Uniform Asymptotic Risk of Averaging GMM Estimator Robust to Misspecification*

Xu Cheng† Zhipeng Liao‡ Ruoyao Shi§

First Version: August 2013; This Version: March, 2015

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. It is an alternative to pre-test estimators that switch between the conservative and aggressive estimators based on model specification tests. This averaging estimator is robust in the sense that it uniformly dominates the conservative estimator by reducing the risk under any degree of misspecification, whereas the pre-test estimators reduce the risk in parts of the parameter space and increase it in other parts.

To establish uniform dominance of one estimator over another, we establish asymptotic theories on uniform approximations of the finite-sample risk differences between two estimators. These asymptotic results are developed along drifting sequences of data generating processes (DGPs) that model various degrees of local misspecification as well as global misspecification. Extending seminal results on the James-Stein estimator, the uniform dominance is established in non-Gaussian semiparametric nonlinear models. The proposed averaging estimator is applied to estimate the human capital production function in a life-cycle labor supply model.

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

*We thank Donald Andrews, Denis Chetverikov, Patrik Guggenberger, Jinyong Hahn, Jia Li, Rosa Matzkin, Hyunsik Roger Moon, Ulrich Mueller, Joris Pinkse, Frank Schorfheide, Shuyang Sheng, and seminar participants at Brown University, Duke University, University of Pennsylvania, Pennsylvania State University, University of California Los Angeles, Yale University, and 2014 New York Area Econometrics Colloquium for helpful comments.

†Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, USA. Email: xucheng@econ.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, UCLA, Mail Stop: 147703, Los Angeles, CA 90095. Email: shiruoyao@ucla.edu

1 Introduction

The generalized method of moments (GMM) estimator (Hansen, 1982) is one of the most popular methods for estimating moment-based models in economics and finance. Properties of the GMM estimator rely on the quality of the moment conditions. While it is appealing to use more moment restrictions for a more efficient estimator, the validity of some moment conditions may be subject to empirical examination. Various specification tests and model selection criteria are available for testing the validity of moment conditions. However, such data-dependent decisions on model specification do not always improve the estimator. For example, consider the comparison between a pre-test GMM estimator that only uses some additional moment restrictions if a specification test (e.g., the J-test) suggests their validity and a conservative GMM estimator that never uses these additional moment restrictions.[1] Measured by the mean squared error (MSE), this pre-test estimator does better than the conservative estimator in parts of the parameter space and worse than the latter in other parts of the parameter space. Post-model-selection estimators also exhibit this type of non-uniform behavior (Leeb and Pötscher, 2008).

This paper aims to *uniformly* reduce the risk of a GMM estimator by utilizing potentially misspecified moment restrictions with data-dependent averaging. Instead of using tests or model-selection criteria to switch between the "conservative" estimator that never uses additional moments and the "aggressive" estimator that always uses additional moments, we consider an averaging estimator that combines the two with a smooth data-dependent weight. The averaging weight is derived as the sample analog of an infeasible optimal weight. This paper establishes "uniform dominance" in the sense that in large sample the risk of this averaging estimator is smaller than or equal to that of the conservative estimator for any DGP in a given parameter space and the former is strictly smaller than the latter for some DGPs. For DGPs in this parameter space, the additional moment conditions may be correctly specified or misspecified to any degrees[2]. The uniform dominance result insures the averaging estimator against any efficiency loss, even if the additional moments are misspecified and the degree of misspecification is unknown. Constructing uniformly valid tests in non-standard problems is an active research area in econometrics in recent years, including models with weak identification, partial identification, local to unit root, post-model-selection inference, etc. This paper focuses on the risk of a point estimator rather than hypothesis testing.

To establish uniform dominance of one estimator over another, this paper provides new asymptotic theories on uniform approximations of the finite-sample risk differences between two es-

[1] Throughout the paper, we assume that the GMM estimator is constructed with the optimal weighting matrice, which is defined in footnote 4.

[2] These DGPs include the $n^{-1/2}$ local misspecification (Newey 1985), the global misspecification (Hall and Inoue, 2003), as well as many other DGPs studied in this paper.

timators. These asymptotic results are developed along drifting sequences of DGPs with different degrees of misspecification. This class of DGPs include the crucial $n^{-1/2}$ local sequences that are considered by Hjort and Claeskens (2003), Saleh (2006), Liu (2013), Hansen (2014a,b,c), DiTraglia (2014) for various averaging estimators, as well as some more distant sequences. The theoretical results glue all sequences together and show that they are sufficient to provide a uniform approximation of the finite-sample risk differences. The proof uses the techniques developed in Andrews and Guggenberger (2010) and Andrews, Cheng, and Guggenberger (2011) for uniformly valid tests and applies them to uniform risk comparison in moment-based models.

This uniform dominance result is related to the Stein's phenomenon (Stein, 1956) in parametric models. The James-Stein (JS) estimator (James and Stein, 1961) is shown to dominate the maximum likelihood estimator in exact normal sampling. Hansen (2014a) considers local asymptotic analysis of the JS-type averaging estimator in general parametric models and substantially extends its application in econometrics. The present paper focuses on the uniformity issue and studies the Stein's phenomenon in non-Gaussian semiparametric nonlinear models. The weight we suggest is different from a JS-type extension for semiparametric models. We find the suggested weight compares favorably to the latter in finite-sample experiments.

For moments constructed by instrumental variables (IVs), the misspecification may come from two sources. One is additional IVs whose validity is questionable. The other is the set of endogenous variables, where a Hausman test (Hausman, 1978) is widely applied to check whether they are actually exogenous. Recently, Hansen (2014b) and DiTraglia (2014) both consider averaging estimators that combine the ordinary least squares (OLS) estimator and the two-stage-least-squares (2SLS) estimator in linear IV models. In linear IV models with homoskedastic errors, our conservative estimator becomes the 2SLS estimator, and our aggressive estimator using both the IVs and the endogenous variables becomes the OLS estimator[3]. However, when applied to linear IV models, the averaging weight we considered is different from those in Hansen (2014b) and DiTraglia (2014).

The estimator proposed in this paper is a frequentist model averaging (FMA) estimator. FMA estimators have received much attention in recent years. Buckland, Burnham, and Augustin (1997) and Burnham and Anderson (2002) suggest model averaging weights based on the AIC or BIC scores. Hjort and Claeskens (2003) study the asymptotic distribution and asymptotic risk of the FMA estimator in locally misspecified parametric models. The results of Hjort and Claeskens (2003) are extended to the Cox's proportional hazards models by Hjort and Claeskens (2006), general semiparametric models by Claeskens and Carroll (2007), and generalized additive partially linear models by Zhang and Liang (2011). Hansen (2007, 2008) and Wan et al. (2010) study the FMA estimator

[3]Consider the linear IV model $Y_i = X_i'\beta + u_i$ with instruments Z_i. The aggressive estimator is equivalent to the OLS estimator because $(X'P_{[X,Z]}X)^{-1}X'P_{[X,Z]}Y = (X'X)^{-1}X'Y$, where $Y = (Y_1, \ldots, Y_n)'$, $X = (X_1, \ldots, X_n)'$, $Z = (Z_1, \ldots, Z_n)'$ and $P_{[X,Z]} = (X, Z)\left[(X, Z)'(X, Z)\right]^{-1}(X, Z)'$ denotes the projection matrix.

with the Mallows' averaging weight. Liang et al. (2011) introduce a general random weight that includes smoothed AIC, smoothed BIC, and many other weights as special cases. Hansen and Racine (2012) investigate the FMA estimator with the cross-validation averaging weight. The estimator by Hansen and Racine (2012) are extended to time series models by Zhang et al. (2013) and to quantile regressions by Lu and Su (2015). Cheng and Hansen (2014) study FMA estimators in factor-augmented regressions. Our paper contributes to this literature by studying the uniform asymptotic risk of the FMA estimator in moment-based semiparametric models and providing an asymptotic framework to show uniform dominance.

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, 2001; Hong, Preston and Shum, 2003), and the penalized estimation methods (see, e.g., Liao, 2013; Cheng and Liao, 2014; Caner, Han and Lee, 2014; Kang, Zhang, Cai and Small, 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 (2012), Guggenberger (2012), Nevo and Rosen (2012), and Kolesar, Chetty, Friedman, Glaeser, Imbens (2014), among others. Moon and Schorfheide (2009) explore over-identifying moment inequalities to reduce the MSE. This paper contributes to this literature by providing new uniform results for potentially misspecified semiparametric models.

The rest of the paper is organized as follows. Section 2 introduces the model and the averaging estimator. Section 3 establishes some general results on the asymptotic risk and the uniform dominance of one estimator over another. Section 4 defines the averaging estimator and uses the general results in Section 3 to show that the averaging GMM estimator uniformly dominates the conservative estimator. Section 5 investigates the finite sample performance of our averaging estimator in different simulation experiments. Section 6 applies the averaging estimator to estimate the human capital production function in a life-cycle labor supply model. Section 7 concludes. Proofs and technical arguments are given in the Appendix.

2 Model and Averaging Estimator

The observations $\{W_i \in \mathbb{R}^{d_W} : i = 1, ..., n\}$ are i.i.d. or stationary with joint distribution $F_0 \in \mathcal{F}$. For some known functions $g_1(\cdot, \theta) \in \mathbb{R}^{r_1}$ and $g^*(\cdot, \theta) \in \mathbb{R}^{r^*}$, we consider estimation of a finite-dimensional parameter $\theta_0 (\in \Theta \subset \mathbb{R}^{d_\theta})$ that satisfies the moment conditions

$$\mathbb{E}_{F_0} [g_1(W_i, \theta_0)] = \mathbf{0}_{r_1} \tag{2.1}$$

and

$$\mathbb{E}_{F_0}\left[g^*(W_i,\theta_0)\right]=\delta_0, \tag{2.2}$$

where $\mathbf{0}_{r_1}$ denotes the $r_1\times 1$ zero vector, the slackness parameter δ_0 is unknown and $\mathbb{E}_F[\cdot]$ denotes the expectation taken with respect to the DGP F. We assume that the moment conditions in (2.1) uniquely identify θ_0 for any $F_0\in\mathcal{F}$. Although a consistent estimator of θ_0 follows from the moment conditions in (2.1), it is desirable to explore the information in (2.2) to improve efficiency.

Because δ_0 is unknown, a data-dependent decision typically is made to switch between the "conservative" estimator that only uses (2.1), and the "aggressive" estimator that uses the moment conditions in both (2.1) and (2.2) with δ_0 imposed to be $\mathbf{0}_{r^*}$. Write

$$g_2(W,\theta)=\begin{pmatrix} g_1(W,\theta)\\ g^*(W,\theta)\end{pmatrix}\in\mathbb{R}^{r_2}. \tag{2.3}$$

The conservative and aggressive GMM estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are defined by

$$\widehat{\theta}_k\equiv\underset{\theta\in\Theta}{\arg\min}\left[n^{-1}\sum_{i=1}^n g_k(W_i,\theta)\right]'\mathcal{W}_{k,n}\left[n^{-1}\sum_{i=1}^n g_k(W_i,\theta)\right] \tag{2.4}$$

where $\mathcal{W}_{k,n}$ is a $r_k\times r_k$ optimal weighting matrix for $k=1$ and 2.[4]

Below is a linear IV example to illustrate the notations introduced in the general GMM framework.

Example. Consider the structural equations

$$Y=X_1'\theta_1+X_2'\theta_2+u, \tag{2.5}$$

$$X_1=\Pi_0X_2+\Pi_1Z_1+\Pi_2Z_2+v, \tag{2.6}$$

where Y is a scalar response variable, X_1 is a vector of endogenous regressors, X_2 is a vector of exogenous regressors, Z_1 and Z_2 are vectors of IVs, u and v are residual terms. We are interested in the coefficients $\theta=(\theta_1',\theta_2')'$. The coefficients Π_j $(j=0,1,2)$ are nuisance parameters. Let F_0 denote the joint distribution of $W=(Y,X_1',X_2',Z_1',Z_2')'$.

In the structural equation (2.5), X_1 is endogenous in the sense that each element of $\mathbb{E}_{F_0}\left[X_1u\right]$

[4]The optimal weighting matrix is

$$\mathcal{W}_{k,n}=(n^{-1}\sum_{i=1}^n g_k(Z_i,\widetilde{\theta}_1)g_k(Z_i,\widetilde{\theta}_1)'-\overline{g}_{k,n}(Z,\widetilde{\theta}_1)\overline{g}_{k,n}(Z,\widetilde{\theta}_1)')^{-1},$$

where $\overline{g}_{k,n}(Z,\widetilde{\theta}_1)=n^{-1}\sum_{i=1}^n g_k(Z_i,\widetilde{\theta}_1)$ and $\widetilde{\theta}_1$ is a preliminary consistent GMM estimator based on $g_1(Z_i,\theta)$ and the identity weighting matrix.

is non-zero and X_2 is exogenous in the sense that $\mathbb{E}_{F_0}[X_2 u] = \mathbf{0}_{d_{x_2}}$. To identify θ, suppose we have valid IVs Z_1 that satisfy the exogenous condition $\mathbb{E}_{F_0}[Z_1 u] = \mathbf{0}_{d_{z_1}}$.The number of valid IVs Z_1 is no smaller than the number of endogenous variables X_1. We also have additional IVs Z_2, but their validity is uncertain, i.e., $\mathbb{E}_{F_0}[Z_2 u] = \delta_0$ and δ_0 may not be a zero vector.

In this example,

$$g_1(W, \theta) = \begin{pmatrix} (Y - X_1'\theta_1 - X_2'\theta_2) X_2 \\ (Y - X_1'\theta_1 - X_2'\theta_2) Z_1 \end{pmatrix} \tag{2.7}$$

and

$$g^*(W, \theta) = \left(Y_1 - X_1'\theta_1 - X_2'\theta_2\right) Z_2. \tag{2.8}$$

GMM estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ follow from (2.4). □

Many estimators considered in the literature fall in the class

$$\widehat{\theta}(\widetilde{\omega}) = (1 - \widetilde{\omega})\widehat{\theta}_1 + \widetilde{\omega}\widehat{\theta}_2 \tag{2.9}$$

where $\widetilde{\omega} \in R$ could be deterministic or random. By definition, $\widehat{\theta}(0) = \widehat{\theta}_1$ and $\widehat{\theta}(1) = \widehat{\theta}_2$. A pre-test estimator takes the form $\widehat{\theta}(\widetilde{\omega}_{\alpha,p})$, where $\widetilde{\omega}_{\alpha,p} = 1\{T_n \leq c_\alpha\}$ for some test statistic T_n with the critical value c_α at the significance level α. Post-model selection estimator also follows this binary decision rule and allows c_α to change with the sample size. For averaging estimators, $\widetilde{\omega}$ typically is a data-dependent weight that is not restricted to 0 or 1 (see, e.g., Hjort and Claeskens, 2003 and Hansen, 2007).

Although various data-dependent choices of $\widetilde{\omega}$ in the literature all aim to improve upon $\widehat{\theta}_1$ by exploring the information in (2.2), it remains to establish an asymptotic framework to show one estimator dominates the other *uniformly*. Uniformity is important because $\widetilde{\omega}$ is data-dependent and the finite-sample risk of $\widehat{\theta}(\widetilde{\omega})$ is sensitive to the degree of misspecification measured by δ_0. In a pointwise asymptotic framework where the DGP is fixed as the sample size increases, a pre-test estimator has smaller asymptotic risk than the conservative estimator $\widehat{\theta}_1$. However, it does not dominates $\widehat{\theta}_1$ uniformly over all the DGPs. As such, we first establish some general asymptotic results that enable one to evaluate the uniform asymptotic risk of an estimator and the risk differences between two estimators over a class of distributions. These uniform asymptotic results aim to provide good approximations to the finite-sample properties. Then, we propose a new averaging estimator in (4.10) and show that it uniformly dominates the conservative estimator $\widehat{\theta}_1$.

3 Asymptotic Risk and Risk Differences

Let $\widehat{\theta} \in \Theta$ be the generic notation of an estimator of θ_0. Let $\ell(\cdot) : \Theta \to \mathbb{R}_+ \cup \{\infty\}$ be a generic loss function. The finite-sample and asymptotic risks of $\widehat{\theta}$ are defined as

$$R_n(\widehat{\theta}) \equiv \sup_{F \in \mathcal{F}} \mathbb{E}_F[\ell(\widehat{\theta})] \text{ and } AsyR(\widehat{\theta}) \equiv \limsup_{n\to\infty} R_n(\widehat{\theta}), \tag{3.1}$$

respectively. The asymptotic risk builds the uniformity over $F \in \mathcal{F}$ into the definition by taking $\sup_{F\in\mathcal{F}}$ before $\limsup_{n\to\infty}$. This uniform asymptotic risk is different from a pointwise asymptotic risk which is either obtained under a fixed DGP or a particular sequence of drifting DGP. It is comparable to the asymptotic size of a test, which is the limit of the finite-sample size defined as the supremum of the finite-sample rejection probabilities.

To compare two estimators $\widehat{\theta}$ and $\widetilde{\theta}$, we consider the finite-sample and asymptotic minimal and maximal risk difference (RD):

$$\begin{aligned} \underline{RD}_n(\widehat{\theta}, \widetilde{\theta}) &\equiv \inf_{F\in\mathcal{F}} \mathbb{E}_F[\ell(\widehat{\theta}) - \ell(\widetilde{\theta})], \ Asy\underline{RD}(\widehat{\theta}, \widetilde{\theta}) \equiv \liminf_{n\to\infty} \underline{RD}_n, \\ \overline{RD}_n(\widehat{\theta}, \widetilde{\theta}) &\equiv \sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell(\widehat{\theta}) - \ell(\widetilde{\theta})], \ Asy\overline{RD}(\widehat{\theta}, \widetilde{\theta}) \equiv \limsup_{n\to\infty} \overline{RD}_n. \end{aligned} \tag{3.2}$$

The objects of interest are the finite-sample risk differences, approximated by their asymptotic counterparts. One estimator $\widehat{\theta}$ uniformly dominates the other estimator $\widetilde{\theta}$ if

$$Asy\underline{RD}(\widehat{\theta}, \widetilde{\theta}) < 0 \text{ and } Asy\overline{RD}(\widehat{\theta}, \widetilde{\theta}) \leq 0. \tag{3.3}$$

In (3.1) and (3.2), the uniformity over $F \in \mathcal{F}$ is crucial for the asymptotic results to give a good approximation to their finite-sample counterparts. The value of F at which the supremum or the infimum are attained often varies with the sample size. Therefore, to determine the asymptotic risk of an estimator and to show one estimator dominates another, one has to derive the asymptotic distributions of these estimators under various sequences $\{F_n\}$. In the subsection below, we provide a sufficiently large class of sequences $\{F_n\}$ such that the pointwise limits along these sequences can combine to represent the uniform asymptotic risk and risk differences.[5]

[5] The metric on $\mathcal{F}$ induces weak convergence of the bivariate distributions (Z_i, Z_j) for all $i, j \geq 1$, such as the Kolmogorov metric or the Prokhorov metric.

3.1 A Sufficient Class of Sequences

Let $\theta(F) \in \Theta$ be the unique value of θ identified by the moments in (2.1), i.e., $\mathbb{E}_F\left[g_1(W, \theta(F))\right] = \mathbf{0}_{r_1}$. Define

$$\delta(F) \equiv \mathbb{E}_F\left[g^*(W, \theta(F))\right], \tag{3.4}$$

which measures the slackness of the additional moments for any F. For $k = 1$ and 2, we define the Jacobian and the variance-covariance matrices of the moment functions by

$$\begin{aligned} G_k(F) &\equiv \mathbb{E}_F\left[g_{k,\theta}(W, \theta(F))\right], \text{ where } g_{k,\theta}(W, \theta) \equiv \frac{g_k(W, \theta)}{\partial \theta'}, \\ \Omega_k(F) &\equiv \lim_{n \to \infty} \mathrm{Var}_F\left[n^{-1/2} \sum_{i=1}^{n} g_k(W_i, \theta(F))\right]. \end{aligned} \tag{3.5}$$

Note that $G_1(F) = S_1 G_2(F)$ and $\Omega_1(F) = S_1 \Omega_2(F) S_1'$, where S_1 is a selector matrix that selects $g_1(W, \theta)$ out of $g_2(W, \theta)$. For the averaging GMM estimator studied below, let

$$v(F) \equiv \left(\mathrm{vec}[G_2(F)]', \mathrm{vech}[\Omega_2(F)]', M_2(\theta; F)'\right)', \tag{3.6}$$

where $M_2(\cdot; F) \equiv \mathbb{E}_F[g_2(W, \cdot)]$ is the moment function indexed by θ for any F, $\mathrm{vec}\,(\cdot)$ denotes vectorization, and $\mathrm{vech}\,(\cdot)$ denotes the half vectorization of a symmetric matrix.

