n \in \mathcal{F} : n \geq 1\}$ such that

$$n^{1/2}\delta_{F_n} \to d \in \mathbb{R}^{r^*} \text{ and } v_{F_n} \to v_F \tag{B.39}$$

by Lemma B.11.(i). The sequence $\mathbb{E}_{F_n}[\ell_\zeta(\widehat{\theta}, \theta_{F_n})]$ may not be convergent, but there exists a subsequence $\{p_n\}$ of n such that $\mathbb{E}_{F_{pn}}[\ell_\zeta(\widehat{\theta}, \theta_{F_{pn}})]$ is convergent and

$$\lim_{n\to\infty} \mathbb{E}_{F_{pn}}[\ell(\widehat{\theta}, \theta_{F_{pn}})] = \limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widehat{\theta}, \theta_{F_n})]. \tag{B.40}$$

As $\{p_n\}$ is a subsequence of $\{n\}_{n\in\mathbb{N}}$, by (B.39)

$$(p_n)^{1/2}\delta_{F_{p_n}} \to d \in \mathbb{R}^{r^*} \text{ and } v_{F_{p_n}} \to v_F. \tag{B.41}$$

By Condition B.1.(iv), we have that

$$\lim_{n\to\infty} \mathbb{E}_{F_{p_n}}[\ell(\widehat{\theta}, \theta_{F_{p_n}})] = R_\zeta(h_{F,d}), \tag{B.42}$$

which combined with (B.40) and the definition of $AsyR_\zeta(\widehat{\theta})$ gives

$$AsyR_\zeta(\widehat{\theta}) = \limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}, \theta_F)] \geq \limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widehat{\theta}, \theta_{F_n})] = R_\zeta(h_{F,d}). \tag{B.43}$$

Second, consider any $h_{d,F} = (d', \mathrm{vec}(G_{2,F})', \mathrm{vech}(\Omega_{2,F})') \in H^*_\infty$. By Lemma B.11.(ii), there exists a sequence of DGPs $\{F_n\}$ such that

$$||n^{1/2}\delta_{F_n}|| \to \infty \text{ and } v_{F_n} \to v_F. \tag{B.44}$$

Using the same arguments in proving (B.40) to (B.42), we can show that for some subsequence $\{p_n\}$ of $\{n\}_{n\in\mathbb{N}}$,

$$|p_n^{1/2}\delta_{F_{p_n}}|| \to \infty \text{ and } v_{p_n} \to v_F \tag{B.45}$$

and

$$\limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widehat{\theta}, \theta_{F_n})] = \lim_{n\to\infty} \mathbb{E}_{F_{p_n}}[\ell(\widehat{\theta}, \theta_{F_{p_n}})] = R_\zeta(h_{F,d}). \tag{B.46}$$

for $||d|| = \infty$ by Conditions B.1.(vi). By the definition of $AsyR_\zeta(\widehat{\theta})$ and (B.46),

$$AsyR_\zeta(\widehat{\theta}) = \limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}, \theta_F)] \geq \limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widehat{\theta}, \theta_{F_n})] = R_\zeta(h_{F,d}). \tag{B.47}$$

Combining the results in (B.43) and (B.47), we immediately get (B.37). ■

Lemma B.13 *Under Conditions B.1.(i) - B.1.(iv), the upper and lower bounds of the asymptotic risk difference between* $\widehat{\theta}$ *and* $\widetilde{\theta}$ *satisfy*

$$AsyRD\overline{RD}(\widehat{\theta}, \widetilde{\theta}) \leq \lim_{\zeta\to\infty} \left(\max\left\{ \sup_{h\in H} \left[R_\zeta(h) - \widetilde{R}_\zeta(h) \right], \sup_{h\in H^*_\infty} \left[R_\zeta(h) - \widetilde{R}_\zeta(h) \right] \right\} \right), \tag{B.48}$$

$$Asy\underline{RD}(\widehat{\theta}, \widetilde{\theta}) \geq \lim_{\zeta\to\infty} \left(\min\left\{ \inf_{h\in H} \left[R_\zeta(h) - \widetilde{R}_\zeta(h) \right], \inf_{h\in H^*_\infty} \left[R_\zeta(h) - \widetilde{R}_\zeta(h) \right] \right\} \right), \tag{B.49}$$

where

$$\widetilde{R}_{\zeta}(h) \equiv \mathbb{E}[\min\{\xi_{1,F}'\Upsilon\xi_{1,F},\zeta\}] \text{ and } R_{\zeta}(h) \equiv \begin{cases} \mathbb{E}\left[\min\left\{\overline{\xi}_F'\Upsilon\overline{\xi}_F,\zeta\right\}\right], & \|d\|<\infty \\ \mathbb{E}\left[\min\left\{\xi_{1,F}'\Upsilon\xi_{1,F},\zeta\right\}\right], & \|d\|=\infty \end{cases}$$

for any $h \in H \cup H_{\infty}^{*}$.

Proof of Lemma B.13. Define

$$\overline{R}_{\zeta}(H,H_{\infty}^{*}) \equiv \max\left\{\sup_{h\in H}\left[R_{\zeta}(h)-\widetilde{R}_{\zeta}(h)\right], \sup_{h\in H_{\infty}^{*}}\left[R_{\zeta}(h)-\widetilde{R}_{\zeta}(h)\right]\right\}, \tag{B.50}$$

$$\underline{R}_{\zeta}(H,H_{\infty}^{*}) \equiv \min\left\{\inf_{h\in H}\left[R_{\zeta}(h)-\widetilde{R}_{\zeta}(h)\right], \inf_{h\in H_{\infty}^{*}}\left[R_{\zeta}(h)-\widetilde{R}_{\zeta}(h)\right]\right\}. \tag{B.51}$$

By the definition of $Asy\overline{RD}(\widehat{\theta},\widetilde{\theta})$, to show (B.48) it is sufficient to show that for any $\zeta>0$

$$\limsup_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{E}_F[\ell_{\zeta}(\widehat{\theta},\theta_F)-\ell_{\zeta}(\widetilde{\theta},\theta_F)] \leq \overline{R}_{\zeta}(H,H_{\infty}^{*}), \tag{B.52}$$

which can be proved using the same arguments in the proof of Lemma B.10 (but replacing $\ell_{\zeta}(\widehat{\theta},\theta_F)$ and $R_{\zeta}(h)$ by $\ell_{\zeta}(\widehat{\theta},\theta_F)-\ell_{\zeta}(\widetilde{\theta},\theta_F)$ and $R_{\zeta}(h)-\widetilde{R}_{\zeta}(h)$ respectively). Similarly by the definition of $Asy\underline{RD}(\widehat{\theta},\widetilde{\theta})$, for (B.49) it is sufficient to show that for any $\zeta>0$

$$\liminf_{n\to\infty}\inf_{F\in\mathcal{F}}\mathbb{E}_F[\ell_{\zeta}(\widehat{\theta},\theta_F)-\ell_{\zeta}(\widetilde{\theta},\theta_F)] \geq \underline{R}_{\zeta}(H,H_{\infty}^{*}), \tag{B.53}$$

which can be proved using the same arguments in the proof of Lemma B.10 (but replacing $\limsup_n$, $\sup_{F\in\mathcal{F}}$, $\ell_{\zeta}(\widehat{\theta},\theta_F)$ and $R_{\zeta}(h)$ by $\liminf_n$, $\inf_{F\in\mathcal{F}}$, $\ell_{\zeta}(\widehat{\theta},\theta_F)-\ell_{\zeta}(\widetilde{\theta},\theta_F)$ and $R_{\zeta}(h)-\widetilde{R}_{\zeta}(h)$ respectively). ■

Lemma B.14 *Under Condition B.1, the upper and lower bounds of the asymptotic risk difference between* $\widehat{\theta}$ *and* $\widetilde{\theta}$ *have the following representations:*

$$Asy\overline{RD}(\widehat{\theta},\widetilde{\theta}) = \lim_{\zeta\to\infty}\left(\max\left\{\sup_{h\in H}\left[R_{\zeta}(h)-\widetilde{R}_{\zeta}(h)\right], \sup_{h\in H_{\infty}^{*}}\left[R_{\zeta}(h)-\widetilde{R}_{\zeta}(h)\right]\right\}\right), \tag{B.54}$$

$$Asy\underline{RD}(\widehat{\theta},\widetilde{\theta}) = \lim_{\zeta\to\infty}\left(\min\left\{\inf_{h\in H}\left[R_{\zeta}(h)-\widetilde{R}_{\zeta}(h)\right], \inf_{h\in H_{\infty}^{*}}\left[R_{\zeta}(h)-\widetilde{R}_{\zeta}(h)\right]\right\}\right). \tag{B.55}$$

Proof of Lemma B.14. By Lemma B.13, it is sufficient to show that

$$\limsup_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{E}_F[\ell_\zeta(\widehat{\theta},\theta_F)-\ell_\zeta(\widetilde{\theta},\theta_F)] \geq \overline{R}_\zeta(H,H^*_\infty), \tag{B.56}$$

$$\liminf_{n\to\infty}\inf_{F\in\mathcal{F}}\mathbb{E}_F[\ell_\zeta(\widehat{\theta},\theta_F)-\ell_\zeta(\widetilde{\theta},\theta_F)] \leq \underline{R}_\zeta(H,H^*_\infty), \tag{B.57}$$

for any $\zeta>0$. (B.56) can be proved using the same arguments in the proof of Lemma B.12 by replacing $\ell_\zeta(\widehat{\theta},\theta_F)$ and $R_\zeta(h)$ by $\ell_\zeta(\widehat{\theta},\theta_F)-\ell_\zeta(\widetilde{\theta},\theta_F)$ and $R_\zeta(h)-\widetilde{R}_\zeta(h)$ respectively. Similarly, (B.57) can be proved using the same arguments in the proof of Lemma B.12 by replacing $\limsup_n$, $\sup_{F\in\mathcal{F}}$, $\ell_\zeta(\widehat{\theta},\theta_F)$ and $R_\zeta(h)$ by $\liminf_n$, $\inf_{F\in\mathcal{F}}$, $\ell_\zeta(\widehat{\theta},\theta_F)-\ell_\zeta(\widetilde{\theta},\theta_F)$ and $R_\zeta(h)-\widetilde{R}_\zeta(h)$ respectively. ■

Lemma B.15 *Under Assumptions 3.2.(ii) and 3.2.(iv), we have*

$$\sup_{h\in H}\mathbb{E}[(\xi_{1,F}'\Upsilon\xi_{1,F})^2]\leq C \text{ and } \sup_{h\in H}\mathbb{E}[(\overline{\xi}_F'\Upsilon\overline{\xi}_F)^2]\leq C. \tag{B.58}$$

Lemma B.16 *Let* $g_\zeta(h)\equiv\mathbb{E}\left[\min\{\overline{\xi}_F'\Upsilon\overline{\xi}_F,\zeta\}-\min\left\{\xi_{1,F}'\Upsilon\xi_{1,F},\zeta\right\}\right]$. *Under Assumptions 3.2.(ii) and 3.2.(iv), we have*

$$\lim_{\zeta\to\infty}\sup_{h\in H}\left[|g_\zeta(h)-g(h)|\right]=0 \tag{B.59}$$

where $\sup_{h\in H}\left[|g(h)|\right]\leq C$.

Proof of Theorem 5.1. The proof consists of two steps. The first step is to apply Lemma B.14 to show (B.60) and (B.61) below, and the second step is to apply Lemma B.16 to show (B.75) and (B.76) below.

In the first step, we apply Lemma B.14 with $\widehat{\theta}=\widehat{\theta}_{eo}$ and $\widetilde{\theta}=\widehat{\theta}_1$ to show that

$$Asy\overline{RD}(\widehat{\theta}_{eo},\widehat{\theta}_1) = \lim_{\zeta\to\infty}\max\left\{\sup_{h\in H}\left[g_\zeta(h)\right],0\right\} \text{ and} \tag{B.60}$$

$$Asy\underline{RD}(\widehat{\theta}_{eo},\widehat{\theta}_1) = \lim_{\zeta\to\infty}\min\left\{\inf_{h\in H}\left[g_\zeta(h)\right],0\right\}. \tag{B.61}$$

To prove (B.60) and (B.61), we now verify Condition B.1 under Assumptions 3.1-3.3. Condition B.1.(i) is verified by Lemma B.7 under Assumptions 3.2.(ii) and 3.3.(ii). Condition B.1.(ii) is implied by Assumptions 3.1.(i) and 3.1.(ii). Condition B.1.(iii) is implied by Assumptions 3.2.(i)-(ii) as a result of Lemma B.2. Condition B.1.(v) is assumed in Assumption 3.3.(ii). We next verify Conditions B.1.(iv) and B.1.(vi).

Consider any sequence of DGPs $\{F_{p_n}\}$ with

$$(p_n)^{1/2}\delta_{F_{p_n}} \to d \text{ for } d \in \mathbb{R}^{r^*}_{\infty} \text{ and } v_{F_{p_n}} \to v_F \tag{B.62}$$

for some $F \in \mathcal{F}$, where $\{p_n\}$ is a subsequence of $\{n\}_{n\in\mathbb{N}}$. First, we consider the case that $d \in \mathbb{R}^{r^*}$. By Lemma 4.1.(a) and 4.2.(a),

$$(p_n)^{1/2}(\widehat{\theta}_1 - \theta_{F_{p_n}}) \to_D \xi_{1,F} \text{ and } (p_n)^{1/2}(\widehat{\theta}_{eo} - \theta_{F_{p_n}}) \to_D \overline{\xi}_F \tag{B.63}$$

which combined with the continuous mapping theorem implies that

$$\ell(\widehat{\theta}_1, \theta_{F_{p_n}}) \to_D \xi_{1,F}'\Upsilon\xi_{1,F} \text{ and } \ell(\widehat{\theta}_{eo}, \theta_{F_{p_n}}) \to_D \overline{\xi}_F'\Upsilon\overline{\xi}_F. \tag{B.64}$$

Since Υ is positive semi-definite, $\xi_{1,F}'\Upsilon\xi_{1,F}$ and $\overline{\xi}_F'\Upsilon\overline{\xi}_F$ are both non-negative. The function $f_\zeta(x) = \min\{x, \zeta\}$ is a bounded continuous function for $x \geq 0$. By (B.64) and the Portmanteau Lemma (see Lemma 2.2 in van der Vaart (1998)),

$$\mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta}_{eo}, \theta_{F_{p_n}})] \to \mathbb{E}\left[\min\{\overline{\xi}_F'\Upsilon\overline{\xi}_F, \zeta\}\right] \text{ and } \mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta}_1, \theta_{F_{p_n}})] \to \mathbb{E}\left[\min\{\xi_{1,F}'\Upsilon\xi_{1,F}, \zeta\}\right]. \tag{B.65}$$

Second, we consider the case that $\|d\| = \infty$. Then under Lemma 4.1.(b) and 4.2.(b),

$$(p_n)^{1/2}(\widehat{\theta}_1 - \theta_{F_{p_n}}) \to_D \xi_{1,F} \text{ and } (p_n)^{1/2}(\widehat{\theta}_{eo} - \theta_{F_{p_n}}) \to_D \xi_{1,F}. \tag{B.66}$$

Using the same arguments in showing (B.65), we get

$$\mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta}_{eo}, \theta_{F_{p_n}})] \to \mathbb{E}\left[\min\{\xi_{1,F}'\Upsilon\xi_{1,F}, \zeta\}\right] \text{ and } \mathbb{E}_{F_{p_n}}[\ell_\zeta(\widehat{\theta}_1, \theta_{F_{p_n}})] \to \mathbb{E}\left[\min\{\xi_{1,F}'\Upsilon\xi_{1,F}, \zeta\}\right]. \tag{B.67}$$

Define

$$\widetilde{R}_\zeta(h_{F,d}) = \mathbb{E}\left[\min\{\xi_{1,F}'\Upsilon\xi_{1,F}, \zeta\}\right] \text{ and } R_\zeta(h_{F,d}) = \begin{cases} \mathbb{E}[\min\{\overline{\xi}_F'\Upsilon\overline{\xi}_F, \zeta\}], & \|d\| < \infty \\ \mathbb{E}\left[\min\{\xi_{1,F}'\Upsilon\xi_{1,F}, \zeta\}\right], & \|d\| = \infty \end{cases}. \tag{B.68}$$

Collecting the results in (B.65) and (B.67), we deduce that under the sequence of DGPs $\{F_{p_n}\}$

satisfying (B.62),

$$\mathbb{E}_{F_{pn}}[\ell_\zeta(\widehat{\theta}_{eo},\theta_{F_{pn}})] \to R_\zeta(h_{F,d}) \text{ and } \mathbb{E}_{F_{pn}}[\ell_\zeta(\widehat{\theta}_1,\theta_{F_{pn}})] \to \widetilde{R}_\zeta(h_{F,d}), \tag{B.69}$$

where $R_\zeta(h_{F,d})$ and $\widetilde{R}_\zeta(h_{F,d})$ are non-negative and bounded from above by ζ for any $d \in \mathbb{R}^{r^*}_\infty$ and any $F \in \mathcal{F}$. This verifies Condition B.1.(iv).

By definition, $\widetilde{R}_\zeta(h_{F,d})$ in (B.68) does not depend on d for any F. Moreover, for any d and $\widetilde{d}$ with $||d|| = \infty$ and $||\widetilde{d}|| = \infty$, by the definition of $R_\zeta(h_{F,d})$ in (B.69),

$$R_\zeta(h_{F,d}) = \mathbb{E}\left[\min\{\xi_{1,F}'\Upsilon\xi_{1,F},\zeta\}\right] = R_\zeta(h_{F,\widetilde{d}}). \tag{B.70}$$

Hence, Condition B.1.(vi) is also verified.

We next apply Lemma B.14 to get (B.60) and (B.61) above. By (B.68),

$$R_\zeta(h) - \widetilde{R}_\zeta(h) = \mathbb{E}[\min\{\overline{\xi}_F'\Upsilon\overline{\xi}_F,\zeta\}] - \mathbb{E}[\min\{\xi_{1,F}'\Upsilon\xi_{1,F},\zeta\}] \text{ for any } h \in H \tag{B.71}$$

and

$$R_\zeta(h) - \widetilde{R}_\zeta(h) = \mathbb{E}\left[\min\{\xi_{1,F}'\Upsilon\xi_{1,F},\zeta\}\right] - \mathbb{E}\left[\min\{\xi_{1,F}'\Upsilon\xi_{1,F},\zeta\}\right] = 0 \text{ for any } h \in H^*_\infty. \tag{B.72}$$

By Lemma B.14, (B.71) and (B.72), we have

$$\begin{aligned} Asy\overline{RD}(\widehat{\theta}_{eo},\widehat{\theta}_1) &= \lim_{\zeta\to\infty} \max\left\{\sup_{h\in H}\left[R_\zeta(h) - \widetilde{R}_\zeta(h)\right], \sup_{h\in H^*_\infty}\left[R_\zeta(h) - \widetilde{R}_\zeta(h)\right]\right\} \\ &= \lim_{\zeta\to\infty} \max\left\{\sup_{h\in H}\mathbb{E}\left[\min\{\overline{\xi}_F'\Upsilon\overline{\xi}_F,\zeta\} - \min\{\xi_{1,F}'\Upsilon\xi_{1,F},\zeta\}\right], 0\right\} \end{aligned} \tag{B.73}$$

and

$$\begin{aligned} Asy\underline{RD}(\widehat{\theta}_{eo},\widehat{\theta}_1) &= \lim_{\zeta\to\infty} \min\left\{\inf_{h\in H}\left[R_\zeta(h) - \widetilde{R}_\zeta(h)\right], \inf_{h\in H^*_\infty}\left[R_\zeta(h) - \widetilde{R}_\zeta(h)\right]\right\} \\ &= \lim_{\zeta\to\infty} \min\left\{\inf_{h\in H}\mathbb{E}\left[\min\{\overline{\xi}_F'\Upsilon\overline{\xi}_F,\zeta\} - \min\{\xi_{1,F}'\Upsilon\xi_{1,F},\zeta\}\right], 0\right\}, \end{aligned} \tag{B.74}$$

which proves (B.60) and (B.61).

In the second step, we show that

$$\lim_{\zeta\to\infty} \max\left\{\sup_{h\in H}\left[g_\zeta(h)\right],0\right\} = \max\left\{\sup_{h\in H}\left[g(h)\right],0\right\}, \text{ and} \tag{B.75}$$

$$\lim_{\zeta\to\infty} \min\left\{\inf_{h\in H}\left[g_\zeta(h)\right],0\right\} = \min\left\{\inf_{h\in H}\left[g(h)\right],0\right\}. \tag{B.76}$$

By Lemma B.16,

$$\lim_{\zeta\to\infty}\sup_{h\in H}\left[g_\zeta(h)\right] = \sup_{h\in H}\left[g(h)\right] \text{ and } \lim_{\zeta\to\infty}\inf_{h\in H}\left[g_\zeta(h)\right] = \inf_{h\in H}\left[g(h)\right], \tag{B.77}$$

where $\sup_{h\in H}\left[g(h)\right]$ and $\inf_{h\in H}\left[g(h)\right]$ are finite real numbers. Let $\overline{f}(x) = \max(x,0)$ and $\underline{f}(x) = \min(x,0)$. It is clear that $\overline{f}(x)$ and $\underline{f}(x)$ are continuos function on $\mathbb{R}$. The asserted results in (B.75) and (B.76) follow by (B.77), and the continuity of $\overline{f}(x)$ and $\underline{f}(x)$. ■

Proof of Theorem 5.2. For any $F\in\mathcal{F}$, define

$$B_F = (\Gamma_{2,F}-\Gamma^*_{1,F})'\Upsilon(\Gamma_{2,F}-\Gamma^*_{1,F}) \text{ and } D_F = (\Gamma_{2,F}-\Gamma^*_{1,F})'\Upsilon\Gamma^*_{1,F}. \tag{B.78}$$

Recall that we have defined $A_F = \Upsilon\left(\Sigma_{1,F}-\Sigma_{2,F}\right)$ in Theorem 5.2. By the definition of $\overline{\xi}_F$,

$$\mathbb{E}[\overline{\xi}'_F\Upsilon\overline{\xi}_F] = \text{tr}(\Upsilon\Sigma_{1,F}) + 2\text{tr}(A_F)J_{1,F} + \text{tr}(A_F)^2 J_{2,F} \tag{B.79}$$

where

$$J_{1,F} = \mathbb{E}\left[\frac{\mathcal{Z}'_{d,2,F}D_F\mathcal{Z}_{d,2,F}}{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\text{tr}(A_F)}\right] \text{ and } J_{2,F} = \mathbb{E}\left[\frac{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}}{(\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\text{tr}(A_F))^2}\right]. \tag{B.80}$$

We provide a upper bound for $J_{1,F}$ defined in (B.80). Define a function

$$\eta(x) \equiv \frac{x}{x'B_Fx+\text{tr}(A_F)} \text{ for any } x\in\mathbb{R}^{r_2}. \tag{B.81}$$

Its derivative is

$$\frac{\partial\eta(x)'}{\partial x} = \frac{1}{x'B_Fx+\text{tr}(A_F)}I_{r_2} - \frac{2B_F}{\left(x'B_Fx+\text{tr}(A_F)\right)^2}xx'. \tag{B.82}$$

Then $J_{1,F} = \mathbb{E}\left[\eta(\mathcal{Z}_{d,2,F})'D_F\mathcal{Z}_{d,2,F}\right]$. Note that $D_F\mathcal{Z}_{d,2,F} = D_F\mathcal{Z}_{2,F}$ by construction because the last r^* columns of $\Gamma_{1,F}^*$ are zeros. Applying Lemma B.9 yields

$$\begin{aligned}\operatorname{tr}(D_F\Omega_{2,F}) &= \operatorname{tr}\left((\Gamma_{2,F}-\Gamma_{1,F}^*)'\Upsilon\Gamma_{1,F}^*\Omega_{2,F}\right)\\ &= \operatorname{tr}(\Upsilon\left(\Gamma_{1,F}^*\Omega_{2,F}\Gamma_{2,F}'-\Gamma_{1,F}^*\Omega_{2,F}\Gamma_{1,F}^*\right))\\ &= \operatorname{tr}(\Upsilon(\Sigma_{2,F}-\Sigma_{1,F})) = -\operatorname{tr}(A_F). \end{aligned} \tag{B.83}$$

By Lemma 1 of Hansen (2016), which is a matrix version of the Stein's Lemma (Stein, 1981),

$$J_{1,F} = \mathbb{E}\left(\eta(\mathcal{Z}_{d,2,F})'D_F\mathcal{Z}_{d,2,F}\right) = \mathbb{E}\left[\operatorname{tr}\left(\frac{\partial\eta(\mathcal{Z}_{d,2,F})'}{\partial x}D_F\Omega_{2,F}\right)\right]. \tag{B.84}$$

Plugging (B.81)-(B.83) into (B.84), we have

$$\begin{aligned}J_{1,F} &= \mathbb{E}\left[\frac{\operatorname{tr}(D_F\Omega_{2,F})}{\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}+\operatorname{tr}(A_F)}\right] - 2\mathbb{E}\left[\frac{\operatorname{tr}\left(B_F\mathcal{Z}_{d,2,F}\mathcal{Z}_{d,2,F}'D_F\Omega_{2,F}\right)}{\left(\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}+\operatorname{tr}(A_F)\right)^2}\right]\\ &= \mathbb{E}\left[\frac{-\operatorname{tr}(A_F)}{\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}+\operatorname{tr}(A_F)}\right] + 2\mathbb{E}\left[\frac{-\mathcal{Z}_{d,2,F}'D_F\Omega_{2,F}B_F\mathcal{Z}_{d,2,F}}{\left(\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}+\operatorname{tr}(A_F)\right)^2}\right]\end{aligned} \tag{B.85}$$

where the second equality is by (B.83). By definition and Lemma B.9

$$\begin{aligned}&-\mathcal{Z}_{d,2,F}'D_F\Omega_{2,F}B_F\mathcal{Z}_{d,2,F}\\ =\;& -\mathcal{Z}_{d,2,F}'(\Gamma_{2,F}-\Gamma_{1,F}^*)'\Upsilon\Gamma_{1,F}^*\Omega_{2,F}(\Gamma_{2,F}-\Gamma_{1,F}^*)'\Upsilon(\Gamma_{2,F}-\Gamma_{1,F}^*)\mathcal{Z}_{d,2,F}\\ =\;& \mathcal{Z}_{d,2,F}'(\Gamma_{2,F}-\Gamma_{1,F}^*)'\Upsilon(\Sigma_{1,F}-\Sigma_{2,F})\Upsilon(\Gamma_{2,F}-\Gamma_{1,F}^*)\mathcal{Z}_{d,2,F}\\ \leq\;& \rho_{\max}(\Upsilon^{1/2}(\Sigma_{1,F}-\Sigma_{2,F})\Upsilon^{1/2})(\mathcal{Z}_{d,2,F}'(\Gamma_{2,F}-\Gamma_{1,F}^*)'\Upsilon(\Gamma_{2,F}-\Gamma_{1,F}^*)\mathcal{Z}_{d,2,F})\\ =\;& \rho_{\max}(A_F)\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F},\end{aligned} \tag{B.86}$$

where the last equality is by $\rho_{\max}(\Upsilon^{1/2}(\Sigma_{1,F}-\Sigma_{2,F})\Upsilon^{1/2}) = \rho_{\max}(\Upsilon(\Sigma_{1,F}-\Sigma_{2,F}))$. Combining the results in (B.85) and (B.86), we get

$$\begin{aligned}
J_{1,F} \leq & \mathbb{E}\left[\frac{-\mathrm{tr}(A_F)}{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)}\right] + 2\mathbb{E}\left[\frac{\rho_{\max}(A_F)\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}}{\left(\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)\right)^2}\right] \\
= & \mathbb{E}\left[\frac{-\mathrm{tr}(A_F)}{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)}\right] \\
& + 2\mathbb{E}\left[\frac{\left[\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A)\right]\rho_{\max}(A_F) - \mathrm{tr}(A_F)\rho_{\max}(A_F)}{\left(\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)\right)^2}\right] \\
= & \mathbb{E}\left[\frac{2\rho_{\max}(A_F)-\mathrm{tr}(A_F)}{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)}\right] - \mathbb{E}\left[\frac{2\rho_{\max}(A_F)\mathrm{tr}(A_F)}{\left(\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)\right)^2}\right]. \qquad \text{(B.87)}
\end{aligned}$$

Next, note that

$$\begin{aligned}
J_{2,F} = & \mathbb{E}\left[\frac{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}}{\left|\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)\right|^2}\right] \\
= & \mathbb{E}\left[\frac{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)-\mathrm{tr}(A_F)}{\left|\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)\right|^2}\right] \\
= & \mathbb{E}\left[\frac{1}{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)}\right] - \mathbb{E}\left[\frac{\mathrm{tr}(A_F)}{\left|\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)\right|^2}\right]. \qquad \text{(B.88)}
\end{aligned}$$

Combining (B.79), (B.87), (B.88) and the definition of $g(h)$ (in Theorem 5.1), we obtain that