Example (Cont.) In the linear IV example, $\theta(F) = (\theta_1(F), \theta_2(F))$ is the solution to the linear equations

$$\mathbf{0}_{r_1} = \mathbb{E}_F\left[g_1(W, \theta(F))\right] = \mathbb{E}_F\left[\left(Y - X_1'\theta_1(F) - X_2'\theta_2(F)\right)\begin{pmatrix} X_2 \\ Z_1 \end{pmatrix}\right]. \tag{3.7}$$

Given $\theta(F)$, $\delta(F)$ in this example is defined as

$$\delta(F) = \mathbb{E}_F\left[g^*(W, \theta(F))\right] = \mathbb{E}_F\left[\left(Y - X_1'\theta_1(F) - X_2'\theta_2(F)\right) Z_2\right]. \tag{3.8}$$

As the moment functions are linear in θ, $G_k(F)$ $(k = 1, 2)$ have simple expressions:

$$G_1(F) = -\mathbb{E}_F\left[\begin{pmatrix} X_2X_1' & X_2X_2' \\ Z_1X_1' & Z_1X_2' \end{pmatrix}\right] \text{ and } G_2(F) = -\mathbb{E}_F\left[\begin{pmatrix} ZX_1' & ZX_2' \end{pmatrix}\right], \tag{3.9}$$

where $Z = (X_2', Z_1', Z_2')'$. In addition, $\Omega_k(F)$ and $M_2(\cdot; F)$ are defined using the moment functions $g_1(W, \theta)$, $g^*(W, \theta)$ and $\theta(F)$ respectively. □

We consider sequences of DGPs $\{F_n\}$ such that $\delta(F_n)$ satisfies

$$\text{(i) } n^{1/2}\delta(F_n) \to d \in \mathbb{R}^{r^*} \text{ or (ii) } ||n^{1/2}\delta(F_n)|| \to \infty. \tag{3.10}$$

and $v(F_n)$ satisfies

$$v(F_n) \to v_0 \equiv \left(\text{vec}[G_2]', \text{vech}[\Omega_2]', M_2(\theta)'\right)', \tag{3.11}$$

where $G_2 \in \mathbb{R}^{r_2 \times d_\theta}$, $\Omega_2 \in \mathbb{R}^{r_2 \times r_2}$, and $M_2(\cdot)$ is a non-random function of θ. Case (ii) in (3.10) includes the intermediate case in which $\delta(F_n) \to \mathbf{0}_{r^*}$ and $||n^{1/2}\delta(F_n)|| \to \infty$ as well the case in which $\delta(F_n)$ is bounded away from $\mathbf{0}_{r^*}$. We collect the sequences $\{F_n\}$ that satisfy (3.10) and (3.11) into two sets

$$\begin{aligned}
\mathcal{S}(d, v_0) &\equiv \left\{\{F_n\} : F_n \in \mathcal{F},\ n^{1/2}\delta(F_n) \to d \in \mathbb{R}^{r^*} \text{ and } v(F_n) \to v_0\right\} \text{ and} \\
\mathcal{S}(\infty, v_0) &\equiv \left\{\{F_n\} : F_n \in \mathcal{F},\ ||n^{1/2}\delta(F_n)|| \to \infty \text{ and } v(F_n) \to v_0\right\}.
\end{aligned} \tag{3.12}$$

The DGPs in $\mathcal{S}(d, v_0)$ model correct specification and local misspecification up to the magnitude of $n^{-1/2}$, whereas the DGPs in $\mathcal{S}(\infty, v_0)$ model more severe misspecification, including the conventional global misspecification case where $\delta(F_n)$ is a fixed non-zero value as well as the intermediate case where $\delta(F_n)$ converges to $\mathbf{0}_{r^*}$ slower than $n^{-1/2}$.

In this model, for each sample size n, the true values of F, θ and δ are denoted as F_n, $\theta_n = \theta(F_n)$, and $\delta_n = \delta(F_n)$, respectively. These true values satisfy the model specified in (2.1) and (2.2) with the subscript 0 replaced by 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.

3.2 Representation of the Asymptotic Risk and Asymptotic Risk Differences

Now we show that pointwise results along sequences in (3.12) combine to yield a uniform result. For two estimators $\widehat{\theta}$ and $\widetilde{\theta}$, we assume that $\mathbb{E}_{F_n}[\ell(\widehat{\theta})]$ and $\mathbb{E}_{F_n}[\ell(\widetilde{\theta})]$ satisfy the following high-level assumptions along a sequence $\{F_n\}$. These high-level assumptions are verified below for the averaging estimator and the pre-test estimator.

Assumption 3.1 *The following results hold under* $\{F_n\}$.
(i) *If* $\{F_n\} \in \mathcal{S}(d, v_0)$ *for* $d \in \mathbb{R}^{r^*}$,

$$\lim_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widehat{\theta})] = R(d, v_0) \in \mathbb{R}_+ \text{ and } \lim_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widetilde{\theta})] = \widetilde{R}(d, v_0) \in \mathbb{R}_+.$$

(ii) *If* $\{F_n\} \in \mathcal{S}(\infty, v_0)$,

$$\lim_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widehat{\theta})] = R(\infty, v_0) \in \mathbb{R}_+ \cup \{\infty\} \text{ and } \lim_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widetilde{\theta})] = \widetilde{R}(\infty, v_0) \in \mathbb{R}_+ \cup \{\infty\}.$$

Assumption 3.1 considers the pointwise limit of the finite-sample risk along $\{F_n\}$. The key requirement is that the limit of the finite-sample risk under $\{F_n\}$ does not depend on the limit of F_n directly. Instead, it depends on the limits of $n^{1/2}\delta(F_n)$ and $v(F_n)$. Moreover, for any sequence $\{F_n\} \in \mathcal{S}(d, v_0)$, the limit of the finite-sample risk must be the same, indexed by (d, v_0). The same requirement applies to a sequence $\{F_n\} \in \mathcal{S}(\infty, v_0)$.

When $\widetilde{\theta}$ is the conservative estimator, we can write $\widetilde{R}(v_0) = \widetilde{R}(d, v_0) = \widetilde{R}(\infty, v_0)$ because its asymptotic risk dose not depend on the degree of misspecification.

Let $\Lambda = \{(\delta(F), v(F)) : F \in \mathcal{F}\}$. The following assumption provides a set of regularity assumptions on the set $\mathcal{F}$ on which we build the uniform results.

Assumption 3.2 (i) $V_{\mathcal{F}} = \{v(F) : F \in \mathcal{F}\}$ *is a compact set.*
(ii) $\delta(F_1) = 0$ *for some* $F_1 \in \mathcal{F}$ *and* $\delta(F_2) \neq 0$ *for some* $F_2 \in \mathcal{F}$.
(iii) *For some* $\varepsilon > 0$, *if* $||\delta|| < \varepsilon$ *and* $(\delta, v) \in \Lambda$ *then* $(a\delta, v) \in \Lambda$ $\forall a \in (0, 1]$.

Assumption 3.2(i) requires that the image of $v(F)$ is a compact set. Assumption 3.2(ii) states that the parameter space contains both correctly specified models and misspecified models. Assumption 3.2(iii) states that the space $\mathcal{F}$ includes some continuous perturbations from a correctly specified model.

For sequences in (3.12), we define parameter spaces:

$$\begin{aligned} H_R &\equiv \{(d, v_0) : \text{there exists some sequence } \{F_n\} \in \mathcal{S}(d, v_0)\}, \\ H_\infty &\equiv \{v_0 : \text{there exists some sequence } \{F_n\} \in \mathcal{S}(\infty, v_0)\}. \end{aligned} \tag{3.13}$$

The set H_R corresponds to the correctly specified and "mildly" misspecified models. The set H_∞ corresponds to the "severely" misspecified models.

Theorem 3.1 *Suppose Assumptions 3.1 and 3.2 hold. Then:*
(a) The asymptotic risk satisfies

$$AsyR(\widehat{\theta}) = \max\left\{ \sup_{(d,v_0)\in H_R} R(d, v_0), \sup_{v_0 \in H_\infty} R(\infty, v_0) \right\}.$$

(b) The asymptotic minimal and maximal risk differences satisfy

$$Asy\underline{RD}(\widehat{\theta},\widetilde{\theta}) = \min\left\{\inf_{(d,v_0)\in H_R}\left[R(d,v_0)-\widetilde{R}(d,v_0)\right], \inf_{v_0\in H_\infty}\left[R(\infty,v_0)-\widetilde{R}(\infty,v_0)\right]\right\},$$

$$Asy\overline{RD}(\widehat{\theta},\widetilde{\theta}) = \max\left\{\sup_{(d,v_0)\in H_R}\left[R(d,v_0)-\widetilde{R}(d,v_0)\right], \sup_{v_0\in H_\infty}\left[R(\infty,v_0)-\widetilde{R}(\infty,v_0)\right]\right\}.$$

Comment 3.1 Theorem 3.1 links the uniform asymptotic risk and risk differences with the pointwise limits of $\mathbb{E}_{F_n}[\ell(\widehat{\theta})]$ and $\mathbb{E}_{F_n}[\ell(\widetilde{\theta})]$ under the sequences considered in Assumption 3.1. It shows that the sequences in $\mathcal{S}(d,v_0)$ and $\mathcal{S}(\infty,v_0)$ form a sufficient class to study the uniform asymptotic risk and asymptotic risk differences. This class is larger than the class of convergent sequences that satisfy $F_n \to F_0$ for some $F_0 \in \mathcal{F}$. Theorem 3.1 is proved by the techniques used to establish the asymptotic size of non-standard tests, see Andrews and Guggenberger (2010), Andrews, Cheng, and Guggenberger (2011), and Andrews and Cheng (2012).[6]

Comment 3.2 The two estimators $\widehat{\theta}$ and $\widetilde{\theta}$ are compared under all DGPs in $\mathcal{F}$ to establish uniform dominance in the sense of (3.3). The smallest and largest differences between their risks are approximated by $Asy\underline{RD}(\widehat{\theta},\widetilde{\theta})$ and $Asy\overline{RD}(\widehat{\theta},\widetilde{\theta})$, respectively. They are different from what one would obtain by simply comparing the individual asymptotic risks of the two estimators.

Comment 3.3 Theorem 3.1 also applies to other non-standard estimation problems where the asymptotic distribution is discontinuous at parts of the parameter space. It is key to verify Assumption 3.1 after specifying $\delta(F)$ and $v(F)$.

3.3 Asymptotic Risk with Truncation

The high-level conditions in Assumption 3.1 typically are verified by first obtaining the asymptotic distribution of $\widehat{\theta}$ and $\widetilde{\theta}$ under $\{F_n\}$, then taking expectations of the limits by assuming uniform integrability. If uniform integrability is not a reasonable assumption, one may consider the truncated loss function $\ell_\zeta(\widehat{\theta}) \equiv \min\{\ell(\widehat{\theta}),\zeta\}$ for some $\zeta \in \mathbb{R}_+$ following Hansen (2014a) and generalize the asymptotic risk to

$$AsyR^*(\widehat{\theta}) \equiv \lim_{\zeta\to\infty}\limsup_{n\to\infty}\sup_{F\in\mathcal{F}}\mathbb{E}_F[\ell_\zeta(\widehat{\theta})]. \tag{3.14}$$

In this case, Assumption 3.1 can be replaced by Assumption 3.3 below.

Assumption 3.3 *The following results hold under* $\{F_n\}$.

[6] In an uncirculated working paper, Andrews and Guggenberger (2006) also considered the asymptotic risk representation of a non-standard estimator.

(i) If $\{F_n\} \in \mathcal{S}(d, v_0)$ *for* $d \in \mathbb{R}^{r^*}$*, then for any* $\zeta \in \mathbb{R}_+$:

$$\lim_{n\to\infty} \mathbb{E}_{F_n}[\ell_\zeta(\widehat{\theta})] = R_\zeta(d, v_0) \in \mathbb{R}_+ \text{ and } \lim_{n\to\infty} \mathbb{E}_{F_n}[\ell_\zeta(\widetilde{\theta})] = \widetilde{R}_\zeta(d, v_0) \in \mathbb{R}_+.$$

(ii) If $\{F_n\} \in \mathcal{S}(\infty, v_0)$*, then for any* $\zeta \in \mathbb{R}_+$:

$$\lim_{n\to\infty} \mathbb{E}_{F_n}[\ell_\zeta(\widehat{\theta})] = R_\zeta(\infty, v_0) \in \mathbb{R}_+ \text{ and } \lim_{n\to\infty} \mathbb{E}_{F_n}[\ell_\zeta(\widetilde{\theta})] = \widetilde{R}_\zeta(\infty, v_0) \in \mathbb{R}_+.$$

For the truncated loss, the asymptotic minimal and maximal risk differences are generalized to

$$\begin{aligned} Asy\underline{RD}^*(\widehat{\theta}, \widetilde{\theta}) &\equiv \lim_{\zeta\to\infty} \liminf_{n\to\infty} \inf_{F\in\mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}) - \ell_\zeta(\widetilde{\theta})], \\ Asy\overline{RD}^*(\widehat{\theta}, \widetilde{\theta}) &\equiv \lim_{\zeta\to\infty} \limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta}) - \ell_\zeta(\widetilde{\theta})]. \end{aligned} \tag{3.15}$$

Corollary 3.2 *Suppose Assumptions 3.2 and 3.3 hold.*

(a) The asymptotic risk satisfies

$$\begin{aligned} AsyR^*(\widehat{\theta}) &= \lim_{\zeta\to\infty} AsyR^*_\zeta(\widehat{\theta}), \text{ where} \\ AsyR^*_\zeta(\widehat{\theta}) &\equiv \max\left\{ \sup_{(d,v_0)\in H_R} R_\zeta(d, v_0), \sup_{v_0\in H_\infty} R_\zeta(\infty, v_0) \right\} \in \mathbb{R}_+ \cup \{\infty\}. \end{aligned}$$

(b) The asymptotic minimal and maximal risk differences satisfy

$$\begin{aligned} Asy\underline{RD}^*(\widehat{\theta}, \widetilde{\theta}) &= \lim_{\zeta\to\infty} Asy\underline{RD}^*_\zeta(\widehat{\theta}, \widetilde{\theta}) \text{ and} \\ Asy\overline{RD}^*(\widehat{\theta}, \widetilde{\theta}) &= \lim_{\zeta\to\infty} Asy\overline{RD}^*_\zeta(\widehat{\theta}, \widetilde{\theta}), \text{ where} \\ Asy\underline{RD}^*_\zeta(\widehat{\theta}, \widetilde{\theta}) &\equiv \min\left\{ \inf_{(d,v_0)\in H_R} \left[R_\zeta(d, v_0) - \widetilde{R}_\zeta(d, v_0)\right], \inf_{v_0\in H_\infty} \left[R_\zeta(\infty, v_0) - \widetilde{R}_\zeta(\infty, v_0)\right] \right\}, \\ Asy\overline{RD}^*_\zeta(\widehat{\theta}, \widetilde{\theta}) &\equiv \max\left\{ \sup_{(d,v_0)\in H_R} \left[R_\zeta(d, v_0) - \widetilde{R}_\zeta(d, v_0)\right], \sup_{v_0\in H_\infty} \left[R_\zeta(\infty, v_0) - \widetilde{R}_\zeta(\infty, v_0)\right] \right\}. \end{aligned}$$

Comment 3.4 In the formula of $AsyR^*(\widehat{\theta})$ in part (a), the supremum is taken before $\zeta \to \infty$ to control the truncation effect uniformly over the parameter space. The order of supremum and $\zeta \to \infty$ should not be switched. Similarly, when comparing two estimators in part (b), we take into account the truncation effect on both estimators uniformly over the parameter space.

4 Averaging GMM Estimator

In this section, we propose an averaging estimator and use the asymptotic risk difference representation in Section 3 to show it uniformly dominates the conservative estimator. To verify the high-level conditions in Assumption 3.1 or Assumption 3.3, we provide primitive regularity assumptions on the moment conditions and derive pointwise asymptotic properties of the averaging estimator along the sequences specified in (3.12).

To introduce this averaging estimator, we first study the asymptotic properties of the conservative and the aggressive GMM estimators under different sequences of DGPs.

4.1 Asymptotic Properties of the GMM Estimator under Misspecification

For the aggressive GMM estimator $\widehat{\theta}_2$, the population criterion function is

$$Q_F(\theta) \equiv \mathbb{E}_F[g_2(W_i,\theta)]'\Omega_2^{-1}(F)\mathbb{E}_F[g_2(W_i,\theta)]. \tag{4.1}$$

Let $\theta^*(F)$ denote the pseudo-true value that minimizes $Q_F(\theta)$ over $\theta \in \Theta$. If all moment conditions are correctly specified, i.e., $\mathbb{E}_F[g_2(W_i,\theta(F)] = 0$, this pseudo-true value is equivalent to the true value, i.e., $\theta^*(F) = \theta(F)$. If some moment conditions in (2.2) are misspecified, they could be different. The identification conditions for $\theta(F)$ and $\theta^*(F)$ are specified in Assumption 4.1 below.

Assumption 4.1 (i) *For any $\varepsilon > 0$, there exists a constant $\delta_\varepsilon > 0$ such that $\forall F \in \mathcal{F}$,*

$$\inf_{\{\theta\in\Theta:\ \|\theta-\theta(F)\|\geq\varepsilon\}} \|\mathbb{E}_F\left[g_1(W_i,\theta(F))\right]\| > \delta_\varepsilon,$$

$$\inf_{\{\theta\in\Theta:\ \|\theta-\theta^*(F)\|\geq\varepsilon\}} \left[Q_F(\theta) - Q_F(\theta^*(F))\right] > \delta_\varepsilon.$$

(ii) *$\theta(F)$ and $\theta^*(F)$ are both in the interior of Θ $\forall F \in \mathcal{F}$.*

For any matrix A, we use $\rho_{\min}(A)$ and $\rho_{\max}(A)$ to denote the smallest and largest eigenvalues of A, respectively. Let C denote a generic finite constant.

Assumption 4.2 (i) $\mathbb{E}_F[\sup_{\theta\in\Theta}(||g_2(W_i,\theta)|| + ||g_{2,\theta}(W_i,\theta)||)] \leq C\ \forall F \in \mathcal{F}$.
(ii) *$g_2(W,\theta)$ is continuously differentiable in θ a.s., and its partial derivative $g_{2,\theta}(W,\theta)$ satisfies*

$$|\mathbb{E}_F[g_{2,\theta}(W_i,\theta_1) - g_{2,\theta}(W_i,\theta_2)]| \leq C||\theta_1-\theta_2||\ \forall\theta_1,\theta_2\in\Theta, \forall F\in\mathcal{F}.$$

(iii) *For $k = 1$ and 2, $C^{-1} \leq \rho_{\min}(\Omega_k(F)) \leq \rho_{\max}(\Omega_k(F)) \leq C\ \forall F \in \mathcal{F}$.*
(iv) *For $k = 1$ and 2, $C^{-1} \leq \rho_{\min}(G_k'(F)G_k(F)) \leq \rho_{\max}(G_k'(F)G_k(F)) \leq C\ \forall F \in \mathcal{F}$.*

(v) $\mathcal{W}_{k,n} \to_p \Omega_k^{-1}$ *under any* $\{F_n\}$ *such that* $\Omega_k(F_n) \to \Omega_k$, *for* $k = 1$ *and* 2.
(vi) $v(F)$ *is continuous in* F $\forall F \in \mathcal{F}$.

We assume the following uniform law of large numbers, uniform central limit theorem, and stochastic equicontinuity of the empirical processes for the triangular array of observations. Let

$$\xi_n(g_2(\theta)) \equiv n^{-1/2} \sum_{i=1}^{n} (g_2(W_i, \theta) - \mathbb{E}_{F_n}[g_2(W_i, \theta)]). \tag{4.2}$$

and let $\theta_n \equiv \theta(F_n)$.

Assumption 4.3 *For any* $\varepsilon_n \to 0$ *and under any sequence* $\{F_n \in \mathcal{F}\}$,
(i) $\sup_{\theta \in \Theta} ||n^{-1} \sum_{i=1}^n g_2(W_i, \theta) - \mathbb{E}_{F_n} g_2(W_i, \theta)|| = o_p(1)$;
(ii) $\sup_{\theta \in \Theta} ||n^{-1} \sum_{i=1}^n g_{2,\theta}(W_i, \theta) - \mathbb{E}_{F_n} g_{2,\theta}(W_i, \theta)|| = o_p(1)$;
(iii) $\xi_n(g_2(\theta_n)) \to_d N(0, \Omega_2)$ *if* $\Omega_2(F_n) \to \Omega_2$;
(iv) $\sup_{\{\theta_1, \theta_2 \in \Theta:\ ||\theta_1 - \theta_2|| \leq \varepsilon_n\}} \xi_n[g_2(\theta_1) - g_2(\theta_2)] = o_p(1)$.

Sufficient conditions of Assumption 4.3 for triangular arrays of i.i.d. and strong mixing observations are available in Assumptions 11.3-11.5 of Andrews and Cheng (2013).

Let $\mathcal{Z}_2$ denote a normal random vector with mean zero and variance-covariance matrix Ω_2. Recall that S_1 is a selector matrix such that $\mathcal{Z}_1 \equiv S_1 \mathcal{Z}_2$ is the first r_1 rows of $\mathcal{Z}_2$. To describe the asymptotic distributions of $\widehat{\theta}_1$ and $\widehat{\theta}_2$, we define

$$\Gamma_k \equiv -\left(G_k' \Omega_k^{-1} G_k\right)^{-1} G_k' \Omega_k^{-1}, \text{ for } k = 1 \text{ and } 2. \tag{4.3}$$

Lemma 4.1 *Under Assumptions 4.1-4.3, the following results hold under* $\{F_n\}$.
(a) *If* $\{F_n\} \in \mathcal{S}(d, v_0) \cup \mathcal{S}(\infty, v_0)$, $n^{1/2}(\widehat{\theta}_1 - \theta_n) \to_d \Gamma_1 \mathcal{Z}_1$.
(b) *If* $\{F_n\} \in \mathcal{S}(d, v_0)$ *for some* $d \in \mathbb{R}^{r^*}$, $n^{1/2}(\widehat{\theta}_2 - \theta_n) \to_d \Gamma_2 \mathcal{Z}_{d,2}$, *where* $\mathcal{Z}_{d,2} = \mathcal{Z}_2 + d_0$ *and* $d_0 = (\mathbf{0}_{1 \times r_1}, d')'$.
(c) *If* $\{F_n\} \in \mathcal{S}(\infty, v_0)$, $M_2(\theta)' \Omega_2^{-1} M_2(\theta)$ *has a unique minimizer* $\theta^*(v_0)$, $\widehat{\theta}_2 \to_p \theta^*(v_0)$ *and* $|n^{1/2}(\widehat{\theta}_2 - \theta_n)| \to_p \infty$.

Comment 4.1 Our results under drifting DGPs complement Hall and Inoue (2003) on the asymptotic distribution of $\widehat{\theta}_2$ under global misspecification with a fixed DGP.

Comment 4.2 When the moment conditions in (2.2) are severely misspecified, i.e., $||n^{1/2}\delta(F_n)|| \to \infty$, it is sufficient to show that $|n^{1/2}(\widehat{\theta}_2 - \theta_n)|$ diverges in probability in order to investigate the asymptotic risk of the averaging GMM estimator. In this case, $\widehat{\theta}_2$ is either inconsistent or consistent but with a convergence rate slower than $n^{-1/2}$.

4.2 Non-Random Optimal Weight

In this subsection, we study the asymptotic risk of the averaging GMM estimator with a non-random weight $\omega \in [0,1]$. The sample analog of this non-random optimal weight is used to construct the averaging estimator proposed in this paper. We consider the weighted quadratic loss function

$$\ell(\widehat{\theta}) = n(\widehat{\theta} - \theta_n)' H (\widehat{\theta} - \theta_n), \tag{4.4}$$

where H is a $d_\theta \times d_\theta$ positive semi-definite matrix. We next derive the non-random optimal weight that minimizes the asymptotic risk.

For $k = 1$ and 2, define

$$\Sigma_k(F) \equiv \left[G_k'(F) \Omega_k^{-1}(F) G_k(F) \right]^{-1}. \tag{4.5}$$

If $v(F_n) \to v_0$ for v_0 defined in (3.11), Σ_k is the limit of $\Sigma_k(F_n)$ given by

$$\Sigma_k \equiv \left(G_k' \Omega_k^{-1} G_k \right)^{-1}. \tag{4.6}$$

Define

$$A_{v_0} \equiv H (\Sigma_1 - \Sigma_2) \text{ and } B_{v_0} \equiv (\Gamma_2 - \Gamma_1^*)' H (\Gamma_2 - \Gamma_1^*), \tag{4.7}$$

where $\Gamma_1^* = [\Gamma_1, \mathbf{0}_{d_\theta \times r^*}]$ and the subscript v_0 indicates that A_{v_0} and B_{v_0} are matrix-valued functions of v_0. For any v_0, the matrix A_{v_0} is positive semi-definite following Lemma 2.1 in Cheng and Liao (2014).

Lemma 4.2 *Under Assumptions 4.1-4.3, the following results hold under $\{F_n\}$.*
(a) *If $\{F_n\} \in \mathcal{S}(d, v_0)$, $\ell(\widehat{\theta}(\omega)) \to_d \lambda_{(d,v_0)}(\omega)$, where $\lambda_{(d,v_0)}(\omega)$ is a random variable with*

$$\mathbb{E}[\lambda_{(d,v_0)}(\omega)] = \mathrm{tr}(H\Sigma_1) - 2\omega \mathrm{tr}\left(A_{v_0}\right) + \omega^2 \left[d_0' B_{v_0} d_0 + \mathrm{tr}\left(A_{v_0}\right) \right] \ \forall \omega \in R.$$

(b) $\mathbb{E}[\lambda_{(d,v_0)}(\omega)]$ *is minimized at*

$$\omega^*(d, v_0) = \frac{\mathrm{tr}\left(A_{v_0}\right)}{d_0' B_{v_0} d_0 + \mathrm{tr}\left(A_{v_0}\right)} \text{ for } d \in \mathbb{R}^{r^*}, \text{ where } d_0 = (\mathbf{0}_{1 \times r_1}, d')'.$$

(c) *If $\{F_n\} \in \mathcal{S}(\infty, v_0)$, $\ell(\widehat{\theta}(\omega)) \to_p \infty$ when $\omega > 0$, and $\ell(\widehat{\theta}(\omega)) \to_d \mathcal{Z}_1' \Gamma_1' H \Gamma_1 \mathcal{Z}_1$ when $\omega = 0$.*

Comment 4.2 Although Lemma 4.2 is derived under the weighted quadratic loss function, the general theory established in the previous section applies to other loss functions as well. The quadratic loss function is attractive because it produces a non-random optimal weight with an explicit analytical solution. The averaging estimator proposed in this paper is a sample analog of

this non-random optimal weight and its analytical form is useful for analyzing the asymptotic risk of the averaging estimator.

Comment 4.3 The optimal weight in Lemma 4.2(b) is infeasible in practice because it depends on unknown parameters. One may consider estimating these unknown parameters and plugging their estimators into the optimal weight formula. The matrices Γ_2, Σ_1, and Σ_2 can be consistently estimated based on $\widehat{\theta}_1$. However, the location parameter d_0 is not consistently estimable. As a result, when d_0 is replaced by its sample analog, one has to account for this estimation error when evaluating the risk of the resulting averaging estimator.

4.3 GMM Averaging Estimator with Empirical Optimal Weight

We propose to use a sample analog of $\omega^*(d, v_0)$ to construct the averaging estimator. This sample analog is called the empirical optimal weight, which takes the form

$$\widetilde{\omega}_{eo} = \frac{tr\left[H(\widehat{\Sigma}_1 - \widehat{\Sigma}_2)\right]}{n(\widehat{\theta}_2 - \widehat{\theta}_1)'H(\widehat{\theta}_2 - \widehat{\theta}_1) + tr\left[H(\widehat{\Sigma}_1 - \widehat{\Sigma}_2)\right]} \tag{4.8}$$

where $\widehat{\Sigma}_k$ is a consistent estimator of Σ_k for $k = 1$ and 2. Lemma 4.1 shows that under a sequence $\{F_n\} \in \mathcal{S}(d, v_0)$,

$$n^{1/2}(\widehat{\theta}_2 - \widehat{\theta}_1) \to_d (\Gamma_2 - \Gamma_1^*)(\mathcal{Z}_2 + d_0). \tag{4.9}$$

The empirical optimal weight $\widetilde{\omega}_{eo}$ is a sample analog of the non-random optimal weight $\omega^*(d, v_0)$ with $(\Gamma_2 - \Gamma_1^*)d_0$ replaced by its asymptotically unbiased estimator $n^{1/2}(\widehat{\theta}_2 - \widehat{\theta}_1)$, and Σ_k replaced by its consistent estimator $\widehat{\Sigma}_k$ for $k = 1, 2$. Under a sequence $\{F_n\} \in \mathcal{S}(\infty, v_0)$, $\widetilde{\omega}_{eo} \to_p 0$ because $|n^{1/2}(\widehat{\theta}_2 - \widehat{\theta}_1)| \to_p \infty$, which means that $\widetilde{\omega}_{eo}$ is asymptotically optimal following Lemma 4.2(c).

The averaging GMM estimator proposed takes the form

$$\widehat{\theta}_{eo} = (1 - \widetilde{\omega}_{eo})\widehat{\theta}_1 + \widetilde{\omega}_{eo}\widehat{\theta}_2. \tag{4.10}$$

By the consistency of $\widehat{\Sigma}_k$ and Lemma 2.1 in Cheng and Liao (2014), we know that $\text{tr}[H(\widehat{\Sigma}_1 - \widehat{\Sigma}_2)] \geq 0$ with probability approaching 1 (w.p.a.1), which together with the form of $\widetilde{\omega}_{eo}$ in (4.8) implies that $\widetilde{\omega}_{eo} \in [0, 1]$ w.p.a.1.

Assumption 4.4 *Under* $\{F_n\} \in \mathcal{S}(d, v_0) \cup \mathcal{S}(\infty, v_0)$, $\widehat{\Sigma}_k \to_p \Sigma_k$ *for* $k = 1$ *and* 2.

Next, we define some notations for the asymptotic distribution of the empirical optimal aver-

aging weight, the averaging GMM estimator, and the loss function:

$$\begin{aligned}
&\widetilde{\omega}_{(d,v_0)} \equiv \frac{\operatorname{tr}(A_{v_0})}{\mathcal{Z}_{d,2}' B_{v_0} \mathcal{Z}_{d,2} + \operatorname{tr}(A_{v_0})}, \\
&\phi_{(d,v_0)} \equiv \Gamma_1^* \mathcal{Z}_{d,2} + \widetilde{\omega}_{(d,v_o)} (\Gamma_2 - \Gamma_1^*) \mathcal{Z}_{d,2}, \quad \phi_{(\infty,v_0)} \equiv \Gamma_1 \mathcal{Z}_1, \\
&\lambda_{(d,v_0)} \equiv \phi_{(d,v_0)}' H \phi_{(d,v_0)}, \quad \lambda_{(\infty,v_0)} \equiv \phi_{(\infty,v_0)}' H \phi_{(\infty,v_0)}.
\end{aligned} \tag{4.11}$$

Lemma 4.3 *Under Assumptions 4.1-4.4, we have the following results.*
(a) *If* $\{F_n\} \in \mathcal{S}(d, v_0)$, $\widetilde{\omega}_{eo} \to_d \widetilde{\omega}_{(d,v_0)}$, $n^{1/2}(\widehat{\theta}_{eo} - \theta_n) \to_d \phi_{(d,v_0)}$, *and* $\ell(\widehat{\theta}_{eo}) \to_d \lambda_{(d,v_0)}$.
(b) *If* $\{F_n\} \in \mathcal{S}(\infty, v_0)$, $\widetilde{\omega}_{eo} \to_p 0$, $n^{1/2}(\widehat{\theta}_{eo} - \theta_n) \to_d \phi_{(\infty,v_0)}$, *and* $\ell(\widehat{\theta}_{eo}) \to_d \lambda_{(\infty,v_0)}$.

Lemma 4.3 shows that $\widetilde{\omega}_{eo}$ converges to a non-degenerate random variable under $\{F_n\} \in \mathcal{S}(d, v_0)$. The formula in Lemma 4.2(a) is derived for non-random weight. In consequence, it cannot be used to justify the averaging estimator $\widehat{\theta}_{eo}$ in (4.10) with a random weight. To study the asymptotic risk of $\widehat{\theta}_{eo}$, it is important to take into account the data-dependent nature of $\widetilde{\omega}_{eo}$ and its uniform property under different degrees of misspecification.

4.4 Uniform Dominance

In this subsection, we show that the averaging GMM estimator based on the empirical optimal weight uniformly dominates the conservative GMM estimator. Without assuming the estimators are uniformly integrable, we consider the truncated loss function and show uniform dominance by applying the general results in Corollary 3.2.[7]

Lemmas 4.1 and 4.3 imply that the high-level conditions in Assumption 3.3 hold for $\widehat{\theta} = \widehat{\theta}_{eo}$ and $\widetilde{\theta} = \widehat{\theta}_1$ with

$$\begin{aligned}
&R_\zeta(d, v_0) = \mathbb{E}\left[\min\{\lambda_{(d,v_0)}, \zeta\}\right], R_\zeta(\infty, v_0) = \mathbb{E}\left[\min\{\lambda_{(\infty,v_0)}, \zeta\}\right], \\
&\widetilde{R}_\zeta(d, v_0) = \widetilde{R}_\zeta(\infty, v_0) = \mathbb{E}\left[\min\{\lambda_{(\infty,v_0)}, \zeta\}\right].
\end{aligned} \tag{4.12}$$

To study the maximal and minimal risk differences, we define

$$g_\zeta(d, v_0) \equiv \mathbb{E}\left[\min\{\lambda_{(d,v_0)}, \zeta\}\right] - \mathbb{E}\left[\min\{\lambda_{(\infty,v_0)}, \zeta\}\right] \tag{4.13}$$

under the truncation value ζ. As $\zeta \to \infty$, its limit is

$$g(d, v_0) \equiv \mathbb{E}[\lambda_{(d,v_0)}] - \mathbb{E}\left[\lambda_{(\infty,v_0)}\right]. \tag{4.14}$$

[7] Under the assumption of uniform integrability, the uniform dominance results also hold and the arguments are simplified.

By the definitions of $\lambda_{(d,v_0)}$ and $\lambda_{(\infty,v_0)}$ in (4.11), some simple algebra gives

$$g(d,v_0)=\mathbb{E}\left[\frac{2\text{tr}(A_{v_0})\mathcal{Z}_{d,2}'D_{v_0}\mathcal{Z}_{d,2}}{\mathcal{Z}_{d,2}'B_{v_0}\mathcal{Z}_{d,2}+\text{tr}(A_{v_0})}\right]+\mathbb{E}\left[\frac{\text{tr}(A_{v_0})^2Z_{d,2}'B_{v_0}\mathcal{Z}_{d,2}}{(\mathcal{Z}_{d,2}'B_{v_0}\mathcal{Z}_{d,2}+\text{tr}(A_{v_0}))^2}\right], \tag{4.15}$$

where A_{v_0} and B_{v_0} are defined in (4.7) and $D_{v_0}=(\Gamma_2-\Gamma_1^*)'H\Gamma_1^*$. Note that the formula for $g(d,v_0)$ in (4.15) can be simulated easily for given values of d and v_0.

Theorem 4.1 *Suppose that Assumptions 3.2 and 4.1-4.4 hold.*
(a) *The averaging GMM estimator $\widehat{\theta}_{eo}$ satisfies*

$$Asy\underline{RD}^*(\widehat{\theta}_{eo},\widehat{\theta}_1)=\lim_{\zeta\to\infty}\min\left\{\inf_{(d,v_0)\in H_R}[g_\zeta(d,v_0)],0\right\},$$
$$Asy\overline{RD}^*(\widehat{\theta}_{eo},\widehat{\theta}_1)=\lim_{\zeta\to\infty}\max\left\{\sup_{(d,v_0)\in H_R}[g_\zeta(d,v_0)],0\right\}.$$

(b) *For large $\zeta\in\mathbb{R}_+$, we have*

$$\inf_{(d,v_0)\in H_R}g_\zeta(d,v_0)\le\inf_{(d,v_0)\in H_R}g(d,v_0),\quad\sup_{(d,v_0)\in H_R}g_\zeta(d,v_0)\le\sup_{(d,v_0)\in H_R}g(d,v_0),\ \textit{and}$$

$$g(d,v_0)\le\text{tr}(A_{v_0})\mathbb{E}\left[\frac{4\lambda_{\max}(A_{v_0})-\text{tr}(A_{v_0})}{\mathcal{Z}_{d,2}'B_{v_0}\mathcal{Z}_{d,2}+\text{tr}(A_{v_0})}\right]-\text{tr}(A_{v_0})^2\mathbb{E}\left[\frac{\text{tr}(A_{v_0})+4\lambda_{\max}(A_{v_0})}{(\mathcal{Z}_{d,2}'B_{v_0}\mathcal{Z}_{d,2}+\text{tr}(A_{v_0}))^2}\right].$$

(c) *If* $\text{tr}(A_{v_0})>0$ *and* $\text{tr}(A_{v_0})\ge4\lambda_{\max}(A_{v_0})$ $\forall v_0\in V_{\mathcal{F}}$, $\widehat{\theta}_{eo}$ *uniformly dominates* $\widehat{\theta}_1$, *i.e.,*

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

Comments 4.4 Part (a) follows from Corollary 3.2 and the pointwise limits in Lemma 4.3. Part (b) provides upper bounds for the infimum and supremum of the truncated risk difference $g_\zeta(d,v_0)$ for a large truncated value ζ. This upper bound is represented by $g(d,v_0)$, which has a closed form representation in (4.15). We derive an analytical upper bound for $g(d,v_0)$ using the Stein's Lemma in part (b). This analytical upper bound leads to the sufficient condition in part (c) for uniform dominance. It is worth noting that the condition in part (c) is sufficient but not necessary.

Comments 4.5 To control the truncation effect uniformly over the parameter space, we cannot automatically replace $g_\zeta(d,v_0)$ with $g(d,v_0)$ in part (a) by switching the order of inf/sup with $\zeta\to\infty$. However, part (b) of the theorem proves that replacing $g_\zeta(d,v_0)$ with $g(d,v_0)$ only provides higher upper bounds, which can be used to show the uniform dominance results by analyzing the analytical upper bound for $g(d,v_0)$.

Comments 4.6 Instead of relying on the sufficient condition in part (c), we can investigate the two upper bounds in part (b) $\inf_{(d,v_0)\in H_R} g(d,v_0)$ and $\sup_{(d,v_0)\in H_R} g(d,v_0)$ by simulating $g(d,v_0)$ in (4.15). In practice, one can replace v_0 by its consistent estimator and plot $g(d,v_0)$ as a function of d. This provides a uniform comparison between the averaging estimator and the conservative estimator. One can also simulate the asymptotic risk of other non-standard estimators after deriving their asymptotic distributions like those in Lemma 4.3. As an illustration, we use simulation model 1 in the next section to show the simulated asymptotic risk based on $g(d,v_0)$ is close to the finite-sample risk for two non-standard estimators. One is the averaging GMM estimator based on $\widetilde{\omega}_{eo}$ and the other one is the pre-test GMM estimator based on the over-identification J-test with significance level 0.01. The asymptotic risk for this pre-test estimator is given by (A.75) in the Appendix. The finite sample risks are calculated using 100,000 simulated samples and the asymptotic risks are simulated by drawing 10,000 normal random vectors with mean zero and variance-covariance $\widehat{\Omega}_2$ in each simulated sample.[8] The simulation results are reported in Figure 1, where the risk of the conservative estimator is normalized to be 1 in all cases.[9] It is clear that the finite sample risk and the simulated asymptotic risk are fairly close and the averaging GMM estimator uniformly dominates the conservatives estimator while the pre-test estimator does not.