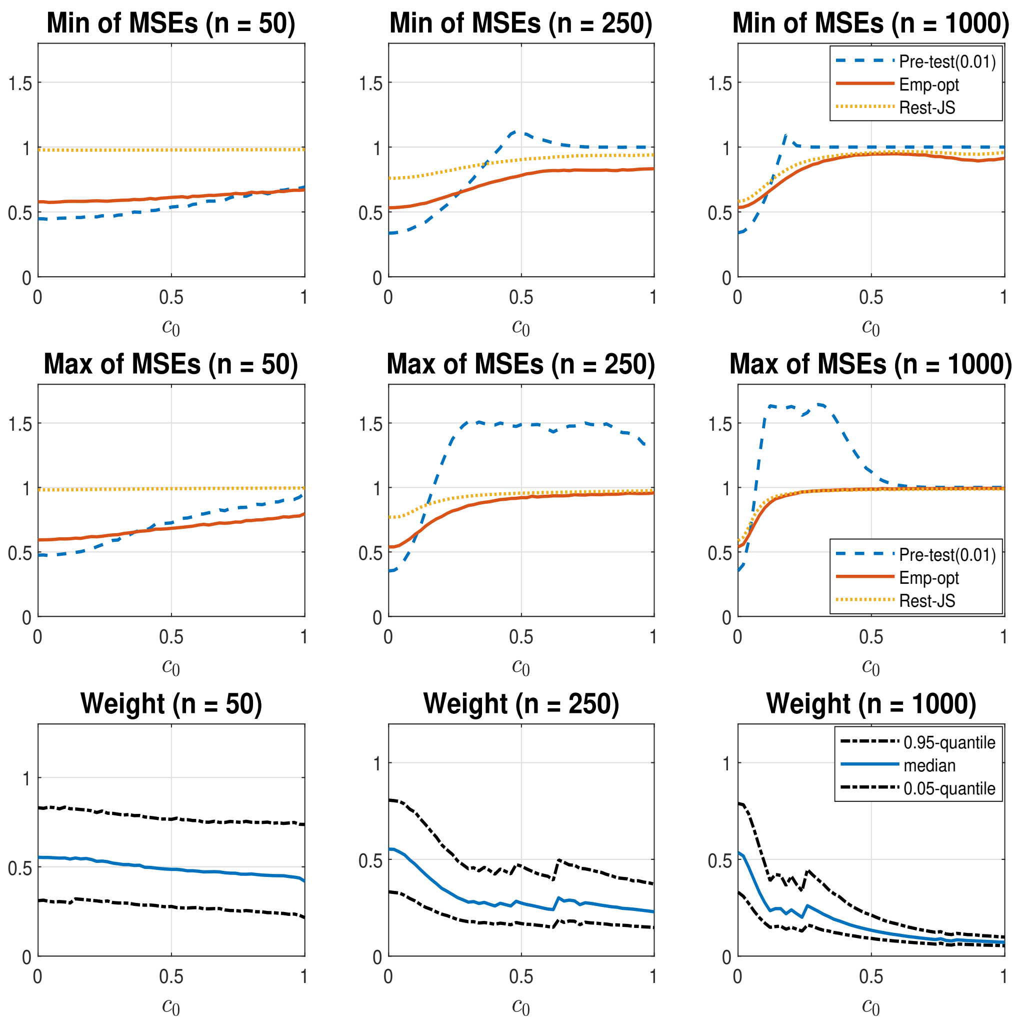
$$\begin{aligned}
g(h_{d,F}) = & 2\mathrm{tr}(A_F)J_{1,F} + \mathrm{tr}(A_F)^2 J_{2,F} \\
\leq & 2\mathrm{tr}(A_F)\left(\mathbb{E}\left[\frac{2\rho_{\max}(A_F)-\mathrm{tr}(A_F)}{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)}\right] - \mathbb{E}\left[\frac{2\mathrm{tr}(A_F)\rho_{\max}(A_F)}{\left|\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)\right|^2}\right]\right) \\
& + \mathrm{tr}(A)^2\left(\mathbb{E}\left[\frac{1}{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)}\right] - \mathbb{E}\left[\frac{\mathrm{tr}(A_F)}{\left|\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)\right|^2}\right]\right) \\
= & \mathbb{E}\left[\frac{\mathrm{tr}(A_F)\left(4\rho_{\max}(A_F)-\mathrm{tr}(A_F)\right)}{\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)}\right] - \mathbb{E}\left[\frac{\mathrm{tr}(A_F)^2\left(4\rho_{\max}(A_F)+\mathrm{tr}(A_F)\right)}{\left|\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)\right|^2}\right]. \qquad \text{(B.89)}
\end{aligned}$$

For all G_2 and Ω_2 such that $h = (d, \mathrm{vec}(G_2)', \mathrm{vech}(\Omega_2)') \in H$, we have $G_2 = G_{2,F}$ and $\Omega_2 =$

$\Omega_{2,F}$ for some $F \in \mathcal{F}$ by the definition of H. If $\text{tr}(A_F) > 0$, then $\rho_{\max}(A_F) > 0$ and thus the second term in the right-hand side of the last equality of (B.89) will be negative. If in addition $\text{tr}(A_F) \geq 4\rho_{\max}(A_F)$, then the first term in the right-hand side of the last equality of (B.89) will be non-negative. As a result, when $\text{tr}(A_F) > 0$ and $4\rho_{\max}(A_F) - \text{tr}(A_F) \leq 0$ for $\forall F \in \mathcal{F}$, we have $\sup_{h \in H}[g(h)] < 0$. This combined with Theorem 5.1 implies the results of this theorem. ■

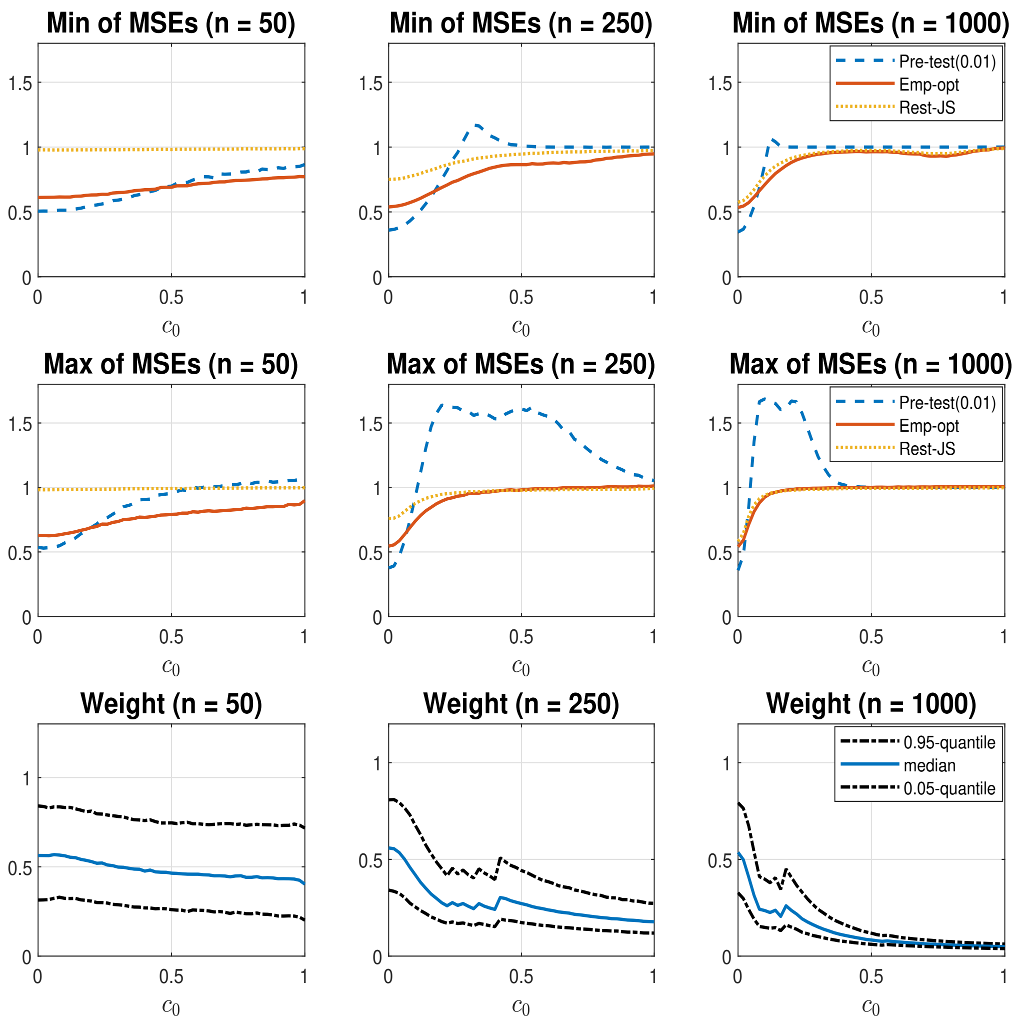
C Supplementary Simulation Results

Figure C.1: Finite Sample MSEs of the Pre-test and Averaging GMM Estimators in S1



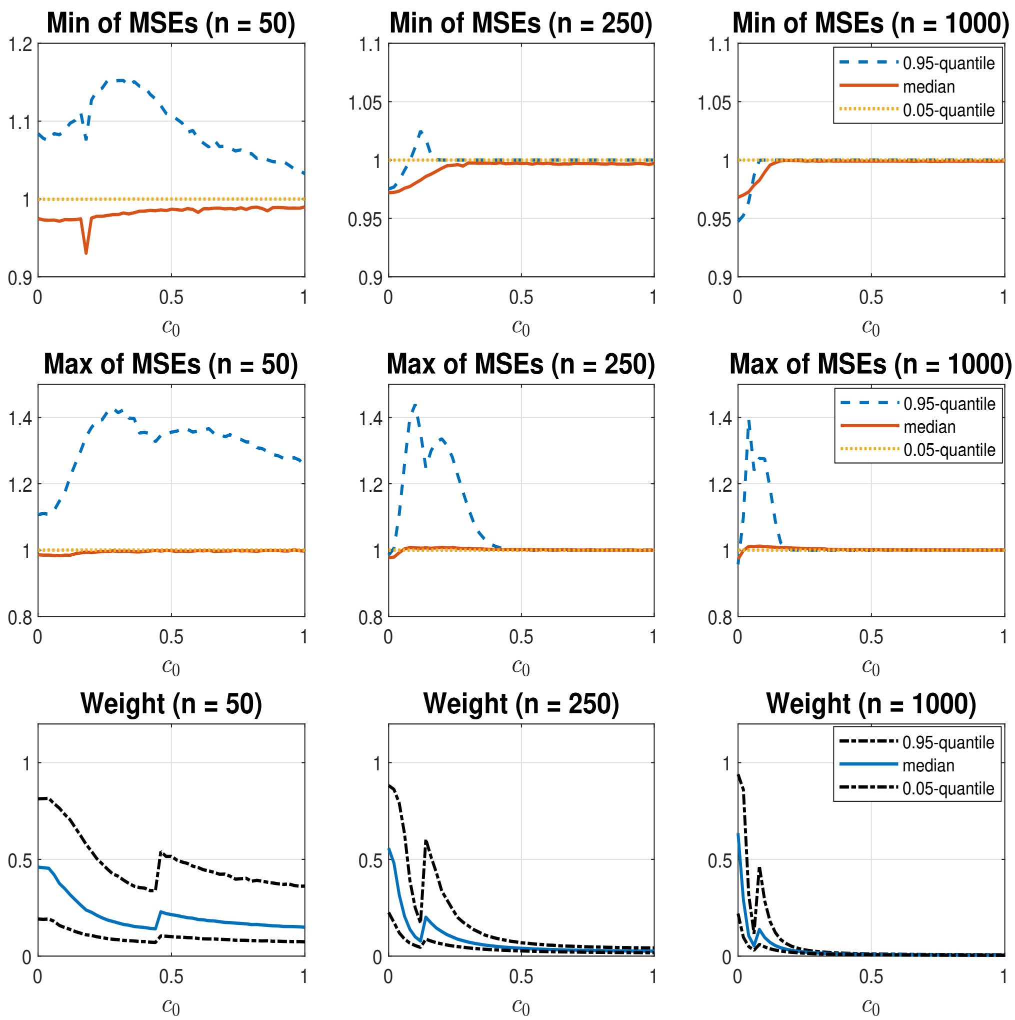
Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively.

Figure C.2: Finite Sample MSEs of the Pre-test and Averaging GMM Estimators in S2



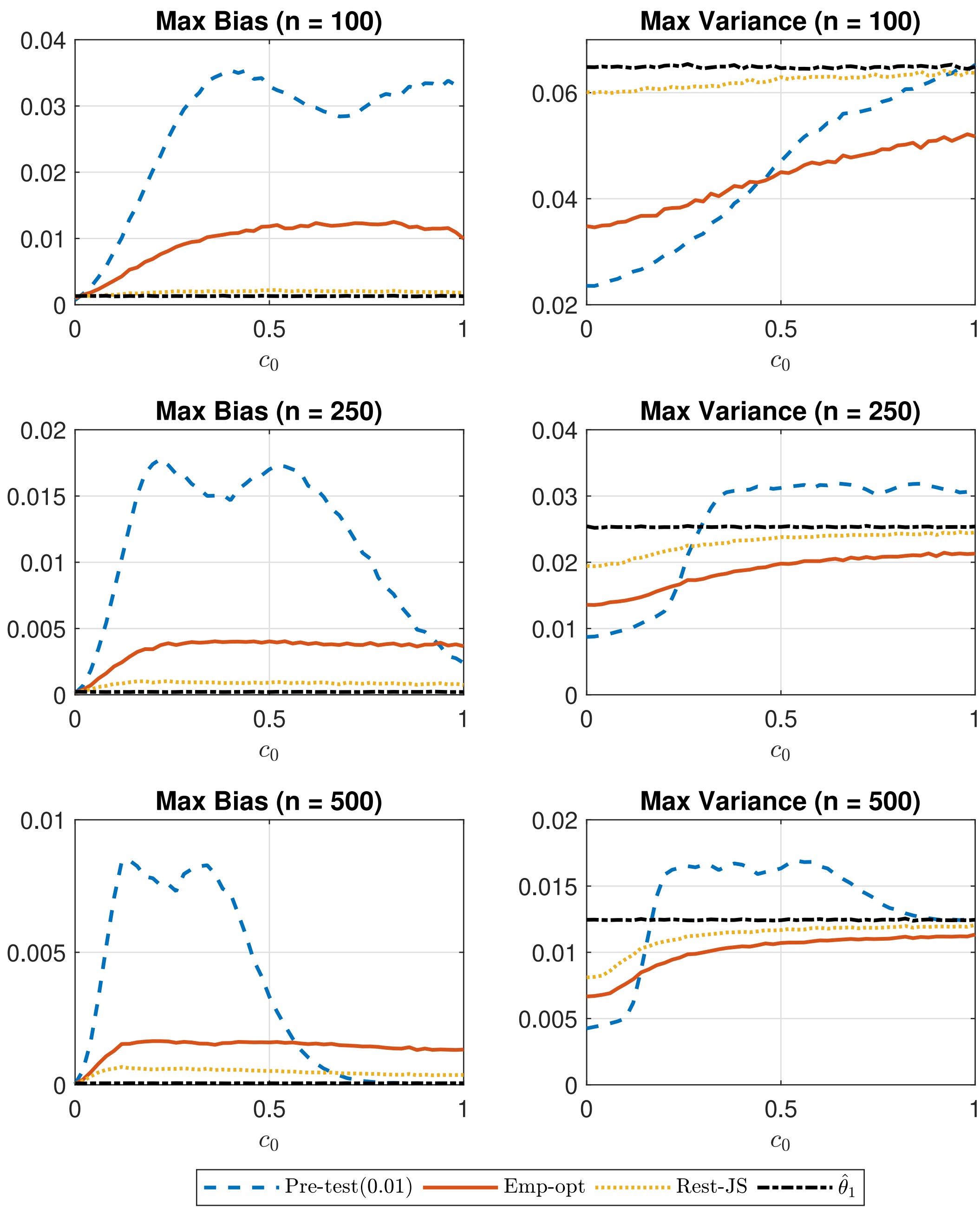
Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively.

Figure C.3: Finite Sample MSEs of the Pre-test and Averaging GMM Estimators in S3



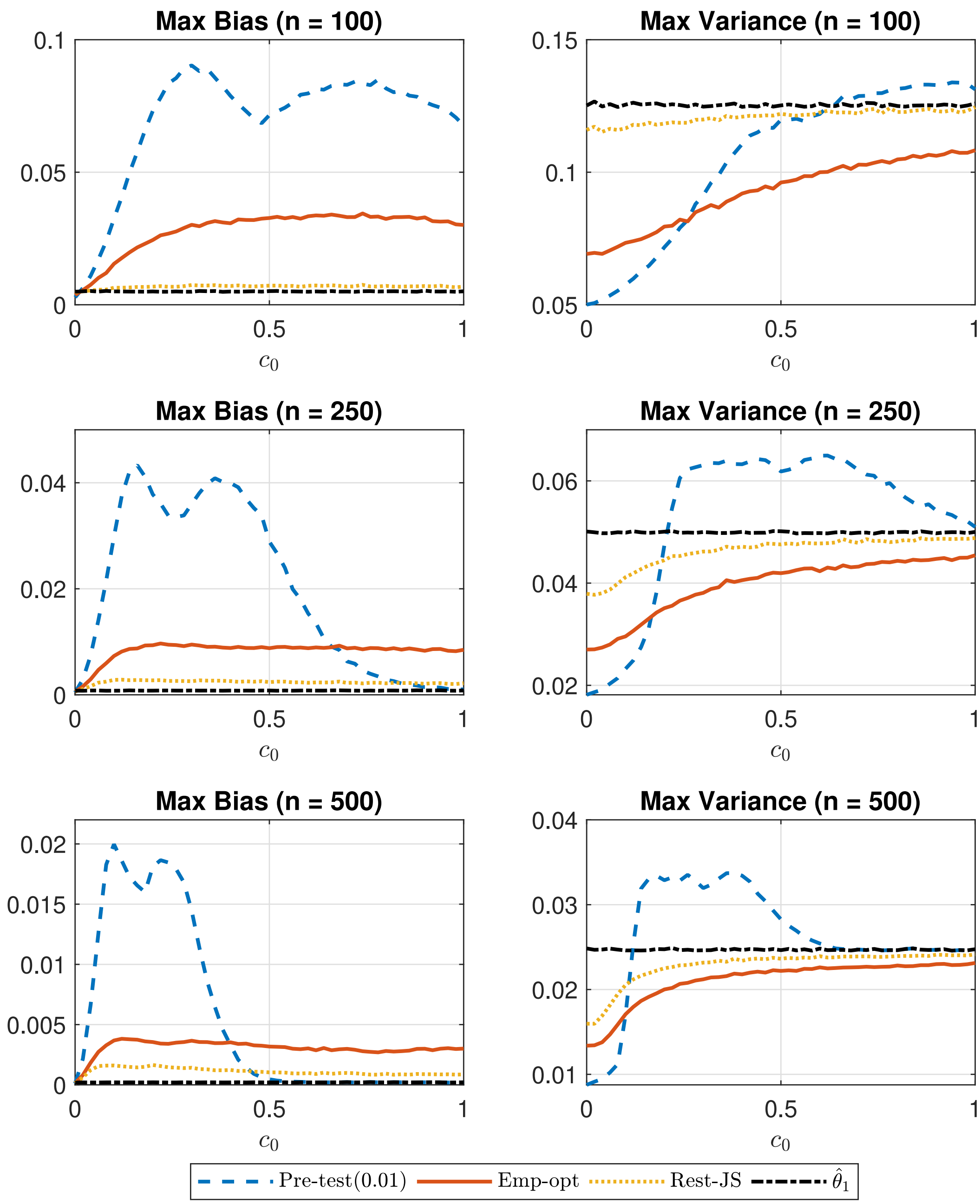
Note: “Pre-test(0.01)” refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; “Emp-opt” refers to the averaging GMM estimator based on the empirical optimal weight; “Rest-JS” refers to the averaging estimators based on the restricted James-Stein weight, respectively.

Figure C.4: Finite Sample Biases and Variances in S1



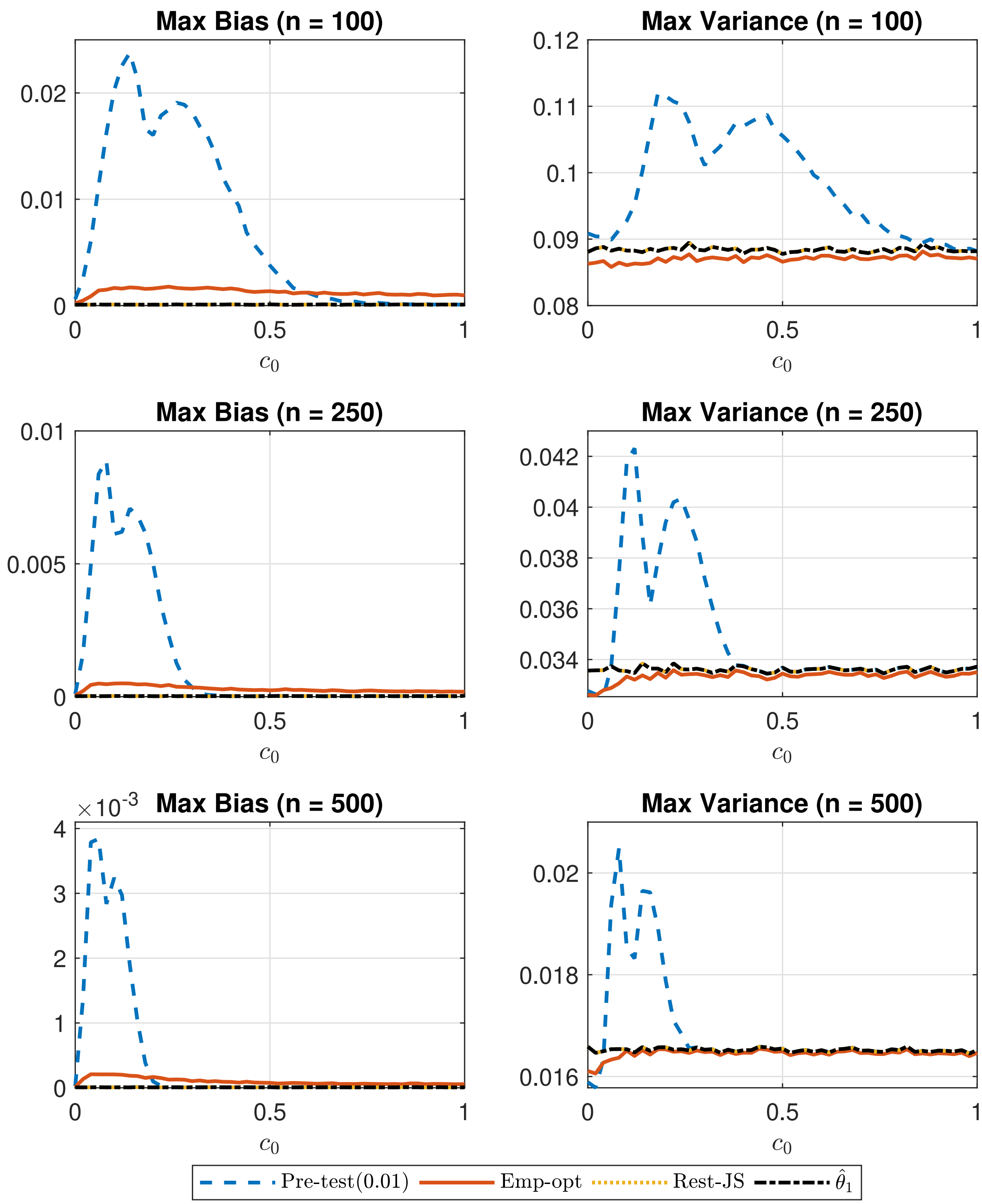
Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively.

Figure C.5: Finite Sample Biases and Variances in S2



Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively.

Figure C.6: Finite Sample Biases and Variances in S3



Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively.

References

[1] Andrews, D. W. K. (1999): "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 67(3), 543-563.

[2] Andrews, D. W. K. and X. Cheng (2013): "Maximum Likelihood Estimation and Uniform Inference with Sporadic Identification Failure," *Journal of Econometrics,* 173(1), 36-56.

[3] Andrews, D. W. K., X. Cheng and P. Guggenberger (2011): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," *Cowles Foundation Discussion Paper, No.1813.*

[4] Andrews, D. W. K. and P. Guggenberger (2010): "On the Asymptotic Maximal Risk of Estimators and the Hausman Pre-Test Estimator," *unpublished manuscript.*

[5] Andrews, D. W. K. and B. Lu (2001): "Consistent Model and Moment Selection Procedures for GMM Estimation with Application to Dynamic Panel Data Models," *Journal of Econometrics*, 101(1), 123-164.

[6] Ashley, R. (2009): "Assessing the Credibility of Instrumental Variables Inference with Imperfect Instruments via Sensitivity Analysis," *Journal of Applied Econometrics*, 24, 325–337.

[7] Berkowitz, D., M. Caner and Y. Fang (2012): "The Validity of Instruments Revisited," *Journal of Econometrics*, 166(2), 255-266.

[8] Buckland, S. T., K. P. Burnham, and N. H. Augustin (1997): "Model Selection: An Integral Part of Inference," *Biometrics*, 53, 603-618.

[9] Charkhi, A., G. Claeskens and Hansen, B. (2016): "Minimum Mean Squared Error Model Averaging in Likelihood Models," *Statistica Sinica*, Vol. 26, 809-840.

[10] Cheng, X. and B. Hansen (2015): "Forecasting with Factor-Augmented Regression: A Frequentist Model Averaging Approach," *Journal of Econometrics*, 186(2), 280-293.

[11] Cheng, X. and Z. Liao (2015): "Select the Valid and Relevant Moments: An Information-Based LASSO for GMM with Many Moments," *Journal of Econometrics*, 186(2), 443-464.

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

[13] Conley T. G., C. B. Hansen, and P. E. Rossi (2012): "Plausibly Exogenous," *Review of Economics and Statistics*, 94(1), 260-272.

[14] DiTraglia, F. (2016): "Using Invalid Instruments on Purpose: Focused Moment Selection and Averaging for GMM," *Journal of Econometrics,* 195(2), 187-208.

[15] Doko Tchatoka, F. and J.-M. Dufour (2008): "Instrument Endogeneity and Identification-robust Tests: Some Analytical Results," *Journal of Statistical Planning and Inference*, 138(9), 2649–2661.

[16] Doko Tchatoka, F. and J.-M. Dufour (2014): "Identification-robust Inference for Endogeneity Parameters in Linear Structural Models," *Econometrics Journal*, 17, 165–187.

[17] Eichenbaum, M., L. Hansen and K. Singleton (1988): "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," *Quarterly Journal of Economics*, 103(1), 51-78.

[18] Guggenberger, P. (2012): "On the Asymptotic Size Distortion of Tests when Instruments Locally Violate the Exogeneity Assumption," *Econometric Theory*, 28(2), 387-421.

[19] Green, E. J. and W. E. Strawderman (1991): "A James-Stein Type Estimator for Combining Unbiased and Possibly Biased Estimators," *Journal of the American Statistical Association*, 86(416), 1001-1006.

[20] Hansen, B. E. (2007): "Least Squares Model Averaging," *Econometrica*, 75, 1175-1189.

[21] Hansen, B. E. (2015): "Shrinkage Efficiency Bounds," *Econometric Theory*, 31, 860-879.

[22] Hansen, B. E. (2016): "Efficient Shrinkage in Parametric Models," *Journal of Econometrics*, 190(1), 115–132.

[23] Hansen, B. E. (2017): "A Stein-Like 2SLS Estimator," *Econometric Reviews*, 36, 840-852.

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

[25] Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

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

[27] Hjort, N. L. and G. Claeskens (2006): "Focused Information Criteria and Model Averaging for the Cox Hazard Regression Model," *Journal of the American Statistical Association*, 101, 1449-1464.

[28] Hong, H., B. Preston and M. Shum (2003): "Generalized Empirical Likelihood-Based Model Selection Criteria for Moment Condition Models," *Econometric Theory*, 19(6), 923-943.

[29] Judge G. C. and R. C. Mittelhammer (2004): "A Semiparametric Basis for Combining Estimating Problems under Quadratic Loss," *Journal of American Statistical Association*, 99, 479-487.

[30] Judge G. C. and R. C. Mittelhammer (2007): "Estimation and Inference in the Case of Competing Sets of Estimating Equations," *Journal of Econometrics*, 138, 513–531.

[31] James W. and C. Stein (1961): "Estimation with Quadratic Loss," *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, vol(1), 361-380.

[32] Kabaila, P. (1995): "The Effect of Model Selection on Confidence Regions and Prediction Regions," *Econometric Theory*, 11, 537–549.

[33] Kabaila, P. (1998): "Valid Confidence Intervals in Regression After Variable Selection," *Econometric Theory*, 14, 463–482.

[34] Kang, H., A. Zhang, T. T. Cai and Small, D. S. (2016): "Instrumental Variables Estimation with Some Invalid Instruments and Its Application to Mendelian Randomization," *Journal of the American Statistical Association*, Vol. 111, 132-144.

[35] Kim, T.H., and White, H. (2001): "James–Stein Type Estimators in Large Samples with Application to the Least Absolute Deviation Estimator," *Journal of the American Statistical Association*, 96, 697–705.