5 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 two other averaging estimators based on the JS type of weights. The first one is based on the positive part of the JS weight[10]:

$$\omega_{P,JS} = 1 - \left(1 - \frac{\mathrm{tr}(\widehat{A}_{v_0}) - 2\lambda_{\max}(\widehat{A}_{v_0})}{n(\widehat{\theta}_2 - \widehat{\theta}_1)' H (\widehat{\theta}_2 - \widehat{\theta}_1)}\right)_+ \tag{5.1}$$

where $(x)_+ = \max\{0, x\}$ and $\widehat{A}_{v_0}$ is the estimator of A_{v_0} using $\widehat{\theta}_1$. The second one uses the restricted JS weight

$$\omega_{R,JS} = (\omega_{P,JS})_+ . \tag{5.2}$$

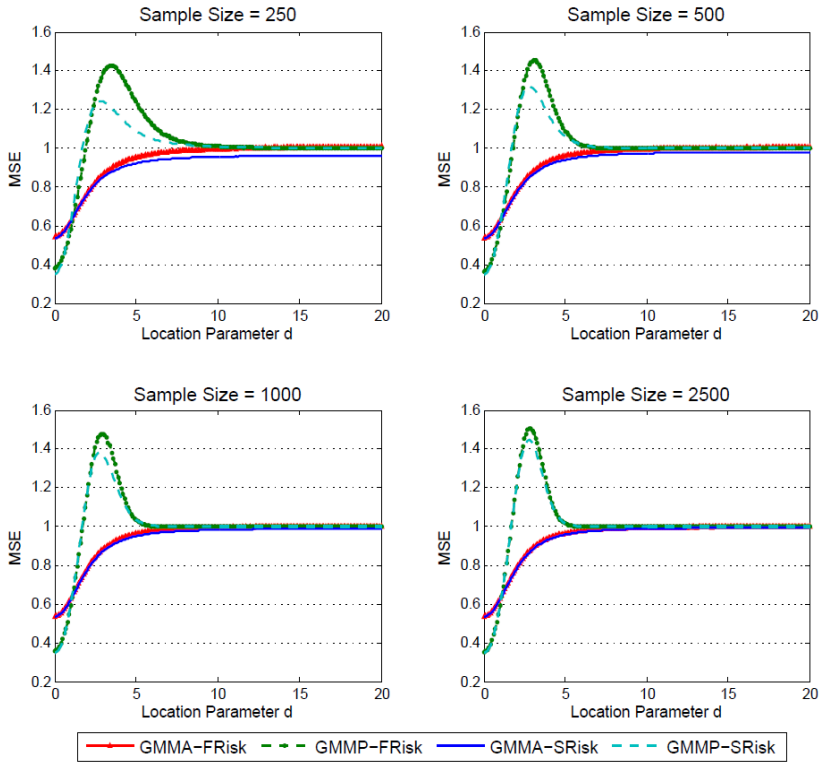
By construction, $\omega_{P,JS} \leq 1$ and $0 \leq \omega_{R,JS} \leq 1$. We compare the finite-sample risks of these three averaging estimators, the conservative GMM estimator $\widehat{\theta}_1$, and the pre-test GMM estimator based on the J-test. The finite-sample risk of the conservative GMM estimator is normalized to be 1.

[8] No truncation is applied to the finite-sample risk.

[9] The finite-sample and simulated asymptotic risk of the averaging GMM estimator are represented by "GMMA-FRisk" and "GMMA-SRisk", respectively. The finite-sample and simulated asymptotic risk of the pre-test GMM estimator are represented by "GMMP-FRisk" and "GMMP-SRisk", respectively.

[10] This formula is a GMM analog of the generalized JS type shrinkage estimator in Hansen (2014a) for parametric models. The shrinkage scalar τ is set to $\mathrm{tr}(\widehat{A}_{v_0}) - 2\lambda_{\max}(\mathrm{tr}(\widehat{A}_{v_0}))$ in a fashion similar to the original JS estimator.

Figure 1. The Finite Sample Risk and the Simulated Asymptotic Risk



In Theorem 4.1(c), we derive a sufficient condition for the uniform dominance: $\text{tr}(A_{v_0}) \geq 4\lambda_{\max}(A_{v_0})$. When this condition is not satisfied, however, it is still possible that our averaging GMM estimator has a smaller risk than the conservative GMM estimator. Therefore we consider three models in simulation studies. In the first model, $\text{tr}(A_{v_0}) \geq 4\lambda_{\max}(A_{v_0})$ and hence the sufficient condition in Theorem 4.1(c) is satisfied. In the second and the third models, $2\lambda_{\max}(A_{v_0}) < \text{tr}(A_{v_0}) < 4\lambda_{\max}(A_{v_0})$ and $\text{tr}(A_{v_0}) < 2\lambda_{\max}(A_{v_0})$, respectively, which means that the sufficient condition in Theorem 4.1(c) does not hold.[11] In each model, we consider four sample sizes, $n = 250, 500, 1000, 2500$, and use 100,000 simulation repetitions.

5.1 Simulation in Model 1

Our first simulation model is

$$Y_i = \sum_{j=1}^{6} \theta_j X_{j,i} + \epsilon_i, \tag{5.3}$$

[11] We differentiate these two cases for a thorough comparison with the JS-type estimators, where the value of $\text{tr}(A_{v_0}) - 2\lambda_{\max}(\text{tr}(A_{v_0}))$ is important.

where $X_{j,i}$ are generated by

$$X_{j,i} = \beta_j(Z_{j,i} + Z_{j+6,i}) + Z_{j+12,i} + u_{j,i} \text{ for } j = 1, ..., 6. \tag{5.4}$$

We draw i.i.d. random vectors $(Z_{1,i}, ..., Z_{18,i}, u_{1,i}, ..., u_{6,i}, \epsilon_i)'$ from normal distributions with mean zero and variance-covariance matrix $\text{diag}(I_{18\times 18}, \Sigma_{7\times 7})$, where

$$\Sigma_{7\times 7} = \begin{pmatrix} I_{6\times 6} & 0.25 \times \mathbf{1}_{6\times 1} \\ 0.25 \times \mathbf{1}_{1\times 6} & 1 \end{pmatrix}. \tag{5.5}$$

We set $(\theta_1, ..., \theta_6) = 2.5 \times \mathbf{1}_{1\times 6}$ and $(\beta_1, ..., \beta_6) = 0.5 \times \mathbf{1}_{1\times 6}$. The observed data are $W_i = (Y_i, X_{1,i}, ..., X_{6,i}, Z_{1,i}, ..., Z_{12,i}, \tilde{Z}_{13,i}, ..., \tilde{Z}_{18,i})'$, where

$$\tilde{Z}_{j,i} = Z_{j,i} + n^{-1/2} d_j \epsilon_i, \text{ for } j = 13, ..., 18. \tag{5.6}$$

In the main regression equation (5.3), all regressors are endogenous because $\mathbb{E}(X_{j,i}\epsilon_i) = 0.25$ for $j = 1, ..., 6$. The instruments $(Z_{1,i}, ..., Z_{12,i})'$ are valid and $(\tilde{Z}_{13,i}, ..., \tilde{Z}_{18,i})'$ are misspecified because $\mathbb{E}(\tilde{Z}_{j,i}\epsilon_i) = n^{-1/2} d_j$ for $j = 13, ..., 18$. In the simulation studies, we consider $(d_{13}, ..., d_{18}) = d \times \mathbf{1}_{1\times 6}$ where d is a scalar that takes values on the grid points between 0 and 20 with the grid length 0.1.

Figure 2. Finite Sample Risks of the Averaging Estimators in Model 1

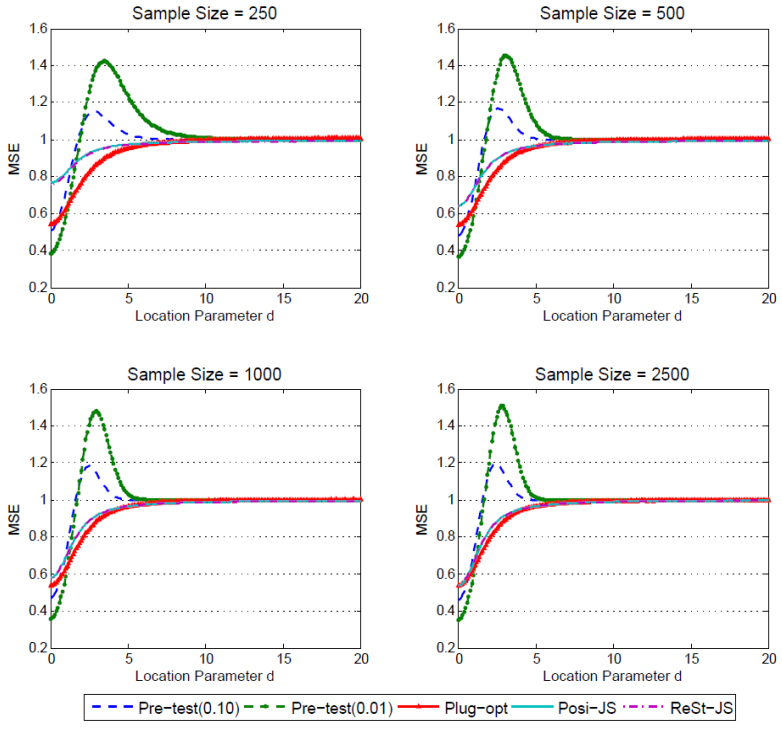


Figure 2 presents the MSEs of all 6 parameters in (5.3). In all figures, "Pre-test(0.10)"and "Pre-test(0.01)" refer to the pre-test GMM estimators based on the J-test with nominal size 0.10 and 0.01, respectively; "Plug-opt" refers to the averaging GMM estimator based on the empirical optimal weight $\widetilde{\omega}_{eo}$; "Posi-JS" and "ReSt-JS" refer to the averaging estimators based on the positive part of the JS weight and the restricted JS weight, respectively.[12]

Our findings in model 1 are summarized as follows. First, the GMM averaging estimators have smaller MSE than $\widehat{\theta}_1$ uniformly over d, which is predicted by our theory because the key sufficient condition is satisfied in this model. Second, the pre-test GMM estimators do not dominate the conservative GMM estimator. When the location parameter d is close to zero, the pre-test GMM estimators have relative MSEs as low as 0.4. However, their relative MSEs are above 1 when d is around 5. Third, the pre-test GMM estimators associated with different nominal sizes display different behaviors. The smaller the size of the over-identification test is, the larger the supremum of the risk is. Fourth, among the three averaging estimators, the one based on $\widetilde{\omega}_{eo}$ has the smallest MSE. The positive JS averaging estimator and the restricted JS averaging estimator have almost identical finite-sample MSE even when the sample size is small, e.g., $n = 250$. Fifth, it is interesting to see that as the sample size grows, the finite sample MSEs of the positive and restricted JS averaging estimators converge to that of the averaging estimator based on $\widetilde{\omega}_{eo}$.

5.2 Simulation in Model 2

The second model is

$$Y_i = \sum_{j=1}^{6} \theta_j X_{j,i} + \epsilon_i, \tag{5.7}$$

where $X_{1,i}$, $X_{2,i}$ and $X_{3,i}$ are exogenous variables generated by

$$X_{1,i} = 3^{-\frac{1}{2}}(Z_{1,i} + Z_{2,i} + Z_{4,i}),\ X_{2,i} = 3^{-\frac{1}{2}}(Z_{2,i} + Z_{3,i} + Z_{6,i}),\ X_{3,i} = 3^{-\frac{1}{2}}(Z_{3,i} + Z_{1,i} + Z_{8,i}), \tag{5.8}$$

and $X_{j,i}$ $(j = 4, 5, 6)$ are generated by

$$X_{j,i} = \beta_j(Z_{j,i} + Z_{j+3,i}) + Z_{j+6,i} + u_{j,i} \text{ for } j = 4, 5, 6. \tag{5.9}$$

We draw i.i.d. random vectors $(Z_{1,i}, ..., Z_{12,i}, u_{4,i}, ..., u_{6,i}, \epsilon_i)'$ from normal distributions with mean zero and variance-covariance matrix $\text{diag}(I_{12\times 12}, \Sigma_{4\times 4})$, where

$$\Sigma_{4\times 4} = \begin{pmatrix} I_{3\times 3} & 0.25 \times \mathbf{1}_{3\times 1} \\ 0.25 \times \mathbf{1}_{1\times 3} & 1 \end{pmatrix}. \tag{5.10}$$

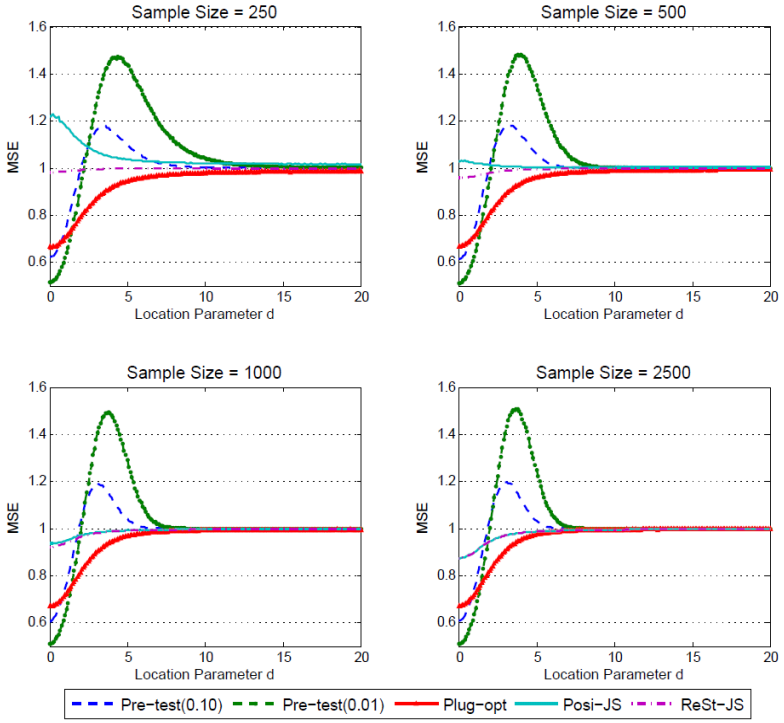
[12] The same notations are used in Figures 3-8.

The observed data are $W_i = (Y_i, X_{1,i}, ..., X_{6,i}, Z_{4,i}, ..., Z_{9,i}, \tilde{Z}_{4,i}, ..., \tilde{Z}_{6,i})'$, where

$$\tilde{Z}_{j,i} = Z_{j+6,i} + n^{-1/2} d_j \epsilon_i \text{ for } j = 4, 5, 6. \tag{5.11}$$

We set $(\theta_1, ..., \theta_6) = 2.5 \times \mathbf{1}_{1\times 6}$ and $(\beta_4, ..., \beta_6) = 0.5 \times \mathbf{1}_{1\times 3}$. In this model, $X_{j,i}$ $(j = 4, 5, 6)$ are endogenous regressors, $(Z_{4,i}, ..., Z_{9,i})'$ are valid IVs, and $(\tilde{Z}_{4,i}, ..., \tilde{Z}_{6,i})'$ are misspecified IVs. In the simulation, we consider $(d_4, ..., d_6) = d \times \mathbf{1}_{1\times 3}$ where d is a scalar that takes values on the grid points between 0 and 20 with grid length 0.1.

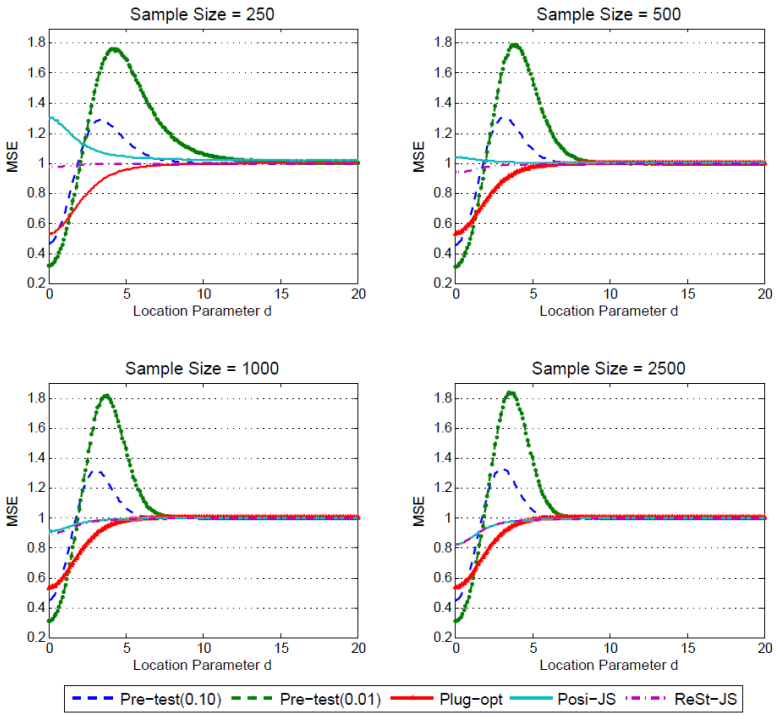
Figure 3. Finite Sample Risks of Averaging Estimators in Model 2



The simulation results in model 2 are depicted in Figures 3, 4 and 5. Figure 3 presents the MSEs of all 6 parameters in (5.7). Figure 4 contains the MSEs of the estimators of $(\theta_4, \theta_5, \theta_6)'$, the coefficients of the endogenous regressors in the main equation (5.7). Figure 5 provides the MSEs of the estimators of $(\theta_1, \theta_2, \theta_3)'$, the coefficients of the exogenous regressors in the main equation (5.7).

Our findings in Figure 3 are summarized as follows. First, even though the sufficient condition in Theorem 4.1(c) is not satisfied, the averaging estimator based on $\widetilde{\omega}_{eo}$ has a smaller MSE than $\widehat{\theta}_1$ uniformly over d. Moreover, its MSE is much smaller than that of the other two averaging estimators. Second, the properties of the pre-test estimators are similar to those in model 1. That

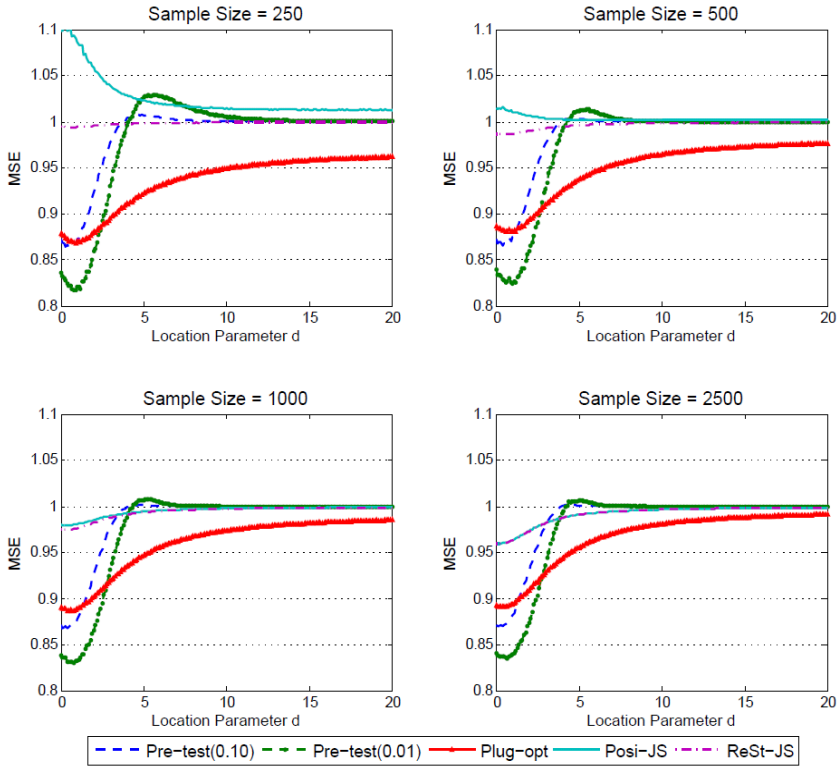
Figure 4. Finite Sample Risks of the Averaging Estimators in Model 2 – Endogenous Subvector



is, they do not dominate the conservative estimator and the behavior changes with the nominal size of the test. Third, the averaging estimator using $\omega_{P,JS}$ has a larger MSE than $\widehat{\theta}_1$ when the location parameter d is close to zero and the sample size is small (e.g., $n = 250$ and 500). When the sample size becomes large (e.g., $n = 1000$ and 2500), its MSE dominates $\widehat{\theta}_1$ uniformly over d. Fourth, the averaging estimator using $\omega_{R,JS}$ dominates $\widehat{\theta}_1$ uniformly over d in all the sample sizes we considered. Fifth, with the growth of the sample size, the finite sample MSE of the averaging estimator using $\omega_{P,JS}$ converges to that of the averaging estimator using $\omega_{R,JS}$. Moreover, the finite sample MSEs of these two averaging estimators converge to that of the GMM averaging estimator based on $\widetilde{\omega}_{eo}$.

In Figures 4 and 5, the averaging estimators based on $\widetilde{\omega}_{eo}$ and $\omega_{R,JS}$ both uniformly dominate $\widehat{\theta}_1$ and the former is better than the latter. In particular, for the coefficients of the exogenous regressors $(\theta_1, \theta_2, \theta_3)'$, the averaging estimator based on $\widetilde{\omega}_{eo}$ demonstrates a substantial advantage uniformly over d.

Figure 5. Finite Sample Risks of the Averaging Estimators in Model 2 – Exogenous Subvector



5.3 Simulation in Model 3

The third simulation model is

$$Y_i = \sum_{j=1}^{6} \theta_j X_{j,i} + \epsilon_i, \tag{5.12}$$

where $X_{j,i}$ $(j = 1, ..., 5)$ are exogenous variables generated by

$$\begin{aligned} X_{j,i} &= 3^{-\frac{1}{2}}(Z_{j,i} + Z_{j+1,i} + Z_{j+8,i}), \text{for } j = 1, ..., 4, \\ X_{5,i} &= 3^{-\frac{1}{2}}(Z_{5,i} + Z_{1,i} + Z_{13,i}), \end{aligned} \tag{5.13}$$

and $X_{6,i}$ is generated by

$$X_{6,i} = \sum_{j=6}^{8} \beta_j Z_{j,i} + \sum_{j=9}^{13} Z_{j,i} + u_{6,i} \ . \tag{5.14}$$

We draw i.i.d. random vectors $(Z_{1,i}, ..., Z_{13,i}, u_{6,i}, \epsilon_i)'$ from normal distributions with mean zero and variance-covariance matrix $diag(I_{13\times 13}, \Sigma_{2\times 2})$, where

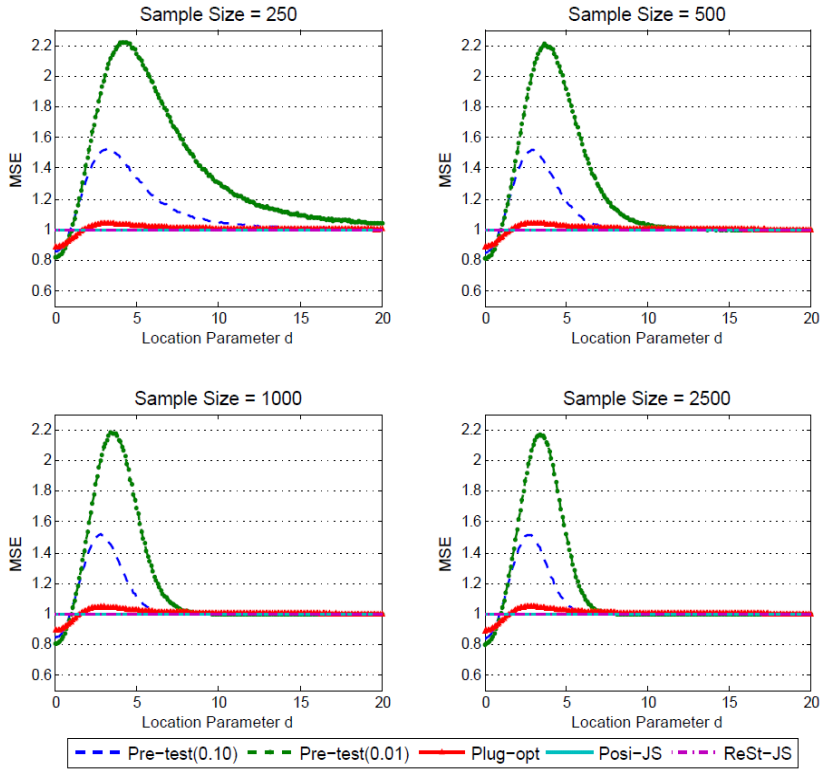
$$\Sigma_{2\times 2} = \begin{pmatrix} 1 & 0.25 \times \mathbf{1}_{2\times 1} \\ 0.25 \times \mathbf{1}_{1\times 2} & 1 \end{pmatrix}. \tag{5.15}$$

The observed data are $W_i = (Y_i, X_{1,i}, ..., X_{6,i}, Z_{6,i}, ..., Z_{8,i}, \tilde{Z}_{9,i}, ..., \tilde{Z}_{13,i})'$, where

$$\tilde{Z}_{j,i} = Z_{j,i} + n^{-1/2} d_j \epsilon_i \text{ for } j = 9, ..., 13. \tag{5.16}$$

We set $(\theta_1, ..., \theta_6) = 2.5 \times \mathbf{1}_{1\times 6}$ and $(\beta_6, ..., \beta_8) = 0.5 \times \mathbf{1}_{1\times 3}$. In this model, $X_{6,i}$ is an endogenous regressor, $(Z_{6,i}, ..., Z_{8,i})'$ are valid IVs, and $(\tilde{Z}_{9,i}, ..., \tilde{Z}_{13,i})'$ are misspecified IVs. We consider $(d_9, ..., d_{13}) = d \times \mathbf{1}_{1\times 5}$ where d is a scalar that takes values on the grid points between 0 and 20 with grid length 0.1.

Figure 6. Finite Sample Risks of the Averaging Estimators in Model 3

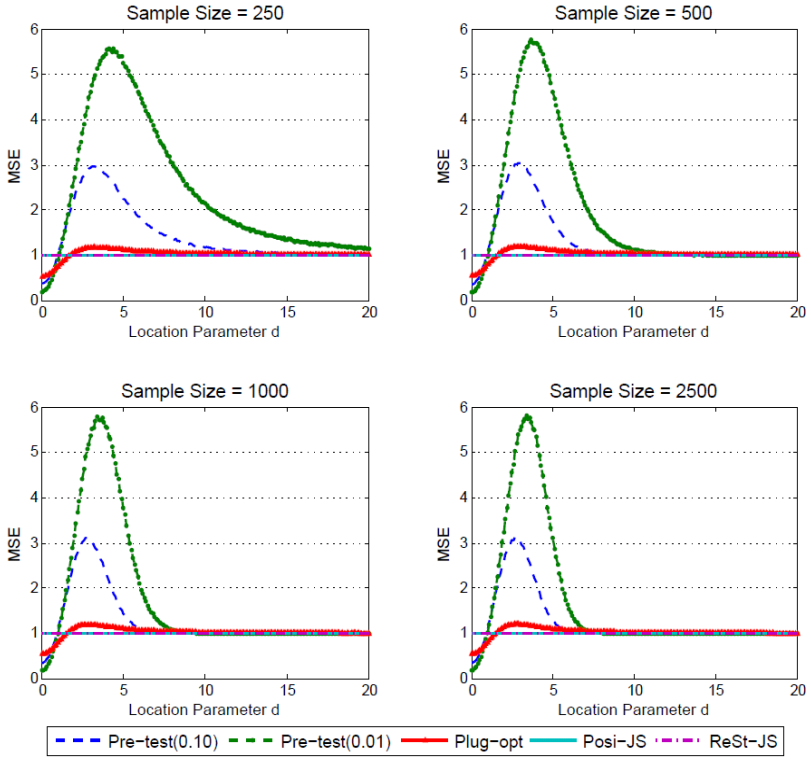


The simulation results in model 3 are depicted in Figures 6, 7, and 8. Figure 6 presents the MSEs of all 6 parameters in (5.12). Figure 7 shows the MSEs of the estimators of θ_6, the coefficient of the endogenous regressor in the main equation (5.12). Figure 8 provides the MSEs of the estimators of

$(\theta_1, ..., \theta_5)'$, the coefficients of the exogenous regressors in the main equation (5.7).

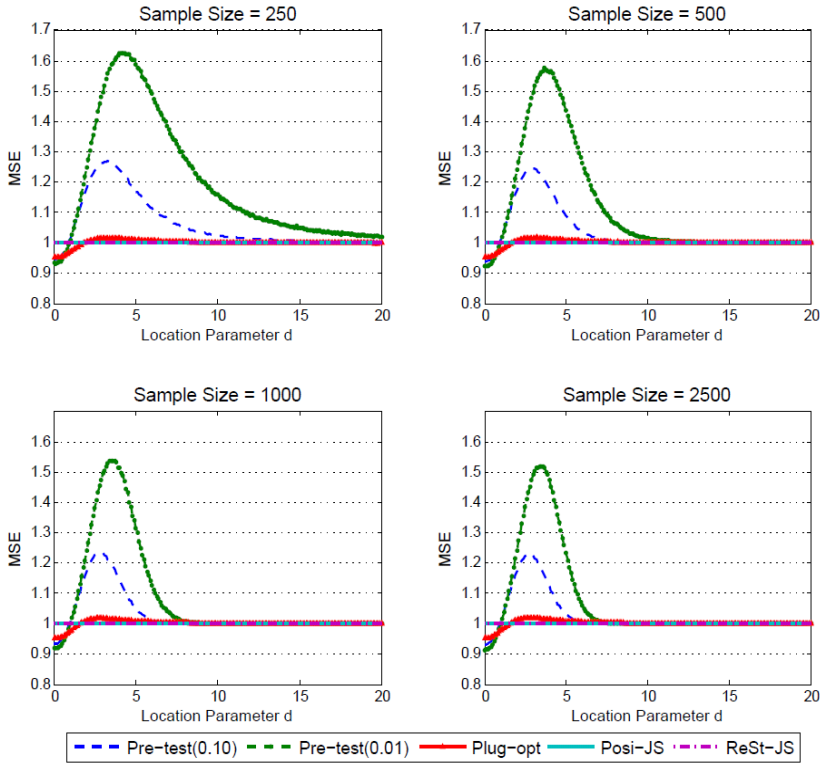
Our findings in Figure 6 are summarized as follows. First, the properties of the pre-test estimators are very similar to those in models 1 and 2. Second, the MSEs of the averaging estimators based on $\omega_{P,JS}$ and $\omega_{R,JS}$ becomes identical to that of $\widehat{\theta}_1$. Third, the averaging estimator based on $\widetilde{\omega}_{eo}$ does not dominate $\widehat{\theta}_1$. It has much smaller MSE when d is close to zero, while its MSE is inflated slightly above that of $\widehat{\theta}_1$ when d moves away from zero, and then converges to 1.

Figure 7. Finite Sample Risks of the Averaging Estimators in Model 3 – Endogenous Subvector



Comparing the results in Figures 7 and 8, we see that pre-testing and model averaging have stronger effect on the coefficient of the endogenous regressor. In Figures 7 and 8, the averaging estimators based on $\omega_{R,JS}$ and $\omega_{P,JS}$ are almost identical to $\widehat{\theta}_1$ for all sample sizes considered. Finally, for θ_6, the averaging estimator based on $\widetilde{\omega}_{eo}$ does not dominate $\widehat{\theta}_1$, although its MSE is only inflated slightly for d around 3. On the other hand, for $(\theta_1, ..., \theta_5)'$, the averaging estimator based on $\widetilde{\omega}_{eo}$ dominates $\widehat{\theta}_1$ when the sample size is small (e.g., $n = 250, 500, 1000$). Its MSE is slightly inflated when d is around 3 and the sample size is large (e.g., $n = 2500$).

Figure 8. Finite Sample Risks of the Averaging Estimators in Model 3 – Exogenous Subvector



6 An Empirical Application

One important issue in the empirical analysis of life cycle labor supply is to estimate the individual human capital production function. The knowledge about the human capital function allows researchers to estimate the household's utility function, and hence to evaluate how changes in policies, such as tax reduction, affect consumption, labor market outcomes, and welfare (see, e.g., Heckman, 1976; Shaw, 1989; and Imai and Keane, 2004). This section applies the averaging GMM to estimate the human capital production function.

We follow the literature (see, e.g., Shaw, 1989) to specify the human capital production function as a quadratic function of $k_{i,t}$, log of the human capital stock $K_{i,t}$, and $h_{i,t}$, log of the hours of work $H_{i,t}$:

$$f(k_{i,t}, h_{i,t}, \theta) = \gamma_1 h_{i,t} + \gamma_2 h_{i,t}^2 + \gamma_3 h_{i,t} k_{i,t} + \gamma_4 k_{i,t} + \gamma_5 k_{i,t}^2, \tag{6.1}$$

where $\theta = (\gamma_1, \ldots, \gamma_5)$ are unknown parameters. Denote the regressors by

$$X_{i,t} = \left(h_{i,t},\ h_{i,t}^2,\ h_{i,t}k_{i,t},\ k_{i,t},\ k_{i,t}^2\right)'. \tag{6.2}$$

The log human capital stock $k_{i,t}$ is accumulated through the equation

$$k_{i,t+1} = f(k_{i,t}, h_{i,t}, \theta) + \varepsilon_{i,t} \tag{6.3}$$

where $\varepsilon_{i,t} = \alpha_i + u_{i,t}$ is the unobservable residual term that contains an individual heterogeneity component α_i and a random shock $u_{i,t}$. To avoid unnecessary complications, we follow Shaw (1989) to specify the real wage as $w_{i,t} = R_{i,t}K_{i,t}$, and follow Hokayem and Ziliak (2014) to assume $R_{i,t} = 1$ for all i and all t.[13]

We use the same data set as in Hokayem and Ziliak (2014) from the Panel Study of Income Dynamics (PSID). The sample includes biennial observations for 1654 men from 1999 to 2009. We further narrow the sample to individuals with at least three consecutive periods of observations, which gives us a data set with 5774 individual-year observations.

To eliminate the individual effect, we take first difference on equation (6.3):

$$\Delta k_{i,t+1} = \Delta f(k_{i,t}, h_{i,t}, \theta) + \Delta u_{i,t} \tag{6.4}$$

where "Δ" denotes the first order difference operator. The unknown parameter θ can be estimated by GMM estimator $\widehat{\theta}_1$ with the moment functions

$$g_1(\Delta k_{i,t+1}, \Delta X_{i,t}, Z_{1,t}, \theta) = [\Delta k_{i,t+1} - \Delta f(k_{i,t}, h_{i,t}, \theta)] \otimes Z_{1,t} \tag{6.5}$$

where $Z_{1,t} = (X'_{i,t-1}, Z'_{*,t})$ is a set of IVs including $X_{i,t-1}$ and

$$Z_{*,t} = \left(c_{i,t-1},\, c^2_{i,t-1},\, c_{i,t-1}l_{i,t-1},\, l_{i,t-1},\, l^2_{i,t-1}\right)', \tag{6.6}$$

where $c_{i,t-1} = \log C_{i,t-1}$, $l_{i,t-1} = \log L_{i,t-1}$, and $C_{i,t-1}$ and $L_{i,t-1}$ are, respectively, the consumption and leisure of individual i at period $t-1$. The lagged consumption and leisure variables are included to provide extra identification restrictions for the human capital function.

In equation (6.4), the regressors $\Delta X_{i,t}$ may be endogenous because: (i) $k_{i,t}$ is correlated with $u_{i,t-1}$ and hence $\Delta u_{i,t}$ in view of equation (6.3); and (ii) $h_{i,t}$ is partly determined by $k_{i,t}$ through the individual's labor decision. As a result, the LS estimator based on the following moment function

$$g^*(\Delta k_{i,t+1}, \Delta X_{i,t}, \theta) = [\Delta k_{i,t+1} - \Delta f(k_{i,t}, h_{i,t}, \theta)] \otimes \Delta X_{i,t} \tag{6.7}$$

may be inconsistent. The aggressive GMM estimator $\widehat{\theta}_2$ is constructed using the moment conditions in both (6.5) and (6.7).

[13] Another way to think of this specification is to use real wage rate w_t as a proxy for the human capital stock K_t.

Table 1. Estimator of Human Capital Production Function

	γ_1	γ_2	γ_3	γ_4	γ_5	J-test
$\widehat{\theta}_1$	0.0236	-0.0070	0.0310	0.0656	-0.0381	0.8427
	(0.0571)	(0.0444)	(0.0626)	(0.0621)	(0.0447)	——
$\widehat{\theta}_2$	0.0009	0.0265	-0.0113	-0.2232	-0.0925	0
	(0.0328)	(0.0240)	(0.0496)	(0.0529)	(0.0247)	——

(i) Numbers in the brackets are the standard errors; (ii) Numbers in the last column are the p-values of the J-tests; (iii) GMM estimators are based on the sample from PSID in year 2003, 2005, 2007 and 2009; (iv) Four year dummy variables are included in the moment functions and they are used as their own IVs in the GMM estimation.

Table 1 reports the estimation results on the conservative and the aggressive estimators. The conservative and aggressive GMM estimators of θ differ substantially. The J-test strongly rejects the validities of the moment conditions in (6.7), while it supports the validities of the moment conditions in (6.5). On the other hand, the aggressive GMM estimator $\widehat{\theta}_2$ has much smaller standard error than the conservative estimator $\widehat{\theta}_1$.

Next, we consider the averaging GMM estimator under the quadratic loss function with $H = I_{d_\theta}$. The empirical weight $\widehat{\omega}_{eo}$ on the aggressive GMM estimator is 0.0770. It is interesting that the averaging estimator assigns nontrivial weight to $\widehat{\theta}_2$, even though the J-test indicates misspecification of the moment conditions in (6.7).

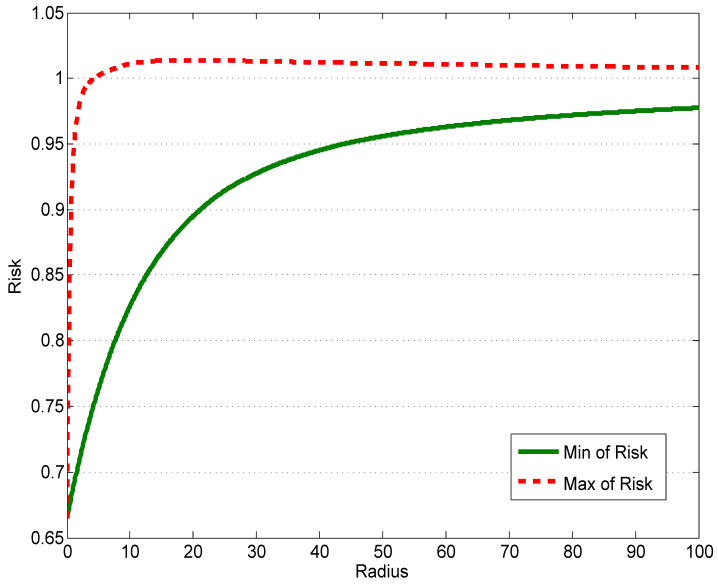
To evaluate the performance of the averaging GMM estimator, we simulate its asymptotic risk following the formula in (4.15). This exercise is the same as that for Figure 1, which shows that this simulated asymptotic risk is a good approximation to the finite-sample risk. As there are 5 moment conditions in (6.7), the risk of the averaging GMM estimator is a function of a 5-dimensional vector of location parameters $d = (d_1, d_2, d_3, d_4, d_5) \in \mathbb{R}^5$. We parameterize it as

$$\begin{aligned} d_1 &= \sqrt{r}\cos\alpha_1, \\ d_2 &= \sqrt{r}\sin\alpha_1\sin\alpha_2\sin\alpha_3, d_3 = \sqrt{r}\sin\alpha_1\sin\alpha_2\cos\alpha_3, \\ d_4 &= \sqrt{r}\sin\alpha_1\cos\alpha_2\sin\alpha_4, d_5 = \sqrt{r}\sin\alpha_1\cos\alpha_2\cos\alpha_4 \end{aligned} \tag{6.8}$$

for some $r \in [0, +\infty)$ and $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in [0, 2\pi]$ such that $\sum_{k=1}^5 d_k^2 = r$. To simulate the risk, we consider 1001 equally spaced grid points for r between 0 and 100, and for each grid point of r, we consider 30 equally spaced grid points for α_1, α_2, α_3 and α_4, respectively, between 0 and 2π (starting at 0). For each grid point of r, this gives 30^4 values for the simulated risk and we record the minimum and maximum values. As in the Monte Carlo simulation studies, the risk of the conservative GMM estimator is normalized to be 1.

The minimum and maximum risks for each grid point of r are depicted in Figure 9. Figure 9 shows that the averaging GMM estimator $\widehat{\theta}_{eo}$ compares favorably to the conservative GMM

Figure 9. Simulated Asymptotic Risk of the Averaging Estimator
of the Human Capital Function



estimator $\widehat{\theta}_1$. The risk of $\widehat{\theta}_{eo}$ is around 1.02 in the least favorable case and is around 0.66 in the most favorable case. As r goes to 100, the maximum and the minimum risks both converge to 1.

7 Conclusion

This paper studies the asymptotic risk of 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 and risk differences. Using this asymptotic approximation, we show that the proposed averaging GMM estimator uniformly dominates the conservative GMM estimator.

Inference based on the averaging estimator is an interesting and challenging problem. 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.

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 (2012): "Estimation and Inference With Weak, Semi-Strong and Strong Identification," *Econometrica*, 80, 2153-2211.

[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*.

[4] Andrews, D. W. K., P. Guggenberger (2006): "On the Asymptotic Maximal Risk of Estimators and the Hausman Pre-Test Estimator," *Working Paper*.

[5] Andrews, D. W. K. and P. Guggenberger (2009): "Hybrid and Size-Corrected Subsampling Methods," *Econometrica*, 77(3), 721–762.

[6] Andrews, D. W. K. and P. Guggenberger (2010): "On the Asymptotic Maximal Risk of Estimators and the Hausman Pre-test Estimator," *Working Paper*.

[7] 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.

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

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

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

[11] Burnham, K. P., and D. R. Anderson (2002): *Model Selection and Multimodel Inference: A Practical Information—Theoretic Approach*. Berlin: Springer-Verlag.

[12] Caner, M., X. Han and Y. Lee (2014): "Bias-Corrected Semiparametrically Efficient High Dimensional GMM Estimator with Many Invalid Moment Conditions: An Application to Dynamic Panel Data Models," *Working Paper.*

[13] Cheng X. and B. Hansen (2014): "Forecasting with Factor-Augmented Regression: A Frequentist Model Averaging Approach," *Journal of Econometrics*, forthcoming.