[36] Kolesar, M., R. Chetty, J. Friedman, E. Glaeser and G. Imbens (2015): "Identification and Inference with Many Invalid Instruments," *Journal of Business Economics and Statistics,* 33(4) 474-484.

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

[38] Leeb, H. and B. M. Pötscher (2008): "Sparse Estimators and the Oracle Property, or the Return of the Hodge's Estimator," *Journal of Econometrics*, 142(1), 201-211.

[39] Lehmann, E. L. and G. Casella (1998): *Theory of Point Estimation*, Springer-Verlag New York, Inc.

[40] Leung, G. and A. Barron (2006): "Information Theory and Mixing Least-square Regressions," *IEEE Transactions on Information Theory*, 52, 3396–3410.

[41] Liao, Z. (2013): "Adaptive GMM Shrinkage Estimation with Consistent Moment Selection," *Econometric Theory*, 29, 1–48.

[42] Liu, C-A. (2015): "Distribution Theory of the Least Squares Averaging Estimator," *Journal of Econometrics,* 186(1), 142–159.

[43] Lu, X. and L. Su (2015): "Jackknife Model Averaging for Quantile Regressions," *Journal of Econometrics*, forthcoming.

[44] Mittelhammer R. C., and G. C. Judge (2005): "Combining Estimators to Improve Structural Model Estimation and Inference under Quadratic Loss," *Journal of Econometrics*, 128, 1–29.

[45] Moon. H. R. and F. Schorfheide (2009): "Estimation with Overidentifying Inequality Moment Conditions," *Journal of Econometrics,* 153(2), 136–154.

[46] Nevo, A. and A. Rosen (2012): "Identification with Imperfect Instruments," *Review of Economics and Statistics*, 93(3), 659-671.

[47] Pötscher, B. M. (1991): "Effects of Model Selection on Inference," *Econometric Theory*, 7 163–185.

[48] Pötscher, B. M. (2006): "The Distribution of Model Averaging Estimators and an Impossibility Result Regarding Its Estimation," in: H.-C. Ho, C.-K. Ing and T.-L. Lai (eds.), Time Series and Related Topics: In Memory of Ching-Zong Wei. *IMS Lecture Notes and Monograph Series*, Vol. 52., 2006, 113-129.

[49] Sargan, J. (1958): "The estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26(3), 393-415.

[50] Shibata, R. (1986): "Consistency of Model Selection and Parameter Estimation," *Journal of Applied Probability*, special volume 23A, 127–141.

[51] Small, D. S. (2007): "Sensitivity Analysis for Instrumental Variables Regression with Overidentifying Restrictions," *Journal of the American Statistical Association*, 102(479), 1049–1058.

[52] Small, D. S., Cai, T. T., Zhang, A. and Kang, H. (2015): "Instrumental Variables Estimation with Some Invalid Instruments and Its Application to Mendelian Randomization," *Journal of the American Statistical Association*, forthcoming.

[53] Stein, C. M. (1956): "Inadmissibility of the Usual Estimator for the Mean of A Multivariate Normal Distribution," *Proc. Third Berkeley Symp. Math. Statist. Probab.*, vol(1), 197-206.

[54] Van der Vaart, A. W. (1998): *Asymptotic Statistics*, Cambridge University Press.

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

[56] Yang, Y. (2003): "Regression with Multiple Candidate Models: Selecting or Mixing?" *Statistica Sinica*, 13, 783–809.

[57] Yang, Y. (2004): "Combining Forecasting Procedures: Some Theoretical Results," *Econometric Theory*, 20, 176–222.

[58] Yang, Y. (2005): "Can the Strengths of AIC and BIC be Shared?–A Conflict Between Model Identification and Regression Estimation," *Biometrika*, 92, 937–950.

Supplemental Appendix of
"An Averaging GMM Estimator Robust to Misspecification"

Xu Cheng, Zhipeng Liao, Ruoyao Shi

In this supplemental appendix, we present supporting materials for Cheng, Liao and Shi (2018) (cited as CLS hereafter in this Appendix):

- Section D provides primitive conditions for Assumptions 3.1, 3.2 and 3.3 and the proof of Lemma 3.1 of CLS.
- Section E provides the proof of (4.3) in Section 4 and the proof of some Lemmas in Appendix B.1 of CLS. The proof of Lemma A.1 in Appendix A of CLS is also included in this section.
- Section F studies the bounds of asymptotic risk difference of the pre-test GMM estimator.
- Section G contains simulation results under the truncated risk for the simulation designs in Section 6 of CLS.
- Section H includes extra simulation studies.

D Primitive Conditions for Assumptions 3.1, 3.2 and 3.3 and Proof of Lemma 3.1 of CLS

In this section, we provide primitive conditions for Assumption Assumptions 3.1, 3.2 and 3.3 in the linear IV model presented in Example 3.1 of CLS.

We first provide a set of sufficient conditions without imposing the normal distribution assumption on $(X', Z_1', V', U)'$ in Lemma D.1. Then, we impose the normality assumptions and show that these conditions can be simplified to those in Lemma 3.1 of CLS under normality.

For ease of notations, we define $\Gamma_{z_1vu^2} \equiv \mathbb{E}_{F^*}[Z_1V'U^2]$, $\Omega_{z_1z_1u^2} \equiv \mathbb{E}_{F^*}[Z_1Z_1'U^2]$ and $\Omega_{vvu^2} \equiv \mathbb{E}_{F^*}[VV'U^2]$. The Jacobian matrices are

$$G_{1,F} = -\mathbb{E}_F[Z_1X'] \text{ and } G_{2,F} = \begin{pmatrix} -\mathbb{E}_F[Z_1X'] \\ -\mathbb{E}_F[Z^*X'] \end{pmatrix}. \tag{D.1}$$

Let $Z_2 = (Z_1', Z^{*\prime})'$. The variance-covariance matrix of the moment conditions is

$$\Omega_{2,F} = \mathbb{E}_F[Z_2Z_2'(Y - X'\theta_0)^2] - \mathbb{E}_F[(Y - X'\theta_0)Z_2]\mathbb{E}_F[(Y - X'\theta_0)Z_2']. \tag{D.2}$$

By definition, $\Omega_{1,F}$ is the leading $r_1 \times r_1$ submatrix of $\Omega_{2,F}$.

Let F denote the joint distribution of $W = (Y, Z_1', Z^{*\prime}, X')'$ induced by θ_0, δ_0 and F^*. By definition, we can write

$$\delta_F = \Omega_{uu}\delta_0,\; G_{2,F} = \begin{pmatrix} -\Gamma_{z_1x} \\ -\delta_0\Gamma_{ux} - \Gamma_{vx} \end{pmatrix},\; \Omega_{2,F} = \begin{pmatrix} \Omega_{z_1z_1u^2} & \Omega_{2,1r,F} \\ \Omega_{2,r1,F} & \Omega_{2,rr,F} \end{pmatrix} \tag{D.3}$$

where

$$\begin{aligned} \Omega_{2,1r,F} &= \Gamma_{z_1u^3}\delta_0' + \Gamma_{z_1vu^2} = \Omega_{2,r1,F}', \text{ and} \\ \Omega_{2,rr,F} &= \Omega_{u^2u^2}\delta_0\delta_0' + \delta_0\Gamma_{u^3v} + \Gamma_{vu^3}\delta_0' + \Omega_{vvu^2}. \end{aligned} \tag{D.4}$$

Therefore, the parameter v_F defined in (3.4) depends on F through F^* and δ_0, and its dependence on F^* is through v_{*,F^*}, where

$$v_{*,F^*} = \begin{pmatrix} \Omega_{u^2u^2}, \Omega_{uu}, \text{vec}(\Gamma_{z_1x})', \text{vec}(\Gamma_{ux})', \text{vec}(\Gamma_{vx})', \text{vech}(\Omega_{z_1z_1u^2})', \\ \text{vec}(\Gamma_{z_1u^3})', \text{vec}(\Gamma_{z_1vu^2})', \text{vec}(\Gamma_{u^3v})', \text{vech}(\Omega_{vvu^2})' \end{pmatrix}. \tag{D.5}$$

Define

$$\begin{aligned} \rho_{2,\max} &\equiv \max\{\sup_{F\in\mathcal{F}} \rho_{\max}(\Omega_{2,F}), \sup_{F\in\mathcal{F}} \rho_{\max}(G_{2,F}G_{2,F}')\}, \\ \rho_{2,\min} &\equiv \min\{\inf_{F\in\mathcal{F}} \rho_{\min}(\Omega_{2,F}), \inf_{F\in\mathcal{F}} \rho_{\min}(G_{2,F}G_{2,F}')\}, \\ C_W &\equiv \sup_{F^*\in\mathcal{F}^*} \mathbb{E}_{F^*}[||(X', Z_1', V', U)||^2] \text{ and } C_\Delta \equiv \sup_{\delta_0\in\Delta_\delta} \|\delta_0\|^2. \end{aligned} \tag{D.6}$$

In the proof of Lemma D.1 below, we show that $\rho_{2,\max} < \infty$ (see (*D.14*) and (*D.18*)). Moreover, we have $\rho_{2,\min} > 0$, $C_W < \infty$ and $C_\Delta < \infty$ by Assumptions *D.1.*(iii), *D.1.*(ii) and *D.1.*(vii) respectively. Define

$$B_{\rho_2}^c \equiv \{\delta \in \mathbb{R}^{r^*} : \|\delta\| \geq \rho_{2,\min}\rho_{2,\max}^{-1}C_\Delta^{-1/2}\}. \tag{D.7}$$

Let Θ_0 be a non-empty set in $\mathbb{R}^{d_\theta}$. Define

$$B_{\Theta_0} \equiv \{\theta \in \mathbb{R}^{d_\theta} : \|\theta - \theta_0\| \leq \rho_{2,\min}^{-4}\rho_{2,\max}^3 C_\Delta C_W^2 \text{ for any } \theta_0 \in \Theta_0\}. \tag{D.8}$$

Let $\{c_{j,\Delta}, C_{j,\Delta}\}_{j=1}^{r^*}$ be a set of finite constants. We next provide the low-level sufficient conditions for Assumptions *3.1, 3.2* and *3.3*.

Assumption D.1 *The following conditions hold:*

(i) $\mathbb{E}_{F^*}[V] = 0$, $\mathbb{E}_{F^*}[U] = 0$, $\mathbb{E}_{F^*}[Z_1U] = 0_{r_1\times 1}$ *and* $\mathbb{E}_{F^*}[VU] = 0_{r^*\times 1}$ *for any* $F^* \in \mathcal{F}^*$*;*

(ii) $\sup_{F^*\in\mathcal{F}^*} \mathbb{E}_{F^*}[||X||^{4+\gamma} + ||Z_1||^{4+\gamma} + ||V||^{4+\gamma} + U^6] < \infty$ *for some* $\gamma > 0$*;*

(iii) $\inf_{F^*\in\mathcal{F}^*} \mathbb{E}_{F^*}[U^2] > 0$, $\inf_{F^*\in\mathcal{F}^*} \rho_{\min}(\Gamma_{xz_1}\Gamma_{z_1x}) > 0$ *and* $\inf_{F\in\mathcal{F}} \rho_{\min}(\Omega_{2,F}) > 0$*;*

(iv) $\inf_{F^*\in\mathcal{F}^*}\inf_{\delta\in B^c_{\rho_2}}\|\delta\|^{-1}\,||(\Gamma_{xz_1}\Omega^{-1}_{z_1z_1u^2}\Gamma_{z_1vu^2}-\Gamma_{xv})\delta+\Gamma_{xz_1}\Omega^{-1}_{z_1z_1u^2}\Gamma_{z_1u^3}-\Gamma_{xu}||>0$;

(v) *the set* $\{v_{*,F^*}:F^*\in\mathcal{F}^*\}$ *is closed;*

(vi) $\theta_0\in\Theta_0$, $B_{\Theta_0}\subset int(\Theta)$ *and* Θ *is compact;*

(vii) $\Delta_\delta=[c_{1,\Delta},C_{1,\Delta}]\times\cdots\times[c_{r^*,\Delta},C_{r^*,\Delta}]$ *where* $c_{j,\Delta}<0<C_{j,\Delta}$ *for* $j=1,\ldots,r^*$.

Lemma D.1 *Suppose that* $\{W_i\}_{i=1}^n$ *are i.i.d. and generated by the linear model (3.6) and (3.8) in CLS. Then under Assumption* D.1, $\mathcal{F}$ *satisfies Assumptions* 3.1, 3.2 *and* 3.3.

For the linear IV model, Lemma D.1 provides simple conditions on θ_0, δ_0 and $\mathcal{F}^*$ on which uniformity results are subsequently established.

Proof of Lemma D.1. By Assumption *D.1.*(i) and the definition of $G_{1,F}$,

$$\mathbb{E}_F\left[g_1(W,\theta)\right]=\mathbb{E}_{F^*}[Z_1(U-X'(\theta-\theta_0))]=G_{1,F}(\theta-\theta_0), \tag{D.9}$$

which together with Assumption *D.1.*(iii) implies that $\theta_F=\theta_0$ and hence $\mathbb{E}_F\left[g_1(W,\theta_F)\right]=0_{r_1\times1}$. Also $\theta_F\in int(\Theta)$ holds by $\theta_F=\theta_0$ and Assumption *D.1.*(vi). This verifies Assumption 3.1.(i).

By (D.9) for any $\theta\in\Theta$ with $||\theta-\theta_F||\geq\varepsilon$ and any $F\in\mathcal{F}$

$$\|\mathbb{E}_F\left[g_1(W,\theta)\right]\|\geq\rho^{1/2}_{\min}(G'_{1,F}G_{1,F})\,\|\theta_F-\theta\|\geq\varepsilon\rho^{1/2}_{\min}(G'_{1,F}G_{1,F}) \tag{D.10}$$

which combined with Assumption *D.1.*(iii) and $G_{1,F}=-\Gamma'_{xz_1,F^*}$ implies that

$$\inf_{F\in\mathcal{F}}\inf_{\theta\in B^c_\varepsilon(\theta_F)}||\mathbb{E}_F\left[g_1(W,\theta)\right]||>0. \tag{D.11}$$

This verifies Assumption 3.1.(ii).

Next, we show Assumption 3.1.(iii). Let $Z_2\equiv(Z_1',Z^{*\prime})'$. By the Lyapunov inequality, Assumptions *D.1.*(i)-(ii) and *D.1.*(vii),

$$\begin{aligned}\sup_{F\in\mathcal{F}}\mathbb{E}_F[||Z_2||^2] \quad\leq\quad &\sup_{F^*\in\mathcal{F}^*}\mathbb{E}_{F^*}[||Z_1||^2]+2\sup_{F^*\in\mathcal{F}^*}\mathbb{E}_{F^*}[||V||^2]\\ &+2\sup_{\delta_0\in\Delta_\delta}\|\delta_0\|^2\sup_{F^*\in\mathcal{F}^*}\mathbb{E}_{F^*}[U^2]<\infty.\end{aligned} \tag{D.12}$$

By (*D.12*), the Hölder inequality, the Lyapunov inequality and Assumption *D.1.*(ii),

$$\sup_{F\in\mathcal{F}}\|G_{2,F}\|=\sup_{F\in\mathcal{F}}\left\|\mathbb{E}_F[Z_2X']\right\|\leq\sup_{F\in\mathcal{F}}(\mathbb{E}_F[||Z_2||^2])^{1/2}\sup_{F^*\in\mathcal{F}^*}(\mathbb{E}_{F^*}[||X||^2])^{1/2}<\infty, \tag{D.13}$$

which together with the definition of $G_{2,F}$ and the Cauchy-Schwarz inequality implies that

$$\sup_{F\in\mathcal{F}}\left\|G'_{2,F}G_{2,F}\right\|<\infty. \tag{D.14}$$

Similarly by the Cauchy-Schwarz inequality, the Lyapunov inequality, Assumptions *D.1.*(ii) and *D.1.*(vii), we have

$$\begin{aligned}
\sup_{F\in\mathcal{F}} \mathbb{E}_F[||Z_2||^4] &= \sup_{F\in\mathcal{F}} \mathbb{E}_F[(||Z_1||^2 + ||Z^*||^2)^2] \\
&\leq 2 \sup_{F^*\in\mathcal{F}^*} \mathbb{E}_{F^*}[||Z_1||^4] + 2 \sup_{F\in\mathcal{F}} \mathbb{E}_F[||Z^*||^4] \\
&\leq 2 \sup_{F^*\in\mathcal{F}^*} \mathbb{E}_{F^*}[||Z_1||^4] + 8 \sup_{F^*\in\mathcal{F}^*} \mathbb{E}_{F^*}[||V||^4] \\
&\quad +8 \sup_{\delta_0\in\Delta_\delta} \|\delta_0\|^4 \sup_{F^*\in\mathcal{F}^*} \mathbb{E}_{F^*}[U^4] < \infty. \qquad \text{(D.15)}
\end{aligned}$$

By (*D.12*), (*D.15*), Assumption *D.1.*(ii), the Lyapunov inequality and the Hölder inequality, we have

$$\begin{aligned}
& \sup_{F\in\mathcal{F}} \left\|\mathbb{E}_F[Z_2 Z_2'(Y - X'\theta_0)^2]\right\| \\
\leq\ & \sup_{F\in\mathcal{F}} \mathbb{E}_F[||Z_2||^2(Y - X'\theta_0)^2] \\
\leq\ & \sup_{F\in\mathcal{F}} (\mathbb{E}_F[||Z_2||^4])^{1/2} \sup_{F^*\in\mathcal{F}^*} (\mathbb{E}_{F^*}[U^4])^{1/2} < \infty, \qquad \text{(D.16)}
\end{aligned}$$

and

$$\sup_{F\in\mathcal{F}} \left\|\mathbb{E}_F[(Y - X'\theta_0)Z_2]\right\| \leq \sup_{F\in\mathcal{F}} (\mathbb{E}_F[||Z_2||^2])^{1/2} \sup_{F^*\in\mathcal{F}^*} (\mathbb{E}_{F^*}[U^2])^{1/2} < \infty. \qquad \text{(D.17)}$$

By the definition of $\Omega_{2,F}$, the triangle inequality, the Cauchy-Schwarz inequality and the results in (D.16) and (D.17),

$$\sup_{F\in\mathcal{F}} \|\Omega_{2,F}\| < \infty. \qquad \text{(D.18)}$$

We then show that $\theta_F^* \in int(\Theta)$. By the triangle inequality, the Cauchy-Schwarz inequality and the Hölder inequality,

$$\begin{aligned}
\|G_{2,F}\| &\leq \|\Gamma_{xz_1}\| + \|\delta_0\| \, \|\Gamma_{xu}\| + \|\Gamma_{xv}\| \\
&\leq (\mathbb{E}_{F^*}[||X||^2])^{1/2}(\mathbb{E}_{F^*}[||Z_1||^2])^{1/2} \\
&\quad + \|\delta_0\| \, (\mathbb{E}_{F^*}[||X||^2])^{1/2}(\mathbb{E}_{F^*}[U^2])^{1/2} \\
&\quad +(\mathbb{E}_{F^*}[||X||^2])^{1/2}(\mathbb{E}_{F^*}[||V||^2])^{1/2} \\
&\leq C_W(2 + C_\Delta^{1/2}), \qquad \text{(D.19)}
\end{aligned}$$

for any $F \in \mathcal{F}$, where $C_W < \infty$ by Assumptions *D.1.*(ii) and (vii). Since $G_{2,F}' = (G_{1,F}', G_{r^*,F}')$ where $G_{r^*,F} = -\delta_0\mathbb{E}_{F^*}[UX'] - \mathbb{E}_{F^*}[VX']$, we have

$$G_{2,F}'G_{2,F} = G_{1,F}'G_{1,F} + G_{r^*,F}'G_{r^*,F}, \qquad \text{(D.20)}$$

which implies that for any $F \in \mathcal{F}$,

$$\rho_{\min}(G'_{2,F}G_{2,F}) \geq \rho_{\min}(G'_{1,F}G_{1,F}). \tag{D.21}$$

To show Assumption 3.1.(iii), we write

$$\begin{aligned} Q_F(\theta) &= \mathbb{E}_F[Z_2(Y - X'\theta)]'\Omega_{2,F}^{-1}\mathbb{E}_F[Z_2(Y - X'\theta)] \\ &= \theta' G'_{2,F}\Omega_{2,F}^{-1}G_{2,F}\theta + 2\theta' G'_{2,F}\Omega_{2,F}^{-1}C_F + C'_F\Omega_{2,F}^{-1}C_F, \end{aligned} \tag{D.22}$$

where $C_F = \mathbb{E}_F[Z_2Y]$. Since $G'_{2,F}\Omega_{2,F}^{-1}G_{2,F}$ is non-singular by (D.18), (D.21) and Assumption D.1.(iii), $Q_F(\theta)$ is minimized at $\theta_F^* = -(G'_{2,F}\Omega_{2,F}^{-1}G_{2,F})^{-1}G'_{2,F}\Omega_{2,F}^{-1}C_F$ for any $F \in \mathcal{F}$. Therefore,

$$\begin{aligned} \|\theta_F^* - \theta_0\|^2 &= \left\|(G'_{2,F}\Omega_{2,F}^{-1}G_{2,F})^{-1}G'_{2,F}\Omega_{2,F}^{-1}\mathbb{E}_F[Z_2U]\right\|^2 \\ &\leq \frac{\rho_{\max}^2(\Omega_{2,F})}{\rho_{\min}^2(G'_{2,F}G_{2,F})}\mathbb{E}_F\left[UZ'_2\right]\Omega_{2,F}^{-1}G_{2,F}G'_{2,F}\Omega_{2,F}^{-1}\mathbb{E}_F[Z_2U] \\ &\leq \frac{\rho_{\max}^2(\Omega_{2,F})\rho_{\max}(G'_{2,F}G_{2,F})\Gamma_{uu}^2}{\rho_{\min}^2(\Omega_{2,F})\rho_{\min}^2(G'_{2,F}G_{2,F})}\|\delta_0\|^2 \\ &\leq \rho_{2,\min}^{-4}\rho_{2,\max}^3 C_\Delta C_W^2 \end{aligned} \tag{D.23}$$

for any $F \in \mathcal{F}$. By Assumption D.1.(vi), $\theta_F^* \in int(\Theta)$. Moreover for any $\theta \in \Theta$ with $||\theta - \theta_F^*|| \geq \varepsilon$,

$$\begin{aligned} Q_F(\theta) - Q_F(\theta_F^*) &\geq \rho_{\min}(G'_{2,F}\Omega_{2,F}^{-1}G_{2,F})\|\theta - \theta_F^*\|^2 \\ &\geq \varepsilon^2\rho_{\min}(G'_{2,F}\Omega_{2,F}^{-1}G_{2,F}) \\ &\geq \varepsilon^2\rho_{\max}^{-1}(\Omega_{2,F})\rho_{\min}(G'_{2,F}G_{2,F}), \end{aligned} \tag{D.24}$$

which together with (D.18), (D.21) and Assumption D.1.(iii) implies that

$$\inf_{F\in\mathcal{F}}\inf_{\theta\in B_\varepsilon^c(\theta_F^*)}[Q_F(\theta) - Q_F(\theta_F^*)] > 0. \tag{D.25}$$

This verifies Assumption 3.1.(iii).

Next, we verify Assumption 3.1.(iv). Let $\Omega_{2,F}^{(22)} = (\Omega_{2,rr,F} - \Omega'_{2,r1,F}\Omega_{z_1z_1u^2}^{-1}\Omega_{2,1r,F})^{-1}$, where

$\Omega_{2,1r,F}$ and $\Omega_{2,rr,F}$ are defined in (D.4). Then

$$\begin{aligned}
& G_{2,F}'\Omega_{2,F}^{-1}\delta_{2,F} \\
= & -(\Gamma_{xz_1}, \Gamma_{xv} + \Gamma_{xu}\delta_0') \begin{pmatrix} -\Omega_{z_1z_1u^2}^{-1}\Omega_{2,1r,F} \\ I_{r^*} \end{pmatrix} \Omega_{2,F}^{(22)}\Omega_{uu}\delta_0 \\
= & \Omega_{uu}[(\Gamma_{xz_1}\Omega_{z_1z_1u^2}^{-1}\Gamma_{z_1u^3} - \Gamma_{xu})\delta_0' + \Gamma_{xz_1}\Omega_{z_1z_1u^2}^{-1}\Gamma_{z_1vu^2} - \Gamma_{xv}]\Omega_{2,F}^{(22)}\delta_0 \\
= & \Omega_{uu}\delta_0'\Omega_{2,F}^{(22)}\delta_0(\Gamma_{xz_1}\Omega_{z_1z_1u^2}^{-1}\Gamma_{z_1u^3} - \Gamma_{xu}) \\
& +\Omega_{uu}(\Gamma_{xz_1}\Omega_{z_1z_1u^2}^{-1}\Gamma_{z_1vu^2} - \Gamma_{xv})\Omega_{2,F}^{(22)}\delta_0, \qquad \text{(D.26)}
\end{aligned}$$

by the formula of the inverse of partitioned matrix. For any $\delta_0 \in \Delta_\delta$ with $\|\delta_0\| > 0$, we have

$$\frac{\delta_0'(\Omega_{2,F}^{(22)})^2\delta_0}{(\delta_0'\Omega_{2,F}^{(22)}\delta_0)^2} \geq \frac{(\rho_{\min}(\Omega_{2,F}^{(22)}))^2}{(\rho_{\max}(\Omega_{2,F}^{(22)}))^2}\frac{1}{\delta_0'\delta_0} \geq \frac{\rho_{2,\min}^2}{C_\Delta\rho_{2,\max}^2} \qquad \text{(D.27)}$$

and

$$\delta_0'\Omega_{2,F}^{(22)}\delta_0 = \frac{\delta_0'\Omega_{2,F}^{(22)}\delta_0(\delta_0'(\Omega_{2,F}^{(22)})^2\delta_0)^{1/2}}{(\delta_0'(\Omega_{2,F}^{(22)})^2\delta_0)^{1/2}} \geq \frac{\|\delta_0\|}{\rho_{2,\max}}\frac{\delta_0'\Omega_{2,F}^{(22)}\delta_0}{(\delta_0'(\Omega_{2,F}^{(22)})^2\delta_0)^{1/2}} \qquad \text{(D.28)}$$

where the last inequality in (D.27) and the inequality in (D.28) are due to

$$\rho_{\min}(\Omega_{2,F}^{(22)}) \geq \rho_{\min}(\Omega_{2,F}^{-1}) = \rho_{2,\max}^{-1}$$

and

$$\rho_{\max}(\Omega_{2,F}^{(22)}) \leq \rho_{\max}(\Omega_{2,F}^{-1}) = \rho_{2,\min}^{-1}.$$

Therefore, for any $F \in \mathcal{F}$ with $\delta_{2,F} = \Omega_{uu}(0_{1\times r_1}, \delta_0')'$ and $\|\delta_0\| > 0$,

$$\begin{aligned}
\frac{\left\|G_{2,F}'\Omega_{2,F}^{-1}\delta_{2,F}\right\|}{\|\delta_{2,F}\|} &= \frac{\delta_0'\Omega_{2,F}^{(22)}\delta_0}{\|\delta_0\|}\left\|\begin{array}{c}(\Gamma_{xz_1}\Omega_{z_1z_1u^2}^{-1}\Gamma_{z_1vu^2} - \Gamma_{xv})\frac{\Omega_{2,F}^{(22)}\delta_0}{\delta_0'\Omega_{2,F}^{(22)}\delta_0} \\ +(\Gamma_{xz_1}\Omega_{z_1z_1u^2}^{-1}\Gamma_{z_1u^3} - \Gamma_{xu})\end{array}\right\| \\
&\geq \frac{1}{\rho_{2,\max}}\frac{\delta_0'\Omega_{2,F}^{(22)}\delta_0}{(\delta_0'(\Omega_{2,F}^{(22)})^2\delta_0)^{1/2}}\left\|\begin{array}{c}(\Gamma_{xz_1}\Omega_{z_1z_1u^2}^{-1}\Gamma_{z_1vu^2} - \Gamma_{xv})\frac{\Omega_{2,F}^{(22)}\delta_0}{\delta_0'\Omega_{2,F}^{(22)}\delta_0} \\ +(\Gamma_{xz_1}\Omega_{z_1z_1u^2}^{-1}\Gamma_{z_1u^3} - \Gamma_{xu})\end{array}\right\| \\
&= \frac{1}{\rho_{2,\max}}\frac{1}{||\widetilde{\delta}_0||}\left\|\begin{array}{c}(\Gamma_{xz_1}\Omega_{z_1z_1u^2}^{-1}\Gamma_{z_1vu^2} - \Gamma_{xv})\widetilde{\delta}_0 \\ +(\Gamma_{xz_1}\Omega_{z_1z_1u^2}^{-1}\Gamma_{z_1u^3} - \Gamma_{xu})\end{array}\right\| \qquad \text{(D.29)}
\end{aligned}$$

where $\widetilde{\delta}_0 \equiv \Omega_{2,F}^{(22)}\delta_0/\delta_0'\Omega_{2,F}^{(22)}\delta_0$ and the inequality is by (D.28). By (D.28) and the definition of $B_{\rho_2}^c$,

$\widetilde{\delta}_0 \in B^c_{\rho_2}$. Therefore, (D.29) implies that

$$\frac{\left\|G'_{2,F}\Omega^{-1}_{2,F}\delta_{2,F}\right\|}{\|\delta_{2,F}\|} \geq \frac{1}{\rho_{2,\max}} \inf_{\delta\in B^c_{\rho_2}} \|\delta\|^{-1} \left\| \begin{array}{c} (\Gamma_{xz_1}\Omega^{-1}_{z_1z_1u^2}\Gamma_{z_1vu^2} - \Gamma_{xv})\delta \\ +(\Gamma_{xz_1}\Omega^{-1}_{z_1z_1u^2}\Gamma_{z_1u^3} - \Gamma_{xu}) \end{array} \right\|. \tag{D.30}$$

Collecting the results in (D.18) and (D.30) and then applying Assumption D.1.(iv), we get

$$\inf_{\{F\in\mathcal{F}\colon\, \|\delta_F\|>0\}} \frac{\left\|G'_{2,F}\Omega^{-1}_{2,F}\delta_{2,F}\right\|}{\|\delta_{2,F}\|} > 0 \tag{D.31}$$

which shows Assumption 3.1.(iv) with $\tau = 1$.

Assumption 3.1.(v) is implied by Assumption D.1.(vii). This finishes the verification of Assumption 3.1.

To verify Assumption 3.2, note that $g_2(W,\theta) = Z_2(U - X'(\theta - \theta_0))$, $g_{2,\theta}(W,\theta) = -Z_2X'$ and $g_{2,\theta\theta}(W,\theta) = 0_{(r_2d_\theta)\times d_\theta}$. Therefore, Assumption 3.2.(i) holds automatically. Moreover Assumption 3.2.(ii) is implied by Assumption D.1.(ii) and the assumption that Θ is bounded. Assumptions 3.2(iii)-(iv) follow from Assumption D.1.(iii).

We next verify Assumption 3.3. By definition,

$$v_F = \left(\text{vec}(G_{2,F})', \text{vech}(\Omega_{2,F})', \delta_F\right). \tag{D.32}$$

Let $\Lambda_* = \{v_{*,F^*} : F^* \in \mathcal{F}^*\}$. From the expressions in (D.3), we see that $\Lambda = \{v_F\colon\, F \in \mathcal{F}\}$ is the image of $\Lambda_* \times \Delta_\delta$ under a continuous mapping. By Assumption D.1.(ii) and the Hölder inequality, Λ_* is bounded which together with Assumption D.1.(v) implies that Λ_* is compact. Since Δ_δ is also a compact set by Assumption D.1.(vii), we know that $\Lambda_* \times \Delta_\delta$ is compact. Therefore, Λ is compact and hence closed. This verifies Assumption 3.3.(ii).

Let $\varepsilon_F = \Omega_{uu}c_\Delta$ where $c_\Delta = \min\{\min_{j\leq r^*}|c_{j,\Delta}|, \min_{j\leq r^*}|C_{j,\Delta}|\}$. Below we show that for any $\widetilde{\delta} \in \mathbb{R}^{r^*}$ with $0 \leq ||\widetilde{\delta}|| \leq \varepsilon_F$, there is $\widetilde{F} \in \mathcal{F}$ such that

$$\widetilde{\delta}_{\widetilde{F}} = \widetilde{\delta},\ ||G_{2,\widetilde{F}} - G_{2,F}|| \leq C_1||\widetilde{\delta}_F||^{1/4} \text{ and } ||\Omega_{2,\widetilde{F}} - \Omega_{2,F}|| \leq C_2||\widetilde{\delta}||^{1/4} \tag{D.33}$$

for some fixed constants C_1 and C_2. This verifies Assumption 3.3.(i) with $\kappa = 1/4$.

First if $\widetilde{\delta} = 0_{r^*\times 1}$, then we set $\widetilde{F}$ to be F which is induced by δ_0, θ_0 and F^* with $\delta_0 = 0_{r^*\times 1}$. By definition $G_{2,\widetilde{F}} = G_{2,F}$, $\Omega_{2,\widetilde{F}} = \Omega_{2,F}$ and $\widetilde{\delta}_{\widetilde{F}} = \delta_F = \delta_0\Omega_{uu} = 0 = \widetilde{\delta}$ which implies that (D.33) holds.

Second consider any $\widetilde{\delta} \in \mathbb{R}^{r^*}$ with $0 < ||\widetilde{\delta}|| < \varepsilon_F$. Define $\widetilde{\delta}_0 = \widetilde{\delta}\Omega^{-1}_{uu}$. Since $||\widetilde{\delta}|| < \varepsilon_F$ and $\varepsilon_F = \Omega_{uu}c_\Delta$,

$$||\widetilde{\delta}_0|| = ||\widetilde{\delta}\Omega^{-1}_{uu}|| = ||\widetilde{\delta}||\Omega^{-1}_{uu} < c_\Delta, \tag{D.34}$$

which combined with the definition of Δ_δ implies that $\widetilde{\delta}_0 \in \Delta_\delta$. Let $\widetilde{F}$ be the joint distribution induced by $\widetilde{\delta}_0$, θ_0 and F^*. By the definition of $\mathcal{F}$, we have $\widetilde{F} \in \mathcal{F}$. Moreover,

$$\widetilde{\delta}_{\widetilde{F}} = \widetilde{\delta}_0 \Omega_{uu} = \widetilde{\delta} \tag{D.35}$$

which verifies the equality in (D.33). By definition,

$$G_{2,\widetilde{F}} = \begin{pmatrix} -\mathbb{E}_{F^*}[Z_1 X'] \\ -\widetilde{\delta}_0 \mathbb{E}_{F^*}[UX'] - \mathbb{E}_{F^*}[VX'] \end{pmatrix} \text{ and } G_{2,F} = \begin{pmatrix} -\mathbb{E}_{F^*}[Z_1 X'] \\ -\mathbb{E}_{F^*}[VX'] \end{pmatrix} \tag{D.36}$$

which together with the Cauchy-Schwarz inequality and the Hölder inequality implies that

$$\begin{aligned} ||G_{2,\widetilde{F}} - G_{2,F}|| &= ||\widetilde{\delta}_0 \mathbb{E}_{F^*}[UX']|| \leq ||\widetilde{\delta}_0||(\Omega_{uu}\mathbb{E}_{F^*}[||X||^2])^{1/2} \\ &= ||\widetilde{\delta}_0||^{3/4} \Omega_{uu}^{1/4} (\mathbb{E}_{F^*}[||X||^2])^{1/2} ||\widetilde{\delta}_0 \Omega_{uu}||^{1/4}. \end{aligned} \tag{D.37}$$

By Assumption D.1.(ii),

$$\sup_{F^* \in \mathcal{F}^*} \mathbb{E}_{F^*}[||X||^2] < \infty \text{ and } \sup_{F^* \in \mathcal{F}^*} \Omega_{uu} < \infty \tag{D.38}$$

which together with (D.34), (D.37) and the definition of $\widetilde{\delta}$ implies that

$$||G_{2,\widetilde{F}} - G_{2,F}|| \leq C_1 ||\widetilde{\delta}||^{1/4}, \tag{D.39}$$

where $C_1 = c_{\Delta}^{3/4} \sup_{F^* \in \mathcal{F}^*} (\mathbb{E}_{F^*}[||X||^2])^{1/2} \sup_{F^* \in \mathcal{F}^*} \Omega_{uu}^{1/4}$ is finite.

To show the last inequality in (D.33), note that by definition $\theta_{\widetilde{F}} = \theta_0 = \theta_F$ and hence

$$\mathbb{E}_{\widetilde{F}}[Z_1 Z_1'(Y - X'\theta_{\widetilde{F}})^2] = \mathbb{E}_{F^*}[Z_1 Z_1' U^2] = \mathbb{E}_F[Z_1 Z_1'(Y - X'\theta_F)^2]. \tag{D.40}$$

Under $\widetilde{F}$,

$$\mathbb{E}_{\widetilde{F}}[Z_1 Z^{*\prime}(Y - X'\theta_{\widetilde{F}})^2] = \mathbb{E}_{F^*}[Z_1(U\widetilde{\delta}_0 + V)'U^2] = \mathbb{E}_{F^*}[U^3 Z_1]\widetilde{\delta}_0' + \mathbb{E}_{F^*}[U^2 Z_1 V'], \tag{D.41}$$

and

$$\begin{aligned} & \mathbb{E}_{\widetilde{F}}[Z^* Z^{*\prime}(Y - X'\theta_{\widetilde{F}})^2] \\ = \ & \mathbb{E}_{F^*}[(U\widetilde{\delta}_0 + V)(U\widetilde{\delta}_0 + V)'U^2] \\ = \ & \mathbb{E}_{F^*}[U^4]\widetilde{\delta}_0\widetilde{\delta}_0' + \widetilde{\delta}_0 \mathbb{E}_{F^*}[U^3 V'] + \mathbb{E}_{F^*}[U^3 V]\widetilde{\delta}_0' + \mathbb{E}_{F^*}[U^2 V V']. \end{aligned} \tag{D.42}$$

Under F,

$$\mathbb{E}_F[Z_1 Z^{*\prime}(Y - X'\theta_F)^2] = \mathbb{E}_{F^*}[U^2 Z_1 V'] \text{ and } \mathbb{E}_F[Z^* Z^{*\prime}(Y - X'\theta_F)^2] = \mathbb{E}_{F^*}[U^2 V V']. \quad \text{(D.43)}$$

Collecting the results in (D.40), (D.41), (D.42) and (D.43), and applying the triangle inequality, we get

$$\begin{aligned}
& \left\|\mathbb{E}_{\widetilde{F}}[Z_2 Z_2'(Y - X'\theta_{\widetilde{F}})^2] - \mathbb{E}_F[Z_2 Z_2'(Y - X'\theta_F)^2]\right\| \\
\leq\ & \left\|\mathbb{E}_{F^*}[U^3 Z_1]\widetilde{\delta}_0'\right\| + \left\|\mathbb{E}_{F^*}[U^4]\widetilde{\delta}_0\widetilde{\delta}_0'\right\| \\
& + \left\|\widetilde{\delta}_0\mathbb{E}_{F^*}[U^3 V']\right\| + \left\|\mathbb{E}_{F^*}[U^3 V]\widetilde{\delta}_0'\right\|. \quad \text{(D.44)}
\end{aligned}$$

By Assumption D.1.(ii) and the Lyapunov inequality

$$\sup_{F^* \in \mathcal{F}^*} \mathbb{E}_{F^*}[|U|^5] < \infty, \ \sup_{F^* \in \mathcal{F}^*} \mathbb{E}_{F^*}[||Z_1||^4] < \infty \text{ and } \sup_{F^* \in \mathcal{F}^*} \mathbb{E}_{F^*}[||V||^4] < \infty. \quad \text{(D.45)}$$

By the Hölder inequality,

$$\begin{aligned}
\left\|\mathbb{E}_{F^*}[U^3 Z_1]\right\| &\leq (\mathbb{E}_{F^*}[|U|||Z_1||^2]\mathbb{E}_{F^*}[|U|^5])^{1/2} \\
&\leq (\mathbb{E}_{F^*}[|U|^5])^{1/2}(\mathbb{E}_{F^*}[||Z_1||^4])^{1/4}(\mathbb{E}_{F^*}[U^2])^{1/4} \\
&= \Omega_{uu}^{1/4}(\mathbb{E}_{F^*}[|U|^5])^{1/2}(\mathbb{E}_{F^*}[||Z_1||^4])^{1/4}. \quad \text{(D.46)}
\end{aligned}$$

Similarly, we can show that

$$\left\|\mathbb{E}_{F^*}[U^3 V']\right\| \leq \Omega_{uu}^{1/4}(\mathbb{E}_{F^*}[|U|^5])^{1/2}(\mathbb{E}_{F^*}[||V||^4])^{1/4} \quad \text{(D.47)}$$

and

$$\mathbb{E}_{F^*}[U^4] \leq (\mathbb{E}_{F^*}[U^2]\mathbb{E}_{F^*}[U^6])^{1/2} = \Omega_{uu}^{1/4} \sup_{F^* \in \mathcal{F}^*} (\mathbb{E}_{F^*}[U^6])^{1/2}. \quad \text{(D.48)}$$

Let $C_{2,0} = \sup_{F^* \in \mathcal{F}^*}\{(\mathbb{E}_{F^*}[|U|^5])^{1/2}[(\mathbb{E}_{F^*}[||Z_1||^4])^{1/4} + (\mathbb{E}_{F^*}[||V||^4])^{1/4}] + (\mathbb{E}_{F^*}[U^6])^{1/2}\}$. Combining the results in (D.44), (D.46), (D.47) and (D.48), and applying the Cauchy-Schwarz inequality, we get

$$\begin{aligned}
& \left\|\mathbb{E}_{\widetilde{F}}[Z_2 Z_2'(Y - X'\theta_{\widetilde{F}})^2] - \mathbb{E}_F[Z_2 Z_2'(Y - X'\theta_F)^2]\right\| \\
\leq\ & 3C_{2,0}\Omega_{uu}^{1/4}||\widetilde{\delta}_0|| + C_{2,0}\Omega_{uu}^{1/4}||\widetilde{\delta}_0||^2 \\
=\ & (3C_{2,0}||\widetilde{\delta}_0||^{3/4} + C_{2,0}||\widetilde{\delta}_0||^{7/4})\Omega_{uu}^{1/4}||\widetilde{\delta}_0||^{1/4} \leq C_{2,1}||\widetilde{\delta}||^{1/4} \quad \text{(D.49)}
\end{aligned}$$

where $C_{2,1} = C_{2,0}(3c_{\Delta}^{3/4} + c_{\Delta}^{7/4})$, the second inequality is by (D.34) and the definition of $\widetilde{\delta}$. By (D.45), Assumption D.1.(ii) and the definition of c_{Δ},

$$C_{2,1} < \infty. \quad \text{(D.50)}$$

Next note that

$$\mathbb{E}_{\widetilde{F}}[Z_2(Y - X'\theta_{\widetilde{F}})] = \begin{pmatrix} \mathbb{E}_{F^*}[Z_1 U] \\ \widetilde{\delta}_0 \Omega_{uu} \end{pmatrix} \text{ and } \mathbb{E}_F[Z_2(Y - X'\theta_F)] = \begin{pmatrix} \mathbb{E}_{F^*}[Z_1 U] \\ 0_{r^* \times 1} \end{pmatrix} \tag{D.51}$$

which implies that

$$\begin{aligned}
& \left\| \begin{array}{c} \mathbb{E}_{\widetilde{F}}[Z_2(Y - X'\theta_{\widetilde{F}})]\mathbb{E}_{\widetilde{F}}[Z_2'(Y - X'\theta_{\widetilde{F}})] \\ -\mathbb{E}_F[Z_2(Y - X'\theta_F)]\mathbb{E}_F[Z_2'(Y - X'\theta_F)] \end{array} \right\| \\
= & \left\| \begin{pmatrix} 0_{r_1 \times r_1} & \Omega_{uu,F^*}\mathbb{E}_{F^*}[Z_1 U]\widetilde{\delta}_0' \\ \widetilde{\delta}_0 \mathbb{E}_{F^*}[Z_1' U]\Omega_{uu} & \widetilde{\delta}_0 \widetilde{\delta}_0' \Omega_{uu}^2 \end{pmatrix} \right\| \\
\leq & \ \Omega_{uu} \left\| \mathbb{E}_{F^*}[Z_1 U]\widetilde{\delta}_0' \right\| + \Omega_{uu} \left\| \widetilde{\delta}_0 \mathbb{E}_{F^*}[Z_1 U] \right\| + \Omega_{uu}^2 ||\widetilde{\delta}_0||^2 \\
\leq & \ 2\Omega_{uu} ||\widetilde{\delta}_0|| \, \|\mathbb{E}_{F^*}[Z_1 U]\| + \Omega_{uu}^2 ||\widetilde{\delta}_0||^2 \\
\leq & \ (2\Omega_{uu}^{5/4} ||\widetilde{\delta}_0||^{3/4} (\mathbb{E}_{F^*}[||Z_1||^2])^{1/2} + \Omega_{uu}^{7/4} ||\widetilde{\delta}_0||^{7/4}) ||\widetilde{\delta}||^{1/4} \leq C_{2,2} ||\widetilde{\delta}||^{1/4}
\end{aligned} \tag{D.52}$$

where $C_{2,2} = \sup_{F^* \in \mathcal{F}^*} \{2\Omega_{uu}^{5/4} (\mathbb{E}_{F^*}[||Z_1||^2])^{1/2} c_\Delta^{3/4} + \Omega_{uu}^{7/4} c_\Delta^{7/4}\}$, the second inequality is by the Cauchy-Schwarz inequality, the third inequality is by the Hölder inequality. By Assumption D.1.(ii) and the definition of c_Δ,

$$C_{2,2} < \infty. \tag{D.53}$$

By the definition of $\Omega_{2,F}$ in (D.2), we can use the triangle inequality and the results in (D.49) and (D.52) to deduce that

$$\left\| \Omega_{2,\widetilde{F}} - \Omega_{2,F} \right\| \leq C_2 ||\widetilde{\delta}||^{1/4} \tag{D.54}$$

where $C_2 = C_{2,1} + C_{2,2}$ and $C_2 < \infty$ by (D.50) and (D.53), which proves the second inequality in (D.33). This verifies Assumption 3.3.(i) with $\kappa = 1/4$. ■

Proof of Lemma *3.1*. Next, we apply Lemma D.1 to prove Lemma 3.1 in the paper. For convenience, the conditions of Lemma 3.1 are stated here. The proof verifies the conditions of Lemma D.1 with the following conditions in a Gaussian model. Let $\mathcal{F}^*$ denote the set of normal distributions which satisfies:
(i) $\phi_u = 0$, $\Gamma_{z_1 u} = 0_{r_1 \times 1}$ and $\Gamma_{vu} = 0_{r^* \times 1}$;
(ii) $\inf_{F^* \in \mathcal{F}^*} \rho_{\min}(\Gamma_{xz_1}\Gamma_{z_1 x}) > 0$, $\sup_{F^* \in \mathcal{F}^*} ||\phi||^2 < \infty$ and
$0 < \inf_{F^* \in \mathcal{F}^*} \rho_{\min}(\Psi) \leq \sup_{F^* \in \mathcal{F}^*} \rho_{\max}(\Psi) < \infty$;
(iii) $\inf_{F^* \in \mathcal{F}^*} \inf_{\{\|\delta\| \geq \varepsilon\}} \|\delta\|^{-1} ||(\Gamma_{xz_1}\Gamma_{z_1 z_1}^{-1}\Gamma_{z_1 v} - \Gamma_{xv})\delta - \Gamma_{xu}|| > 0$ for some $\varepsilon > 0$ that is small enough (where ε is given in (B.3) in the Appendix of CLS);
(iv) $\theta_0 \in int(\Theta)$ and Θ is compact and large enough such that the pseudo-true value $\theta^*(F) \in int(\Theta)$;
(v) $\Delta_\delta = [c_{1,\Delta}, C_{1,\Delta}] \times \cdots \times [c_{r^*,\Delta}, C_{r^*,\Delta}]$ where $\{c_{j,\Delta}, C_{j,\Delta}\}_{j=1}^{r^*}$ is a set of finite constants with $c_{j,\Delta} < 0 < C_{j,\Delta}$ for $j = 1, \ldots, r^*$.

Specifically, we assume that condition (ii) of Lemma 3.1 holds with some constants c_ρ and C_ρ such that $c_\rho \leq \rho_{\min}(\Gamma_{xz_1}\Gamma_{z_1 x})$, $||\phi||^2 \leq C_\rho$ and $c_\rho \leq \rho_{\min}(\Psi) \leq \rho_{\max}(\Psi) \leq C_\rho$; condition (iii) of

Lemma 3.1 holds with

$$\inf_{\delta\in B_{\varepsilon}^{c}}\|\delta\|^{-1}\,\|(\Gamma_{xz_1}\Gamma_{z_1z_1}^{-1}\Gamma_{z_1v}-\Gamma_{xv})\delta-\Gamma_{xu}\|\geq c_{\Gamma} \tag{D.55}$$

for some positive constant c_{Γ} and

$$B_{\varepsilon}^{c}\equiv\{\delta\in\mathbb{R}^{r^*}:\|\delta\|\geq c_{*,\rho}C_{*,\rho}^{-1}C_{\Delta}^{-1}\}, \tag{D.56}$$

where

$$C_{*,W}\equiv 2(d_{\theta}+r_2+1)C_{\rho},\ c_{*,\rho}\equiv\min\{1,c_{\rho}^{2}\}\text{ and }C_{*,\rho}\equiv C_{*,W}^{2}(2+C_{\Delta}^{1/2})^{2} \tag{D.57}$$

and $C_{\Delta}\equiv\sup_{\delta_0\in\Delta_{\delta}}\|\delta_0\|^2$.

Assumption *D.1.*(i) holds under Condition (i) of Lemma 3.1. Since $(X',Z_1',V',U)'$ is a normal random vector, Assumption *D.1.*(ii) holds by $\|\phi\|^2\leq C_{\rho}$ and $\rho_{\max}(\Psi)\leq C_{\rho}$. By $\rho_{\min}(\Psi)\geq c_{\rho}$ and $\phi_u=0$, we have $\mathbb{E}_{F^*}[U^2]\geq c_{\rho}$ for any $F^*\in\mathcal{F}^*$ and hence $\inf_{F^*\in\mathcal{F}^*}\mathbb{E}_{F^*}[U^2]>0$. Let F denote the distribution of W induced by F^* with mean ϕ and variance-covariance matrix Ψ. By definition, $G_{1,F}=\ -\mathbb{E}_{F^*}[Z_1X']=\Gamma_{z_1x}$. Therefore,

$$\inf_{F\in\mathcal{F}}\rho_{\min}(G_{1,F}'G_{1,F})\geq c_{\rho}>0 \tag{D.58}$$

holds by $\rho_{\min}(\Gamma_{xz_1}\Gamma_{z_1x})\geq c_{\rho}>0$ for any $F^*\in\mathcal{F}^*$. Since $\Gamma_{z_1u}=0_{r_1\times 1}$ and $\Gamma_{vu}=0_{r^*\times 1}$ for any $F^*\in\mathcal{F}^*$, U is independent with respect to $(Z_1',V')'$ under the normality assumption. Therefore, by Condition (i) of Lemma 3.1

$$\begin{aligned}\Omega_{2,F} &= \begin{pmatrix}\Omega_{uu}\Gamma_{z_1z_1} & \Omega_{uu}\Gamma_{z_1v}\\ \Omega_{uu}\Gamma_{z_1v}' & 2\Omega_{uu}^{2}\delta_0\delta_0'+\Omega_{uu}\Gamma_{vv}\end{pmatrix}\\ &= \Omega_{uu}\begin{pmatrix}\Omega_{z_1z_1} & \Omega_{z_1v}\\ \Omega_{vz_1} & \Omega_{vv}\end{pmatrix}+\Omega_{uu}\begin{pmatrix}\phi_{z_1}\\ \phi_v\end{pmatrix}\begin{pmatrix}\phi_{z_1}\\ \phi_v\end{pmatrix}'+\begin{pmatrix}0_{r_1\times r_1} & 0_{r_1\times r^*}\\ 0_{r^*\times r_1} & 2\Omega_{uu}^{2}\delta_0\delta_0'\end{pmatrix}\end{aligned} \tag{D.59}$$

which implies that $\rho_{\min}(\Omega_{2,F})\geq\rho_{\min}^{2}(\Psi)$ where F is the distribution of W induced by F^* with mean ϕ and variance-covariance matrix Ψ. Since $\rho_{\min}(\Psi)\geq c_{\rho}>0$, we have

$$\inf_{F\in\mathcal{F}}\rho_{\min}(\Omega_{2,F})\geq c_{\rho}^{2}>0. \tag{D.60}$$

This finishes the proof of Assumption *D.1.*(iii).

By (*D.59*), Conditions (ii) and (v) of Lemma 3.1

$$\sup_{F\in\mathcal{F}}\rho_{\max}(\Omega_{2,F})\leq\rho_{\max}^{2}(\Psi)+\rho_{\max}(\Psi)\|\phi\|^2+2\rho_{\max}^{2}(\Psi)C_{\Delta}\leq 2C_{\rho}^{2}(1+C_{\Delta}). \tag{D.61}$$

By (D.19) in the proof of Lemma D.1,

$$\|G_{2,F}\| \leq 2C_\rho(d_\theta + r_2 + 1)(2 + C_\Delta^{1/2})$$

which implies that

$$\sup_{F\in\mathcal{F}} \rho_{\max}(G'_{2,F}G_{2,F}) \leq 4C_\rho^2(d_\theta + r_2 + 1)^2(2 + C_\Delta^{1/2})^2. \tag{D.62}$$

By (D.58) and (D.60),

$$\min\left\{\inf_{F\in\mathcal{F}} \rho_{\min}(\Omega_{2,F}), \inf_{F\in\mathcal{F}} \rho_{\min}(G'_{2,F}G_{2,F})\right\} \geq \min\{1, c_\rho^2\}. \tag{D.63}$$

By (D.61) and (D.62),

$$\max\left\{\sup_{F\in\mathcal{F}} \rho_{\max}(\Omega_{2,F}), \sup_{F\in\mathcal{F}} \rho_{\max}(G'_{2,F}G_{2,F})\right\} \leq 4C_\rho^2(d_\theta + r_2 + 1)^2(2 + C_\Delta^{1/2})^2. \tag{D.64}$$

From (*D.63*), (*D.64*), the definitions of $c_{*,\rho}$, $C_{*,\rho}$ and $B^c_{N,\rho}$, we have $B^c_{\rho_2} \subset B^c_{N,\rho}$ where $B^c_{\rho_2}$ is defined in (*D.7*). Moreover, by $\phi_u = 0$, the normality assumption and the independence between U and $(Z_1', V')'$, we have $\Omega_{z_1z_1u^2} = \Omega_{uu}\Gamma_{z_1z_1}$, $\Gamma_{z_1vu^2} = \Omega_{uu}\Gamma_{z_1v}$ and $\Gamma_{z_1u^3} = 0_{r_1\times 1}$, which implies that

$$\begin{aligned} & ||(\Gamma_{xz_1}\Omega^{-1}_{z_1z_1u^2}\Gamma_{z_1vu^2} - \Gamma_{xv})\delta + \Gamma_{xz_1}\Omega^{-1}_{z_1z_1u^2}\Gamma_{z_1u^3} - \Gamma_{xu}|| \\ = \; & ||(\Gamma_{xz_1}\Gamma^{-1}_{z_1z_1}\Gamma_{z_1v} - \Gamma_{xv})\delta - \Gamma_{xu}||. \end{aligned} \tag{D.65}$$

Assumption *D.1.*(iv) follows by $B^c_{\rho_2} \subset B^c_{N,\rho}$, (*D.65*) and Condition (iii) of the lemma.