[14] Cheng X. and Z. Liao (2014): "Select the Valid and Relevant Moments: An Information-Based LASSO for GMM with Many Moments," *Journal of Econometrics*, forthcoming.

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

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

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

[18] DiTraglia, F. (2014): "Using Invalid Instruments on Purpose: Focused Moment Selection and Averaging for GMM," University of Pennsylvania, *Working Paper.*

[19] Doko Tchatoka, F. and J.-M. Dufour (2012): "Identification-robust Inference for Endogeneity Parameters in Linear Structural Models," *MPRA Paper 40695*, University Library of Munich, Germany.

[20] Eichenbaum, M., L. P. 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.

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

[22] Hall, A. R. and A. Inoue (2003): "The Large Sample Behavior of the Generalized Method of Moments Estimator in Misspecified Models," *Journal of Econometrics*, 114, 361-394.

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

[24] Hansen, B. E. (2008). "Least Squares Forecast Averaging," *Journal of Econometrics*, 146, 342-350.

[25] Hansen, B. E. (2014a): "Efficient Shrinkage in Parametric Models," *Working Paper*, University of Wisconsin.

[26] Hansen, B. E. (2014b): "A Stein-Like 2SLS Estimator," *Working Paper*, University of Wisconsin.

[27] Hansen, B. E. (2014c): "Model Averaging, Asymptotic Risk, and Regressor Groups," *Quantitative Economics*, 5, 495-530.

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

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

[30] Hausman (1978): "Specification Tests in Econometrics," *Econometrica,* 46, 1251-1271.

[31] Heckman, J. J. (1976): "Estimation of A Human Capital Production Function Embedded in A Life-cycle Model of Labor Supply," In *Household Production and Consumption, NBER*, 225-264.

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

[33] 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.

[34] Hokayem, C. and J. P. Shum (2014): "Health, Human Capital, and Life Cycle Labor Supply," *American Economic Review: Papers & Proceedings 2014*, 104(5), 127-131.

[35] 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.

[36] Imai, S. and M. P. Kean (2004): "Intertemporal Labor Supply and Human Capital Accumulation," *International Economic Review*, 45, 601-641.

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

[38] Kang, H., A. Zhang, T. Cai, D. Small (2014): "Instrumental Variables Estimation With Some Invalid Instruments and its Application to Mendelian Randomization," *Working Paper,* The Wharton School, University of Pennsylvania.

[39] Kolesar, M., R. Chetty, J. Friedman, E. Glaeser, G. Imbens (2014): "Identification and Inference with Many Invalid Instruments," *Journal of Business Economics and Statistics,* forthcoming

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

[41] Liang, H., G. Zou, A. T. Wan, and X. Zhang (2011): "Optimal Weight Choice for Frequentist Model Average Estimators," *Journal of the American Statistical Association*, 106, 1053-1066.

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

[43] Liu, C-A. (2013): "Distribution Theory of the Least Squares Averaging Estimator," *Working Paper*, National University of Singapore.

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

[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] Rudin W (1976): *Principles of Mathematical Analysis*, International Series in Pure and Applied Mathematics, McGraw-Hill, New York.

[48] Sargan, J. (1958): "The estimation of economic relationships using instrumental variables," *Econometrica*, 26(3), 393-415

[49] Saleh, A. K. Md. Ehsanses (2006): *Theory of Preliminary Test and Stein-Type Estimation with Applications*, Hoboken, Wiley.

[50] Shaw, K. (1989): "Life-cycle Labor Supply with Human Capital Accumulation," *International Economic Review*, 30, 431-456.

[51] 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.

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

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

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

A Appendix

A.1 Proofs for the general asymptotic risk results

Proof of Theorem 3.1. The proof uses the subsequence techniques used to show the asymptotic size of a test in Andrews, Cheng, and Guggenberger (2011). We first show that

$$AsyR(\widehat{\theta}) \leq \max\left\{ \sup_{(d,v_0)\in H_R} R(d,v_0), \sup_{v_0\in H_\infty} R(\infty,v_0) \right\}. \tag{A.1}$$

Let $\{F_n\}$ be a sequence such that

$$\limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widehat{\theta})] = \limsup_{n\to\infty} \left(\sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell(\widehat{\theta})] \right) = AsyR(\widehat{\theta}). \tag{A.2}$$

Such a sequence always exists by the definition of supremum. The sequence $\{\mathbb{E}_{F_n}[\ell(\widehat{\theta})] : n \geq 1\}$ may not converge. Now let $\{w_n : n \geq 1\}$ be a subsequence of $\{n\}$ such that $\{\mathbb{E}_{F_{w_n}}[\ell(\widehat{\theta})] : n \geq 1\}$ converges and its limit equals $AsyR(\widehat{\theta})$. Such a subsequence always exists by the definition of limsup. Below we show that there exists a subsequence $\{p_n\}$ of $\{w_n\}$ such that

$$\mathbb{E}_{F_{p_n}}[\ell(\widehat{\theta})] \to R(d,v_0) \text{ for some } (d,v_0) \in H_R \tag{A.3}$$

or

$$\mathbb{E}_{F_{p_n}}[\ell(\widehat{\theta})] \to R(\infty,v_0) \text{ for some } v_0 \in H_\infty. \tag{A.4}$$

Provided (A.3) or (A.4) holds, we obtain the desired result in (A.1).

To show that there exists a subsequence $\{p_n\}$ of $\{w_n\}$ such that either (A.3) or (A.4) holds, it suffices to show claims (1) and (2): (1) for any sequence $\{F_n\}$ and any subsequence $\{w_n\}$ of $\{n\}$, there exists a subsequence $\{p_n\}$ of $\{w_n\}$ for which

$$p_n^{1/2}\delta(F_{p_n}) \to d \in R^{r^*} \text{ and } v(F_{p_n}) \to v_0 \text{ for some } (d,v_0) \in H_R \tag{A.5}$$

or

$$\left\| p_n^{1/2}\delta(F_{p_n}) \right\| \to \infty \text{ and } v(F_{p_n}) \to v_0 \text{ for some } v_0 \text{ such that } v_0 \in H_\infty; \tag{A.6}$$

and (2) for any subsequence $\{p_n\}$ of $\{n\}$ and any sequence $\{F_{p_n} : n \geq 1\}$, (A.5) together with Assumption 3.1(i) implies (A.3), and (A.6) combined with Assumption 3.1(ii) implies (A.4).

To show (1), let $\delta_{w_n,j}$ denote the j-th component of $\delta(F_{w_n})$ and $p_{1,n} = w_n \ \forall n \geq 1$. For $j = 1$, either (i) $\limsup_{n\to\infty} |p_{j,n}^{1/2}\delta_{p_{j,n},j}| < \infty$ or (ii) $\limsup_{n\to\infty} |p_{j,n}^{1/2}\delta_{p_{j,n},j}| = \infty$. If (i) 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 $d_j \in R$. If (ii) 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 $\{p_n^*\}$ of $\{w_n\}$ such that $(p_n^*)^{1/2}\delta_{p_n^*} \to d^* \in \mathbb{R}^{r^*}$ or $(p_n^*)^{1/2}||\delta_{p_n^*}|| \to \infty$. Finally, there exists a subsequence $\{p_n\}$ of $\{p_n^*\}$ such that $v(F_{p_n}) \to v^*$ because $\{v(F) : F \in \mathcal{F}\}$ is a compact set by Assumption 3.2.

We have constructed the subsequence $\{p_n\}$ of $\{n\}$ such that either (i) $(p_n)^{1/2}\delta_{p_n} \to d^* \in \mathbb{R}^{r^*}$

and $v(F_{p_n}) \to v^*$; or (ii) $(p_n)^{1/2}||\delta_{p_n}|| \to \infty$ and $v(F_{p_n}) \to v^*$. To conclude (A.5) holds in case (i), it remains to show $(d^*, v^*) \in H_R$ in case (i). Similarly, to show (A.6) holds in case (ii), it remains to show $v^* \in H_\infty$. This step is necessary because d^* and v^* are the limits along a subsequence, whereas H_R and H_∞ are defined using limits of the full sequence. To close this gap, we show that for the subsequence $\{p_n\}$ constructed above there exists a full sequence with the same limit. For case (i), such a full sequence of DGP $\{F_k^* \in \mathcal{F} : k \geq 1\}$ can be constructed as follows. First, consider the case where $d^* \in \mathbb{R}^{r^*}$. (i) $\forall k = p_n$, define $F_k^* = F_{p_n}$ and (ii) $\forall k \in (p_n, p_{n+1})$, define F_k^* to be a true distribution such that

$$\delta(F_k^*) = (p_n/k)^{1/2}\delta_{p_n} \text{ and } v(F_k^*) = v(F_{p_n}). \tag{A.7}$$

There exists $F_k^* \in \mathcal{F}$ for which (A.7) holds for large n by Assumption 3.2(iii). To see it, we first note that $(\delta(F_{p_n}), v(F_{p_n})) \in \Lambda$ because $F_{p_n} \in \mathcal{F}$. Moreover, we have $p_n/k < 1$, and $||\delta_{p_n}|| < \varepsilon$ for large n because $\delta_{p_n} \to \mathbf{0}_{r^*}$. Hence Assumption 3.2(iii) holds, which ensures the existence of F_k^* for any $k \in (p_n, p_{n+1})$. Along this constructed sequence $\{F_k^* \in \mathcal{F} : k \geq 1\}$, we have $k^{1/2}\delta(F_k^*) \to d^*$ and $v(F_k^*) \to v^*$ as desired. This shows that $(d^*, v^*) \in H_R$ in case (i). For case (ii), define $F_k^* = F_{p_n}$ for $k \in [p_n, p_{n+1})$. Then, $k^{1/2}||\delta(F_k^*)|| \geq (p_n)^{1/2}||\delta_{p_n}|| \ \forall k \in [p_n, p_{n+1})$. In consequence, $(p_n)^{1/2}||\delta_{p_n}|| \to \infty$ as $n \to \infty$ implies that $k^{1/2}||\delta(F_k^*)|| \to \infty$ as $k \to \infty$. In addition, $v(F_k^*) \to v^*$ as $k \to \infty$. Hence, in case (ii), $v^* \in H_\infty$. Combined the results for case (i) and (ii), we have completed the proof of (1).

To show (2), note that we have proved that for any subsequence $\{p_n\}$ of $\{n\}$ and any sequence $\{F_{p_n} : n \geq 1\}$ such that (A.5) holds, there exists a full sequence $\{F_k^* \in \mathcal{F} : k \geq 1\}$ such that $n^{1/2}\delta(F_k^*) \to d^* \in \mathbb{R}^{r^*}$, $v(F_n^*) \to v^*$, and $F_{p_n}^* = F_{p_n}$ $\forall n \geq 1$. Similarly, if (A.6) holds, there exists a full sequence $\{F_k^* \in \mathcal{F} : k \geq 1\}$ such that $n^{1/2}\delta(F_k^*) \to \infty$, $v(F_n^*) \to v^*$, and $F_{p_n}^* = F_{p_n}$ $\forall n \geq 1$. This together with Assumption 3.1(i) and (ii) implies (2). This proves either (A.3) or (A.4) holds, which in turn implies (A.1).

Next, we show that

$$AsyR(\widehat{\theta}) \geq \max\left\{\sup_{(d,v_0)\in H_R} R(d, v_0), \sup_{v_0 \in H_\infty} R(\infty, v_0)\right\}. \tag{A.8}$$

For any $(d, v_0) \in H_R$, there exists a sequence $\{F_n \in \mathcal{F} : n \geq 1\}$ such that $n^{1/2}\delta(F_n) \to d$ and $v(F_n) \to v_0$. Moreover,

$$AsyR(\widehat{\theta}) = \limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell(\widehat{\theta})] \geq \limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widehat{\theta})] = R(d, v_0), \tag{A.9}$$

where the last equality holds by Assumption 3.1(i). Similarly, for any $v_0 \in H_\infty$, there exists a sequence $\{F_n \in \mathcal{F} : n \geq 1\}$ such that $n^{1/2}||\delta(F_n)|| \to \infty$ and $v(F_n) \to v_0$, which together with Assumption 3.1(ii) implies that

$$AsyR(\widehat{\theta}) = \limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell(\widehat{\theta})] \geq \limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell(\widehat{\theta})] = R(\infty, v_0). \tag{A.10}$$

(A.9) combined with (A.10) immediately yields (A.8). Finally, part (a) of the Theorem is implied

by (A.1) and (A.8).

The claim in part (b) follows from the same arguments as those in part (a) with $\mathbb{E}_F[\ell(\widehat{\theta})]$ replaced by $\mathbb{E}_F[\ell(\widehat{\theta}) - \ell(\widetilde{\theta})]$. ■

Proof of Corollary 3.2. For any $\zeta \in \mathbb{R}_+$, Theorem 3.1 together with Assumptions 3.2 and 3.3(i) implies that

$$\begin{aligned} AsyR^*_\zeta(\widehat{\theta}) &\equiv \limsup_{n\to\infty} \left(\sup_{F\in\mathcal{F}} \mathbb{E}_F[\ell_\zeta(\widehat{\theta})] \right) \\ &= \max\left\{ \sup_{(d,v_0)\in H_R} R_\zeta(d,v_0), \sup_{v_0\in H_\infty} R_\zeta(\infty,v_0) \right\}. \end{aligned} \tag{A.11}$$

Part (a) follows from $AsyR^*(\widehat{\theta}) = \lim_{\zeta\to\infty} AsyR^*_\zeta(\widehat{\theta})$ and (A.11). Part (b) follows from part (b) of Theorem 3.1 and the definitions of $Asy\underline{RD}^*(\widehat{\theta},\widetilde{\theta})$ and $Asy\overline{RD}^*(\widehat{\theta},\widetilde{\theta})$. ■

A.2 Proofs for the conservative and aggressive estimators

Lemma A.1 *Under Assumption 4.1, we have the following results for any $v_0 \in H_\infty$.*

(a) *$M_2(\theta)'\Omega_2^{-1}M_2(\theta)$ uniquely identifies $\theta^*(v_0)$.*

(b) *$M_1(\theta) = \mathbf{0}_{r_1}$ uniquely identifies $\theta(v_0)$, where $M_1(\theta)$ denotes the first r_1 rows of $M_2(\theta)$.*

Proof of Lemma A.1. Note that $M_2(\cdot)$ and Ω_2 are the limits of $M_2(\cdot,F_n)$ and $\Omega_2(F_n)$. By Assumption 3.2(i), there exists $F_0 \in \mathcal{F}$ such that $M_2(\theta) = E_{F_0}[g_2(W_i,\theta)]$ and $\Omega_2 = \Omega_2(F_0)$. Following Assumption 4.1, $M_2(\theta)'\Omega_2^{-1}M_2(\theta)$ uniquely identifies $\theta^*(v_0)$ and it only depends on v_0, not on F_0. Similarly, $E_{F_0}[g_1(W_i,\theta)] = M_1(\theta)$, which uniquely identifies $\theta(v_0)$ by Assumption 4.1. ■

For notational simplicity, $\theta(v_0)$ and $\theta^*(v_0)$ defined in Lemma A.1 are abbreviated to θ_0 and θ_0^* in the proof below.

Lemma A.2 *Suppose Assumptions 4.1-4.3 hold. Let $\mathcal{S}_2(v_0) \equiv \mathcal{S}(d,v_0) \cup \mathcal{S}(\infty,v_0)$. Under $\{F_n\} \in \mathcal{S}_2(v_0)$, $\widehat{\theta}_1 - \theta_n \to_p 0$ and $\widehat{\theta}_2 \to_p \theta_0^*$, recall that $\theta_n = \theta(F_n)$.*

Proof of Lemma A.2. We first show the results for $\widehat{\theta}_2$. Note that $\mathbb{E}_{F_n}[g_2(W_i,\theta)] \to M_2(\theta)$ by $v(F_n) \to v_0$, which together with the uniform law of large numbers (ULLN) in Assumption 4.3(i) implies that

$$n^{-1}\sum_{i=1}^n [g_2(W_i,\theta) - M_2(\theta)] = n^{-1/2}\xi_n(g_2(\theta)) + (\mathbb{E}_{F_n}[g_2(W_i,\theta)] - M_2(\theta)) \to_p \mathbf{0}_{r_2}. \tag{A.12}$$

uniformly over $\theta \in \Theta$. Using (A.12) and Assumption 4.2(v), we deduce that

$$\begin{aligned} Q_{F_n}(\theta) &= \frac{[\sum_{i=1}^n g_2(W_i,\theta)]'\mathcal{W}_{2,n}[\sum_{i=1}^n g_2(W_i,\theta)]}{n^2} \\ &= M_2(\theta)'\Omega_2^{-1}M_2(\theta) + o_p(1), \end{aligned} \tag{A.13}$$

uniformly over $\theta \in \Theta$. In addition, $M_2(\theta)'\Omega_2^{-1}M_2(\theta)$ uniquely identifies θ_0^* under Assumption 4.1, which was established in Lemma A.1. Given the uniform convergence of the criterion function

and the identification of θ_0^*, $\widehat{\theta}_2 \to_p \theta_0^*$ follows from standard arguments for the consistency of an extremum estimator.

Similarly, $\widehat{\theta}_1 \to_p \theta_0$ follows from the unique identification of θ_0 established in Lemma A.1, the uniform convergence of the GMM criterion function in (A.13), and $\mathcal{W}_{1,n} \to \Omega_1^{-1}$ by Assumption 4.2(v). In addition, we have $\theta_n \to \theta_0$ because the criterion function has a unique minimizer by Assumption 4.1. Finally, $\widehat{\theta}_1 - \theta_n = (\widehat{\theta}_1 - \theta_0) - (\theta_n - \theta_0) \to_p \mathbf{0}_{d_\theta}$. ■

Proof of Lemma 4.1. We first prove part (b) of the lemma. We start with showing that in this case $\theta_0^* = \theta_0$, where by definition θ_0^* uniquely minimizes $M_2(\theta)\Omega_2^{-1}M_2(\theta)$ and θ_0 is the unique value such that $M_1(\theta_0) = \mathbf{0}_{r_1}$. To this end, it is sufficient to show $M_2(\theta_0) = \mathbf{0}_{r_2}$ given that Ω_2 is positive definite. The condition $\delta(F_n) \to \mathbf{0}_{r^*}$ implies that $\mathbb{E}_{F_n}[g_2(W_i, \theta_n)] \to \mathbf{0}_{r_2}$. Because $\mathbb{E}_{F_n}[g(W_i, \theta)] \to M_2(\theta)$, $\theta_n \to \theta_0$, and $M_2(\theta)$ is continuous, we have $\mathbb{E}_{F_n}[g(W_i, \theta_n)] \to M_2(\theta_0) = \mathbf{0}_{r_2}$ as desired, which proves $\theta_0^* = \theta_0$ in this case. This together with Lemma A.2 implies that $\widehat{\theta}_2$ is consistent because

$$\widehat{\theta}_2 - \theta_n = (\widehat{\theta}_2 - \theta_0^*) + (\theta_0^* - \theta_0) + (\theta_0 - \theta_n) = o_p(1). \tag{A.14}$$

By the consistency of $\widehat{\theta}_2$ in (A.14), the stochastic equicontinuity of $\xi_n(g_2(\theta))$ in Assumption 4.3(iv), and Assumptions 4.2(i) and (ii), we have

$$n^{-1}\sum_{i=1}^n g_2(W_i, \widehat{\theta}_2) = n^{-1}\sum_{i=1}^n g_2(W_i, \theta_n) + [G_2 + o_p(1)]\,(\widehat{\theta}_2 - \theta_n) + o_p(n^{-1/2}). \tag{A.15}$$

Using the consistency of $\widehat{\theta}_2$ in (A.14), Assumption 4.2(ii) and Assumption 4.3(ii), we get

$$n^{-1}\sum_{i=1}^n g_{2,\theta}(W_i, \widehat{\theta}_2) = G_2 + o_p(1). \tag{A.16}$$

From the first order condition for the GMM estimator $\widehat{\theta}_2$, we deduce that

$$\begin{aligned}
0 &= \left[n^{-1}\sum_{i=1}^n g_{2,\theta}(W_i, \widehat{\theta}_2)\right]' \mathcal{W}_{2,n}\left[n^{-1}\sum_{i=1}^n g_2(W_i, \widehat{\theta}_2)\right] \\
&= [G_2 + o_p(1)]'\left[\Omega_2^{-1} + o_p(1)\right]\left\{n^{-1}\sum_{i=1}^n g_2(W_i, \theta_n) + [G_2 + o_p(1)]\,(\widehat{\theta}_2 - \theta_n) + o_p(n^{-1/2})\right\} \\
&= \left[G_2'\Omega_2^{-1} + o_p(1)\right]\left\{n^{-1}\sum_{i=1}^n g_2(W_i, \theta_n) + [G_2 + o_p(1)]\,(\widehat{\theta}_2 - \theta_n)\right\} + o_p(n^{-1/2})
\end{aligned} \tag{A.17}$$

where the second equality follows from (A.15), (A.16) and $\mathcal{W}_{2,n} - \Omega_2^{-1} \to_p \mathbf{0}_{r_2 \times r_2}$ by Assumption

4.2(v). By (A.17) and the regularity conditions in Assumption 4.2,

$$
\begin{aligned}
&n^{1/2}(\widehat{\theta}_2 - \theta_n) \\
&= -\left[\left(G_2'\Omega_2^{-1}G_2\right)^{-1} + o_p(1)\right]' \left[G_2'\Omega_2^{-1} + o_p(1)\right] \left[n^{-1/2}\sum_{i=1}^{n} g_2(W_i,\theta_n)\right] + o_p(1) \qquad \text{(A.18)} \\
&= -\left[\left(G_2'\Omega_2^{-1}G_2\right)^{-1} G_2'\Omega_2^{-1} + o_p(1)\right] \left\{\xi_n(g_2(\theta_n)) + n^{1/2}\mathbb{E}_{F_n}\left[g_2(W_i,\theta_n\right]\right\} + o_p(1).
\end{aligned}
$$

If $n^{1/2}\delta(F_n) \to d \in \mathbb{R}^{r^*}$, we have $n^{1/2}\mathbb{E}_{F_n}\left[g_2(W_i,\theta_n)\right] \to d_0 = [\mathbf{0}_{1\times r_1}, d']'$. Then, (A.18) implies that

$$
n^{1/2}(\widehat{\theta}_2 - \theta_n) \to_d -\left(G_2'\Omega_2^{-1}G_2\right)^{-1} G_2'\Omega_2^{-1}\left(\mathcal{Z}_2 + d_0\right), \text{ where } \mathcal{Z}_2 \sim N(\mathbf{0}_{r_2\times 1}, \Omega_2), \qquad \text{(A.19)}
$$

by the Slutzky's theorem and the CLT in Assumption 4.3(iii). This proves Part (b).