We next show that Assumption *D.1.*(v) holds. Define

$$\overline{v}_{*,F^*} = \begin{pmatrix} \Omega_{uu}, \text{vec}(\Gamma_{xz_1})', \text{vec}(\Gamma_{xu})', \text{vec}(\Gamma_{xv})', \\ \text{vec}(\Gamma_{z_1v})', \text{vech}(\Gamma_{z_1z_1})', \text{vech}(\Gamma_{vv})' \end{pmatrix}.$$

Under Condition (i) of Lemma 3.1 and the normality assumption, $\Gamma_{u^2u^2} = 3\Omega^2_{uu}$, $\Gamma_{z_1u^3} = 0_{r_1\times 1}$, $\Gamma_{vu^3} = 0_{r^*\times 1}$, $\Omega_{z_1z_1u^2} = \Omega_{uu}\Gamma_{z_1z_1}$, $\Gamma_{z_1vu^2} = \Omega_{uu}\Gamma_{z_1v}$ and $\Omega_{vvu^2} = \Omega_{uu}\Gamma_{vv}$. Therefore to verify Assumption *D.1.*(v), it is sufficient to show that the set $\{\overline{v}_{*,F^*} : F^* \in \mathcal{F}^*\}$ is compact because the set $\{v_{*,F^*} : F^* \in \mathcal{F}^*\}$ is the image of the set $\{\overline{v}_{*,F^*} : F^* \in \mathcal{F}^*\}$ under a continuous mapping. Let $\{(\phi_n, \Psi_n)\}_n$ be a convergent sequence where (ϕ_n, Ψ_n) satisfies Conditions (i)-(iii) of Lemma 3.1 for any n. Let $\widetilde{\phi}$ and $\widetilde{\Psi}$ denote the limits of ϕ_n and Ψ_n under the Euclidean norm respectively. We first show that Conditions (i)-(iii) of Lemma 3.1 hold for $(\widetilde{\phi}, \widetilde{\Psi})$. Since $\phi_{u,n} = 0$, $\Gamma_{z_1u,n} = 0_{r_1\times 1}$ and $\Gamma_{vu,n} = 0_{r^*\times 1}$ for any n, we have $\widetilde{\phi}_u = 0$, $\widetilde{\Gamma}_{z_1u} = 0_{r_1\times 1}$ and $\widetilde{\Gamma}_{vu} = 0_{r^*\times 1}$ which shows that $(\widetilde{\phi}, \widetilde{\Psi})$ satisfies Condition (i) of Lemma 3.1. Since $\phi_n \to \widetilde{\phi}$ and $\|\phi_n\|^2 \leq C_\rho$ for any n, we have $||\widetilde{\phi}||^2 \leq C_\rho$. By the convergence of (ϕ_n, Ψ_n), $\Gamma_{xz_1,n} \to \widetilde{\Gamma}_{xz_1}$. Since the roots of a polynomial continuously depends on its coefficients, we have

$$\rho_{\min}(\Gamma_{xz_1,n}\Gamma'_{xz_1,n}) \to \rho_{\min}(\widetilde{\Gamma}_{xz_1}\widetilde{\Gamma}'_{xz_1}),\ \rho_{\min}(\Psi_n) \to \rho_{\min}(\widetilde{\Psi}) \text{ and } \rho_{\max}(\Psi_n) \to \rho_{\max}(\widetilde{\Psi})$$

which together with the assumption that $\Gamma_{xz_1,n}$ and Ψ_n satisfy Condition (ii) of Lemma 3.1 implies that

$$c_\rho \leq \rho_{\min}(\widetilde{\Gamma}_{xz_1}\widetilde{\Gamma}'_{xz_1}) \text{ and } c_\rho \leq \rho_{\min}(\widetilde{\Psi}) \leq \rho_{\max}(\widetilde{\Psi}) \leq C_\rho.$$

This shows that Condition (ii) of Lemma 3.1 holds for $(\widetilde{\phi}, \widetilde{\Psi})$. For any $\delta \in B^c_{N,\rho}$, by the triangle inequality, the Cauchy-Schwarz inequality and $\|\delta\| \geq c_\rho^2 C_\rho^{-2} C_\Delta^{-1}(1+C_\Delta)^{-1}2^{-1}$,

$$\begin{aligned} & \|\delta\|^{-1} \|(\widetilde{\Gamma}_{xz_1}\widetilde{\Gamma}^{-1}_{z_1z_1}\widetilde{\Gamma}_{z_1v} - \widetilde{\Gamma}_{xv})\delta - \widetilde{\Gamma}_{xu}\| \\ \geq\ & \|\delta\|^{-1} \|(\Gamma_{xz_1,n}\Gamma^{-1}_{z_1z_1,n}\Gamma_{z_1v,n} - \Gamma_{xv,n})\delta - \Gamma_{xu,n}\| \\ & -\|\widetilde{\Gamma}_{xz_1}\widetilde{\Gamma}^{-1}_{z_1z_1}\widetilde{\Gamma}_{z_1v} - \Gamma_{xz_1,n}\Gamma^{-1}_{z_1z_1,n}\Gamma_{z_1v,n}\| \\ & -\|\widetilde{\Gamma}_{xv} - \Gamma_{xv,n}\| - 2C_\rho^2 C_\Delta(1+C_\Delta)c_\rho^{-2}\|\Gamma_{xu,n} - \widetilde{\Gamma}_{xu}\| \end{aligned}$$

which together with the convergence of (ϕ_n, Ψ_n) and Conditions (ii)-(iii) of Lemma 3.1 implies that

$$\begin{aligned} & \|\delta\|^{-1} \|(\widetilde{\Gamma}_{xz_1}\widetilde{\Gamma}^{-1}_{z_1z_1}\widetilde{\Gamma}_{z_1v} - \widetilde{\Gamma}_{xv})\delta - \widetilde{\Gamma}_{xu}\| \\ \geq\ & c_\Gamma - \|\widetilde{\Gamma}_{xz_1}\widetilde{\Gamma}^{-1}_{z_1z_1}\widetilde{\Gamma}_{z_1v} - \Gamma_{xz_1,n}\Gamma^{-1}_{z_1z_1,n}\Gamma_{z_1v,n}\| \\ & -\|\widetilde{\Gamma}_{xv} - \Gamma_{xv,n}\| - 2C_\rho^2 C_\Delta(1+C_\Delta)c_\rho^{-2}\|\Gamma_{xu,n} - \widetilde{\Gamma}_{xu}\| \end{aligned}$$

for any n. Let n go to infinity, we get

$$\|\delta\|^{-1} \|(\widetilde{\Gamma}_{xz_1}\widetilde{\Gamma}^{-1}_{z_1z_1}\widetilde{\Gamma}_{z_1v} - \widetilde{\Gamma}_{xv})\delta - \widetilde{\Gamma}_{xu}\| \geq c_\Gamma$$

for any $\delta \in B^c_\varepsilon$. This shows that Condition (iii) of Lemma 3.1 also holds for $(\widetilde{\phi}, \widetilde{\Psi})$. Hence the set of (ϕ, Ψ) which satisfies Conditions (i)-(iii) of Lemma 3.1 is closed. By Conditions (i)-(ii) of the Lemma, we know that this set is compact because it is also bounded. Let F^* denote the normal distribution with mean ϕ and variance-covariance matrix Ψ. Then $\overline{v}_{*,F^*}$ is the image of (ϕ, Ψ) under a continuous mapping, which implies that $\{\overline{v}_{*,F^*} : F^* \in \mathcal{F}^*\}$ is compact. Therefore the set $\{v_{*,F^*} : F^* \in \mathcal{F}^*\}$ is compact and hence closed. This proves Assumption *D.1.*(v).

Assumption *D.1.*(vi) is used to show that $\theta_F \in int(\Theta)$ and $\theta^*_F \in int(\Theta)$ for any $F \in \mathcal{F}$. By $\theta_F = \theta_0$ and Condition (iv) of Lemma 3.1, we have $\theta_F \in int(\Theta)$ and $\theta^*_F \in int(\Theta)$.

Finally, Assumption *D.1.*(vii) is the same as Condition (v) of Lemma 3.1. ■

E Proof of Some Auxiliary Results in Sections 4 and 5 of CLS

Proof of Lemma B.2. (i) Let $g_{2,j}(w,\theta)$ denote the j-th ($j = 1,\ldots,r_2$) component of $g_2(w,\theta)$. By the mean value expansion,

$$g_{2,j}(w,\theta_1) - g_{2,j}(w,\theta_2) = g_{2,j,\theta}(w,\widetilde{\theta}_{1,2})(\theta_1 - \theta_2) \tag{E.1}$$

for any $j = 1,\ldots,r_2$, where $\widetilde{\theta}_{1,2}$ is some vector between θ_1 and θ_2. By (E.1) and the Cauchy-Schwarz inequality

$$|\mathbb{E}_F[g_{2,j}(w,\theta_1) - g_{2,j}(w,\theta_2)]| \leq \mathbb{E}_F\left[\sup_{\theta\in\Theta}\|g_{2,\theta}(W,\theta)\|\right]\|\theta_1 - \theta_2\|, \tag{E.2}$$

for any $j = 1,\ldots,r_2$. By (E.2), we deduce that

$$\begin{aligned}\|M_{2,F}(\theta_1) - M_{2,F}(\theta_2)\| &\leq \sqrt{r_2}\mathbb{E}_F\left[\sup_{\theta\in\Theta}\|g_{2,\theta}(W,\theta)\|\right]\|\theta_1 - \theta_2\| \\ &\leq C_{M,1}\sqrt{r_2}\,\|\theta_1 - \theta_2\| \end{aligned} \tag{E.3}$$

for any $F \in \mathcal{F}$, where $C_{M,1} \equiv \sup_{F\in\mathcal{F}} \mathbb{E}_F[\sup_{\theta\in\Theta}\|g_{2,\theta}(W,\theta)\|]$ and $C_{M,1} < \infty$ by Assumption 3.2.(ii). This immediately proves the claim in (i). The claim in (ii) follows by similar argument and its proof is omitted.

(iii) By the mean value expansion,

$$\begin{aligned}& g_{2,j_1}(w,\theta_1)g_{2,j_2}(w,\theta_1) - g_{2,j_1}(w,\theta_2)g_{2,j_2}(w,\theta_2) \\ =\; & \left[g_{2,j_1,\theta}(w,\widetilde{\theta}_{1,2})g_{2,j_2}(w,\widetilde{\theta}_{1,2}) + g_{2,j_1}(w,\widetilde{\theta}_{1,2})g_{2,j_2,\theta}(w,\widetilde{\theta}_{1,2})\right](\theta_1 - \theta_2)\end{aligned} \tag{E.4}$$

for any $j_1, j_2 = 1,\ldots,r_2$, where $\widetilde{\theta}_{1,2}$ is some vector between θ_1 and θ_2 and may take different values from the $\widetilde{\theta}_{1,2}$ in (E.1). By (E.4), the triangle inequality and the Cauchy-Schwarz inequality

$$\begin{aligned}& |\mathbb{E}_F[g_{2,j_1}(w,\theta_1)g_{2,j_2}(w,\theta_1) - g_{2,j_1}(w,\theta_2)g_{2,j_2}(w,\theta_2)]| \\ \leq\; & 2\mathbb{E}_F\left[\sup_{\theta\in\Theta}\|g_2(W,\theta)\|\,\|g_{2,\theta}(W,\theta)\|\right]\|\theta_1 - \theta_2\| \\ \leq\; & \mathbb{E}_F\left[\sup_{\theta\in\Theta}(\|g_2(W,\theta)\|^2 + \|g_{2,\theta}(W,\theta)\|^2)\right]\|\theta_1 - \theta_2\| \end{aligned} \tag{E.5}$$

for any $j_1, j_2 = 1,\ldots,r_2$, where the second inequality is by the simple inequality that $|ab| \leq (a^2+b^2)/2$. By (E.5)

$$\begin{aligned}& \left\|\mathbb{E}_F\left[g_2(W,\theta_1)g_2(W,\theta_1)' - g_2(W,\theta_2)g_2(W,\theta_2)'\right]\right\| \\ \leq\; & r_2\mathbb{E}_F\left[\sup_{\theta\in\Theta}(\|g_2(W,\theta)\|^2 + \|g_{2,\theta}(W,\theta)\|^2)\right]\|\theta_1 - \theta_2\| \\ \leq\; & r_2 C_{M,2}\,\|\theta_1 - \theta_2\| \end{aligned} \tag{E.6}$$

for any $F \in \mathcal{F}$, where $C_{M,2} \equiv \sup_{F\in\mathcal{F}} \mathbb{E}_F \left[\sup_{\theta\in\Theta}(\|g_2(W,\theta)\|^2 + \|g_{2,\theta}(W,\theta)\|^2)\right]$ and $C_{M,2} < \infty$ by Assumption 3.2.(ii). Using the triangle inequality, and the inequality in (E.2), we deduce that

$$\begin{aligned} & |\mathbb{E}_F[g_{2,j_1}(w,\theta_1)]\mathbb{E}_F[g_{2,j_2}(w,\theta_1)] - \mathbb{E}_F[g_{2,j_1}(w,\theta_2)]\mathbb{E}_F[g_{2,j_2}(w,\theta_2)]| \\ \leq\ & |\mathbb{E}_F[g_{2,j_1}(w,\theta_1) - g_{2,j_1}(w,\theta_2)]\mathbb{E}_F[g_{2,j_2}(w,\theta_1)]| \\ & + |\mathbb{E}_F[g_{2,j_1}(w,\theta_2)]\mathbb{E}_F[g_{2,j_2}(w,\theta_2) - g_{2,j_2}(w,\theta_1)]| \\ \leq\ & 2\mathbb{E}_F\left[\sup_{\theta\in\Theta}\|g_2(W,\theta)\|\right]\mathbb{E}_F\left[\sup_{\theta\in\Theta}\|g_{2,\theta}(W,\theta)\|\right]\|\theta_1-\theta_2\| \end{aligned} \tag{E.7}$$

for any $j_1, j_2 = 1, \ldots, r_2$. By (E.7)

$$\left\|\mathbb{E}_F[g_2(w,\theta_1)]\mathbb{E}_F[g_2(w,\theta_1)'] - \mathbb{E}_F[g_2(w,\theta_2)]\mathbb{E}_F[g_2(w,\theta_2)']\right\| \leq r_2 C_{M,3}\|\theta_1-\theta_2\| \tag{E.8}$$

for any $F \in \mathcal{F}$, where $C_{M,3} \equiv 2\sup_{F\in\mathcal{F}} \mathbb{E}_F[\sup_{\theta\in\Theta}\|g_2(W,\theta)\|]\,\mathbb{E}_F[\sup_{\theta\in\Theta}\|g_{2,\theta}(W,\theta)\|]$ and $C_{M,3} < \infty$ by Assumption 3.2.(ii).

By the definition of $\Omega_{2,F}(\theta)$, the triangle inequality and the results in (E.6) and (E.8)

$$\|\Omega_{2,F}(\theta_1) - \Omega_{2,F}(\theta_2)\| \leq r_2(C_{M,2} + C_{M,3})\|\theta_1-\theta_2\|, \tag{E.9}$$

which immediately proves the claim in (iii). ■

Proof of Lemma B.3. By Lemma B.1.(i),

$$\overline{g}_2(\theta) = M_{2,F_n}(\theta) + \left[n^{-1}\sum_{i=1}^n g_2(W_i,\theta) - M_{2,F_n}(\theta)\right] = M_{2,F_n}(\theta) + o_p(1), \tag{E.10}$$

uniformly over $\theta \in \Theta$. As $g_1(W,\theta)$ is a subvector of $g_2(W,\theta)$, by (E.10) and Assumption 3.2.(ii),

$$\overline{g}_1(\theta)'\overline{g}_1(\theta) = M_{1,F_n}(\theta)'M_{1,F_n}(\theta) + o_p(1) \tag{E.11}$$

uniformly over $\theta \in \Theta$. By Assumptions 3.1.(i)-(ii) and $F_n \in \mathcal{F}$, $M_{1,F_n}(\theta)'M_{1,F_n}(\theta)$ is uniquely minimized at θ_{F_n}, which together with the uniform convergence in (E.11) implies that

$$\widetilde{\theta}_1 - \theta_{F_n} \to_p 0. \tag{E.12}$$

To show the consistency of $\overline{\Omega}_2$, note that

$$\begin{aligned} \overline{\Omega}_2 &= n^{-1}\sum_{i=1}^n g_2(W_i,\widetilde{\theta}_1)g_2(W_i,\widetilde{\theta}_1)' - \overline{g}_2(\widetilde{\theta}_1)\overline{g}_2(\widetilde{\theta}_1)' \\ &= \mathbb{E}_{F_n}[g_2(W,\widetilde{\theta}_1)g_2(W,\widetilde{\theta}_1)'] - M_{2,F_n}(\widetilde{\theta}_1)'M_{2,F_n}(\widetilde{\theta}_1) + o_p(1) \\ &= \Omega_{2,F_n}(\widetilde{\theta}_1) + o_p(1) = \Omega_{2,F_n} + o_p(1), \end{aligned} \tag{E.13}$$

where the first equality is by the definition of $\overline{\Omega}_2$, the second equality holds by (E.10), Lemma B.1.(ii) and Assumption 3.2.(ii), the third equality follows from the definition of $\Omega_{2,F_n}(\theta)$, and the last equality holds by Lemma B.2.(iii) and (E.12). This shows the consistency of $\overline{\Omega}_2$. ∎

In the rest of the Supplemental Appendix, we use C denote a generic fixed positive finite constant whose value does not depend on F or n.

Proof of Lemma B.4. As $\overline{g}_1(\theta)$ is a subvector of $\overline{g}_2(\theta)$, and $\overline{\Omega}_{1,n}$ is a submatrix of $\overline{\Omega}_{2,n}$, using (E.10), (E.13) and Assumptions 3.2.(ii)-(iii), we have

$$\overline{g}_1(\theta)'(\overline{\Omega}_1)^{-1}\overline{g}_1(\theta) = M_{1,F_n}(\theta)'\Omega_{1,F_n}^{-1}M_{1,F_n}(\theta) + o_p(1), \tag{E.14}$$

uniformly over Θ. By Assumptions 3.2.(ii)-(iii),

$$C^{-1} \leq \rho_{\min}(\Omega_{1,F_n}^{-1}) \leq \rho_{\max}(\Omega_{1,F_n}^{-1}) \leq C \tag{E.15}$$

which together with Assumptions 3.1.(i)-(ii) implies that $M_{1,F_n}(\theta)'\Omega_{1,F_n}^{-1}M_{1,F_n}(\theta)$ is uniquely minimized at θ_{F_n}. By the standard arguments for the consistency of an extremum estimator, we have

$$\widehat{\theta}_1 - \theta_{F_n} = o_p(1). \tag{E.16}$$

Using (E.16), Lemma B.1.(iv) and Assumption 3.2.(ii), we have

$$\begin{aligned}\overline{g}_1(\widehat{\theta}_1) &= \overline{g}_1(\theta_{F_n}) + \left[M_{1,F_n}(\widehat{\theta}_1) - M_{1,F_n}(\theta_{F_n})\right] + o_p(n^{-1/2}) \\ &= \overline{g}_1(\theta_{F_n}) + \left[G_{1,F_n}(\theta_{F_n}) + o_p(1)\right](\widehat{\theta}_1 - \theta_{F_n}) + o_p(n^{-1/2}).\end{aligned} \tag{E.17}$$

Similarly,

$$n^{-1}\sum_{i=1}^{n} g_{1,\theta}(W_i,\widehat{\theta}_1) = G_{1,F_n}(\widehat{\theta}_1) + o_p(1) = G_{1,F_n} + o_p(1), \tag{E.18}$$

where the first equality follows from Lemma B.1.(iii) and the second equality follows by (E.16) and Lemma B.2.(ii). From the first order condition for the GMM estimator $\widehat{\theta}_1$, we deduce that

$$\begin{aligned}0 &= \left[n^{-1}\sum_{i=1}^{n} g_{1,\theta}(W_i,\widehat{\theta}_1)\right]'(\overline{\Omega}_1)^{-1}\overline{g}_1(\widehat{\theta}_1) \\ &= (G_{1,F_n}'\Omega_{1,F_n}^{-1} + o_p(1))\left[\overline{g}_1(\theta_{F_n}) + (G_{1,F_n} + o_p(1))(\widehat{\theta}_1 - \theta_{F_n}) + o_p(n^{-1/2})\right]\end{aligned} \tag{E.19}$$

where the second equality follows from Assumptions 3.2.(ii)-(iii), (E.13), (E.17) and (E.18). By (E.19), $\mathbb{E}_{F_n}\left[g_1(W,\theta_{F_n})\right] = 0$ and Assumption 3.2,

$$n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) = (\Gamma_{1,F_n} + o_p(1))\mu_n(g_1(W,\theta_{F_n})) + o_p(1). \tag{E.20}$$

By Assumptions 3.2 and Lemma B.1.(v), $\Gamma_{1,F_n} = O(1)$ and $\mu_n(g_1(W, \theta_{F_n})) = o_p(1)$, which together with (E.20) implies that

$$n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) = \Gamma_{1,F_n}\mu_n(g_1(W, \theta_{F_n})) + O_p(1),$$

where $\Gamma_{1,F_n}\mu_n(g_1(W, \theta_{F_n})) = O_p(1)$. This finishes the proof. ■

Proof of Lemma B.5. By (E.10), (E.13) and Assumptions 3.2.(ii)-(iii), we have

$$\overline{g}_2(\theta)'(\overline{\Omega}_2)^{-1}\overline{g}_2(\theta) = M_{2,F_n}(\theta)'\Omega^{-1}_{2,F_n}M_{2,F_n}(\theta) + o_p(1) = Q_{F_n}(\theta) + o_p(1) \tag{E.21}$$

uniformly over Θ. By Assumption 3.1.(iii), $Q_{F_n}(\theta)$ is uniquely minimized at $\theta^*_{F_n}$. The consistency result $\widehat{\theta}_2 - \theta^*_{F_n} \rightarrow_p 0$ follows from standard arguments for the consistency of an extremum estimator. ■

Proof of Lemma B.6. By the definition of $\widehat{\theta}_2$,

$$\overline{g}_2(\widehat{\theta}_2)'(\overline{\Omega}_2)^{-1}\overline{g}_2(\widehat{\theta}_2) \leq \overline{g}_2(\theta_{F_n})'(\overline{\Omega}_2)^{-1}\overline{g}_2(\theta_{F_n}), \tag{E.22}$$

which implies that

$$||\overline{g}_2(\widehat{\theta}_2)||^2 \leq \rho_{\max}(\overline{\Omega}_2)\rho^{-1}_{\min}(\overline{\Omega}_2) \left\|\overline{g}_2(\theta_{F_n})\right\|^2. \tag{E.23}$$

By (E.13) and Assumptions 3.2.(ii)-(iii),

$$C^{-1} \leq \rho_{\min}(\overline{\Omega}_2) \leq \rho_{\max}(\overline{\Omega}_2) \leq C \tag{E.24}$$

with probability approaching 1. By Lemma B.1.(i), $M_{1,F_n}(\theta_{F_n}) = 0_{r_1 \times 1}$ and $\delta_{F_n} = o(1)$,

$$\left\|\overline{g}_2(\theta_{F_n})\right\|^2 = o_p(1) \tag{E.25}$$

which combined with (E.23) and (E.24) implies that

$$||\overline{g}_2(\widehat{\theta}_2)|| = o_p(1). \tag{E.26}$$

Moreover, by (E.26), Lemma B.1.(i) and the triangle inequality,

$$||M_{2,F_n}(\widehat{\theta}_2)|| \leq ||\overline{g}_2(\widehat{\theta}_2) - M_{2,F_n}(\widehat{\theta}_2)|| + ||\overline{g}_2(\widehat{\theta}_2)|| = o_p(1) \tag{E.27}$$

which immediately implies that

$$||M_{1,F_n}(\widehat{\theta}_2)|| = o_p(1). \tag{E.28}$$

The first result in Lemma B.6 follows by (E.28) and the unique identification of θ_{F_n} maintained by Assumptions 3.1.(i)-(ii).

Using $\widehat{\theta}_2 - \theta_{F_n} = o_p(1)$, Lemma B.1.(iv) and Assumption 3.2.(ii), we have

$$\begin{aligned}\overline{g}_2(\widehat{\theta}_2) &= \overline{g}_2(\theta_{F_n}) + \left[M_{2,F_n}(\widehat{\theta}_2) - M_{2,F_n}(\theta_{F_n})\right] + o_p(n^{-1/2}) \\ &= \overline{g}_2(\theta_{F_n}) + [G_{2,F_n}(\theta_{F_n}) + o_p(1)]\,(\widehat{\theta}_2 - \theta_{F_n}) + o_p(n^{-1/2}). \end{aligned} \tag{E.29}$$

Similarly,

$$n^{-1}\sum_{i=1}^{n} g_{2,\theta}(W_i,\widehat{\theta}_2) = G_{2,F_n}(\widehat{\theta}_2) + o_p(1) = G_{2,F_n}(\theta_{F_n}) + o_p(1), \tag{E.30}$$

where the first equality follows from Lemma B.1.(iii) and the second equality follows by $\widehat{\theta}_2 - \theta_{F_n} = o_p(1)$ and Lemma B.2.(ii). From the first order condition for the GMM estimator $\widehat{\theta}_2$, we deduce that

$$\begin{aligned} 0 &= \left[n^{-1}\sum_{i=1}^{n} g_{2,\theta}(W_i,\widehat{\theta}_2)\right]' (\overline{\Omega}_2)^{-1}\overline{g}_2(\widehat{\theta}_2) \\ &= (G'_{2,F_n}\Omega^{-1}_{2,F_n} + o_p(1))\left[\overline{g}_2(\theta_{F_n}) + (G_{2,F_n} + o_p(1))(\widehat{\theta}_2 - \theta_{F_n}) + o_p(n^{-1/2})\right] \end{aligned} \tag{E.31}$$

where the second equality follows from Assumptions 3.2.(ii)-(iii), (E.13), (E.29) and (E.30). By (E.31) and Assumption 3.2,

$$n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) = (\Gamma_{2,F_n} + o_p(1))\left\{\mu_n(g_2(W,\theta_{F_n})) + n^{1/2}\mathbb{E}_{F_n}\left[g_2(W,\theta_{F_n})\right]\right\} + o_p(1), \tag{E.32}$$

where $\Gamma_{2,F_n} = -\left(G'_{2,F_n}\Omega^{-1}_{2,F_n}G_{2,F_n}\right)^{-1} G'_{2,F_n}\Omega^{-1}_{2,F_n}$. ■

Proof for the claim in equation (4.3).

Consider the case $n^{1/2}\delta_{F_n} \to d \in \mathbb{R}^{r^*}$. By Lemma 4.1,

$$\begin{aligned} n^{1/2}\left[\widehat{\theta}(\omega) - \theta_{F_n}\right] &= n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) + \omega\left[n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) - n^{1/2}(\widehat{\theta}_1 - \theta_{F_n})\right] \\ &\to_D \Gamma^*_{1,F}\mathcal{Z}_{d,2,F} + \omega(\Gamma_{2,F} - \Gamma^*_{1,F})\mathcal{Z}_{d,2,F}, \end{aligned} \tag{E.33}$$

where $\mathcal{Z}_{d,2,F}$ has the same distribution as $\mathcal{Z}_{2,F} + d_0$. This implies that

$$\ell(\widehat{\theta}(\omega)) = n\left[\widehat{\theta}_n(\omega) - \theta_{F_n}\right]' \Upsilon \left[\widehat{\theta}_n(\omega) - \theta_{F_n}\right] \to_D \lambda_F(\omega) \tag{E.34}$$

where

$$\begin{aligned} \lambda_F(\omega) &= \mathcal{Z}'_{d,2,F}\Gamma^{*\prime}_{1,F}\Upsilon\Gamma^*_{1,F}\mathcal{Z}_{d,2,F} + 2\omega\mathcal{Z}'_{d,2,F}(\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon\Gamma^*_{1,F}\mathcal{Z}_{d,2,F} \\ &\quad + \omega^2\mathcal{Z}'_{d,2,F}(\Gamma_{2,F} - \Gamma^*_{1,F})'\Upsilon(\Gamma_{2,F} - \Gamma^*_{1,F})\mathcal{Z}_{d,2,F}. \end{aligned}$$

Now we consider $\mathbb{E}[\lambda_F(\omega)]$ using the equalities in Lemma B.9 below. First,

$$\mathbb{E}[\mathcal{Z}_{d,2,F}'\Gamma_{1,F}^{*\prime}\Upsilon\Gamma_{1,F}^{*}\mathcal{Z}_{d,2,F}] = \text{tr}(\Upsilon\Sigma_{1,F}) \tag{E.35}$$

because $\Gamma_1^*\mathcal{Z}_{d,2,F} = \Gamma_{1,F}\mathcal{Z}_{1,F}$ and $\Gamma_{1,F}\mathbb{E}[\mathcal{Z}_{1,F}\mathcal{Z}_{1,F}']\Gamma_{1,F}' = \Sigma_{1,F}$ by definition. Second,

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{Z}_{d,2,F}'(\Gamma_{2,F}-\Gamma_{1,F}^*)'\Upsilon\Gamma_{1,F}^*\mathcal{Z}_{d,2,F}\right] \\
&= \text{tr}(\Upsilon\Gamma_{1,F}^*\mathbb{E}\left[\mathcal{Z}_{d,2,F}\mathcal{Z}_{d,2,F}'\right](\Gamma_{2,F}-\Gamma_{1,F}^*)') \\
&= \text{tr}(\Upsilon\Gamma_{1,F}^*\left[d_0d_0'+\Omega_{2,F}\right](\Gamma_{2,F}-\Gamma_{1,F}^*)') \\
&= \text{tr}(\Upsilon(\Sigma_{2,F}-\Sigma_{1,F})),
\end{aligned} \tag{E.36}$$

where the last equality holds by Lemma B.9. Third,

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{Z}_{d,2,F}'(\Gamma_{2,F}-\Gamma_{1,F}^*)'\Upsilon(\Gamma_{2,F}-\Gamma_{1,F}^*)\mathcal{Z}_{d,2,F}\right] \\
&= \text{tr}(\Upsilon(\Gamma_{2,F}-\Gamma_{1,F}^*)\left[d_0d_0'+\Omega_{2,F}\right](\Gamma_{2,F}-\Gamma_{1,F}^*)') \\
&= d_0'\Gamma_{2,F}'\Upsilon\Gamma_{2,F}d_0 - \text{tr}(\Upsilon(\Sigma_{2,F}-\Sigma_{1,F}))
\end{aligned} \tag{E.37}$$

by Lemma B.9. Combining the results in (E.35)-(E.37), we obtain

$$\begin{aligned}
\mathbb{E}[\lambda_F(\omega)] \;=\; & \text{tr}(\Upsilon\Sigma_{1,F}) - 2\omega\text{tr}\left(\Upsilon\left(\Sigma_{1,F}-\Sigma_{2,F}\right)\right) \\
& +\omega^2\left[d_0'\Gamma_{2,F}'\Upsilon\Gamma_{2,F}d_0 + tr\left(\Upsilon\left(\Sigma_{1,F}-\Sigma_{2,F}\right)\right)\right].
\end{aligned} \tag{E.38}$$

Note that $d_0'\Gamma_{2,F}'\Upsilon\Gamma_{2,F}d_0 = d_0'(\Gamma_{2,F}-\Gamma_{1,F}^*)'\Upsilon(\Gamma_{2,F}-\Gamma_{1,F}^*)d_0$ because $\Gamma_{1,F}^*d_0 = \mathbf{0}_{d_\theta}$. It is clear that the optimal weight ω_F^* in (4.3) minimizes the quadratic function of ω in (E.38). ■

Proof of Lemma B.9. By construction, $\Gamma_{1,F}^*d_0 = \mathbf{0}_{d_\theta\times 1}$. For ease of notation, we write $\Omega_{2,F}$ and $G_{2,F}$ as

$$\Omega_{2,F} = \begin{pmatrix} \Omega_{1,F} & \Omega_{1r^*} \\ \Omega_{r^*1,F} & \Omega_{r^*,F} \end{pmatrix} \text{ and } G_{2,F} = \begin{pmatrix} G_{1,F} \\ G_{r^*,F} \end{pmatrix}. \tag{E.39}$$

To prove part (b), we have

$$\begin{aligned}
\Gamma_{1,F}^*\Omega_{2,F}\Gamma_{1,F}^* &= [\Gamma_{1,F},\mathbf{0}_{d_\theta\times r^*}]\begin{pmatrix} \Omega_{1,F} & \Omega_{1r^*} \\ \Omega_{r^*1,F} & \Omega_{r^*,F} \end{pmatrix}[\Gamma_{1,F},\mathbf{0}_{d_\theta\times r^*}] \\
&= \Gamma_{1,F}\Omega_{1,F}\Gamma_{1,F}' = \left(G_{1,F}'\Omega_{1,F}^{-1}G_{1,F}\right)^{-1} = \Sigma_{1,F}.
\end{aligned} \tag{E.40}$$

To show part (c), note that

$$\Gamma_{1,F}^{*}\Omega_{2,F}\Gamma_{2,F}' = -\left[\Gamma_{1,F}, \mathbf{0}_{d_\theta \times r^*}\right]\Omega_{2,F}\Omega_{2,F}^{-1}G_{2,F}\left(G_{2,F}'\Omega_{2,F}^{-1}G_{2,F}\right)^{-1}$$
$$= -\Gamma_{1,F}G_{1,F}\left(G_{2,F}'\Omega_{2,F}^{-1}G_{2,F}\right)^{-1} = \left(G_{2,F}'\Omega_{2,F}^{-1}G_{2,F}\right)^{-1} = \Sigma_{2,F} \tag{E.41}$$

because $-\Gamma_{1,F}G_{1,F} = I_{d_\theta \times d_\theta}$. Part (d) follows from the definition of $\Gamma_{2,F}$. ■

Proof of Lemma 4.2. We first prove the consistency of $\widehat{\Omega}_k$, $\widehat{G}_k$ and $\widehat{\Sigma}_k$ for $k = 1, 2$. By Lemma 4.1, we have $\widehat{\theta}_1 = \theta_{F_n} + o_p(1)$. Using the same arguments in showing (E.13), we can show that

$$\widehat{\Omega}_2 = \Omega_{2,F_n} + o_p(1) = \Omega_{2,F} + o_p(1), \tag{E.42}$$

where the second equality is by the assumption of the lemma that $v_{F_n} \to v_F$ for some $F \in \mathcal{F}$. As $\widehat{\Omega}_1$ is a submatrix of $\widehat{\Omega}_2$, by (E.42) we have

$$\widehat{\Omega}_1 = \Omega_{1,F_n} + o_p(1) = \Omega_{1,F} + o_p(1). \tag{E.43}$$

By the consistency of $\widehat{\theta}_1$ and the same arguments used to show (E.30), we have

$$n^{-1}\sum_{i=1}^{n} g_{2,\theta}(W_i, \widehat{\theta}_1) = G_{2,F_n}(\theta_{F_n}) + o_p(1) = G_{2,F} + o_p(1), \tag{E.44}$$

where the second equality is by (B.10) which is assumed in the lemma. As $n^{-1}\sum_{i=1}^{n} g_{1,\theta}(W_i, \widehat{\theta}_1)$ is a submatrix of $n^{-1}\sum_{i=1}^{n} g_{2,\theta}(W_i, \widehat{\theta}_1)$, by (E.44) we have

$$n^{-1}\sum_{i=1}^{n} g_{1,\theta}(W_i, \widehat{\theta}_1) = G_{1,F_n}(\theta_{F_n}) + o_p(1) = G_{1,F} + o_p(1). \tag{E.45}$$

From Assumption 3.2, (E.42), (E.43), (E.44) and (E.45), we see that $\widehat{\Omega}_k$ and $\widehat{G}_k$ are consistent estimators of $\Omega_{k,F}$ and $G_{k,F}$ respectively for $k = 1, 2$. By the Slutsky theorem and Assumption 3.2, we know that $\widehat{\Sigma}_k$ is a consistent estimator of $\Sigma_{k,F}$ for $k = 1, 2$.

In the case where $n^{1/2}\delta_{F_n} \to d \in \mathbb{R}^{r^*}$, the desired result follows from Lemma 4.1, the consistency of $\widehat{\Sigma}_{1,F}$ and $\widehat{\Sigma}_{2,F}$, and the CMT. In the case where $||n^{1/2}\delta_{F_n}|| \to \infty$, $\widetilde{\omega}_{eo} \to_p 0$ because $n^{1/2}||\widehat{\theta}_2 - \widehat{\theta}_1|| \to_p \infty$ and

$$n^{1/2}(\widehat{\theta}_{eo} - \theta_{F_n}) = n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) + \widetilde{\omega}_{eo}n^{1/2}(\widehat{\theta}_2 - \widehat{\theta}_1)$$
$$= n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) + \frac{n^{1/2}(\widehat{\theta}_2 - \widehat{\theta}_1)tr\left[\Upsilon(\widehat{\Sigma}_1 - \widehat{\Sigma}_2)\right]}{n(\widehat{\theta}_2 - \widehat{\theta}_1)'\Upsilon(\widehat{\theta}_2 - \widehat{\theta}_1) + tr\left[\Upsilon(\widehat{\Sigma}_1 - \widehat{\Sigma}_2)\right]} \to_D \xi_{1,F} \tag{E.46}$$

by Lemma 4.1. ■

Proof of Lemma B.15. By definition,

$$\xi_{1,F}'\Upsilon\xi_{1,F}=\mathcal{Z}_{1,F}'\Gamma_{1,F}'\Upsilon\Gamma_{1,F}\mathcal{Z}_{1,F}=\mathcal{Z}_1'\Omega_{1,F}^{1/2}\Gamma_{1,F}'\Upsilon\Gamma_{1,F}\Omega_{1,F}^{1/2}\mathcal{Z}_1 \tag{E.47}$$

where $\mathcal{Z}_1\sim N(\mathbf{0}_{r_1},I_{r_1\times r_1})$. By Assumptions 3.2.(ii) and 3.2.(iv), and the fact that Υ is a fixed matrix,

$$\sup_{F\in\mathcal{F}}\rho_{\max}(\Omega_{1,F}^{1/2}\Gamma_{1,F}'\Upsilon\Gamma_{1,F}\Omega_{1,F}^{1/2})\leq C. \tag{E.48}$$

By (E.48),

$$\sup_{h\in H}\mathbb{E}[(\xi_{1,F}'\Upsilon\xi_{1,F})^2]\leq\sup_{h\in H}\rho_{\max}^2(\Omega_{1,F}^{1/2}\Gamma_{1,F}'\Upsilon\Gamma_{1,F}\Omega_{1,F}^{1/2})\mathbb{E}[(\mathcal{Z}_1'\mathcal{Z}_1)^2]\leq 3r_1C \tag{E.49}$$

where the second inequality is by $\mathbb{E}[(\mathcal{Z}_1'\mathcal{Z}_1)^2]\leq 3r_1+r_1(r_1-1)=r_1^2+2r_1$ which is implied by the assumption that $\mathcal{Z}_1$ is a r_1-dimensional standard normal random vector. The first inequality of this lemma follows as the upper bound in (E.49) does not depend on F.

For any $F\in\mathcal{F}$, define

$$B_F\equiv(\Gamma_{2,F}-\Gamma_{1,F}^*)'\Upsilon(\Gamma_{2,F}-\Gamma_{1,F}^*).$$

By the Cauchy-Schwarz inequality and the simple inequality $|ab|\leq(a^2+b^2)/2$ (for any real numbers a and b),

$$\begin{aligned}\overline{\xi}_F'\Upsilon\overline{\xi}_F &\leq 2\left(\mathcal{Z}_{d,2,F}'\Gamma_{1,F}^{*\prime}\Upsilon\Gamma_{1,F}^*\mathcal{Z}_{d,2,F}+\overline{\omega}_F^2\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}\right)\\ &= 2\left(\mathcal{Z}_{1,F}'\Gamma_{1,F}'\Upsilon\Gamma_{1,F}\mathcal{Z}_{1,F}+\overline{\omega}_F^2\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}\right)\end{aligned} \tag{E.50}$$

where the equality is by $\Gamma_{1,F}^*d_0=\mathbf{0}_{d_\theta\times 1}$ (which is proved in Lemma B.9). By (E.50) and the simple inequality $(a+b)^2\leq 2(a^2+b^2)$ (for any real numbers a and b),

$$(\overline{\xi}_F'\Upsilon\overline{\xi}_F)^2\leq 8(\mathcal{Z}_{1,F}'\Gamma_{1,F}'\Upsilon\Gamma_{1,F}\mathcal{Z}_{1,F})^2+8(\overline{\omega}_F^2\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F})^2. \tag{E.51}$$

By the first inequality of this lemma, we have $\sup_{h\in H}\mathbb{E}[(\xi_{1,F}'\Upsilon\xi_{1,F})^2]\leq C$. Hence by (E.51), to show the second inequality of this lemma, it is sufficient to prove that

$$\sup_{h\in H}\mathbb{E}[(\overline{\omega}_F^2\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F})^2]\leq C. \tag{E.52}$$

Recall that we have defined $A_F=\Upsilon\left(\Sigma_{1,F}-\Sigma_{2,F}\right)$ in Theorem 5.2. By the definition,

$$\begin{aligned}\overline{\omega}_F^2\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F} &= \frac{(\mathrm{tr}(A_F))^2\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}}{(\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F))^2}\\ &= \mathrm{tr}(A_F)\frac{\mathrm{tr}(A_F)}{\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)}\frac{\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}}{\mathcal{Z}_{d,2,F}'B_F\mathcal{Z}_{d,2,F}+\mathrm{tr}(A_F)}.\end{aligned} \tag{E.53}$$

By Lemma 2.1 in Cheng and Liao (2015), $\text{tr}(A_F) \geq 0$ for any $F \in \mathcal{F}$. This together with $\mathcal{Z}'_{d,2,F} B_F \mathcal{Z}_{d,2,F} \geq 0$ implies that

$$\frac{\text{tr}(A_F)}{\mathcal{Z}'_{d,2,F} B_F \mathcal{Z}_{d,2,F} + \text{tr}(A_F)} \leq 1 \text{ and } \frac{\mathcal{Z}'_{d,2,F} B_F \mathcal{Z}_{d,2,F}}{\mathcal{Z}'_{d,2,F} B_F \mathcal{Z}_{d,2,F} + \text{tr}(A_F)} \leq 1. \tag{E.54}$$

By (E.54) and $\text{tr}(A_F) \geq 0$,

$$\overline{\omega}_F^2 \mathcal{Z}'_{d,2,F} B_F \mathcal{Z}_{d,2,F} \leq \text{tr}(A_F) = \text{tr}(\Upsilon\Sigma_{1,F}) - \text{tr}(\Upsilon\Sigma_{2,F}), \tag{E.55}$$

where the equality is by $A_F = \Upsilon(\Sigma_{1,F} - \Sigma_{2,F})$. By (E.55) and the simple inequality $(a+b)^2 \leq 2(a^2+b^2)$,

$$\mathbb{E}[(\overline{\omega}_F^2 \mathcal{Z}'_{d,2,F} B_F \mathcal{Z}_{d,2,F})^2] \leq 2(\text{tr}(\Upsilon\Sigma_{1,F}))^2 + 2(\text{tr}(\Upsilon\Sigma_{2,F}))^2. \tag{E.56}$$

By Assumptions 3.2.(ii) and 3.2.(iv),

$$\rho_{\min}(G'_{k,F}\Omega_{k,F}^{-1}G_{k,F}) \geq \rho_{\min}(\Omega_{k,F}^{-1})\rho_{\min}(G'_{k,F}G_{k,F}) = \rho_{\min}(G'_{k,F}G_{k,F})/\rho_{\max}(\Omega_{k,F}) \geq C^{-1} \tag{E.57}$$

for any $F \in \mathcal{F}$ and for $k = 1, 2$. By (E.57) and the definition of $\Sigma_{k,F}$ $(k = 1, 2)$,

$$\rho_{\max}(\Sigma_{k,F}) = \rho_{\min}^{-1}(G'_{k,F}\Omega_{k,F}^{-1}G_{k,F}) \leq C \tag{E.58}$$

for any $F \in \mathcal{F}$. As Υ and $\Sigma_{k,F}$ are positive definite symmetric matrix, by the standard trace inequality ($tr(AB) \leq tr(A)\rho_{\max}(B)$ for Hermitian matrices $A \geq 0$ and $B \geq 0$),

$$\text{tr}(\Upsilon\Sigma_{k,F}) \leq \text{tr}(\Upsilon)\rho_{\max}(\Sigma_{k,F}) \leq C \text{ for } k = 1, 2, \tag{E.59}$$

for any $F \in \mathcal{F}$. Collecting the results in (E.56) and (E.59), we immediately get (E.52). This finishes the proof. ∎

Proof of Lemma B.16. First note that

$$\min\{x, \zeta\} - x = (\zeta - x)I\{x > \zeta\}. \tag{E.60}$$

Hence we have

$$\begin{aligned}
& \sup_{h\in H} \left|\mathbb{E}\left[\min\{\overline{\xi}'_F\Upsilon\overline{\xi}_F, \zeta\} - \overline{\xi}'_F\Upsilon\overline{\xi}_F\right]\right| \\
\leq\ & \sup_{h\in H} \mathbb{E}\left[\left|\zeta - \overline{\xi}'_F\Upsilon\overline{\xi}_F\right| I\{\overline{\xi}'_F\Upsilon\overline{\xi}_F > \zeta\}\right] \\
\leq\ & \zeta \sup_{h\in H} \mathbb{E}\left[I\{\overline{\xi}'_F\Upsilon\overline{\xi}_F > \zeta\}\right] + \sup_{h\in H} \mathbb{E}\left[\overline{\xi}'_F\Upsilon\overline{\xi}_F I\{\zeta^{-1} > (\overline{\xi}'_F\Upsilon\overline{\xi}_F)^{-1}\}\right] \\
\leq\ & 2\zeta^{-1} \sup_{h\in H} \mathbb{E}\left[(\overline{\xi}'_F\Upsilon\overline{\xi}_F)^2\right] \leq 2C\zeta^{-1}
\end{aligned} \tag{E.61}$$

where the first inequality is by the Jensen's inequality, the second inequality is by the Markov inequality, the third inequality is by the monotonicity of expectation and the last inequality is by Lemma B.15. Using the same arguments, we can show that

$$\sup_{h\in H}\left|\mathbb{E}\left[\min\{\xi_{1,F}'\Upsilon\xi_{1,F},\zeta\}-\xi_{1,F}'\Upsilon\xi_{1,F}\right]\right|\leq 2C\zeta^{-1}. \tag{E.62}$$

Collecting the results in (E.61) and (E.62), and applying the triangle inequality, we deduce that

$$\sup_{h\in H}\left[|g_{\zeta}(h)-g(h)|\right]\leq 4C\zeta^{-1}. \tag{E.63}$$

The claimed result of this lemma follows by (E.63) as C is a fixed constant.

By the triangle inequality, the Jensen's inequality and Lemma B.15,

$$\begin{aligned}\sup_{h\in H}|g(h)| &= \sup_{h\in H}\left|\mathbb{E}[\overline{\xi}_F'\Upsilon\overline{\xi}_F-\xi_{1,F}'\Upsilon\xi_{1,F}]\right|\\ &\leq \sup_{h\in H}\mathbb{E}[\overline{\xi}_F'\Upsilon\overline{\xi}_F]+\sup_{h\in H}\mathbb{E}[\xi_{1,F}'\Upsilon\xi_{1,F}]\leq C\end{aligned}$$

which finishes the proof of the lemma. ■

Proof of Lemma A.1. By definition

$$\begin{aligned}&\mathbb{E}\left[||\widehat{\theta}_{eo}-\theta||^2\right]-\mathbb{E}\left[||\widehat{\theta}_1-\theta||^2\right]\\ =\ &\mathbb{E}\left[\frac{k^2\sigma^4(Y-X)'(Y-X)}{(2k\sigma^2+(Y-X)'(Y-X))^2}\right]+\mathbb{E}\left[\frac{2k\sigma^2(X-\theta)'(Y-X)}{2k\sigma^2+(Y-X)'(Y-X)}\right]\end{aligned} \tag{E.64}$$

Let

$$J_1\equiv\mathbb{E}\left[\frac{(X-\theta)'(Y-X)}{2k\sigma^2+(Y-X)'(Y-X)}\right]\text{ and }J_2\equiv\mathbb{E}\left[\frac{(Y-X)'(Y-X)}{(2k\sigma^2+(Y-X)'(Y-X))^2}\right]. \tag{E.65}$$

Let $X^*=\sigma^{-1}(X-\theta)$, $Y^*=\sigma^{-1}(Y-\theta)$ and $Z^*=(X^{*\prime},Y^{*\prime})'$. Then we can write

$$\begin{aligned}J_1 &= \mathbb{E}\left[\frac{(X-\theta)'(Y-X)}{2k\sigma^2+(Y-X)'(Y-X)}\right]\\ &= \mathbb{E}\left[\frac{X^{*\prime}(Y^*-X^*)}{2k+(Y^*-X^*)'(Y^*-X^*)}\right]=\mathbb{E}\left[\frac{Z^{*\prime}D_1Z^*}{2k+Z^{*\prime}D_2Z^*}\right]\end{aligned} \tag{E.66}$$

where

$$D_1=\begin{pmatrix}-I_k & 0_k\\ I_k & 0_k\end{pmatrix}\text{ and }D_2=\begin{pmatrix}I_k & -I_k\\ -I_k & I_k\end{pmatrix}. \tag{E.67}$$

Note that

$$\mathbb{E}\left[D_1Z^*Z^{*\prime}D_1'\right]=D_2 \tag{E.68}$$

by definition and the Gaussian assumption. Let $\eta(x) = x/(x'D_2x + 2k)$. Its derivative is

$$\frac{\partial \eta(x)'}{\partial x} = \frac{1}{x'D_2x + 2k} I_k - \frac{2}{(x'D_2x + 2k)^2} D_2xx'. \tag{E.69}$$

By Lemma 1 of Hansen (2016), which is a matrix version of the Stein's Lemma (Stein, 1981),

$$\begin{aligned}
J_1 &= \mathbb{E}\left(\eta(Z^*)'D_1Z^*\right) = \mathbb{E}\left[\operatorname{tr}\left(\frac{\partial \eta(Z^*)'}{\partial x} D_1\right)\right] \\
&= \mathbb{E}\left[\frac{\operatorname{tr}(D_1)}{2k + Z^{*\prime}D_2Z^*}\right] - 2\mathbb{E}\left[\frac{\operatorname{tr}(D_2Z^*Z^{*\prime}D_1)}{(2k + Z^{*\prime}D_2Z^*)^2}\right] \\
&= \mathbb{E}\left[\frac{-k}{2k + Z^{*\prime}D_2Z^*}\right] - 2\mathbb{E}\left[\frac{Z^{*\prime}D_1D_2Z^*}{(2k + Z^{*\prime}D_2Z^*)^2}\right] \\
&= \mathbb{E}\left[\frac{-k}{2k + Z^{*\prime}D_2Z^*}\right] + 2\mathbb{E}\left[\frac{Z^{*\prime}D_2Z^*}{(2k + Z^{*\prime}D_2Z^*)^2}\right] \\
&= \mathbb{E}\left[\frac{2-k}{2k + Z^{*\prime}D_2Z^*}\right] + \mathbb{E}\left[\frac{-4k}{(2k + Z^{*\prime}D_2Z^*)^2}\right]
\end{aligned} \tag{E.70}$$

where the fourth equality follows from

$$D_1D_2 = \begin{pmatrix} -I_k & I_k \\ I_k & -I_k \end{pmatrix} = -D_2. \tag{E.71}$$

Moreover,

$$\begin{aligned}
k^2\sigma^4 J_2 &= \mathbb{E}\left[\frac{k^2\sigma^4(Y-X)'(Y-X)}{(2k\sigma^2 + (Y-X)'(Y-X))^2}\right] \\
&= \mathbb{E}\left[\frac{k^2\sigma^2}{2k + Z^{*\prime}D_2Z^*}\right] - \mathbb{E}\left[\frac{2k^3\sigma^2}{(2k + Z^{*\prime}D_2Z^*)^2}\right]
\end{aligned} \tag{E.72}$$

which together with (E.70) implies that

$$\begin{aligned}
&\mathbb{E}\left[||\widehat{\theta}_{eo} - \theta||^2\right] - \mathbb{E}\left[||\widehat{\theta}_1 - \theta||^2\right] \\
= &\ \sigma^2\mathbb{E}\left[\frac{2k(2-k) + k^2}{2k + Z^{*\prime}D_2Z^*}\right] - \sigma^2\mathbb{E}\left[\frac{2k^3 + 8k^2}{(2k + Z^{*\prime}D_2Z^*)^2}\right] \\
= &\ \sigma^2\mathbb{E}\left[\frac{k(4-k)}{2k + Z^{*\prime}D_2Z^*}\right] - \sigma^2\mathbb{E}\left[\frac{2k^2(k+4)}{(2k + Z^{*\prime}D_2Z^*)^2}\right].
\end{aligned} \tag{E.73}$$

The asserted result follows from the fact that D_2 is positive semi-definite and the second term on the right-hand side of the second equality of (E.73) is always negative. ■

F Asymptotic Risk of the Pre-test GMM Estimator

In this section, we establish similar results in Theorem 5.1 for the pre-test GMM estimator based on the J-test statistic. The pre-test estimator is defined as

$$\widehat{\theta}_{pre} = 1\{J_n > c_\alpha\}\widehat{\theta}_1 + 1\{J_n \leq c_\alpha\}\widehat{\theta}_2, \tag{F.1}$$

where $J_n = n\overline{g}_2(\widehat{\theta}_2)'(\widehat{\Omega}_2)^{-1}\overline{g}_2(\widehat{\theta}_2)$ and c_α is the $100(1-\alpha)$th quantile of the chi-squared distribution with degree of freedom $r_2 - d_\theta$.

Theorem F.1 *Suppose that Assumptions 3.1-3.3 hold. The bounds of the asymptotic risk difference satisfy*

$$AsyRD(\widehat{\theta}_{pre}, \widehat{\theta}_1) = \min\left\{\inf_{h\in H}[g_p(h)], 0\right\},$$
$$Asy\overline{RD}(\widehat{\theta}_{pre}, \widehat{\theta}_1) = \max\left\{\sup_{h\in H}[g_p(h)], 0\right\},$$

where $g_p(h) \equiv \mathbb{E}[\overline{\xi}_{p,F}'\Upsilon\overline{\xi}_{p,F} - \xi_{1,F}'\Upsilon\xi_{1,F}]$ *and* $\overline{\xi}_{p,F}$ *is defined in (F.3) below.*

Proof of Theorem F.1. The two equalities and inequalities in the theorem follow by the same arguments in the proof of Theorem 5.1 with Lemma 4.2 for $\widehat{\theta}_{eo}$ replaced by Lemma F.1 for $\widehat{\theta}_{pre}$, Lemma B.15 replaced by Lemma F.2, and Lemma B.16 replaced by Lemma F.3. Its proof is hence omitted. ■

By Definition,

$$\begin{aligned}\mathbb{E}[\overline{\xi}_{p,F}'\Upsilon\overline{\xi}_{p,F}] &= \mathbb{E}[\mathcal{Z}_{d,2,F}'\Gamma_{1,F}^{*\prime}\Upsilon\Gamma_{1,F}^{*}\mathcal{Z}_{d,2,F}] + 2\mathbb{E}[\overline{\omega}_{p,F}\mathcal{Z}_{d,2,F}'(\Gamma_{2,F} - \Gamma_{1,F}^{*})'\Upsilon\Gamma_{1,F}^{*}\mathcal{Z}_{d,2,F}] \\ &\quad + \mathbb{E}[\overline{\omega}_{p,F}^{2}\mathcal{Z}_{d,2,F}'(\Gamma_{2,F} - \Gamma_{1,F}^{*})'\Upsilon(\Gamma_{2,F} - \Gamma_{1,F}^{*})\mathcal{Z}_{d,2,F}] \\ &= tr(\Upsilon\Sigma_{1,F}) + 2\mathbb{E}[\overline{\omega}_{p,F}\mathcal{Z}_{d,2,F}'(\Gamma_{2,F} - \Gamma_{1,F}^{*})'\Upsilon\Gamma_{1,F}^{*}\mathcal{Z}_{d,2,F}] \\ &\quad + \mathbb{E}[\overline{\omega}_{p,F}^{2}\mathcal{Z}_{d,2,F}'(\Gamma_{2,F} - \Gamma_{1,F}^{*})'\Upsilon(\Gamma_{2,F} - \Gamma_{1,F}^{*})\mathcal{Z}_{d,2,F}] \end{aligned} \tag{F.2}$$

The asymptotic risk of the pre-test estimator $\widehat{\theta}_p$ in Figure 2 is simulated based on the formula in (F.2).

The following lemma provides the asymptotic distribution of the pre-test GMM estimator under various sequence of DGPs, which is used to show Theorem F.1.

Lemma F.1 *Suppose that Assumptions 3.1-3.3 hold. Consider* $\{F_n\}$ *such that* $v_{F_n} \to v_F$ *for some* $F \in \mathcal{F}$.

(a) *If* $n^{1/2}\delta_{F_n} \to d$ *for some* $d \in \mathbb{R}^{r^*}$, *then*

$$J_n \to_D J_\infty(h_{d,F}) \equiv (\mathcal{Z}_{2,F} + d_0)'L_F(\mathcal{Z}_{2,F} + d_0),$$

where $L_F \equiv \Omega_{2,F}^{-1} - \Omega_{2,F}^{-1} G_{2,F} \left(G'_{2,F} \Omega_{2,F}^{-1} G_{2,F} \right)^{-1} G'_{2,F} \Omega_{2,F}^{-1}$ *and* $d_0 = (0_{1\times r_1}, d')'$, *and*

$$n^{1/2}(\widehat{\theta}_{pre} - \theta_{F_n}) \to_D \bar{\xi}_{p,F} \equiv (1 - \overline{\omega}_{p,F})\xi_{1,F} + \overline{\omega}_{p,F}\xi_{2,F} \tag{F.3}$$

where $\overline{\omega}_{p,F} = 1\{J_\infty(h_{d,F}) \le c_\alpha\}$.

(b) *If* $||n^{1/2}\delta_{F_n}|| \to \infty$, *then* $\overline{\omega}_{p,F} \to_p 0$ *and* $n^{1/2}(\widehat{\theta}_{pre} - \theta_{F_n}) \to_D \xi_{1,F}$.

Proof of Lemma F.1. (a) By Assumption 3.2.(ii), (E.29) and (E.32),

$$\begin{aligned} \overline{g}_2(\widehat{\theta}_2) &= \overline{g}_2(\theta_{F_n}) + [G_{2,F_n}(\theta_{F_n}) + o_p(1)](\widehat{\theta}_2 - \theta_{F_n}) + o_p(n^{-1/2}) \\ &= \overline{g}_2(\theta_{F_n}) + G_{2,F_n}\Gamma_{2,F_n}\overline{g}_2(\theta_{F_n}) + o_p(n^{-1/2}) \\ &= (I_{r_2} + G_{2,F_n}\Gamma_{2,F_n})\overline{g}_2(\theta_{F_n}) + o_p(n^{-1/2}), \end{aligned} \tag{F.4}$$

which implies that

$$J_n = n\overline{g}_2(\theta_{F_n})' L_{F_n} \overline{g}_2(\theta_{F_n}) + o_p(1) \tag{F.5}$$

where $L_{F_n} \equiv \Omega_{2,F_n}^{-1} - \Omega_{2,F_n}^{-1} G_{2,F_n} (G'_{2,F_n} \Omega_{2,F_n}^{-1} G_{2,F_n})^{-1} G'_{2,F_n} \Omega_{2,F_n}^{-1}$.

By $n^{1/2}\delta_{F_n} \to d$ and Lemma B.1.(v),

$$n^{1/2}\Omega_{2,F_n}^{-1/2}\overline{g}_2(\theta_{F_n}) = \Omega_{2,F_n}^{-1/2}\mu_n(g_2(W,\theta_{F_n})) + \Omega_{2,F_n}^{-1/2} n^{1/2}\delta_{F_n} \to_D \mathcal{Z} + \Omega_{2,F}^{-1/2} d_0 \tag{F.6}$$

where $d_0' = (0_{1\times r_1}, d')$ and $\mathcal{Z}$ is a $r_2 \times 1$ standard normal random vector. By $v_{F_n} \to v_F$, (F.5), (F.6) and the CMT,

$$J_n \to_D (\mathcal{Z}_{2,F} + d_0)' L_F (\mathcal{Z}_{2,F} + d_0). \tag{F.7}$$

Recall that Lemma 4.1.(a) implies that

$$n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) \to_D \xi_{1,F} \text{ and } n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) \to_D \xi_{2,F}, \tag{F.8}$$

which together with (F.7) and the CMT implies that

$$\begin{aligned} n^{1/2}(\widehat{\theta}_{pre} - \theta_{F_n}) &= 1\{J_n > c_\alpha\} n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) + 1\{J_n \le c_\alpha\} n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) \\ &\to_D (1 - \overline{\omega}_{p,F})\xi_{1,F} + \overline{\omega}_{p,F}\xi_{2,F}, \end{aligned} \tag{F.9}$$

which finishes the proof of the claim in (a).