Part (a) follows from the same arguments as those for part (b) with all components for $\widehat{\theta}_2$ replaced by those for $\widehat{\theta}_1$ and d_0 replaced by 0 because all moments are correctly specified.

Next, we prove part (c). Lemma A.2 implies $\widehat{\theta}_2 \to_p \theta_0^*$. First, if $\theta_0^* = \theta_0$, the arguments for part (b) also applies here. In this case, $||n^{1/2}\mathbb{E}_{F_n}\left[g_2(W_i,\theta_n)\right]|| \to_p \infty$ and (A.18) implies that $|n^{1/2}(\widehat{\theta}_2 - \theta_n)| \to_p \infty$. Second, we consider the case in which $||\theta_0^* - \theta_0|| > 0$ for part (c). By the first order condition of the GMM estimator $\widehat{\theta}_2$,

$$
\begin{aligned}
0 &= \left[n^{-1}\sum_{i=1}^{n} g_{2,\theta}(W_i,\widehat{\theta}_2)\right]' W_{2,n}\left[n^{-1}\sum_{i=1}^{n} g_2(W_i,\widehat{\theta}_2)\right] \qquad \text{(A.20)} \\
&= \left[G_2\left(\theta_0^*\right)'\Omega_2^{-1} + o_p(1)\right]\left\{n^{-1}\sum_{i=1}^{n} g_2(W_i,\theta_0^*) + \left[G_2(\theta_0^*) + o_p(1)\right]\left(\widehat{\theta}_2 - \theta_0^*\right)\right\} + o_p(n^{-1/2})
\end{aligned}
$$

where the second equality is similar to that in (A.17) but is around the pseudo-true value θ_0^*. Then,

$$
\begin{aligned}
&n^{1/2}(\widehat{\theta}_2 - \theta_0^*) \\
&= -\left[\left(G_2(\theta_0^*)'\Omega_2^{-1}G_2(\theta_0^*)\right)^{-1} G_2(\theta_0^*)'\Omega_2^{-1} + o_p(1)\right] n^{-1/2}\sum_{i=1}^{n} g_2(W_i,\theta_0^*) + o_p(1) \\
&= O_p\left(\left\|G_2(\theta_0^*)'\Omega_2^{-1}\left(n^{-1/2}\sum_{i=1}^{n} g_2(W_i,\theta_0^*)\right)\right\|\right) + o_p(n^{-1/2}) = o_p(1), \qquad \text{(A.21)}
\end{aligned}
$$

where the first and second equalities follow from (A.20) and the regularity conditions in Assumption 4.2(iii) and(iv), and the third equality follows from

$$
\begin{aligned}
&G_2(\theta_0^*)'\Omega_2^{-1}n^{-1/2}\sum_{i=1}^{n} g_2(W_i,\theta_0^*) \qquad \text{(A.22)} \\
&= n^{1/2}G_2(\theta_0^*)'\Omega_2^{-1}\left\{\frac{\xi_n(g_2(\theta_0^*))}{n^{1/2}} + \left(\mathbb{E}_{F_n}\left[g_2(W_i,\theta_0^*)\right] - M_2(\theta_0^*)\right) + M_2(\theta_0^*)\right\} = o_p(n^{1/2}).
\end{aligned}
$$

In (A.22), the first equality is a simple decomposition, the second equality follows from the regularity conditions in Assumption 4.2, the ULLN in Assumption 4.3, $\mathbb{E}_{F_n}[g_2(W_i,\theta_0^*)] \to M_2(\theta_0^*)$ following $v(F_n) \to v_0$, and $G_2(\theta_0^*)'\Omega_2^{-1}M_2(\theta_0^*) = \mathbf{0}_{d_\theta}$, which in turn holds because (i) θ_0^* minimizes $M_2(\theta)'\Omega_2^{-1}M_2(\theta)$ and (ii) for some $F_0 \in \mathcal{F}$, $M_2(\theta) = \mathbb{E}_{F_0}[g_2(W_i,\theta)]$ and $G_2(\theta) = \mathbb{E}_{F_0}[g_{2,\theta}(W_i,\theta)] = \partial(\mathbb{E}_{F_0}[g_2(W_i,\theta)])/\partial\theta'$ by the dominated convergence theorem and Assumption 4.2.

In consequence,

$$\begin{aligned} n^{1/2}(\widehat{\theta}_2 - \theta_n) &= n^{1/2}(\widehat{\theta}_2 - \theta_0^*) + n^{1/2}(\theta_0^* - \theta_0) + n^{1/2}(\theta_0 - \theta_n) \\ &= n^{1/2}(\theta_0^* - \theta_0) + o_p(n^{1/2}), \end{aligned} \tag{A.23}$$

following $n^{1/2}(\widehat{\theta}_2 - \theta_0^*) = o_p(n^{1/2})$ and $\theta_n \to \theta_0$. Because $\theta_0^* \neq \theta_0$, it follows that $||n^{1/2}(\widehat{\theta}_2 - \theta_n)|| \to_p \infty$. This completes the proof of part (c). ■

A.3 Proofs for the optimal non-random weights

Proof of Lemma 4.2. We first consider $\{F_n\} \in \mathcal{S}(d, v_0)$ for $d \in \mathbb{R}^{r^*}$. By Lemma 4.1,

$$\begin{aligned} n^{1/2}\left[\widehat{\theta}(\omega) - \theta_n\right] &= n^{1/2}(\widehat{\theta}_1 - \theta_n) + \omega\left[n^{1/2}(\widehat{\theta}_2 - \theta_n) - n^{1/2}(\widehat{\theta}_1 - \theta_n)\right] \\ &\to_d \Gamma_1^*\mathcal{Z}_{d,2} + \omega(\Gamma_2 - \Gamma_1^*)\mathcal{Z}_{d,2}, \end{aligned} \tag{A.24}$$

under $\{F_n\} \in \mathcal{S}(d, v_0)$. This implies that

$$\begin{aligned} \ell(\widehat{\theta}(\omega)) &= n\left[\widehat{\theta}_n(\omega) - \theta_n\right]' H\left[\widehat{\theta}_n(\omega) - \theta_n\right] \to_d \lambda_{(d,v_0)}(\omega), \text{ where} \\ \lambda_{(d,v_0)}(\omega) &= \mathcal{Z}_{d,2}'\Gamma_1^{*\prime}H\Gamma_1^*\mathcal{Z}_{d,2} + 2\omega\mathcal{Z}_{d,2}'(\Gamma_2 - \Gamma_1^*)'H\Gamma_1^*\mathcal{Z}_{d,2} \\ &\quad + \omega^2\mathcal{Z}_{d,2}'(\Gamma_2 - \Gamma_1^*)'H(\Gamma_2 - \Gamma_1^*)\mathcal{Z}_{d,2}. \end{aligned} \tag{A.25}$$

under $\{F_n\} \in \mathcal{S}(d, v_0)$.

Now we consider the expectation of $\lambda_{(d,v_0)}(\omega)$ using the equalities in Lemma A.3 below. First,

$$\mathbb{E}[\mathcal{Z}_{d,2}'\Gamma_1^{*\prime}H\Gamma_1^*\mathcal{Z}_{d,2}] = tr(H\Sigma_1) \tag{A.26}$$

because $\Gamma_1^*\mathcal{Z}_{d,2} = \Gamma_1\mathcal{Z}_1$ and $\Gamma_1\mathbb{E}(\mathcal{Z}_1\mathcal{Z}_1')\Gamma_1' = \Sigma_1$ by definition. Second,

$$\begin{aligned} \mathbb{E}\left[\mathcal{Z}_{d,2}'(\Gamma_2 - \Gamma_1^*)'H\Gamma_1^*\mathcal{Z}_{d,2}\right] &= \mathrm{tr}(H\Gamma_1^*\mathbb{E}\left[\mathcal{Z}_{d,2}\mathcal{Z}_{d,2}'\right](\Gamma_2 - \Gamma_1^*)') \\ &= \mathrm{tr}(H\Gamma_1^*\left[d_0d_0' + \Omega_2\right](\Gamma_2 - \Gamma_1^*)') \\ &= \mathrm{tr}(H(\Sigma_2 - \Sigma_1)), \end{aligned} \tag{A.27}$$

where the last equality holds by Lemma A.3. Third,

$$\begin{aligned}\mathbb{E}\left[\mathcal{Z}_{d,2}'(\Gamma_2-\Gamma_1^*)'H(\Gamma_2-\Gamma_1^*)\mathcal{Z}_{d,2}\right] &= \mathrm{tr}(H(\Gamma_2-\Gamma_1^*)\left[d_0d_0'+\Omega_2\right](\Gamma_2-\Gamma_1^*)') \\ &= d_0'\Gamma_2'H\Gamma_2 d_0+\mathrm{tr}(H(\Sigma_1-\Sigma_2)) \end{aligned} \tag{A.28}$$

by Lemma A.3. Combining the results in (A.26)-(A.28), we obtain

$$\mathbb{E}[\lambda_{(d,v_0)}(\omega)] = \mathrm{tr}(H\Sigma_1)-2\omega\mathrm{tr}\left(H\left(\Sigma_1-\Sigma_2\right)\right)+\omega^2\left[d_0'\Gamma_2'H\Gamma_2d_0+tr\left(H\left(\Sigma_1-\Sigma_2\right)\right)\right]. \tag{A.29}$$

Note that $d_0'\Gamma_2'H\Gamma_2d_0 = d_0'(\Gamma_2-\Gamma_1^*)'H(\Gamma_2-\Gamma_1^*)d_0 = d_0'B_{v_0}d_0$ because $\Gamma_1^*d_0=\mathbf{0}_{d_\theta}$. This shows part (a).

Part (b) follows from part (a) by minimizing the quadratic function of ω.

Part (c) follows from Lemma 4.1 directly. ■

Lemma A.3 (a) $\Gamma_1^*d_0=\mathbf{0}_{d_\theta}$; (b) $\Gamma_1^*\Omega_2\Gamma_1^{*\prime}=\Sigma_1$; (c) $\Gamma_1^*\Omega_2\Gamma_2'=\Sigma_2$; (d) $\Gamma_2\Omega_2\Gamma_2'=\Sigma_2$.

Proof of Lemma A.3. By construction, $\Gamma_1^*d_0=\mathbf{0}_{d_\theta}$. For the ease of notation, we write Ω_2 and G_2 as

$$\Omega_2=\begin{pmatrix}\Omega_1 & \Omega_{1r}\\ \Omega_{r1} & \Omega_r\end{pmatrix} \text{ and } G_2=\begin{pmatrix}G_1\\ G_r\end{pmatrix}. \tag{A.30}$$

To prove part (b), we have

$$\begin{aligned}\Gamma_1^*\Omega_2\Gamma_1^{*\prime} &= [\Gamma_1,\mathbf{0}_{d_\theta\times r^*}]\begin{pmatrix}\Omega_1 & \Omega_{1r}\\ \Omega_{r1} & \Omega_r\end{pmatrix}[\Gamma_1,\mathbf{0}_{d_\theta\times r^*}] \\ &= \Gamma_1\Omega_1\Gamma_1' = \left(G_1'\Omega_1^{-1}G_1\right)^{-1}=\Sigma_1. \end{aligned} \tag{A.31}$$

To show part (c), note that

$$\begin{aligned}\Gamma_1^*\Omega_2\Gamma_2' &= -\left[\Gamma_1,\mathbf{0}_{d_\theta\times r^*}\right]\Omega_2\Omega_2^{-1}G_2\left(G_2'\Omega_2^{-1}G_2\right)^{-1} \\ &= -\Gamma_1G_1\left(G_2'\Omega_2^{-1}G_2\right)^{-1}=\left(G_2'\Omega_2^{-1}G_2\right)^{-1}=\Sigma_2 \end{aligned} \tag{A.32}$$

because $-\Gamma_1G_1=I_{d_\theta\times d_\theta}$. Part (d) follows from the definition of Γ_2. ■

A.4 Proof for the empirical optimal averaging estimator

In the proofs below, we use A, B and D to denote A_{v_0}, B_{v_0} and D_{v_0}, respectively, for notational simplicity.

Proof of Lemma 4.3. We first consider $\{F_n\}\in\mathcal{S}(d,v_0)$. By Lemma 4.1, Assumption 4.4, and

the continuous mapping theorem (CMT),

$$\widetilde{\omega}_{eo} \to_d \widetilde{\omega}_{(d,v_0)} = \frac{\mathrm{tr}\,(A)}{\mathcal{Z}'_{d,2} B \mathcal{Z}_{d,2} + \mathrm{tr}\,(A)}. \tag{A.33}$$

Then,

$$\begin{aligned} n^{1/2}(\widehat{\theta}_{eo} - \theta_n) &= n^{1/2}(\widehat{\theta}_1 - \theta_n) + \widetilde{\omega}_{eo}\left[n^{1/2}(\widehat{\theta}_2 - \theta_n) - n^{1/2}(\widehat{\theta}_1 - \theta_n)\right] \\ &\to_d \phi_{(d,v_0)} = \Gamma_1^* \mathcal{Z}_{d,2} + \widetilde{\omega}_{(d,v_0)}\left(\Gamma_2 - \Gamma_1^*\right)\mathcal{Z}_{d,2}. \end{aligned} \tag{A.34}$$

By the CMT,

$$\ell(\widehat{\theta}_{eo}) = n\left[(\widehat{\theta}_{eo} - \theta_n)' H (\widehat{\theta}_{eo} - \theta_n)\right] \to_d \lambda_{(d,v_0)} = \phi'_{(d,v_0)} H \phi_{(d,v_0)}. \tag{A.35}$$

Under $\{F_n\} \in \mathcal{S}(\infty, v_0)$, $\widetilde{\omega}_{eo} \to_p 0$ because $n^{1/2}|\widehat{\theta}_2 - \widehat{\theta}_1| \to_p \infty$,

$$\begin{aligned} &n^{1/2}(\widehat{\theta}_{eo} - \theta_n) = n^{1/2}(\widehat{\theta}_1 - \theta_n) + \widetilde{\omega}_{eo} n^{1/2}(\widehat{\theta}_2 - \widehat{\theta}_1) \\ &= n^{1/2}(\widehat{\theta}_1 - \theta_n) + \frac{n^{1/2}(\widehat{\theta}_2 - \widehat{\theta}_1) tr\left[H(\widehat{\Sigma}_1 - \widehat{\Sigma}_2)\right]}{n(\widehat{\theta}_2 - \widehat{\theta}_1)' H (\widehat{\theta}_2 - \widehat{\theta}_1) + tr\left[H(\widehat{\Sigma}_1 - \widehat{\Sigma}_2)\right]} \\ &\to_d \phi_{(\infty,v_0)} = \Gamma_1 \mathcal{Z}_1 \end{aligned} \tag{A.36}$$

by Lemma 4.1. Then by the CMT,

$$\ell(\widehat{\theta}_{eo}) \to_d \lambda_{(\infty,v_0)} = \phi'_{(\infty,v_0)} H \phi_{(\infty,v_0)}. \tag{A.37}$$

■

Proof of Theorem 4.1. For any $\zeta \in \mathbb{R}_+$, under $\{F_n\} \in \mathcal{S}(d, v_0)$,

$$\mathbb{E}\left[\ell_\zeta(\widehat{\theta}_{eo})\right] \to \mathbb{E}\left[\min\{\lambda_{(d,v_0)}, \zeta\}\right] \tag{A.38}$$

by the Portmanteau Lemma and Lemma 4.3(a) given that $\ell_\zeta(\widehat{\theta}_{eo})$ is bounded by ζ. Similarly under $\{F_n\} \in \mathcal{S}(\infty, v_0)$,

$$\mathbb{E}\left[\ell_\zeta(\widehat{\theta}_{eo})\right] \to \mathbb{E}\left[\min\{\lambda_{(\infty,v_0)}, \zeta\}\right]. \tag{A.39}$$

for any $\zeta \in \mathbb{R}$. Under $\{F_n\} \in \mathcal{S}_2(v_0)$, the conservative estimator $\widehat{\theta}_1$ satisfies

$$\mathbb{E}\left[\ell_\zeta(\widehat{\theta}_1)\right] \to \mathbb{E}\left[\min\{\mathcal{Z}_1' \Gamma_1' H \Gamma_1 \mathcal{Z}_1, \zeta\}\right] = \mathbb{E}\left[\min\{\lambda_{(\infty,v_0)}, \zeta\}\right]. \tag{A.40}$$

This verifies Assumptions 3.3 with

$$\begin{aligned} R_\zeta(d,v_0) &= \mathbb{E}[\min\{\lambda_{(d,v_0)},\zeta\}] \\ R_\zeta(\infty,v_0) &= \mathbb{E}\left[\min\{\lambda_{(\infty,v_0)},\zeta\}\right], \text{ and} \\ \widetilde{R}_\zeta(d,v_0) &= \mathbb{E}\left[\min\{\lambda_{(\infty,v_0)},\zeta\}\right] = \widetilde{R}(\infty,v_0), \end{aligned} \tag{A.41}$$

for $d \in \mathbb{R}^{r^*}$. Part (a) follows from Corollary 3.2 with

$$\begin{aligned} Asy\underline{RD}^*_\zeta(\widehat{\theta}_{eo},\widehat{\theta}_1) &= \min\left\{\inf_{(d,v_0)\in H_R} g_\zeta(d,v_0), 0\right\}, \\ Asy\overline{RD}^*_\zeta(\widehat{\theta}_{eo},\widehat{\theta}_1) &= \max\left\{\sup_{(d,v_0)\in H_R} g_\zeta(d,v_0), 0\right\}. \end{aligned} \tag{A.42}$$

Next, we show the upper bound for $g_\zeta(d,v_0)$ when ζ is large in part (b). As $\min\{\lambda_{(d,v_0)},\zeta\} \leq \lambda_{(d,v_0)}$ with probability 1,

$$\mathbb{E}\left[\min\{\lambda_{(d,v_0)},\zeta\}\right] \leq \mathbb{E}\left[\lambda_{(d,v_0)}\right]. \tag{A.43}$$

The expectation $\mathbb{E}\left[\lambda_{(d,v_0)}\right]$ exists because

$$\begin{aligned} \mathbb{E}\left[\lambda_{(d,v_0)}\right] &\leq 2\mathbb{E}\left[\mathcal{Z}'_{d,2}\Gamma_1^{*\prime}H\Gamma_1^*\mathcal{Z}_{d,2} + \widetilde{\omega}^2_{(d,v_0)}\mathcal{Z}'_{d,2}\left(\Gamma_2-\Gamma_1^*\right)'H\left(\Gamma_2-\Gamma_1^*\right)\mathcal{Z}_{d,2}\right] \\ &= 2\mathbb{E}\left[\mathcal{Z}_1'\Gamma_1'H\Gamma_1\mathcal{Z}_1 + \mathrm{tr}(A)\frac{\mathrm{tr}(A)}{\mathcal{Z}'_{d,2}B\mathcal{Z}_{d,2}+\mathrm{tr}(A)}\frac{\mathcal{Z}'_{d,2}B\mathcal{Z}_{d,2}}{\mathcal{Z}'_{d,2}B\mathcal{Z}_{d,2}+\mathrm{tr}(A)}\right] \\ &\leq 2\mathbb{E}\left[\mathcal{Z}_1'\Gamma_1'H\Gamma_1\mathcal{Z}_1\right] + 2\mathrm{tr}(A) \leq C \end{aligned} \tag{A.44}$$

where the first inequality is by the Cauchy-Schwarz inequality, the third inequality is by

$$\frac{\mathrm{tr}(A)}{\mathcal{Z}'_{d,2}B\mathcal{Z}_{d,2}+\mathrm{tr}(A)} \leq 1 \text{ and } \frac{\mathcal{Z}'_{d,2}B\mathcal{Z}_{d,2}}{\mathcal{Z}'_{d,2}B\mathcal{Z}_{d,2}+\mathrm{tr}(A)} \leq 1 \text{ with probability 1.} \tag{A.45}$$

And the last inequality is by the regularity conditions in Assumption 4.2(iii) and (iv). Similarly, we also have

$$\mathbb{E}\left[\lambda_{(\infty,v_0)}\right] \leq C_1\mathbb{E}[\,\|\Gamma_1\mathcal{Z}_1\|]^2 \leq C \tag{A.46}$$

for some $C_1, C < \infty$.