(b) There are two cases to consider: (i) $||\delta_{F_n}|| > C^{-1}$; and (ii) $||\delta_{F_n}|| \to 0$. We first consider case (i). As $\overline{g}_1(\widehat{\theta}_2)$ is a subvector of $\overline{g}_2(\widehat{\theta}_2)$,

$$\begin{aligned} J_n &= n\overline{g}_2(\widehat{\theta}_2)'(\widehat{\Omega}_2)^{-1}\overline{g}_2(\widehat{\theta}_2) \\ &\ge n\rho_{\max}^{-1}(\widehat{\Omega}_2)\overline{g}_2(\widehat{\theta}_2)'\overline{g}_2(\widehat{\theta}_2) \\ &\ge n\rho_{\max}^{-1}(\widehat{\Omega}_2)\overline{g}_1(\widehat{\theta}_2)'\overline{g}_1(\widehat{\theta}_2). \end{aligned} \tag{F.10}$$

By (B.22) and (B.23) in the Appendix of CLS

$$\left\|\widehat{\theta}_2 - \theta_{F_n}\right\| \geq C^{-1} \text{ with probability approaching 1,} \tag{F.11}$$

which together with Assumption 3.1.(ii) and Lemma B.1.(i) implies that

$$\overline{g}_1(\widehat{\theta}_2) = M_{1,F}(\widehat{\theta}_2) + o_p(1) \geq C \tag{F.12}$$

with probability approaching 1. By (E.42) and Assumption 3.2.(ii), we have

$$\rho_{\max}(\widehat{\Omega}_2) \leq C \text{ with probability approaching 1.} \tag{F.13}$$

Combining the results in (F.10), (F.12) and (F.13), we deduce that

$$J_n \geq nC^{-1} \text{ with probability approaching 1,} \tag{F.14}$$

which immediately implies that

$$\overline{\omega}_{p,F} = 1\{J_n \leq c_\alpha\} = 0 \tag{F.15}$$

with probability approaching 1, as c_α is a fixed constant. By Lemma 4.1.(b), (F.15) and the assumption that Θ is bounded, we have

$$\begin{aligned} n^{1/2}(\widehat{\theta}_{pre} - \theta_{F_n}) &= 1\{J_n > c_\alpha\}n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) + 1\{J_n \leq c_\alpha\}n^{1/2}(\widehat{\theta}_2 - \theta_{F_n}) \\ &= 1\{J_n > c_\alpha\}n^{1/2}(\widehat{\theta}_1 - \theta_{F_n}) + o_p(1) \to_D \xi_{1,F} \end{aligned} \tag{F.16}$$

where the convergence in distribution is by the CMT.

We next consider the case that $||\delta_{F_n}|| \to 0$ and $||n^{1/2}\delta_{F_n}|| \to \infty$. In the proof of Lemma 4.1, we have shown that $\widehat{\theta}_2 - \theta_{F_n} = o_p(1)$, and that (F.4) and (F.5) hold in this case. It is clear that

$$n^{1/2}\overline{g}_2(\theta_{F_n}) = \mu_n(g_2(W, \theta_{F_n}) + \begin{pmatrix} 0_{r_1\times 1} \\ n^{1/2}\delta_{F_n} \end{pmatrix} \tag{F.17}$$

which implies that

$$\begin{aligned} n\overline{g}_2(\theta_{F_n})' L_{F_n}\overline{g}_2(\theta_{F_n}) &= [\mu_n(g_2(W, \theta_{F_n})]' L_{F_n}[\mu_n(g_2(W, \theta_{F_n})] \\ &\quad +2\begin{pmatrix} 0_{1\times r_1} & n^{1/2}\delta'_{F_n} \end{pmatrix} L_{F_n}[\mu_n(g_2(W, \theta_{F_n})] \\ &\quad +\begin{pmatrix} 0_{1\times r_1} & n^{1/2}\delta'_{F_n} \end{pmatrix} L_{F_n}\begin{pmatrix} 0_{1\times r_1} & n^{1/2}\delta'_{F_n} \end{pmatrix}'. \end{aligned} \tag{F.18}$$

By Lemma B.1.(v) and Assumptions 3.2.(ii)-(iii),

$$[\mu_n(g_2(W,\theta_{F_n})]' L_{F_n}[\mu_n(g_2(W,\theta_{F_n})] = O_p(1). \tag{F.19}$$

In order to bound the third term in (F.18) from below, we shall show that for any $d_0 = (0_{1\times r_1}, d')'$ for $d \in \mathbb{R}^{r_*}$ with $\|d\| = 1$,

$$d_0' L_{F_n} d_0 \geq C^{-1} \tag{F.20}$$

By definition, L_{F_n} has d_θ many zero eigenvalues and $r_2 - d_\theta$ many of eigenvalues of ones. The matrix G_{2,F_n} contains the d_θ many eigenvectors of the zero eigenvalues of L_{F_n}, because

$$L_{F_n} G_{2,F_n} = 0_{r_2\times d_\theta} \text{ and } \rho_{\min}(G_{2,F_n}' G_{2,F_n}) \geq C^{-1}. \tag{F.21}$$

Let $G_{\perp,F_n}$ denote the orthogonal complement of G_{2,F_n} with $G_{\perp,F_n}' G_{\perp,F_n} = I_{r_2-d_\theta}$. Then we have

$$\begin{pmatrix} G_{1,F_n} \\ G_{r^*,F_n} \end{pmatrix} a_1 + \begin{pmatrix} G_{1,\perp,F_n} \\ G_{r^*,\perp,F_n} \end{pmatrix} a_2 = \begin{pmatrix} 0_{r_1\times 1} \\ d \end{pmatrix} \tag{F.22}$$

for some constant vectors $a_1 \in \mathbb{R}^{d_\theta}$ and $a_2 \in \mathbb{R}^{r_2-d_\theta}$. As $\rho_{\min}(G_{1,F_n}' G_{1,F_n}) \geq C^{-1}$ by Assumption 3.2, we have

$$a_1 = -(G_{1,F_n}' G_{1,F_n})^{-1} G_{1,F_n}' G_{1,\perp,F_n} a_2 \tag{F.23}$$

and

$$(G_{r^*,\perp,F_n} - G_{r^*,F_n}(G_{1,F_n}' G_{1,F_n})^{-1} G_{1,F_n}' G_{1,\perp,F_n}) a_2 = d. \tag{F.24}$$

Let $H_{F_n} = G_{r^*,\perp,F_n} - G_{r^*,F_n}(G_{1,F_n}' G_{1,F_n})^{-1} G_{1,F_n}' G_{1,\perp,F_n}$. By $\rho_{\min}(G_{1,F_n}' G_{1,F_n}) \geq C^{-1}$, Assumptions 3.2.(ii), (F.24) and the Cauchy-Schwarz inequality,

$$\|d\|^2 = a_2' H_{F_n} H_{F_n}' a_2 \leq C \|a_2\|^2 \tag{F.25}$$

which together with $\|d\| = 1$ implies that

$$\|a_2\|^2 \geq C^{-1}. \tag{F.26}$$

Using (F.21), (F.22) and (F.26), we deduce that

$$\begin{aligned} d_0' L_{F_n} d_0 &= (G_{2,F_n} a_1 + G_{\perp,F_n} a_2)' L_{F_n} (G_{2,F_n} a_1 + G_{\perp,F_n} a_2) \\ &= a_2' G_{\perp,F_n}' L_{F_n} G_{\perp,F_n} a_2 = \|a_2\|^2 \geq C^{-1} \end{aligned} \tag{F.27}$$

which proves (F.20). By (F.20),

$$\begin{pmatrix} 0_{1\times r_1} & n^{1/2}\delta_{F_n}' \end{pmatrix} L_{F_n} \begin{pmatrix} 0_{1\times r_1} & n^{1/2}\delta_{F_n}' \end{pmatrix}' \geq C^{-1} n ||\delta_{F_n}||^2 \tag{F.28}$$

which together with $n||\delta_{F_n}||^2 \to \infty$ implies that

$$\left(\begin{array}{cc} 0_{1\times r_1} & n^{1/2}\delta'_{F_n} \end{array} \right) L_{F_n} \left(\begin{array}{cc} 0_{1\times r_1} & n^{1/2}\delta'_{F_n} \end{array} \right)' \to \infty. \tag{F.29}$$

Collecting the results in (F.18), (F.19) and (F.19),and by the Cauchy-Schwarz inequality, we deduce that $n\overline{g}_2(\theta_{F_n})'L_{F_n}\overline{g}_2(\theta_{F_n}) \to_p \infty$, which together with (F.5) implies that

$$J_n \to_p \infty. \tag{F.30}$$

Using the same arguments in showing (F.16), we deduce that

$$n^{1/2}(\widehat{\theta}_{pre} - \theta_{F_n}) \to_D \xi_{1,F}. \tag{F.31}$$

This finishes the proof. ■

Lemma F.2 *Under Assumptions 3.2, we have*

$$\sup_{h\in H} \mathbb{E}[(\overline{\xi}'_{p,F}\Upsilon\overline{\xi}_{p,F})^2] \leq C. \tag{F.32}$$

Proof of Lemma F.2. By the same arguments in showing (E.51), we have

$$(\overline{\xi}'_{p,F}\Upsilon\overline{\xi}_{p,F})^2 \leq 8(\mathcal{Z}'_{1,F}\Gamma'_{1,F}\Upsilon\Gamma_{1,F}\mathcal{Z}_{1,F})^2 + 8(\overline{\omega}^2_{p,F}\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F})^2. \tag{F.33}$$

By the first inequality in (B.58) in the Appendix of CLS, we have $\sup_{h\in H}\mathbb{E}[(\xi'_{1,F}\Upsilon\xi_{1,F})^2] \leq C$. Hence by (F.33), to show the inequality in (F.32), it is sufficient to prove that

$$\sup_{h\in H} \mathbb{E}[(\overline{\omega}^2_{p,F}\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F})^2] \leq C. \tag{F.34}$$

By definition,

$$\overline{\omega}_{p,F} = I\{J_\infty(h_{d,F}) \leq c_\alpha\} = I\{\mathcal{Z}'_{d,2,F}L_F\mathcal{Z}_{d,2,F} \leq c_\alpha\}. \tag{F.35}$$

By the simple inequality $(a+b)^2 \geq a^2/2 - 2b^2$,

$$(z+d_0)'L_F(z+d_0) \geq d'_0L_Fd_0/2 - 2z'L_Fz \tag{F.36}$$

for any $z \in \mathbb{R}$, which together with Assumption 3.2 and (F.20) implies that

$$(z+d_0)'L_F(z+d_0) \geq \|d\|^2/C - 2z'L_Fz \geq \|d\|^2/C - C\|z\|^2. \tag{F.37}$$

Under Assumption 3.2, $\|B_F\| \leq C$ for any $F \in \mathcal{F}$ which together with the simple inequality

$(a+b)^2 \leq 2(a^2+b^2)$ implies that

$$(z+d_0)'B_F(z+d_0) \leq 2C(\|d\|^2 + \|z\|^2) \tag{F.38}$$

for any $z \in \mathbb{R}$. Collecting the results in (F.36) and (F.38), we get

$$\begin{aligned} & I\{(z+d_0)'L_F(z+d_0) \leq c_\alpha\}z'B_F z \\ \leq\ & 2CI\{\|d\|^2 \leq c_\alpha C + C^2\|z\|^2\}(\|d\|^2+\|z\|^2) \\ \leq\ & 2C(c_\alpha C + (C^2+1)\|z\|^2) \end{aligned} \tag{F.39}$$

which implies that

$$\begin{aligned} & \sup_{h\in H} \mathbb{E}[(\overline{\omega}^2_{p,F}\mathcal{Z}'_{d,2,F}B_F\mathcal{Z}_{d,2,F})^2] \\ \leq\ & 4C^2\mathbb{E}[(c_\alpha C + (C^2+1)\mathcal{Z}'_{2,F}\mathcal{Z}_{2,F})^2] \\ \leq\ & C(c_\alpha + \mathbb{E}[(\mathcal{Z}'_{2,F}\mathcal{Z}_{2,F})^2]) = C(c_\alpha + 3\rho^2_{\max}(\Omega_2)r_2). \end{aligned} \tag{F.40}$$

This finishes the proof. ■

Lemma F.3 *Let* $g_{p,\zeta}(h) \equiv \mathbb{E}\left[\min\{\overline{\xi}'_{p,F}\Upsilon\overline{\xi}_{p,F},\zeta\} - \min\{\xi'_{1,F}\Upsilon\xi_{1,F},\zeta\}\right]$. *Under Assumptions 3.2, we have*

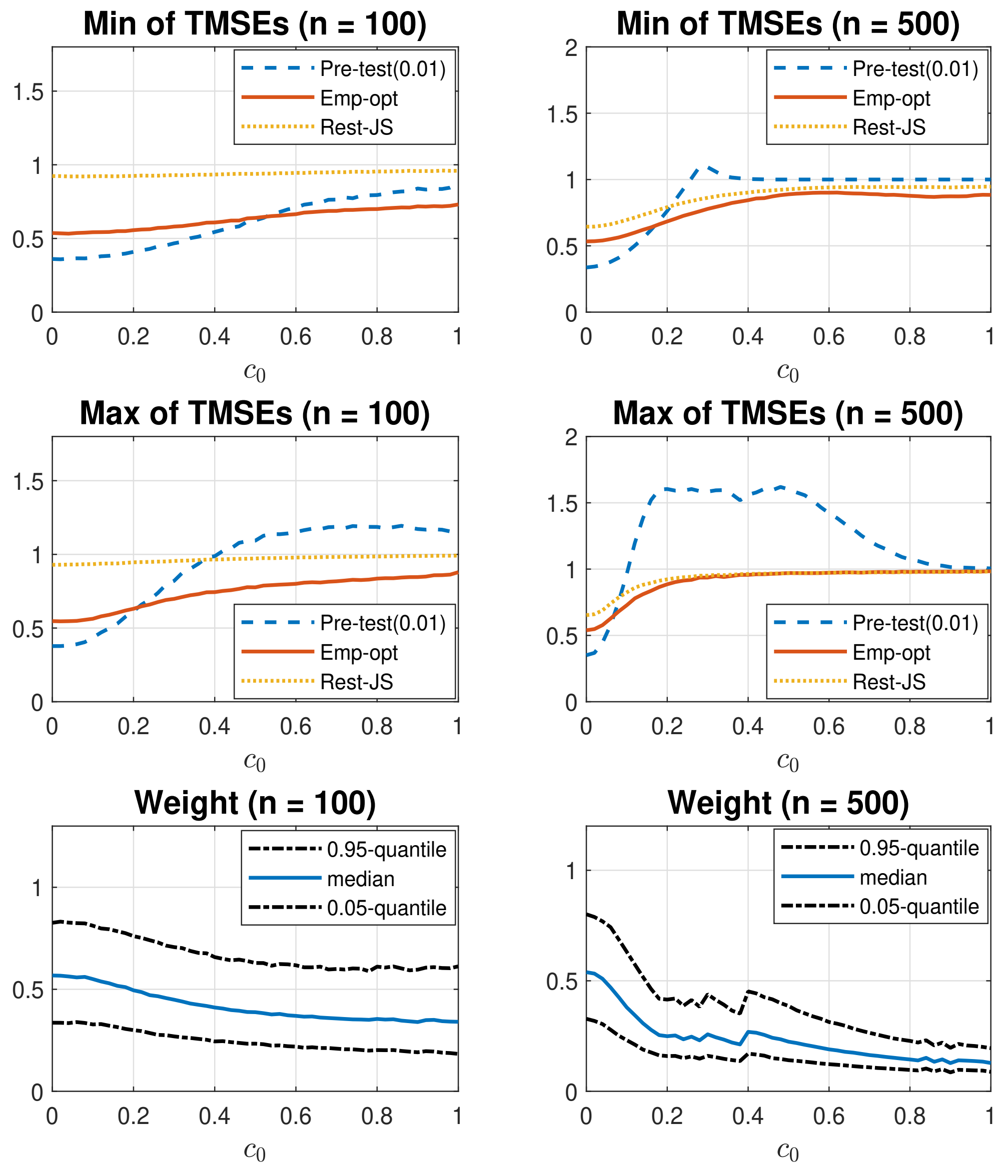
$$\lim_{\zeta\to\infty}\sup_{h\in H}[|g_{p,\zeta}(h) - g_p(h)|] = 0 \tag{F.41}$$

where $\sup_{h\in H}[|g_p(h)|] \leq C$.

Proof of Lemma F.3. The proof follows the same arguments of the proof of Lemma B.16 with the second inequality in (B.58) in the Appendix of CLS replaced by (F.32). ■

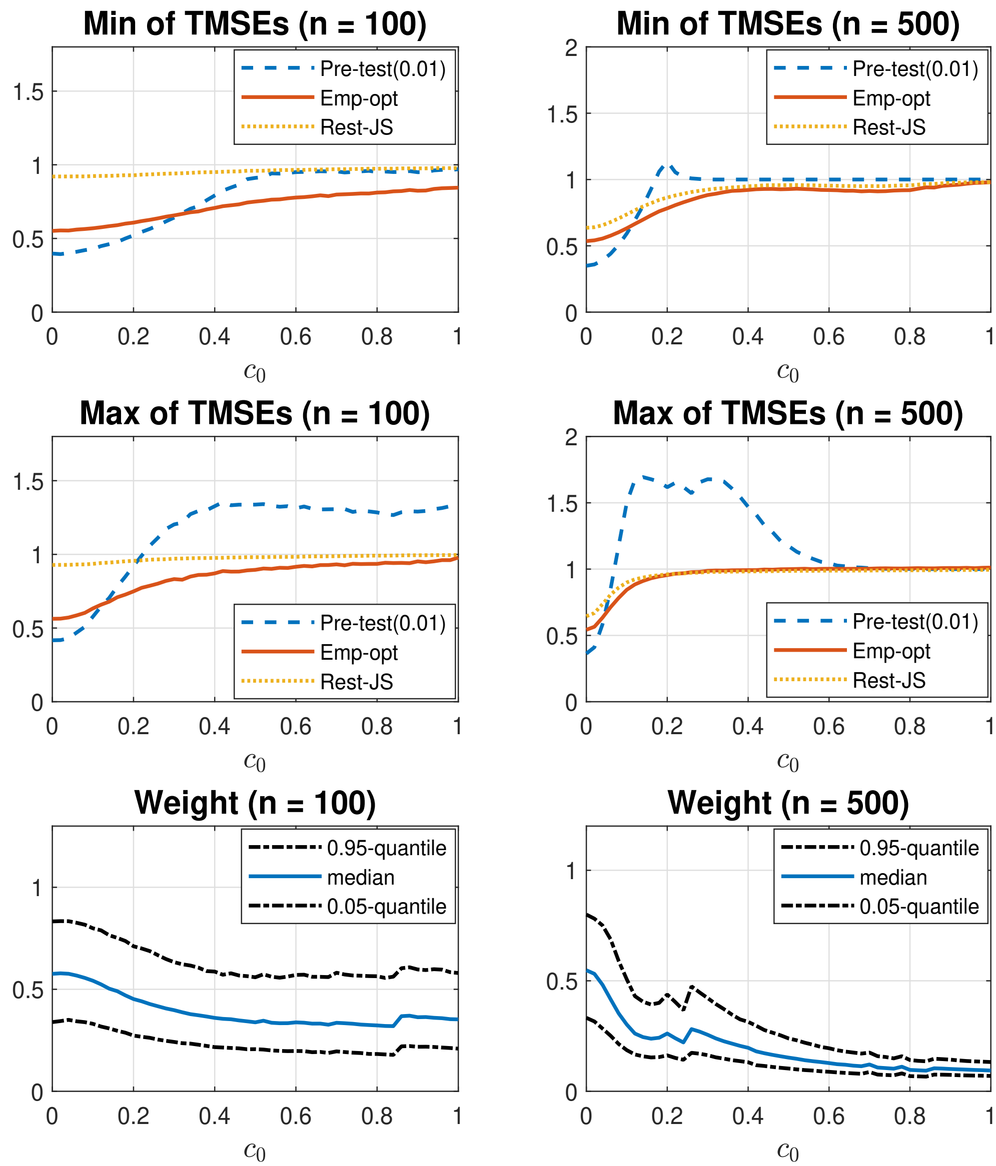
G Simulation Results on Truncated Risk for Section 6

Figure G.1: Finite Sample Truncated MSEs of the Pre-test and Averaging GMM Estimators in S1



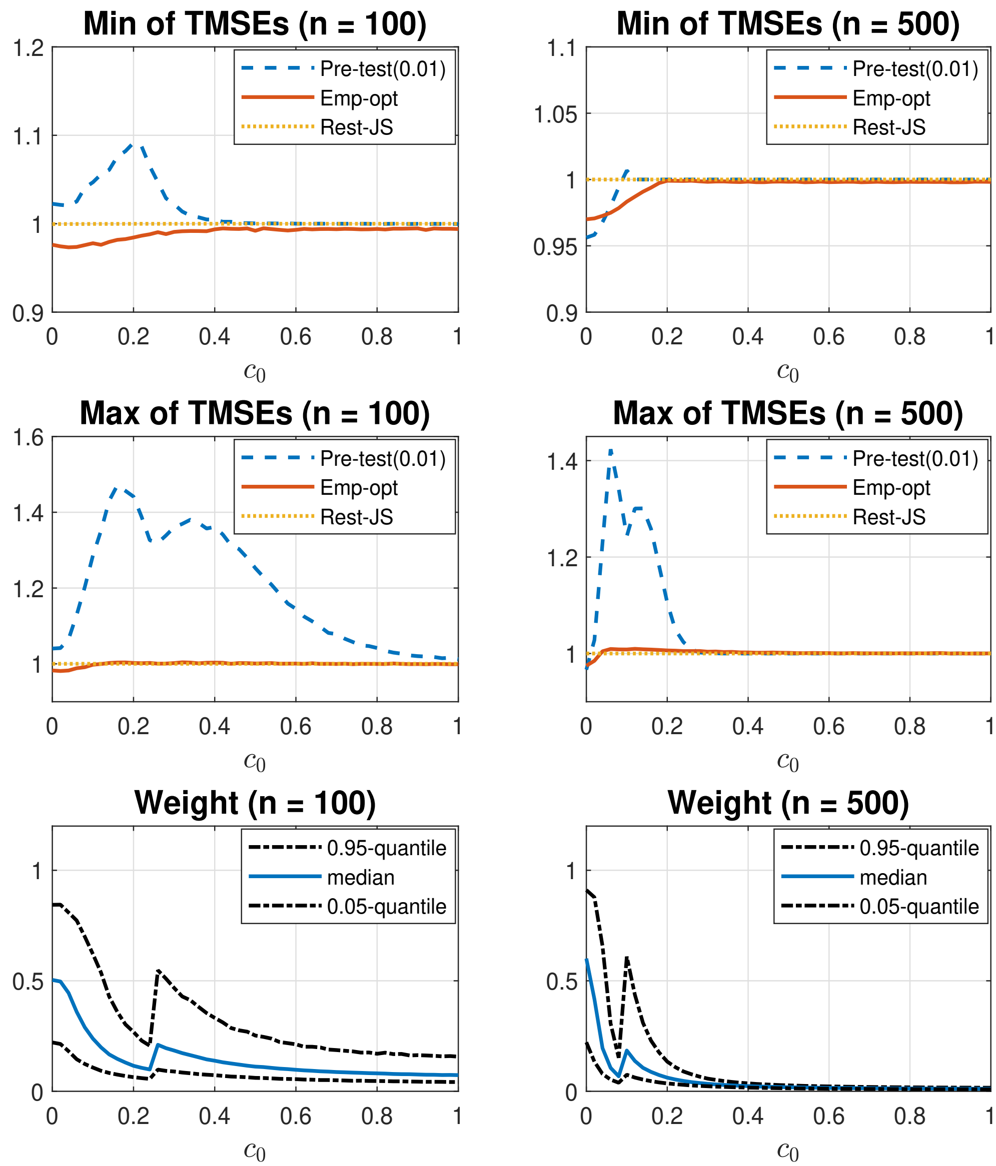
Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively. The truncation parameter for the truncated MSE is $\zeta = 1000$.

Figure G.2: Finite Sample Truncated MSEs of the Pre-test and Averaging GMM Estimators in S2



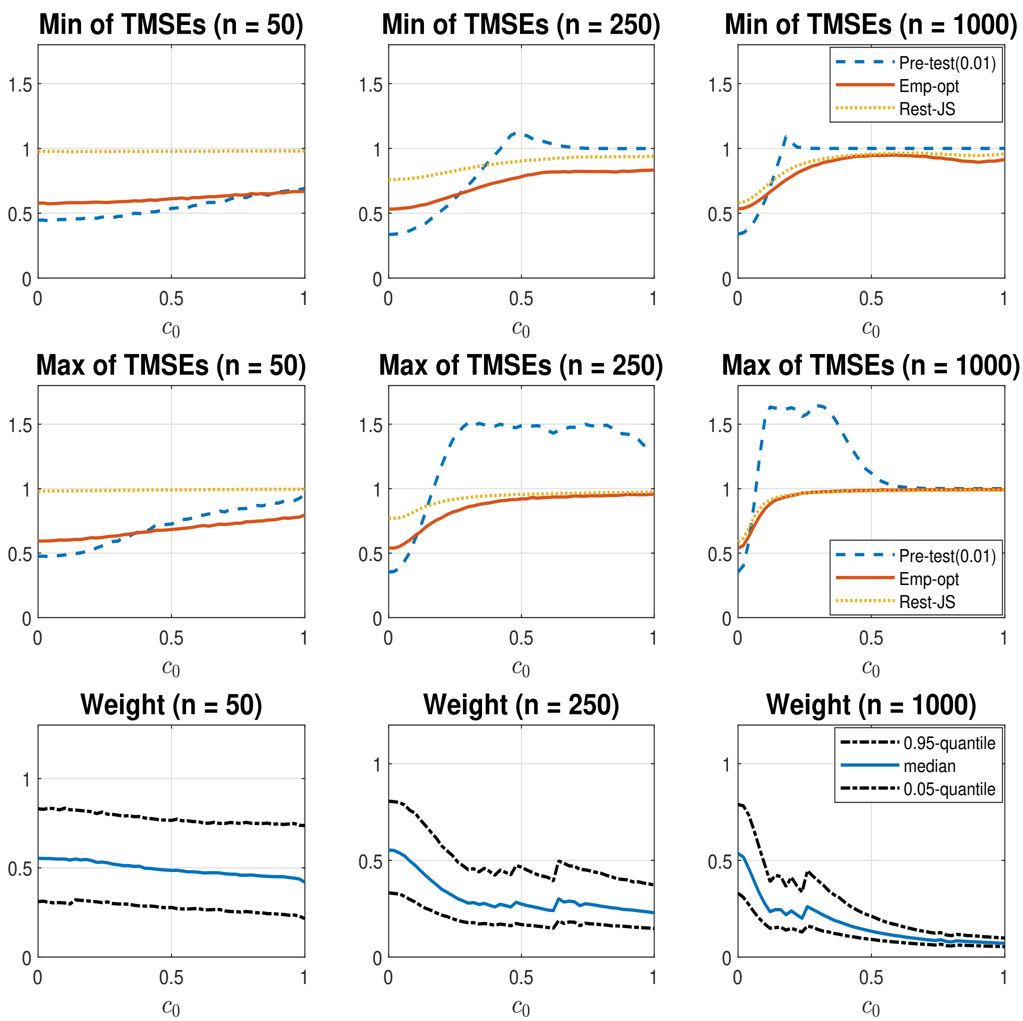
Note: “Pre-test(0.01)” refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; “Emp-opt” refers to the averaging GMM estimator based on the empirical optimal weight; “Rest-JS” refers to the averaging estimators based on the restricted James-Stein weight, respectively. The truncation parameter for the truncated MSE is $\zeta = 1000$.

Figure G.3: Finite Sample Truncated MSEs of the Pre-test and Averaging GMM Estimators in S3



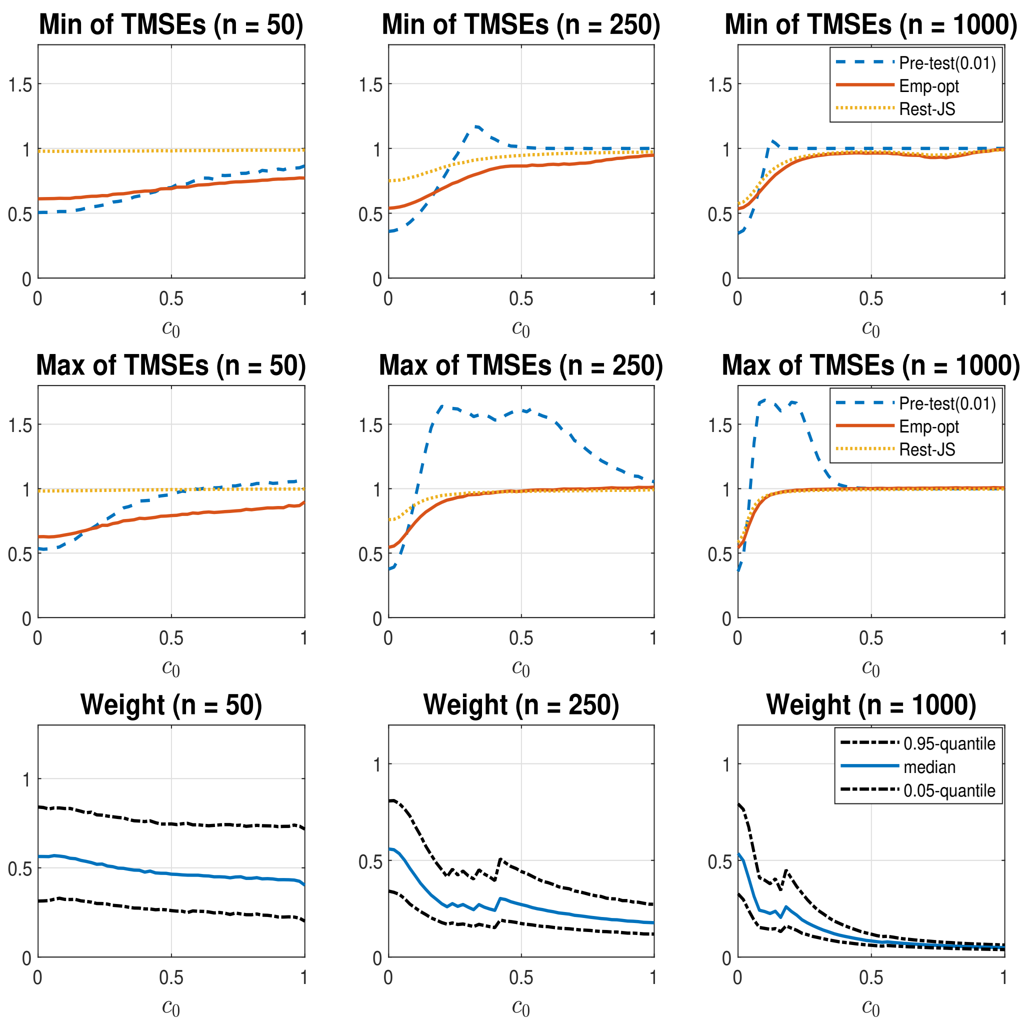
Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively. The truncation parameter for the truncated MSE is $\zeta = 1000$.

Figure G.4: Finite Sample Truncated MSEs of the Pre-test and Averaging GMM Estimators in S1



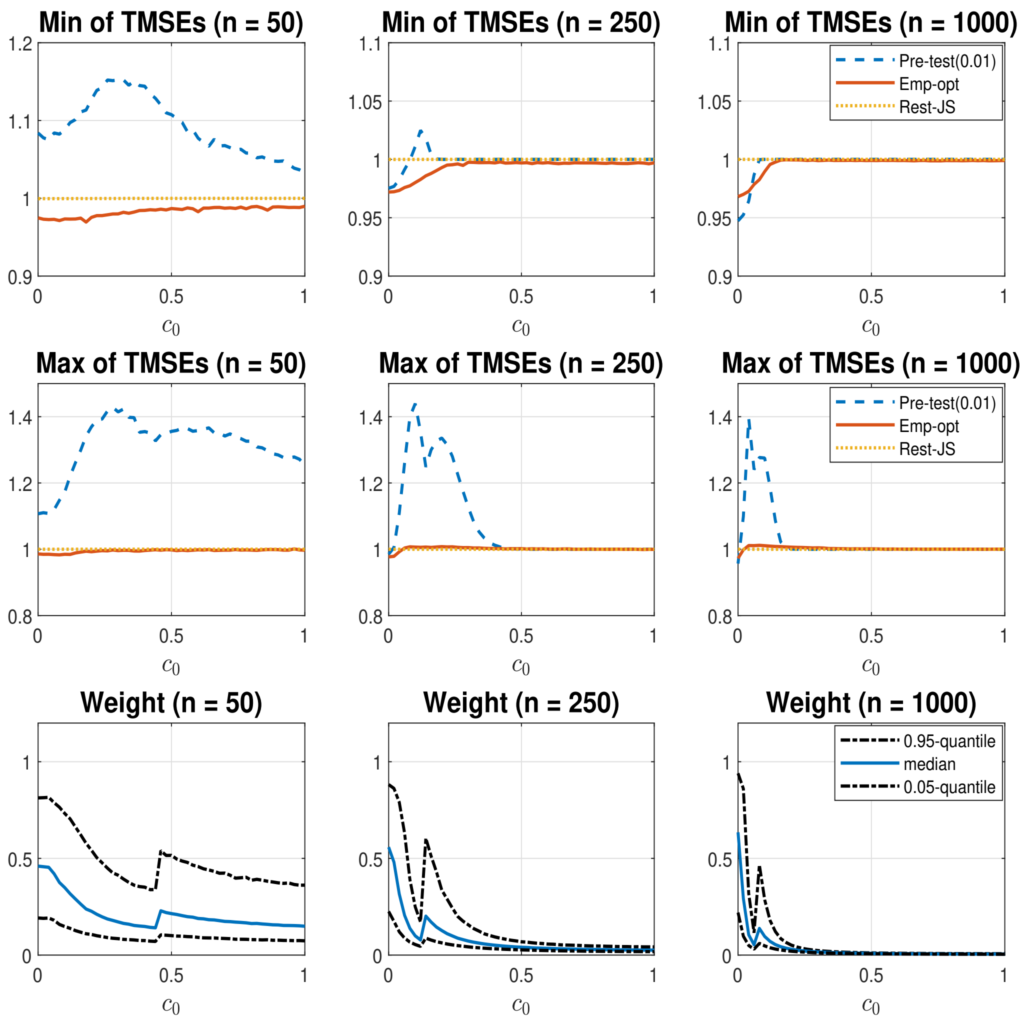
Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively. The truncation parameter for the truncated MSE is $\zeta = 1000$.

Figure G.5: Finite Sample Truncated MSEs of the Pre-test and Averaging GMM Estimators in S2



Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively. The truncation parameter for the truncated MSE is $\zeta = 1000$.

Figure G.6: Finite Sample Truncated MSEs of the Pre-test and Averaging GMM Estimators in S3



Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively. The truncation parameter for the truncated MSE is $\zeta = 1000$.

Table G.1: The Lower and Upper Bounds of the Finite Sample Relative Truncated MSEs

		Design S1		Design S2		Design S3	
		Lower	Upper	Lower	Upper	Lower	Upper
$n = 50$	$\widehat{\theta}_{oe}$	0.5732	0.7968	0.6113	0.8980	0.9694	1.0012
	$\widehat{\theta}_{JS}$	0.9755	0.9959	0.9776	0.9978	0.9995	1.0003
	$\widehat{\theta}_{pret}$	0.4424	0.9574	0.5057	1.0973	1.0324	1.4283
$n = 100$	$\widehat{\theta}_{oe}$	0.5325	0.8789	0.5513	0.9781	0.9733	1.0040
	$\widehat{\theta}_{JS}$	0.9208	0.9911	0.9202	0.9956	0.9996	1.0002
	$\widehat{\theta}_{pret}$	0.3586	1.1940	0.3937	1.3539	0.9990	1.4709
$n = 250$	$\widehat{\theta}_{oe}$	0.5316	0.9587	0.5384	1.0118	0.9720	1.0079
	$\widehat{\theta}_{JS}$	0.7591	0.9787	0.7506	0.9923	0.9999	1.0000
	$\widehat{\theta}_{pret}$	0.3360	1.5106	0.3598	1.6392	0.9753	1.4394
$n = 500$	$\widehat{\theta}_{oe}$	0.5331	0.9846	0.5355	1.0112	0.9700	1.0096
	$\widehat{\theta}_{JS}$	0.6443	0.9823	0.6359	0.9953	1.0000	1.0000
	$\widehat{\theta}_{pret}$	0.3368	1.6196	0.3495	1.6937	0.9562	1.4236
$n = 1,000$	$\widehat{\theta}_{oe}$	0.5335	0.9934	0.5341	1.0082	0.9681	1.0119
	$\widehat{\theta}_{JS}$	0.5803	0.9890	0.5737	0.9978	1.0000	1.0000
	$\widehat{\theta}_{pret}$	0.3395	1.6433	0.3451	1.6864	0.9473	1.3953

Note: 1. $\hat{\theta}_{JS}$ and $\hat{\theta}_{pret}$ denote the GMM averaging estimator based on the weight in (6.1) and the pre-testing GMM estimator based on J-test with nominal size 0.01 respectively; 2. the "Upper" and "Lower" refer to the upper bound and the lower bound of the finite sample relative MSEs among all DGPs considered in the simulation design given the sample size.

H Simulation under the Student-t Distribution

In this subsection, we report the simulation results on the finite sample properties of the pre-test and averaging GMM estimators, when the residual term u in the structure equation (6.4) of CLS is a student-t random variable with degree of freedom 2. The simulation design is the same as the one in Subsection 6.1, except that we generate the structural error u in the following way

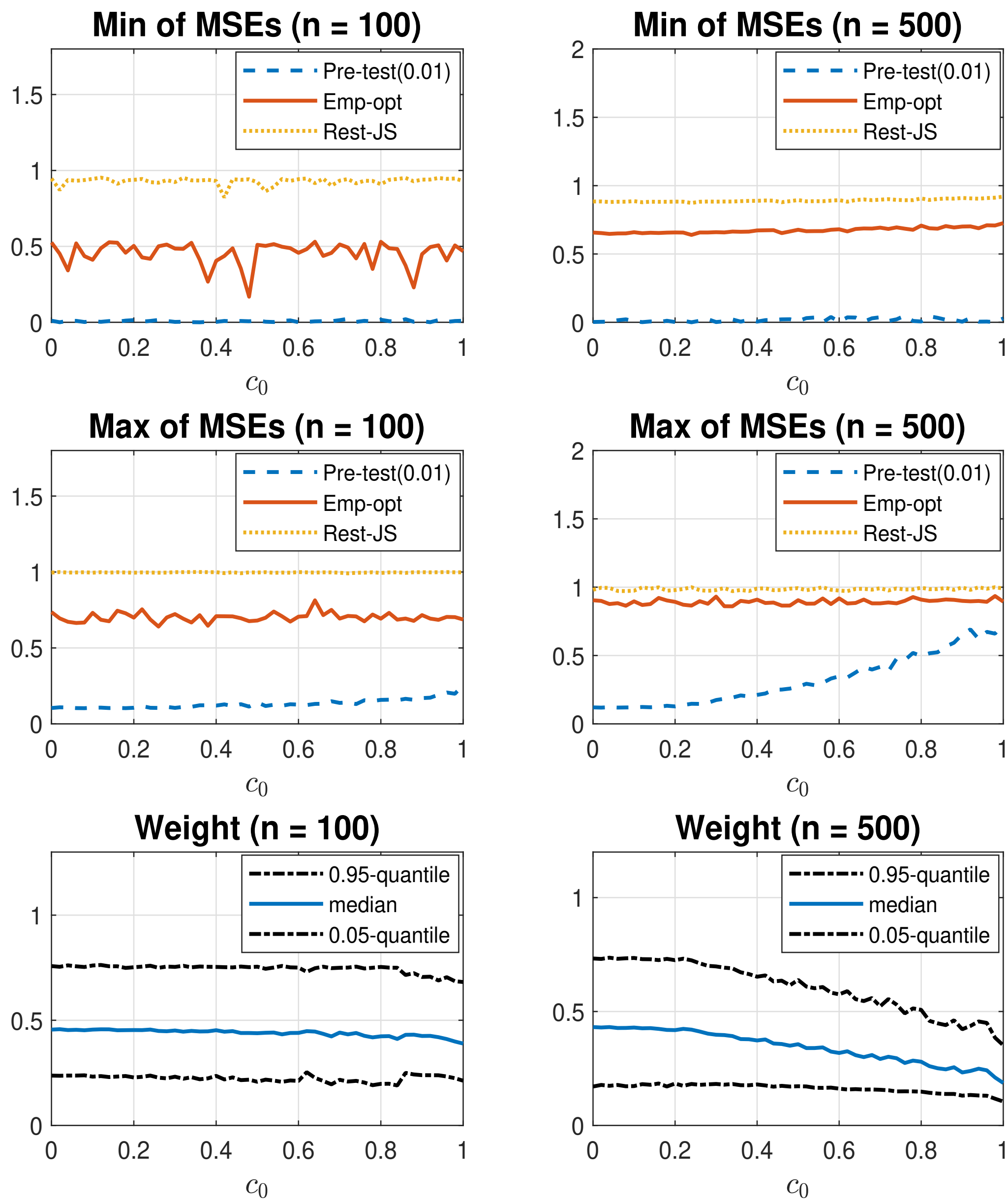
$$u = \frac{u^*}{((\eta_1^2 + \eta_2^2)/2)^{1/2}}$$

where η_1 and η_2 are independent standard normal random variables which are independent with respect to $(Z_1, ..., Z_{18}, \varepsilon_1, \ldots, \varepsilon_6, u^*)$.[30] We call this simulation design as S4.

The finite sample untruncated and truncated MSEs are reported in Figures H.1 - H.5. It is interesting to see that in this simulation design, both the pre-test GMM estimator and the averaging GMM estimator have smaller finite sample MSEs than the conservative GMM estimator. The main reason for this phenomenon is that the residual term u in the structural equation is Student-t with degree of freedom 2, which implies that u has infinite variance and hence the conservative GMM estimator has large variance in finite samples. When the extra IVs Z_j^* $(j = 1, \ldots, 6)$ are used in the GMM estimation, the finite sample variances of the GMM estimator is greatly reduced. Therefore, the finite samples biases of the pre-testing GMM estimator and the averaging GMM estimator introduced by the extra IVs Z_j^* $(j = 1, \ldots, 6)$ are more than offset by the reduced finite sample variances, which enables both estimator have smaller finite sample MSEs.

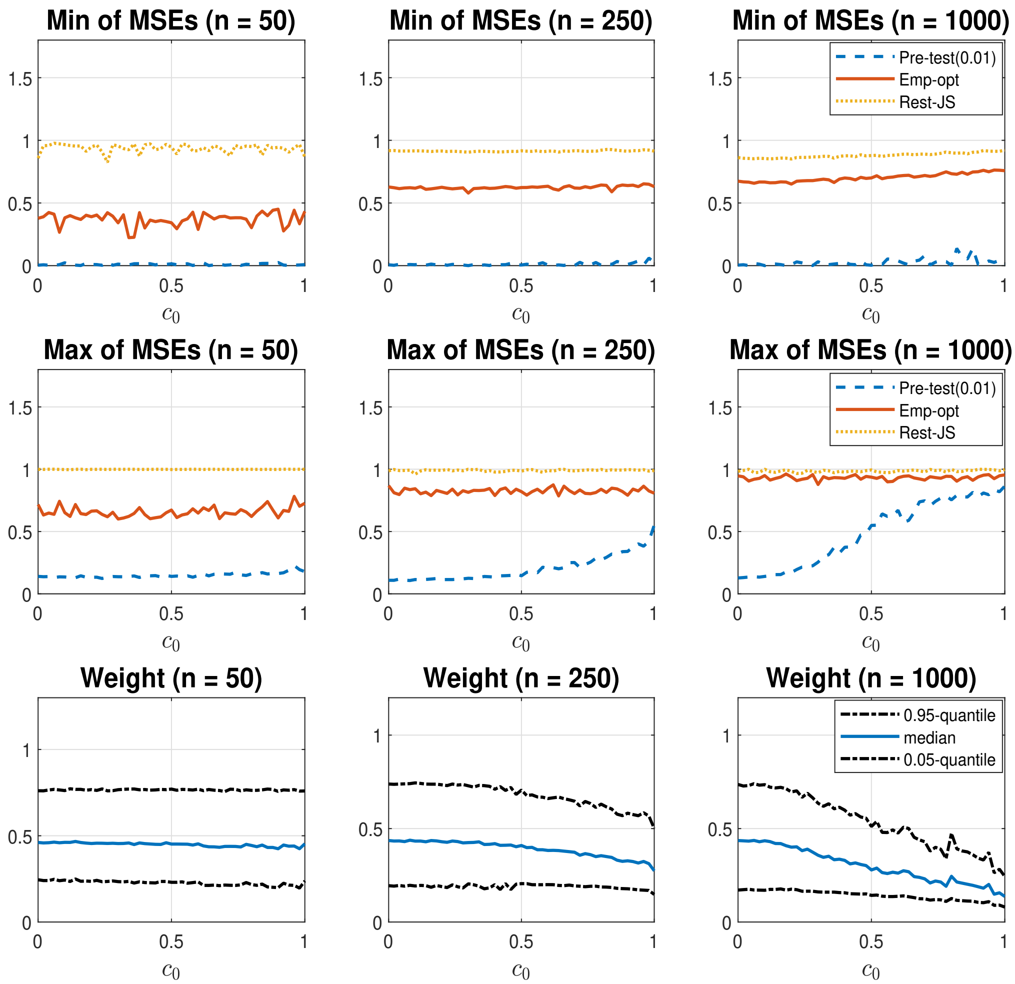
[30] In this design, the structural error u does not enter the possibly invalid IVs (6.4). Therefore the IVs and the regressors are normally distributed. We thank an anonymous who suggested this simulation design.

Figure H.1: Finite Sample MSEs of the Pre-test and Averaging GMM Estimators in S4



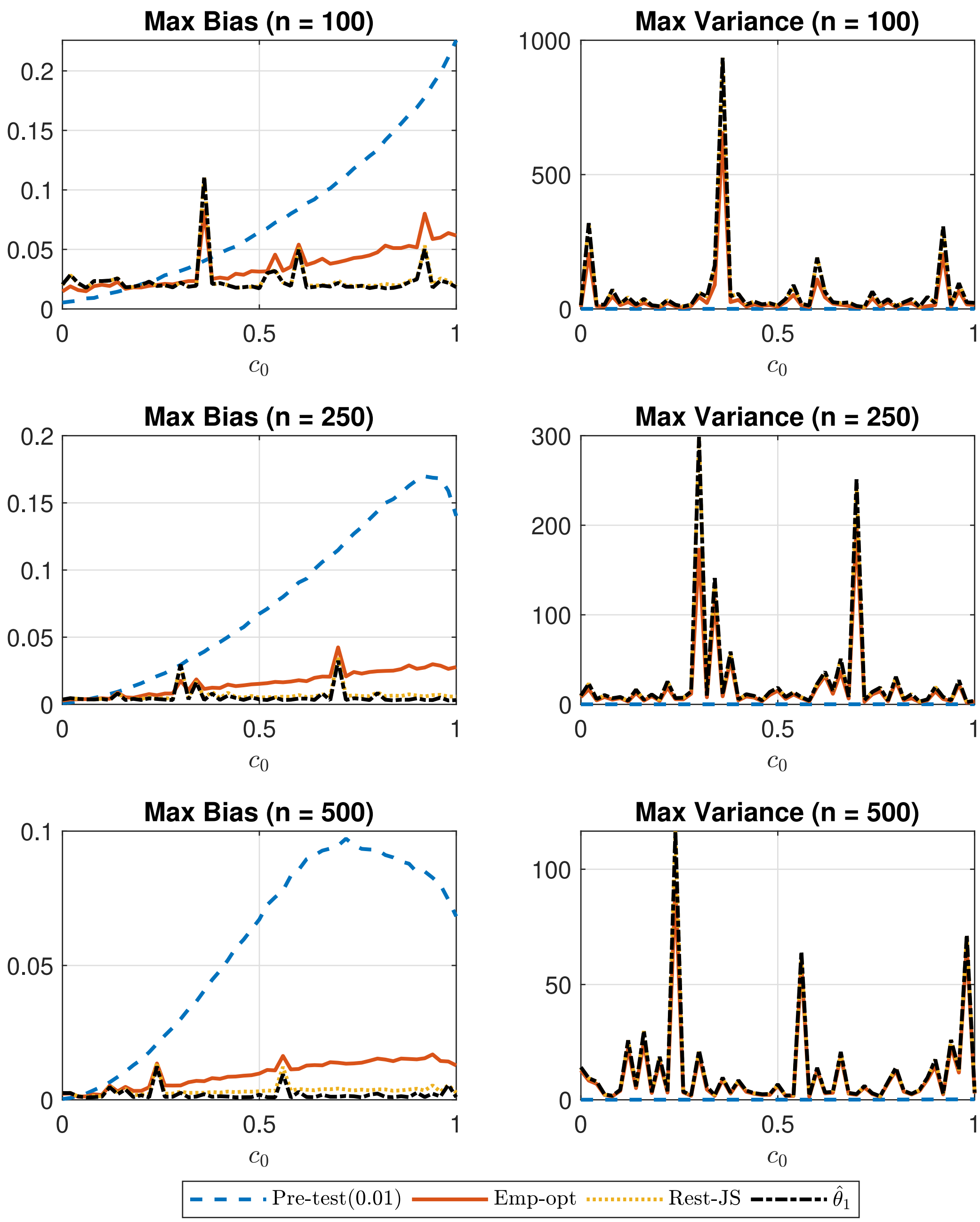
Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively.

Figure H.2: Finite Sample MSEs of the Pre-test and Averaging GMM Estimators in S4



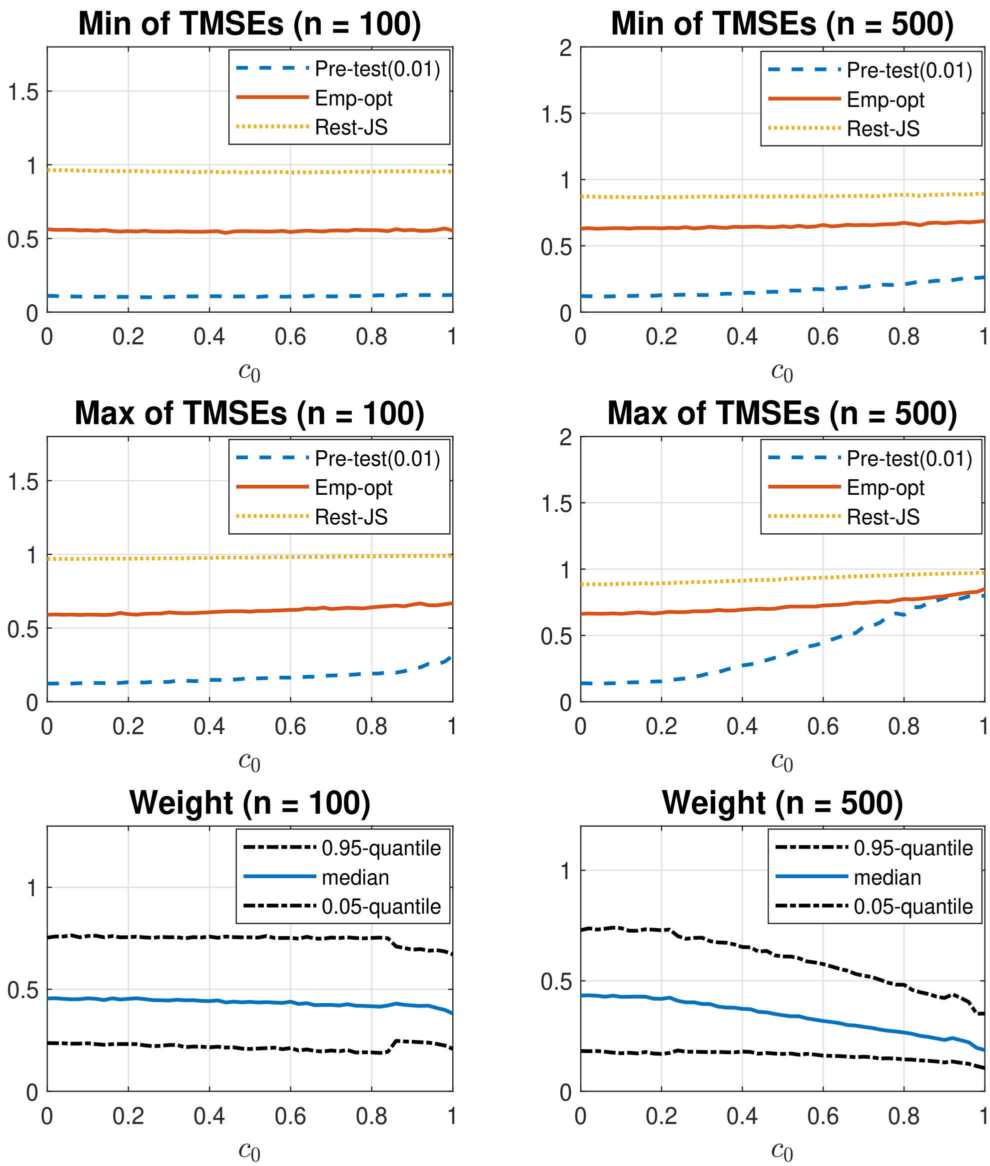
Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively.

Figure H.3: Finite Sample Biases and Variances in S4



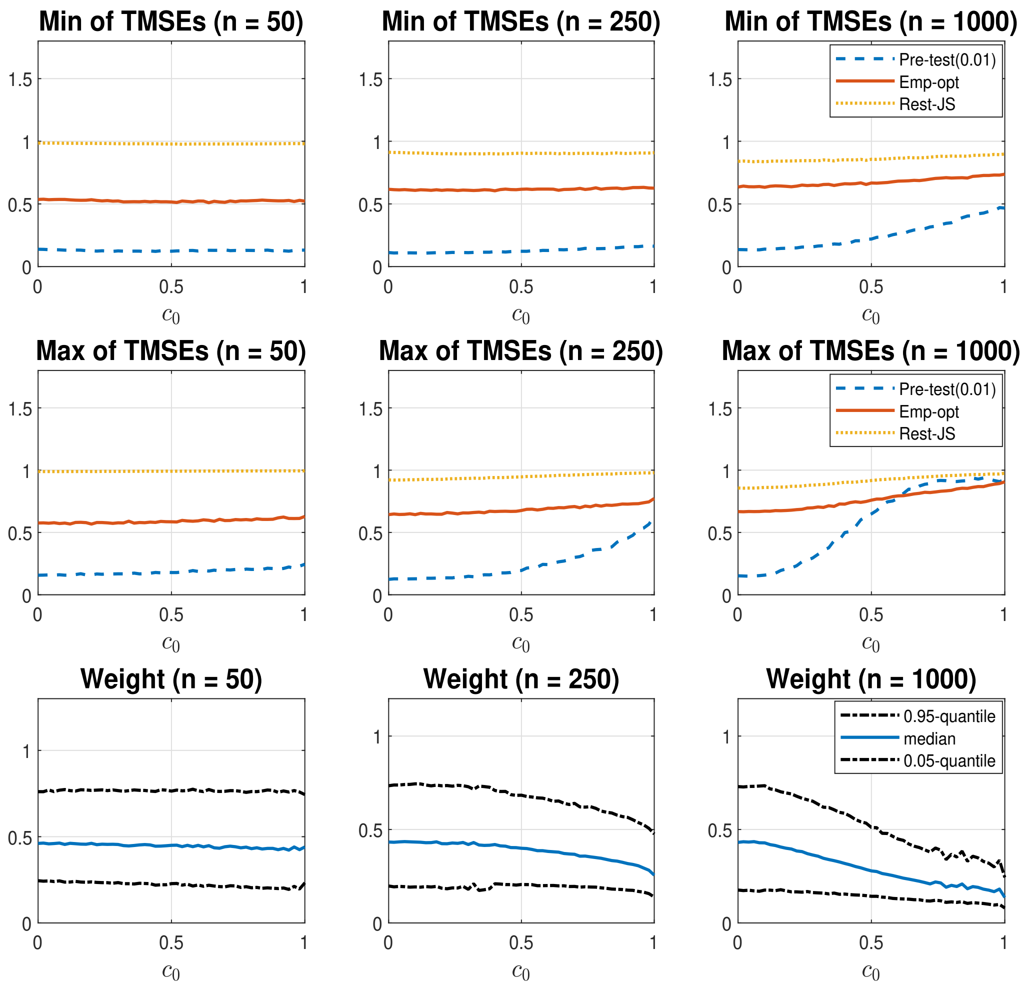
Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively.

Figure H.4: Finite Sample TMSEs of the Pre-test and Averaging GMM Estimators in S4



Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively.The truncation parameter for the truncated MSE is $\zeta = 1000$.

Figure H.5: Finite Sample TMSEs of the Pre-test and Averaging GMM Estimators in S4



Note: "Pre-test(0.01)" refers to the pre-test GMM estimator based on the J-test with nominal size 0.01; "Emp-opt" refers to the averaging GMM estimator based on the empirical optimal weight; "Rest-JS" refers to the averaging estimators based on the restricted James-Stein weight, respectively. The truncation parameter for the truncated MSE is $\zeta = 1000$.