By the definitions of $\lambda_{(d,v_0)}$ and $\lambda_{(\infty,v_0)}$, we can write

$$\begin{aligned} g(d,v_0) &= \mathbb{E}[\lambda_{(d,v_0)}] - \mathbb{E}\left[\lambda_{(\infty,v_0)}\right] = 2\mathrm{tr}(A)J_1 + \mathrm{tr}(A)^2J_2, \text{ where} \\ J_1 &= \mathbb{E}\left[\frac{\mathcal{Z}'_{d,2}D\mathcal{Z}_{d,2}}{\mathcal{Z}'_{d,2}B\mathcal{Z}_{d,2}+\mathrm{tr}(A)}\right] \text{ and } J_2 = \mathbb{E}\left[\frac{\mathcal{Z}'_{d,2}B\mathcal{Z}_{d,2}}{(\mathcal{Z}'_{d,2}B\mathcal{Z}_{d,2}+\mathrm{tr}(A))^2}\right]. \end{aligned} \tag{A.47}$$

From the definition of $g_\zeta(d, v_0)$ and $g(d, v_0)$, we use (A.43) to deduce that

$$\begin{aligned} g_\zeta(d, v_0) &= \mathbb{E}\left[\min\{\lambda_{(d,v_0)}, \zeta\}\right] - \mathbb{E}\left[\min\{\lambda_{(\infty,v_0)}, \zeta\}\right] \\ &\leq \mathbb{E}\left[\lambda_{(d,v_0)}\right] - \mathbb{E}\left[\min\{\lambda_{(\infty,v_0)}, \zeta\}\right] \\ &= \mathbb{E}[\lambda_{(\infty,v_0)} - \min\{\lambda_{(\infty,v_0)}, \zeta\}] + g(d, v_0) \end{aligned} \tag{A.48}$$

for any $(d, v_0) \in H_R$.

Next we show

$$\lim_{\zeta\to\infty} \sup_{(d,v_0)\in H_R} \left\{\mathbb{E}[\lambda_{(\infty,v_0)} - \min\{\lambda_{(\infty,v_0)}, \zeta\}]\right\} = 0. \tag{A.49}$$

Recall that Γ_1 is a function of G_1 and Ω_1. Define

$$q(\mathcal{Z}, G_1, \Omega_1) \equiv \mathcal{Z}'\Omega_1^{1/2}\Gamma_1' H \Gamma_1 \Omega_1^{1/2}\mathcal{Z}, \text{ where } \mathcal{Z} \sim N(\mathbf{0}_{r_1}, I_{r_1\times r_1}). \tag{A.50}$$

Then we can write

$$\begin{aligned} f_\zeta(G_1, \Omega_1) &\equiv \mathbb{E}[\lambda_{(\infty,v_0)} - \min\{\lambda_{(\infty,v_0)}, \zeta\}] \\ &= \mathbb{E}\left[q(\mathcal{Z}, G_1, \Omega_1) - \min\{q(\mathcal{Z}, G_1, \Omega_1), \zeta\}\right] \end{aligned} \tag{A.51}$$

following the definition of $\lambda_{(\infty,v_0)}$. Let

$$\Upsilon_1 = \{(G_1, \Omega_1) : G_1(F_n) \to G_1 \text{ and } \Omega_1(F_n) \to \Omega_1 \text{ for some } \{F_n\} \in \mathcal{F}\}. \tag{A.52}$$

We now have

$$\lim_{\zeta\to\infty} \sup_{(d,v_0)\in H_R} f_\zeta(G_1, \Omega_1) \leq \lim_{\zeta\to\infty} \sup_{(G_1,\Omega_1)\in\Upsilon_1} f_\zeta(G_1, \Omega_1) \tag{A.53}$$

because $(d, v_0) \in H_R$ requires the convergence listed in (A.52) as well as the convergence of some other functions.

It remains to show $\lim_{\zeta\to\infty} \sup_{(G_1,\Omega_1)\in\Upsilon_1} f_\zeta(G_1, \Omega_1) = 0$. First, $\forall (G_1, \Omega_1) \in \Upsilon_1$, $\lim_{\zeta\to\infty} f_\zeta(G_1, \Omega_1) = 0$ by the dominated convergence theorem (DCT) because

$$0 \leq q(\mathcal{Z}, G_1, \Omega_1) - \min\{q(\mathcal{Z}, G_1, \Omega_1), \zeta\} \leq q(\mathcal{Z}, G_1, \Omega_1) \tag{A.54}$$

and $\mathbb{E}\left[q(\mathcal{Z}, G_1, \Omega_1)\right] = \text{tr}(H\Sigma_1) \leq C$. Second, this convergence is uniform in $(G_1, \Omega_1) \in \Upsilon_1$ by the Dini's Theorem (see, Rudin (1976)) because (i) $f_\zeta(G_1, \Omega_1)$ is monotonically decreasing in ζ, (ii) Υ_1 is compact, and (iii) $f_\zeta(G_1, \Omega_1)$ is continuous in (G_1, Ω_1). The set Υ_1 is compact following Assumption 3.2(i). The continuity of $f_\zeta(G_1, \Omega_1)$ in (G_1, Ω_1) is by the DCT because (a) $q(\mathcal{Z}, G_1, \Omega_1)$ is continuous in (G_1, Ω_1) and (b) $\mathbb{E}[\sup_{(G_1,\Omega_1)\in\Upsilon_1} q(\mathcal{Z}, G_1, \Omega_1)] < \infty$. To see (b), note that

$$\sup_{(G_1,\Omega_1)\in\Upsilon_1} q(\mathcal{Z}, G_1, \Omega_1) \leq \left[\sup_{(G_1,\Omega_1)\in\Upsilon_1} \lambda_{\max}\left(\Omega_1^{1/2}\Gamma_1' H \Gamma_1 \Omega_1^{1/2}\right)\right] \mathcal{Z}'\mathcal{Z} \leq C\mathcal{Z}'\mathcal{Z} \tag{A.55}$$

by Assumption 4.2(iii) and (iv).

This completes the verification of (A.49). It follows from (A.49) that for large ζ,

$$\sup_{(d,v_0)\in H_R} g_\zeta(d,v_0) \leq \sup_{(d,v_0)\in H_R} g(d,v_0) \text{ and } \inf_{(d,v_0)\in H_R} g_\zeta(d,v_0) \leq \inf_{(d,v_0)\in H_R} g(d,v_0). \tag{A.56}$$

Next, we provide a upper bound for J_1. Let

$$\eta(x) = \frac{x}{x'Bx + \text{tr}(A)}, \text{ where } x = \mathcal{Z}_{d,2} \text{ and } B = (\Gamma_2 - \Gamma_1^*)'H(\Gamma_2 - \Gamma_1^*). \tag{A.57}$$

Its derivative is

$$\frac{\partial \eta(x)'}{\partial x} = \frac{1}{x'Bx + \text{tr}(A)} I_{r_2} - \frac{2}{(x'Bx + \text{tr}(A))^2} Bxx'. \tag{A.58}$$

Recall that

$$D = (\Gamma_2 - \Gamma_1^*)'H\Gamma_1^*, \tag{A.59}$$

which satisfies $D\mathcal{Z}_{d,2} = D\mathcal{Z}_2$ by construction because the last r^* rows of Γ_1^* are zeros. By Lemma 1 of Hansen (2014a), which is a matrix version of the Stein's Lemma (Stein, 1981),

$$J_1 = \mathbb{E}\left(\eta(\mathcal{Z}_{d,2})'D\mathcal{Z}_{d,2}\right) = \mathbb{E}\left(\eta(\mathcal{Z}_{d,2})'D\mathcal{Z}_2\right) = \mathbb{E}\left[\text{tr}\left(\frac{\partial \eta(\mathcal{Z}_{d,2})'}{\partial x} D\Omega_2\right)\right]. \tag{A.60}$$

Applying Lemma A.3 yields

$$\begin{aligned} \text{tr}\left(D\Omega_2\right) &= \text{tr}\left((\Gamma_2 - \Gamma_1^*)'H\Gamma_1^*\Omega_2\right) = \text{tr}(H\left(\Gamma_1^*\Omega_2\Gamma_2 - \Gamma_1^*\Omega_2\Gamma_1^*\right)) \\ &= \text{tr}(H\left(\Sigma_2 - \Sigma_1\right)) = -\text{tr}(A). \end{aligned} \tag{A.61}$$

Plugging (A.57)-(A.59) into (A.60), we have

$$
\begin{aligned}
J_1 &= \mathbb{E}\left[\frac{\mathcal{Z}_{d,2}' D \mathcal{Z}_{d,2}}{\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)}\right] \\
&= \mathbb{E}\left[\frac{\mathrm{tr}\left(D\Omega_2\right)}{\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)}\right] - 2\mathbb{E}\left[\frac{\mathrm{tr}\left(\left\{B\mathcal{Z}_{d,2}\mathcal{Z}_{d,2}'\right\} D\Omega_2\right)}{\left(\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)\right)^2}\right] \\
&\leq \mathbb{E}\left[\frac{-\mathrm{tr}(A)}{\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)}\right] + 2\mathbb{E}\left[\frac{\left(\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2}\right)\lambda_{\max}(A)}{\left(\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)\right)^2}\right] \\
&= \mathbb{E}\left[\frac{-\mathrm{tr}(A)}{\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)}\right] + 2\mathbb{E}\left[\frac{\left[\left(\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2}\right) + \mathrm{tr}(A)\right]\lambda_{\max}(A) - \mathrm{tr}(A)\lambda_{\max}(A)}{\left(\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)\right)^2}\right] \\
&= \mathbb{E}\left[\frac{2\lambda_{\max}(A) - \mathrm{tr}(A)}{\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)}\right] - \mathbb{E}\left[\frac{2\lambda_{\max}(A)\mathrm{tr}(A)}{\left(\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)\right)^2}\right],
\end{aligned} \tag{A.62}
$$

where the inequality follows from (A.61) and $\mathrm{tr}(CD) \leq \mathrm{tr}(C)\lambda_{\max}(D)$. Next, note that

$$
\begin{aligned}
J_2 &= \mathbb{E}\left[\frac{Z_{d,2}' B \mathcal{Z}_{d,2}}{\left|\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)\right|^2}\right] = \mathbb{E}\left[\frac{\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A) - \mathrm{tr}(A)}{\left|\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)\right|^2}\right] \\
&= \mathbb{E}\left[\frac{1}{\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)}\right] - \mathbb{E}\left[\frac{\mathrm{tr}(A)}{\left(\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)\right)^2}\right].
\end{aligned} \tag{A.63}
$$

Combining (A.62) and (A.63), we obtain that

$$
\begin{aligned}
g(d, v_0) &= 2\mathrm{tr}(A)J_1 + \mathrm{tr}(A)^2 J_2 \\
&\leq 2\mathrm{tr}(A)\left(\mathbb{E}\left[\frac{2\lambda_{\max}(A) - \mathrm{tr}(A)}{\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)}\right] - \mathbb{E}\left[\frac{2\mathrm{tr}(A)\lambda_{\max}(A)}{\left(\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)\right)^2}\right]\right) \\
&\quad + \mathrm{tr}(A)^2\left(\mathbb{E}\left[\frac{1}{\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)}\right] - \mathbb{E}\left[\frac{\mathrm{tr}(A)}{\left(\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)\right)^2}\right]\right) \\
&= \mathbb{E}\left[\frac{\mathrm{tr}(A)\left(4\lambda_{\max}(A) - \mathrm{tr}(A)\right)}{\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)}\right] - \mathbb{E}\left[\frac{\mathrm{tr}(A)^2\left(4\lambda_{\max}(A) + \mathrm{tr}(A)\right)}{\left(\mathcal{Z}_{d,2}' B \mathcal{Z}_{d,2} + \mathrm{tr}(A)\right)^2}\right].
\end{aligned} \tag{A.64}
$$

To show part (c), note that for any v_0 such that $(d, v_0) \in H_R$ for some $d \in \mathbb{R}^{r^*}$, we have $G_2 = G_2(F)$ and $\Omega_2 = \Omega_2(F)$ for some $F \in \mathcal{F}$. This implies that $\Sigma_1 = \Sigma_1(F)$ and $\Sigma_2 = \Sigma_2(F)$ for

some $F \in \mathcal{F}$ for any $(d, v_0) \in H_R$. Therefore,

$$\sup_{(d,v_0)\in H_R} g(d, v_0) \leq 0 \text{ and } \inf_{(d,v_0)\in H_R} g(d, v_0) < 0 \tag{A.65}$$

if $A = H(\Sigma_1(F) - \Sigma_2(F))$ satisfies $\text{tr}(A) > 0$ and $4\lambda_{\max}(A) - \text{tr}(A) \leq 0$ for $\forall F \in \mathcal{F}$. The claim in part (c) follows from (A.65) and part (a). ■

A.5 Asymptotic risk of the pre-test GMM estimator with the J-test statistic

To simulate the asymptotic risk of the pre-test estimator in Figure 1, we consider the asymptotic risk under $\{F_n\} \in \mathcal{S}(d, v_0)$ and $\{F_n\} \in \mathcal{S}(\infty, v_0)$. We first consider $\{F_n\} \in \mathcal{S}(d, v_0)$ for $d \in \mathbb{R}^{r^*}$. Under Assumptions 4.2 and 4.3, by (A.15) and (A.18), it is easy to show that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n g_2(W_i, \widehat{\theta}_2) &= \frac{1}{n}\sum_{i=1}^n g_2(W_i, \theta_n) + G_2(\widehat{\theta}_2 - \theta_n) + o_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^n g_2(W_i, \theta_n) + G_2\Gamma_2\frac{1}{n}\sum_{i=1}^n g_2(W_i, \theta_n) + o_p(n^{-1/2}) \\
&= (I_{r_2} + G_2\Gamma_2)\frac{1}{n}\sum_{i=1}^n g_2(W_i, \theta_n) + o_p(n^{-1/2})
\end{aligned} \tag{A.66}$$

which together with Assumptions 4.2(v) implies that

$$\begin{aligned}
J_n &\equiv \left[n^{-1/2}\sum_{i=1}^n g_2(W_i, \widehat{\theta}_2)\right]' W_{2,n}\left[n^{-1/2}\sum_{i=1}^n g_2(W_i, \widehat{\theta}_2)\right] \qquad \text{(A.67)} \\
&= \left[n^{-1/2}\sum_{i=1}^n g_2(W_i, \theta_n)\right]' \left[\Omega_2^{-1} - \Omega_2^{-1}G_2\left(G_2'\Omega_2^{-1}G_2\right)^{-1}G_2'\Omega_2^{-1}\right]\left[n^{-1/2}\sum_{i=1}^n g_2(W_i, \theta_n)\right] + o_p(1).
\end{aligned}$$

By Assumption 4.3(iii), we have

$$n^{-1/2}\sum_{i=1}^n g_2(W_i, \theta_n) \to_d \mathcal{Z}_2 + d_0 = \mathcal{Z}_{d,2}, \tag{A.68}$$

which combined with the CMT implies that

$$J_n \to_d \mathcal{Z}_{d,2}'\left[\Omega_2^{-1} - \Omega_2^{-1}G_2\left(G_2'\Omega_2^{-1}G_2\right)^{-1}G_2'\Omega_2^{-1}\right]\mathcal{Z}_{d,2} \equiv J_\infty(G_2, \Omega_2, d_0) \tag{A.69}$$

for any $\{F_n\} \in \mathcal{S}(d, v_0)$.

Let α be a prespecified significance level. The critical value c_α is the $1-\alpha$ quantile of the chi-square distribution with $(r_2 - d_\theta)$ degree of freedom, which is the asymptotic distribution of J_n when $d_0 = \mathbf{0}_{r_2}$.

The pre-test estimator based on J_n can be written as

$$\widehat{\theta}_p = (1 - \widetilde{\omega}_{\alpha,p})\widehat{\theta}_1 + \widetilde{\omega}_{\alpha,p}\widehat{\theta}_2 \text{ where } \widetilde{\omega}_{\alpha,p} = I\left\{J_n < c_\alpha\right\}. \tag{A.70}$$

By (A.69) and the CMT, we have

$$\widetilde{\omega}_{\alpha,p} \to_d I\left\{J_\infty(G_2, \Omega_2, d_0) < c_\alpha\right\} \equiv \widetilde{\omega}_{\alpha,\infty} \tag{A.71}$$

and

$$n^{1/2}(\widehat{\theta}_p - \theta_n) = \widehat{\theta}_1 + \widetilde{\omega}_{\alpha,p}(\widehat{\theta}_2 - \widehat{\theta}_1) \to_d \Gamma_1^* \mathcal{Z}_{d,2} + \widetilde{\omega}_{\alpha,\infty}(\Gamma_2 - \Gamma_1^*)\mathcal{Z}_{d,2}, \tag{A.72}$$

under $\{F_n\} \in \mathcal{S}(d, v_0)$. Furthermore, by the CMT,

$$\ell(\widehat{\theta}_p) = n(\widehat{\theta}_p - \theta_n)' H (\widehat{\theta}_p - \theta_n) \to_d \lambda_{(d,v_0)}(\widetilde{\omega}_{\alpha,\infty}), \tag{A.73}$$

where

$$\begin{aligned}\lambda_{(d,v_0)}(\widetilde{\omega}_{\alpha,\infty}) = {} & \mathcal{Z}_{d,2}'\Gamma_1^{*\prime} H \Gamma_1^* \mathcal{Z}_{d,2} + 2\widetilde{\omega}_{\alpha,\infty}\mathcal{Z}_{d,2}'(\Gamma_2 - \Gamma_1^*)' H \Gamma_1^* \mathcal{Z}_{d,2} \\ & + \widetilde{\omega}_{\alpha,\infty}^2 \mathcal{Z}_{d,2}'(\Gamma_2 - \Gamma_1^*)' H (\Gamma_2 - \Gamma_1^*)\mathcal{Z}_{d,2}.\end{aligned} \tag{A.74}$$

and its expectation is

$$\begin{aligned}\mathbb{E}[\lambda_{(d,v_0)}(\widetilde{\omega}_{\alpha,\infty})] = {} & tr(H\Sigma_1) + 2\mathbb{E}\left[\widetilde{\omega}_{\alpha,\infty}\mathcal{Z}_{d,2}'(\Gamma_2 - \Gamma_1^*)' H \Gamma_1^* \mathcal{Z}_{d,2}\right] \\ & + \mathbb{E}\left[\widetilde{\omega}_{\alpha,\infty}\mathcal{Z}_{d,2}'(\Gamma_2 - \Gamma_1^*)' H (\Gamma_2 - \Gamma_1^*)\mathcal{Z}_{d,2}\right].\end{aligned} \tag{A.75}$$

This verifies Assumption 3.1 for the pre-test estimator with $R(d, v_0) = \mathbb{E}[\lambda_{(d,v_0)}(\widetilde{\omega}_{\alpha,\infty})]$, assuming it is uniformly integrable. Otherwise, we can consider the truncated risk. For $\{F_n\} \in \mathcal{S}(\infty, v_0)$, the J-test is consistent and the pretest estimator is the conservative GMM estimator w.p.a.1.

The asymptotic risk of the pre-test estimator $\widehat{\theta}_p$ in Figure 1 is simulated based on the formula in (A.75).